Pontifícia Universidade Católica
do Rio de Janeiro

Felipe Schoemer Jardim

Xbar Chart with Estimated Parameters: The Average Run Length Distribution and Corrections to the Control Limits

Tese de Doutorado

Thesis presented to the Programa de Pós-Graduação em Engenharia de Produção of PUC-Rio in partial fulfillment of the requirements for the degree of Doutor em Engenharia de Produção.

Advisor: Prof. Eugenio Kahn Epprecht

Co-advisor: Prof. Subhabrata Chakraborti

Rio de Janeiro

March 2018

Pontifícia Universidade Católica
do Rio de Janeiro

Felipe Schoemer Jardim

Xbar Chart with Estimated Parameters: The Average Run Length Distribution and Corrections to the Control Limits

Thesis presented to the Programa de Pós-Graduação em Engenharia de Produção of PUC-Rio in partial fulfillment of the requirements for the degree of Doutor em Engenharia de Produção. Approved by the undersigned Examination Committee.

Prof. Eugênio Kahn Epprecht
Advisor
Departamento de Engenharia Industrial - PUC-Rio

Prof. Fernando Luiz Cyrino
Departamento de Engenharia Industrial - PUC-Rio

Prof. Helio Côrtes Vieira Lopes
Departamento de Informática - PUC-Rio

Prof. Maysa Sacramento de Magalhães
Fundação Instituto Brasileiro de Geografia e Estatística – IBGE/ENCE

Prof. Mônica Barros
Pesquisador Autônomo

Prof. Bruno Francisco Teixeira Simões
Universidade Federal do Estado do Rio de Janeiro - UNIRIO

Prof. Márcio da Silveira Carvalho
Vice Dean of Graduate Studies
Centro Técnico Científico – PUC-Rio

Rio de Janeiro, March 23rd, 2018

Felipe Schoemer Jardim

Graduated in Electrical Engineering in 2009 (with emphasis in Decision Support Systems) at the Pontifical Catholic University of Rio de Janeiro (PUC-Rio). Between 2008 and 2011, he was a trainee at Rede Globo Group and, as a graduated engineer, worked in the real estate sector. In 2013, he obtained his M.Sc degree in Production Engineering in PUC-Rio followed by his registration in the Department of Industrial Engineering to obtain his D.Sc. degree with the support of The University of Alabama, where he was a visiting researcher for two periods in 2016.

Bibliographic data

Jardim, Felipe Schoemer

Xbar chart with estimated parameters: the average run length distribution and corrections to the control limits / Felipe Schoemer Jardim ; advisor: Eugenio Kahn Epprecht ; co-advisor: Subhabrata Chakraborti. – 2018.

202 f. : il. ; 30 cm

Tese (doutorado)–Pontifícia Universidade Católica do Rio de Janeiro, Departamento de Engenharia Industrial, 2018.

Inclui bibliografia

1. Engenharia Industrial – Teses. 2. Desempenho do gráfico de controle Xbarra. 3. Desempenho condicional. 4. Ajuste nos limites de controle. 5. Desempenho em controle garantido. 6. Variabilidade de desempenho entre usuários. I. Epprecht, Eugenio Kahn. II. Chakraborti, Subhabrata. III. Pontifícia Universidade Católica do Rio de Janeiro. Departamento de Engenharia Industrial. IV. Título.

CDD: 658.5

Acknowledgements

First, I would like to thank my advisors Prof. Eugenio Kahn Epprecht and Prof. Subhabrata Chakraborti of The University of Alabama for the time they have spent on me. Since the time of my masters, Prof. Epprecht always believed in my potential of doing quality research. He provided me with important knowledge in statistics. During the period of my doctorate research, he introduced me to Prof. Chakraborti who showed me some of his papers that literally revolutionized my work and inspired me to change the topic I was working on back then. Also, he introduced me to some of the topics that are on the rise in the international Statistical Process Control area and presented me with new ideas and statistical techniques. He encouraged me to collaborate and participate in international conferences. All of this has made my research work and outputs achieve international standards. To be honest, words are not enough to express how I am grateful to my advisors.

Also, many thanks to: **(1)** my colleagues Martin Sarmiento and Dey Salvador for the many discussions about our research topics. **(2)** my parents Eduardo Galvão Moura Jardim and Regina Schoemer Jardim that always encouraged, supported and helped me throughout the period of the doctorate. **(3)** Júlia Lorêdo Pereira Leite who stayed by my side throughout all these years supporting me. This was essential for the accomplish of this work. **(4)** my brothers and sisters: Leandro Jardim, Paula Jardim and Henrique Jardim. **(5)** The Faculty and Staff of the Department of Industrial Engineering of PUC-Rio, in particular, to Claudia Teti, Eduardo, Fernanda Lopes, Gilvan Nascimento, Isabel Finoti and Graça. **(6)** The University of Alabama and the Department of Information Systems, Statistics and Management Science by giving me the opportunity to visit the university as a researcher for a total of 6 months and work with Prof. Chakraborti. **(7)** CNPq (Brazilian Council for Scientific and Technological Development) for an international split PhD scholarship and to CAPES (Brazilian Coordination for the Improvement of Higher Education Personnel) for a national PhD scholarship.

Abstract

Jardim, Felipe Schoemer; Epprecht, Eugenio Kahn (Advisor); Chakraborti, Subhabrata (Co-advisor). **Xbar Chart with Estimated Parameters: The Average Run Length Distribution and Corrections to the Control Limits**. Rio de Janeiro, 2018. 202p. Tese de Doutorado – Departamento de Engenharia Industrial, Pontifícia Universidade Católica do Rio de Janeiro.

Control charts are among the indispensable tools for monitoring process performance in various industries. When parameter estimation is needed to design these charts, their performance is affected due to parameter estimation errors. To overcome this problem, in the past, researchers have evaluated the performance of control charts and designed them in terms of the expectation of the realized in-control (IC) average run length ($CARL_0$). But, as pointed recently, this solution does not account for what is known as the practitioner-to-practitioner variability (i.e., the variability of $CARL_0$). So, a recent idea emerged where control chart performance is measured by the probability of the $CARL_0$ being greater than a specified value – which must be close to the nominal desired one. This is called the Exceedance Probability Criterion (*EPC*). To apply the *EPC*, the cumulative distribution function (c.d.f.) of the $CARL_0$ is required. However, for the most well-known control chart, named the two-sided Shewhart Xbar (or simply $\bar{X}$) Chart (under normality assumption), the mathematical c.d.f. expression of the $CARL_0$ is not available in the literature. As a contribution in this respect, the present work presents the derivation of the exact c.d.f. expression of the $CARL_0$ for three cases of parameters estimation: (1) when both the process mean and standard deviation are unknown, (2) when only the mean is unknown and (3) when only the standard deviation is unknown. Using these key results, it was possible to calculate the *exact* minimum number of initial (Phase I) samples (*m*) that guarantees a desired in-control performance in terms of the *EPC*. These results show that *m* can be prohibitively large (such as 3.000 samples). As a solution to this problem, two new equations are derived here to adjust the control limits to guarantee a desired in-control performance in terms of the *EPC* for any given value of *m*. The advantage of these equations (compared to the existing adjustments methods) is that one provides exact results and the other one does not require too many computational resources to perform the calculations.

A further study about the impact of these adjustments on the out-of-control (OOC) performance provides useful tables to decide the appropriate amount of data and the adjustments that corresponds to a user preferred tradeoff between the IC and OOC performances of the chart. Practical recommendations for using these findings are also provided in this research work.

Keywords

Xbar Control Chart Performance, Conditional Performance, Exceedance Probability Criterion, Control Limits Adjustments, Guaranteed In-Control Performance

Resumo

Jardim, Felipe Schoemer; Epprecht, Eugenio Kahn (Orientador); Chakraborti, Subhabrata (Co-orientador). **Gráfico Xbarra com Parâmetros Estimados: A Distribuição da Taxa de Alarmes e Correções nos Limites**. Rio de Janeiro, 2018. 202p. Tese de Doutorado – Departamento de Engenharia Industrial, Pontifícia Universidade Católica do Rio de Janeiro.

Os gráficos de controle estão entre as ferramentas indispensáveis para monitorar o desempenho de um processo em várias indústrias. Quando estimativas de parâmetros são necessárias para projetar esses gráficos, seu desempenho é afetado devido aos erros de estimação. Para resolver esse problema, no passado, pesquisadores avaliavam o desempenho desses métodos em termos do valor esperado do número médio de amostras até um alarme falso condicionado às estimativas dos parâmetros (denotado por $CARL_0$). No entanto, esta solução não considera a grande variabilidade do $CARL_0$ entre usuários. Então, recentemente, surgiu a ideia de medir o desempenho dos gráficos de controle usando a probabilidade de o $CARL_0$ ser maior do que um valor especificado – que deve estar próximo do desejado nominal. Isso é chamado de *Exceedance Probability Criterion (EPC)*. Para aplicar o *EPC*, a função de distribuição acumulada (c.d.f.) do $CARL_0$ é necessária. No entanto, para um dos gráficos de controle mais utilizados, o gráfico Xbarra, também conhecido como gráfico $\bar{X}$ (sob a suposição de distribuição normal), a expressão matemática da c.d.f. não está disponível na literatura. Como contribuição nesse sentido, o presente trabalho apresenta a derivação exata da expressão matemática da c.d.f. do $CARL_0$ para três possíveis casos de estimação de parâmetros: (1) quando a média e o desvio-padrão são desconhecidos, (2) quando apenas a média é desconhecida e (3) quando apenas o desvio-padrão é desconhecido. Assim, foi possível calcular o número mínimo de amostras iniciais, *m*, que garantem um desempenho desejada do gráfico em termos de *EPC*. Esses resultados mostram que *m* pode assumir valores consideravelmente grandes (como, por exemplo, 3.000 amostras). Como solução, duas novas equações são derivadas aqui para ajustar os limites de controle garantindo assim um desempenho desejado

para qualquer valor de *m*. A vantagem dessas equações é que uma delas fornece resultados exatos enquanto a outra dispensa avançados *softwares* de computador para os cálculos. Um estudo adicional sobre o impacto desses ajustes no desempenho fora de controle (OOC) fornece tabelas que ajudam na decisão do melhor *tradeoff* entre quantidade adequada de dados e desempenhos IC e OOC preferenciais do gráfico. Recomendações práticas para uso desses resultados são aqui também fornecidas.

Palavras-Chave

Desempenho do Gráfico de Controle Xbarra, Desempenho Condicional, Ajuste nos Limites de Controle, Desempenho em Controle Garantido

Table of contents

1. Introduction 18
1.1. Motivation and Objectives 18
1.2. Methodology of this work 23
1.3. Organization of the Thesis 24

2. Litereature review and basic concepts 26
2.1. Previous works 26
2.2. The Control Limits of the $\bar{X}$ Chart 31
2.3. The Conditional Probability of a Signal and the Conditional False Alarm Rate 33

3. Derivation of the cumulative distribution functions of the $CARL_0$ and $CFAR$ 39
3.1. C.d.f of $CFAR$ (or CPS_δ) and $CARL_0$ (or $CARL_\delta$) in Case UU 40
3.2. C.d.f of $CFAR$ (or CPS_δ) and $CARL_0$ (or $CARL_\delta$) in Case KU 45
3.3. C.d.f of $CFAR$ (or CPS_δ) and $CARL_0$ (or $CARL_\delta$) in Case UK 50
3.4. The mean and standard deviation of $CFAR$ and $CARL_0$ 54
3.5. Prediction Bounds for $CFAR$ and $CARL_\delta$ 56

4. Number of phase i samples required for a guaranteed in-control performance 61

5. Adjustment of the limits for a guaranteed conditional in-control performance 65
5.1. Adjustment in Case UU 65
5.1.1. Exact Method 66
5.1.2. Approximate Methods 67
5.1.3. Bootstrap Method 69
5.1.4. Adjustment Results and Discussion in Case UU 70

5.2. Adjustment in Case KU 75
5.3. Adjustment in Case UK 77
5.4. Out–of-control Performance Analysis after the Adjustments 79

6. Conclusions and recommendations 85

7. References 89

Appendix A - Extras Plots of the $CARL_0$ Curves 95

Appendix B – The Search Algorithm 96

Appendix C – Derivation of the approximate formula for $F_{CPS_{\delta,UK}}$ 97

Appendix D - Derivation of the Approximate Equations
of $\alpha_{p,UU}$, L^*_{CA2} and L^*_{CA3} for case UU 100

Appendix E - Expressions of $g(L)$ and $g'(L)$ 103

Appendix F - Derivation of Formula (63) for the Bootstrap Method 104

Appendix G – Extra Tables of the Out-of-Control Analysis 105

Appendix H - Codes in R Language 108
R CODES FOR CASE UU 108
R CODES FOR CASE KU 115
R CODES FOR CASE UK 123

Annex A – Paper under Review in the Journal of Quality Technology 134

Annex B – Paper Under Review in The Production
Operation Management 161

Annex C - Paper Accepted in The American Statistician 194

List of figures

Figure 1. Summary of the scope of the present thesis 23
Figure 2. Adjusting the $\bar{X}$ chart control limits
for a guaranteed in-control performance 31
Figure 3. $\mathrm{CFAR_{UU}}$ as function of Z and Y
when $m = 30$, $n = 5$ and $\alpha = 0.0027$ (i.e., $L = 3$)................................ 36
Figure 4.$\mathrm{CFAR_{KU}}$ as function of u f
or $n = 5$, $m = 10, 20, 50,100,500$ and $\alpha = 0.0027$ (i.e., $L = 3$). 37
Figure 5. $\mathrm{CFAR_{UK}}$ as function of u
for $n = 5$, $m = 10, 20, 50,100,500$ and $\alpha = 0.0027$ (i.e., $L = 3$). 38
Figure 6. c.d.f. of $\mathrm{CFAR_{UU}}$
for $n = 5$, $m = 10, 20, 50,100,500$ and $\alpha = 0.0027$ (i.e., $L = 3$). 43
Figure 7. c.d.f. of $\mathrm{CARL_{0,UU}}$
for $n = 5$, $m = 10, 20, 50,100,500$ and $\alpha = 0.0027$ (i.e., $L = 3$). 43
Figure 8. p.d.f. of $\mathrm{CARL_{0,UU}}$
for $n = 5$, $m = 10, 20, 50,100,500$ and $\alpha = 0.0027$ ($L = 3$). 44
Figure 9. p.d.f. of $\mathrm{CFAR_{UU}}$
for $n = 5$, $m = 10, 20, 50,100,500$ and $\alpha = 0.0027$ ($L = 3$). 45
Figure 10. c.d.f. of $\mathrm{CFAR_{KU}}$
for $n = 5$, $m = 10, 20, 50,100,500$ and $\alpha = 0.0027$ (i.e., $L = 3$). 47
Figure 11. c.d.f. of $\mathrm{CARL_{0,KU}}$
for $n = 5$, $m = 10, 20, 50,100,500$ and $\alpha = 0.0027$ (i.e., $L = 3$). 48
Figure 12. p.d.f of of $\mathrm{CFAR_{KU}}$
for $n = 5$, $m = 10, 20, 50,100,500$ and $\alpha = 0.0027$ (i.e., $L = 3$). 49
Figure 13. p.d.f. *of* $\mathrm{CARL_{0,KU}}$
for $n = 5$, $m = 10, 20, 50,100,500$ and $\alpha = 0.0027$ (i.e., $L = 3$) 49
Figure 14. c.d.f. *of* $\mathrm{CARL_{0,UK}}$
for $n = 5$, $m = 10, 20, 50,100,500$ and $\alpha = 0.0027$ (i.e., $L = 3$) 52

Figure 15. c.d.f. *of* $\mathrm{CFAR_{UK}}$
for $\mathrm{n} = 5$, $\mathrm{m} = 10, 20, 50, 100, 500$ and $\alpha = 0.0027$ (i.e., $\mathrm{L} = 3$)...................... 53
Figure 16. p.d.f. *of* $\mathrm{CARL_{0,UK}}$
for $\mathrm{n} = 5$, $\mathrm{m} = 10, 20, 50, 100, 500$ and $\alpha = 0.0027$ (i.e., $\mathrm{L} = 3$)...................... 53
Figure 17. p.d.f. *of* $\boldsymbol{CFAR_{UK}}$
for $\boldsymbol{n = 5}$, $\boldsymbol{m = 10, 20, 50, 100, 500}$ and $\boldsymbol{\alpha = 0.0027}$(i.e.,$\boldsymbol{L = 3}$)……………….54

Figure A. 1. $\mathrm{CARL_{0,KU}}$as function of u
for $\mathrm{n} = 5$, $\mathrm{m} = 10, 20, 50, 100, 500$ and $\alpha = 0.0027$ (i.e., $\mathrm{L} = 3$)...................... 95
Figure A. 2. $\mathrm{CARL_{0,UK}}$ as function of u
for $\mathrm{n} = 5$, $\mathrm{m} = 10, 20, 50, 100, 500$ and $\alpha = 0.0027$ (i.e., $\mathrm{L} = 3$)...................... 95

List of tables

Table 1 - $SDARL_0$ and ARL_0 values for cases UU, KU and UU for several values of values of *m* and *n* and $\alpha = 0.0027$ (i.e., a nominal ARL_0 of 370.4) 56

Table 2. The 0.95 quantile of CFAR, the 0.05 quantile of $CARL_0$ $(p = 0.5)$, the 0.9 quantile of CFAR and the 0.1 quantile of $CARL_0$ $(p = 0.1)$ for $\alpha = 0.0027$ $(L = 3)$. 60

Table 3. Minimum number of in control Phase I samples, m, required for $P(CFAR \leq (1+\varepsilon)\alpha = 1\text{-}p$ or $P(CARL_0 \leq 1(1+\varepsilon)\alpha) = p$ with $\alpha = 0.0027$ $(L = 3)$. 63

Table 4. Minimum total amount of Phase I observations, $(m \times n)$, required for $P(CFAR \leq (1+\varepsilon)\alpha = 1\text{-}p$ or $P(CARL_0 \leq 1(1+\varepsilon)\alpha)$with $\alpha = 0.0027$ $(L = 3)$. 64

Table 5. Values of L*when $L^* = L$, $L^* = LCE^*$, $L^* = LCA1^*$, $L^* = LCA2^*$, $L^* = LCA3^*$, $L^* = Lboot^*$ and their corresponding ARL0, SDARL0 and $P(CFAR \leq (1+\varepsilon)\alpha = 1\text{-}p$ for $L = 3$, $\alpha = 0.0027$, $\varepsilon = 0\%$ and $p = 5\%$ for Case UU 73

Table 6. Values of L*when $L^* = L$, $L^* = LCE^*$, $L^* = LCA1^*$, $L^* = LCA2^*$, $L^* = LCA3^*$, $L^* = Lboot^*$ and their corresponding ARL0, SDARL0 and $P(CFAR \leq (1+\varepsilon)\alpha = 1\text{-}p$ for $L = 3$, $\alpha = 0.0027$, $\varepsilon = 20\%$ and $p = 5\%$ for Case UU 74

Table 7. Values of L*when $L^* = L$, $L^* = LCE^*$, $L^* = LCA1^*$, $L^* = LCA2^*$, $L^* = LCA3^*$, $L^* = Lboot^*$ and their corresponding ARL0, SDARL0 and $PCFAR \leq \alpha 1 + \varepsilon$ for $L = 3$, $\alpha = 0.0027$, $\varepsilon = 20\%$ and $p = 20\%$ for Case UU 75

Table 8. Values of L* to achieve$P(CFAR \leq (1+\varepsilon)\alpha = 1\text{-}p = 1\text{-}p$ or $P(CARL_0 \leq 1(1+\varepsilon)\alpha)$ for $n = 3{,}5{,}9{,}15$, $\varepsilon = 0, 0.10{,}0.20$, $p = 0.05{,}0.1, 0{,}15{,}0.2$ and $m = 25{,}50{,}250{,}500, 1000$ in Case KU 77

Table 9. Values of L* to achieve $P(\mathrm{CFAR} \leq (1+\varepsilon)\alpha = 1$-por $P(\mathrm{CARL}_0 \leq 1(1+\varepsilon)\alpha)$ for $\varepsilon = 0, 0.05, 0.1, 0.15, 0.2$, $p = 0.05, 0.1, 0{,}15, 0.2$ and $m = 25{,}75{,}50{,}100{,}150{,}200{,}250{,}500, 1000$ in Case UK 78

Table 10. The 0.95 and 0.9 quantiles of $\mathrm{CARL}_{\delta,\mathrm{UU}}$ with adjusted limits ($\alpha = 0.0027$, $p = 0.1$ and $\varepsilon = 0$) in grey and unadjusted limits ($L = 3$) in white for different values of m, n and δ (Case UU) 82

Table 11. The 0.95 and 0.9 quantiles of $\mathrm{CARL}_{\delta,\mathrm{KU}}$ with adjusted limits ($\alpha = 0.0027$, $p = 0.1$ and $\varepsilon = 0$) in grey and unadjusted limits ($L = 3$) in white for different values of m, n and δ (Case KU) 82

Table 12. The 0.95 and 0.9 quantiles of $\mathrm{CARL}_{\delta,\mathrm{UK}}$ with adjusted limits ($\alpha = 0.0027$, $p = 0.1$ and $\varepsilon = 0$) in grey and unadjusted limits ($L = 3$) in white for different values of m, n and δ (Case UK) 83

Table G. 1. The 0.95 and 0.9 quantiles of $\mathrm{CARL}_{\delta,\mathrm{UU}}$ with adjusted limits ($\alpha = 0.0027$, $p = 0.1$ and $\varepsilon = 0.2$) in grey and unadjusted limits ($L = 3$) in white for different values of m, n and δ (Case UU)... 105

Table G. 2. The 0.95 and 0.9 quantiles of$\mathrm{CARL}_{\delta,\mathrm{UU}}$ with adjusted limits ($\alpha = 0.0027$, $p = 0.2$ and $\varepsilon = 0.2$) in grey and unadjusted limits ($L = 3$) in white for different values of m, n and δ (Case KU)... 105

Table G. 3. The 0.95 and 0.9 quantiles of $\mathrm{CARL}_{\delta,\mathrm{KU}}$ with adjusted limits ($\alpha = 0.0027$, $p = 0.1$ and $\varepsilon = 0.2$) in grey and unadjusted limits ($L = 3$) in white for different values of m, n and δ (Case KU)... 105

Table G. 4. The 0.95 and 0.9 quantiles of$\mathrm{CARL}_{\delta,\mathrm{KU}}$with adjusted limits ($\alpha = 0.0027$, $p = 0.2$ and $\varepsilon = 0.2$) in grey and unadjusted limits ($L = 3$) in white for different values of m, n and δ (Case KU)... 106

Table G. 5. The 0.95 and 0.9 quantiles of $\mathrm{CARL}_{\delta,\mathrm{UK}}$ with adjusted limits ($\alpha = 0.0027$, $p = 0.1$ and $\varepsilon = 0.2$) in grey and unadjusted limits ($L = 3$) in white for different values of m, n and δ (Case UK)... 106

Table G. 6. The 0.95 and 0.9 quantiles of $\mathrm{CARL}_{\delta,\mathrm{UK}}$ with adjusted limits ($\alpha = 0.0027$, $p = 0.2$ and $\varepsilon = 0.2$) in grey and unadjusted limits ($L = 3$) in white for different values of m, n and δ (Case UK)... 107

Glossary

α	Nominal false alarm rate
α_p	Upper prediction bound to the *CFAR*
α_{TOL}	Tolerated upper bound to the false alarm rate
ARL	Unconditional Average Run Length: *ARL=E(CARL)*
$boot$	Refers to the bootstrap adj. method under the conditional perspective
B	Number of bootstrap simulations
CARL	Conditional (realized) Average Run length
CFAR	Conditional (realized) False Alarm Rate
CPS	Conditional (realized) Probability of a Signal
$c_{4,b}$	The unbiasing constant
Case KK	Case where both the mean and standard deviation are known
Case KU	Case where the mean is known, and the standard deviation is unknown
Case UK	Case where the mean is unknown, and the standard deviation is known
Case UU	Case where both the mean and standard deviation are unknown
CE	Refers to Exact Adjustment Method Under the Conditional Perspective
CA	Refers to Approximate Adj. Method Under the Conditional Perspective
Φ	Cumulative distribution function of a Standard Normal Random Variable
δ	Scaled shift of the mean
EPC	Exceedance Probability Criterion
ε	Percentage of the *CFAR*
FAR	Unconditional False Alarm Rate
F_a	Cumulative distribution function (c.d.f.) of the random variable a

f_a	Probability density function (p.d.f.) of the random variable a
IC	In Control
L	Unadjusted limit factor
L^*	Adjusted Limit Factor
m	Number of Phase I samples
μ_0	In-control process mean
μ_1	Out-of-control process mean
$\hat{\mu}_0$	Estimator for the in-control process mean
μ	Phase II process mean (can be in control or out of control)
n	Size of each Phase I and Phase II sample
OOC	Out of Control
p	Complement of the desired Exceedance Probability
ϕ	Probability density function of a standard normal random variable
$Q_{p_{OOC}}$	$(1 - p_{OOC})$-quantile of the out-of-control $CARL$
σ_0	In-Control process standard deviation
$\hat{\sigma}_0$	An estimator for the in-control process standard deviation
SPC	Statistical Process Control
S_p	Multi-sample estimators of the standard deviation
$SDARL$	Standard Deviation of the *CARL*
t	A possible value of *CFAR*
$t_{(a,b,c)}$	a-quantile of a non-central t distribution with b degrees of freedom and non-centrality parameter c
U	Uniform random variable between 0 and 1
u	A value of U. It also refers here to the order of the quantiles of Y and Z
w	A possible value of *CARL*
$\bar{X}$	Sample mean
$\bar{\bar{X}}$	The grand mean of the m Phase I samples estimator
χ_c^2	Central chi-square random variable with c d.f.
$\chi_{c,[d]}^2$	Non-central chi-square random variable with c d.f. and non-centrality parameter d
Y	Chi-square random variable

Z	Standard Normal Random Variable
Z_1	Standard Normal Random Variable

1
Introduction

1.1.
Motivation and Objectives

Control charts, created by Walter A. Shewhart while working for Bell Labs in the 1920s and first published in a book in 1931, are still one of the most used tools for monitoring the quality characteristics of a process. The most usual control chart to monitor a process mean, namely the $\bar{X}$ control chart, is also widely used in practice in many different areas such as manufacturing industries and medicine. The in-control process mean and standard deviation are important parameters for designing the $\bar{X}$ control chart. Usually these parameters are unknown and must be estimated from m historical samples each of size n collected when process is presumably in control. This is called Phase I Analysis in Statistical Process Control (SPC). For an overview of the Phase I, the reader is referred to Chakraborti et al. (2009) and Jones-Farmer et al. (2014). Then, the chart's control limits may be established using these estimates in the prospective process monitoring (called Phase II), where samples (also of size n) are collected at regular intervals. These samples are used to calculate the plotting statistic [i.e., the sample mean ($\bar{X}$)] to be compared with the control limits. If the plotting statistic is outside the control limits, the probability that something has changed with the process mean must be high and the manager must find and correct the possible problem.

In the past, most of the research involving the development and performance evaluation of the Phase II control charts assumed that the in-control process parameters were known [see, for example the literature review in this topic by Jensen et al. (2016)]. I.e., for the $\bar{X}$ chart, the in-control process mean (denoted μ_0) and the in-control process standard deviation (denoted σ_0) were assumed to be known. This is because this assumption simplifies the design and performance evaluation of the chart. For example, in this situation, the number of samples until an alarm (the so-called *run length*) follows the well-known geometric distribution.

However, in practice, as noted above, these parameters are usually unknown and must be estimated in Phase I. When these estimates are used in place of known parameters, their variability can result in chart performance that differs from that of charts designed with known parameters. For example, the *run length* does not follow a geometric distribution anymore and the actual probability of a false alarm [also named the false alarm rate (*FAR*)] may be larger than the nominal desired one. Many false alarms generate unnecessary costs to the process. For example, the unnecessary waste of time that the manager will have to check for a nonexistent failure pausing the process and decreasing its productivity. Since Shewhart (1939, p. 76), this problem has been pointed out and analyzed by several authors. Thus, when designing a control chart, the manager must consider the effect of parameter estimation on the chart performance to avoid unnecessary costs. Four cases are conceivably possible in the designing of the $\bar{X}$ control chart when parameters are estimated. These cases arise when:

a) the in-control process mean and standard deviation (μ_0 and σ_0) are unknown. This situation is known as "Case UU", which stands for *mean **U**nknown* and *standard deviation **U**nknown* [see Quesenberry (1993)]. This state is also known as the *standard unknown case*. This is the most common case in practice;

b) the process mean is considered known, and the standard deviation is unknown (and needs to be estimated in Phase I). This is Case KU (*mean **K**nown, standard deviation **U**nknown*). This case is less common in practice (compared to case UU), however it appears in some situations, for instance, according to Montgomery (2009; p. 243), "*in processes where the mean of the quality characteristic is controlled by adjustments to the machine, standard or target values of the mean* (i.e., no estimation of the process mean and only estimation of the process standard deviation to calculate the control limits) *are sometimes helpful in achieving management goals with respect to process performance*";

c) the in-control process mean is unknown (and needs to be estimated in Phase I) and the standard deviation is known (Case UK, *mean **U**nknown, standard deviation **K**nown*). This case is not common in practice, however it may appear when the mean may change but the variability around it is known to be considerably stable;

d) the in-control process mean and standard deviation are both known (this case is called "Case KK", *mean **K**nown, standard deviation **K**nown*, or "standard known" case). This case is the simplest case where the performance of he $\bar{X}$ control chart can be easily studied and measured.

Usually the $\bar{X}$ chart performance in the ideal case KK is the practitioner desired chart performance. However, as noted above, the Phase II $\bar{X}$ chart performance in cases UU, KU and UK may be considerable different compared to Case KK.

The realized probability of a signal (*CPS*) [or, if the process is in control, the conditional false alarm rate (*CFAR*)] and the realized average number of samples until a signal [which is usually known as the conditional average run length (*CARL*)] are the most popular performance measures of any control chart. When parameters are estimated, *CFAR* (*or CPS)* and *CARL* are random variables because they are conditioned on the estimated parameters, which are also random variables. So, these measures alone, cannot be used when parameters are estimated because they vary. Then, in cases UU, KU and UK the performance was usually measured using the expectation of $CFAR$ and $CARL$ [i.e., $E(CFAR)$ and $E(CARL)$]. This is known as the "*unconditional perspective*" and it will be explained in more detail in Chapter 2. The fact is that, when parameters are estimated with insufficient Phase I data, $E(CFAR)$ and $E(CARL)$ may be different, respectively, from the false alarm rate (*FAR*) or the average run length (*ARL*) in the ideal case KK. Given this, if more Phase I samples are not available, some authors suggested adjusting the control limits in order to make $E(CFAR)$ and $E(CARL)$ equal to the values of FAR and *ARL* in case KK. However, more recently, a great number of researcher advocated against the use of $E(CFAR)$ and $E(CARL)$ as chart performance measures, because the overall expectation does not account for the practitioner-to-practitioner variability (i.e., each practitioner chart will have one different value of *CFAR* and *CARL* which will most likely not be equal or close to the $E(CFAR)$ and $E(CARL)$, their means. This is because, as noted by them, the variability of *CFAR* and *CARL* is often large even for a relatively large amount of Phase I data, even if the chart is adjusted to have a desired specific value of $E(CFAR)$ and $E(CARL)$.

So, recently, some authors recommended to measure control charts with the probability of *CARL* or *CFAR* be greater than a specified value (close to the nominal desired value). This is known as the Exceedance Probability Criterion (*EPC*) proposed by Albers and Kallenberg (2005) and Albers et al. (2005). These authors recommended adjusting the control limits or finding the appropriate amount of Phase I data using the *EPC* as a performance measure, instead of $E(CFAR)$ and $E(CARL)$. So, note that to use the *EPC*, the knowledge of the cumulative distribution functions of *CARL* or *CFAR* are required. Despite of the long list of works in the literature regarding the performance of the $\bar{X}$ control chart with estimated parameters (under normality assumption), the exact mathematical c.d.f. expressions of *CARL* and *CPS* (or *CFAR*) were unknown in the cases where parameters are estimated. As we will see during this work, the exact expressions of the *CARL* and *CPS* (or *CFAR*) c.d.f´s provides to the practitioners and researchers a better understanding of the effect of parameter estimation on the $\bar{X}$ control chart performance by helping them in the calculation of some important *CARL* and *CPS* (or *CFAR*) properties (such as quantiles, mean, median and variance) and, more important, these c.d.f.´s (of *CARL* and *CPS* [or *CFAR*]) also help the practitioner and researcher to calculate the *exact minimum number of Phase I samples* and the *exact adjustments* to the control limits in order to guarantee an in-control performance in terms of the Exceedance Probability Criterion (*EPC*). Since the exact distributions of *CARL* and *CPS* (or *CFAR*) were unknown, previous authors, for Case UU only, relied on approximations or simulations in order to study the performance of the $\bar{X}$ control chart and design it in terms of the *EPC*.

With this background as motivation, the present work has the following objectives for the cases UU, KU and UK of the $\bar{X}$ chart:

1. Derive the mathematical expressions of the c.d.f. of the conditional average run length (*CARL*) and its reciprocal, the conditional probability of a signal (*CPS*) of the $\bar{X}$ chart for the first time in the literature. Previous authors relied on simulations and approximations to study such distributions;

2. With the aforementioned distributions, calculate the *exact* required numbers of Phase I samples that guarantee a desired conditional in-control

performance in terms of the EPC for the first time in the literature. Previous authors calculated these numbers based on the unconditional perspective or based on the achievement of some desired value for the variability of the $CARL_0$ (i.e., not in terms of the EPC);

3. Derive new simple equations to adjust the control limits to guarantee a desired in-control performance (in terms of the EPC) which provides exact or accurate results. The existing adjustment equations in the literature do not provide exact results and their calculations are rather complicated requiring computers for the calculation of several integrals and derivatives;

4. Study and analyze the effect of the adjustment of the limits on the out-of-control performance of the chart in detail to provide some practical recommendations for the users. Previous authors have tackled this issue only very briefly and focusing mainly on the *unconditional* out-of-control run length. This is an important information for the user, who needs to consider the tradeoff between the number of Phase I samples to consider, the risks of a false alarm rate higher than desired and the possible deterioration in the out-of-control performance regarding the relevant shifts to be detected with minimum delay. Since the adjustments required (and their effects on the OOC performance of the chart, that depend on the size of the relevant shift) are milder with larger numbers of Phase I samples, only a more comprehensive analysis of the OOC CARLs of the chart with adjusted limits for different numbers of Phase I samples and for different shifts in the mean can provide the user with the "big picture" needed in order to make an informed decision on the number and size of Phase I samples and adjustments to adopt;

Figure 1 shows a summary of the basic ideas presented in this thesis.

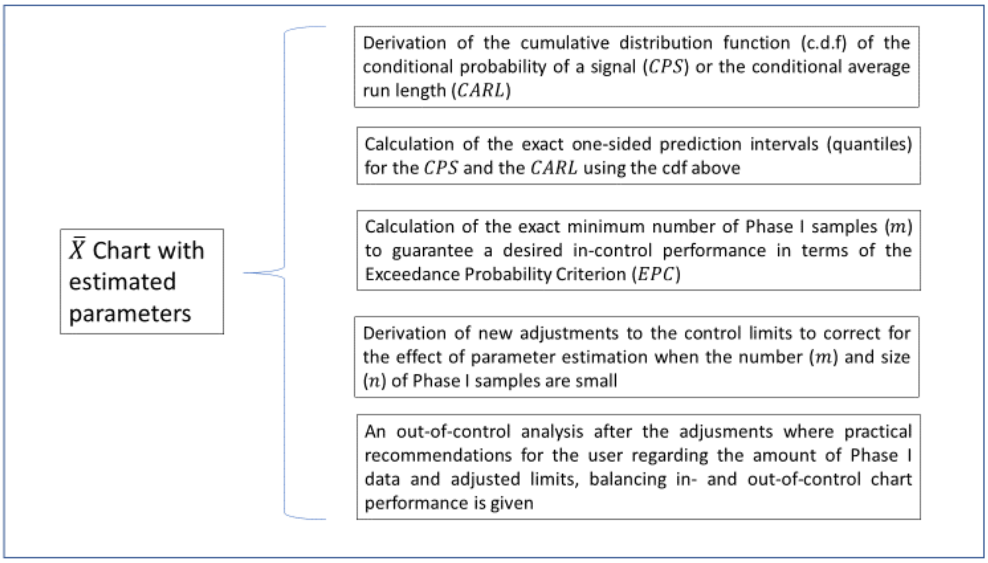


Figure 1. Summary of the scope of the present thesis

1.2. Methodology of this work

To accomplish the objectives of the present work, first a literature review was made where were verified that the mathematical expressions of the c.d.f. of the conditional average run length (*CARL*) and, its reciprocal, the conditional probability of a signal (*CPS*) of the $\bar{X}$ chart under normality assumption were unknown. Given this, with probability and statistical techniques, such as the *conditional-unconditional method* and *the distribution function technique*, these c.d.f.´s were *derived analytically* for the first time in the literature. With these mathematical c.d.f.´s expressions, it was possible to derive *new exact equations* to: (1) calculate the prediction bounds of the in-control *CARL* and *CPS*; (2) calculate the exact minimum amount of Phase I samples to achieve a desired in-control performance in terms of the Exceedance Probability Criterion (*EPC*); and (3) adjust the control limits in order to achieve a desired in-control performance in terms of the *EPC* with a pre-established amount of Phase I data. Some of these exact equations do not have a close-form solution, so in order to find the answers for these cases, two approaches were adopted:

1 – The use of a *search algorithm*, known as the secant method. This method provides extremely accurate result, which can be considered exact. This is because, one can specify the desired accuracy for the search algorithm in order to find a result that is exact up to a specified number of significant digits. Details of this method are in Appendix B.

2 – The *analytical* derivation of new *approximate formulas*. To this end, it was used approximate techniques, such as the *one-step* and *two-steps Taylor approximations* and an approximation for the c.d.f. of a non-central chi-square distribution derived by Cox and Reid (1987). Details of these *approximate formulas* derivations are in Appendixes C and D.

To perform all the above-mentioned calculations, programs were written using the R language. In particularly two R-packages were used: cubature (to compute double integrals) and numDeriv (to compute numerical derivations).

1.3
Organization of the Thesis

In order to achieve the objectives of this research, the remainder of this work is organized as follows:

- In Chapter 2, there is a section of a literature review on the previous works regarding to the effect of parameters estimations on the control charts performance, especially on the $\bar{X}$ control chart and the very recent works in this topic. Another section presents the basic concepts and formulas of the $\bar{X}$ control limits and the parameters estimators used in them. For the three cases (UU, KU and UK), the formulas of the conditional average run length (*CARL*) and conditional probability of a signal (*CPS*) are also derived and explained. Some plots of *CARL* and *CPS* curves are presented along with some analyses.
- In Chapter 3, there are detailed derivations of the cumulative distribution functions (c.d.f.) of *CARL* and *CPS* (or *CFAR*) for all the 3 cases (UU, KU and UK). Plots of the c.d.f. and the probability density function are

presented. In two sections of this chapter, some properties of the distributions of the *CARL* are calculated, such as their means, standard deviations and quantiles (prediction bounds). Some analyses are provided.

- In Chapter 4, equations (based on the c.d.f´s derived in Chapter 3) to calculate the minimum number of Phase I samples in order to guarantee an in-control performance in term of the *EPC* are derived for all three case UU, KU and UK. Some results are tabulated and analyzed.
- In Chapter 5, equations which provide exact and approximate adjustments in the control limits (to guarantee an in-control performance in terms of the *EPC*) are derived. In this chapter, it is shown that these new equations provide accurate results compared to the already existing ones for Case UU. Also, further in this chapter, an out-of-control analysis after the adjustments is made.
- In Chapter 6, the conclusions and some practical recommendations are presented.

To support the understanding of the contents in this work, in appendices some extra derivations to some of the formulas used here are presented along to some extra figures and plots. The codes in R language are also in Appendix. Finally, from some of the contents of this work, three papers were written and submitted to international journals. One of them is already accepted. These papers are in Annexes in the end of this work.

2
Literature review and basic concepts

2.1.
Previous works

The undesired effect of estimating parameters with limited amount of Phase I data on the performance of control charts has long been documented in the literature. Shewhart (1939, p.76) already showed this concern when he wrote "*In the majority of practical instances, the most difficult job of all is to choose the sample that is to be used as the basis for establishing the tolerance range (control limits)*". However, mostly after Queensberry (1993) and Woodall and Montgomery (1999), who emphasized the relevance of this research topic, a large number of papers which studied and proposed solution for the effect of parameter estimation on the performance of control charts have emerged [this is summarized in the literature reviews made by Jensen et al. (2006) and Psarakis et al. (2014)]. The fact is that the performance measures of control charts have been the subject of much debate in the last 20 years or so.

Many researchers who studied the effect of parameter estimation on the performance of control charts have focused on the marginal distribution of the number of observations (or samples) until an alarm (the well-known run length, denoted RL – see, for example, Moskowitz et al. (1994)] and especially on its expected value, the so-called unconditional average run length, denoted ARL [see for example, Quesenberry (1993), Chen (1997), Chakraborti (2000, 2006 and 2007) and Goedhart et al. (2016a)]. They focused almost exclusively on the in-control performance, i.e., they considered the in-control run length (RL_0) distribution, its mean (ARL_0) and its the standard deviation ($SDRL_0$). The RL_0 distribution is obtained by averaging over the distribution of the parameter estimators, thus, a performance given by ARL_0 or $SDRL_0$ is an "average" performance over an infinite number of possible control charts, each one constructed with a possible value of an infinite set of estimated parameters, rather than a performance of a specific control chart. Like noted in the introduction, this is denoted as "*the unconditional perspective*". Thus the ARL_0 and $SDRL_0$ does not account for what is called *the practitioner-to-practitioner variability*. This is a subject of much debate in the literature, see for example, Trietsch and Bischak (1998), Albers and Kallenberg (2004a,b, and 2005), Albers et al. (2005), Bischak and Trietsch (2007), Kumar and Chakraborti (2014), Saleh et al (2015a,b), Epprecht et al. (2015), Faraz et al. (2015), Goedhart et al (2017 and 2018).

In any real application, where the user has just one set of reference samples to estimate the chart parameters and calculate its control limits, the realized average number of samples until a false alarm will not actually be the ARL_0. The realized number of samples until a false alarm will be conditioned on the parameters estimates (this is denoted CRL_0) and so its average, denoted as $CARL_0$. Recently [see, for example, Saleh et al (2015a,b), Epprecht et al. (2015)] authors recognized that the $CARL_0$, in fact, is the "real/actual" average number of samples until a false alarm, and, different from the ARL_0, it is a random variable that will most probably not assume its expectation (which is also the ARL_0, i.e., $E(CARL_0) = ARL_0 = E(RL_0)$), since its variability, as also noted by them, is often large (especially if the amount of Phase I data to estimate the chart parameters is not very large, such as three thousands). This alternative and new point of view is known as "*the conditional perspective*". A detailed comparison between the conditional and unconditional perspectives is presented in Jardim et al. (2017a) – a resulting paper of the present thesis (see Annex A).

As noted in Chapter 1, recognizing that the $CARL_0$ (or its reciprocal, the conditional false alarm rate, denoted, $CFAR$) is a random variable, Albers et al. (2005) proposed to measure and set up control chart limits so as to guarantee that a given tolerated value for the $CARL_0$ had only a large (specified) probability (e.g. 90%) of being exceeded, the Exceedance Probability Criterion (*EPC*). Thus, it is evident that, to use the *EPC*, it is necessary to deal with the distribution of $CARL_0$. When it comes to the possibly most well-known control chart of all, the $\bar{X}$ control chart under normality assumption, the exact c.d.f. expression of the $CARL_0$ was unknown until the present work. So, most authors studying and proposing solutions to the negative effect of parameter estimation on the performance of the $\bar{X}$ charts using the *EPC* relied in simulations, bootstrap methods, and approximations for the distribution of the $CARL_0$ (as it will be seen next).

It has been found by several authors [see, for example, Quesenberry (1993) and Chakraborti (2000) for the $\bar{X}$ chart] that while estimating parameters, the *unconditional perspective* leads to requiring larger amounts of Phase I data for parameters estimation so that some nominal in-control chart performance (in terms of the *ARL* and *SDRL*) can be achieved as in the known parameters case (case KK). This required amount of data is much larger than traditionally recommended which is m = 25 or 30 Phase I subgroups each of size $n = 5$ given by the usual manuals and books in this topic [see, for example, Montgomery (2013)]. On the other hand, focusing on the conditional perspective, Saleh et al. (2015a,b), for the $\bar{X}$, X and the EWMA charts, showed that using the standard deviation of the $CARL_0$ distribution, in addition to the average (ARL_0), as a performance measure, better accounts for practitioner-to-practitioner variability and leads to a requiring even larger amounts of Phase I reference data (i.e., even larger values of m). Although this is technically sound advice, it has been noted that such huge amounts of data that require very large values of m (several hundreds or even some thousands), may be typically infeasible in routine control charting applications. Similar findings were made by Loureiro et al. (2017), Epprecht et al. (2015) and Kumar and Chakraborti (2014) who considered, respectively, the joint $\bar{X}$-S control charts, the one-sided S charts and Shewhart charts for monitoring times between events following an exponential distribution. Using the *EPC*, their findings of the required numbers of Phase I samples to guarantee a specified nominal in-control performance, were much larger than found by previous authors who based their analyses on the unconditional performance measures (*ARL* and *SDRL*). It is important to note that, until the present work, the minimum number of reference sample to achieve the in-control performance of the $\bar{X}$ chart in terms of the *EPC* wasn't presented (calculated) in the literature.

Given the finding that under both perspectives (the unconditional and conditional), large, often impractical, amounts of Phase I data are required to guarantee some in-control performance of the Phase II $\bar{X}$ chart, some authors have considered using adjustments to the control limits to properly compensate for the effects of parameter estimation and to guarantee a desired in-control performance with the available amount of data. Such control limits are called adjusted limits and this adjustment consists of replacing the limit factor (L) (usually equal to 3 in the traditional Shewhart $\bar{X}$ chart), by a new (or corrected or adjusted) limit factor (L^*), which yields a specified nominal in-control performance. For example, in the unconditional perspective, the constant 3 in the traditional "3-sigma limits" may be replaced (or adjusted) by a constant ($L^* = 3.15$, say), to guarantee that the ARL_0 has a desired nominal value. On the other hand, in the conditional perspective, one recognizes that the $CARL_0$ is a random variable with a distribution and thus one uses the *EPC* and replaces the traditional limit factor (L) by an adjusted limit factor (L^*), to guarantee that with a high probability, the $CARL_0$ is greater than a specified value, say, 370.4. Of course, any adjustment to the control limits also impacts the chart's out-of-control performance and one must carefully balance the gains and losses on both fronts. The convention in *SPC* has been to weigh the chart's in-control performance more heavily, so that not too many false alarms are seen relative to what is nominally expected, but this must also be balanced so that the chart's shift detecting ability is not highly compromised.

To underscore the keen interest in this area of research, note that several articles have about adjustments of control limits for the $\bar{X}$ chart have appeared in major journals over the last decade. These include Chakraborti (2006), Gandy and Kvaløy (2013), Saleh et al. (2015b), Goedhart et al. (2016 and 2017), Jardim et al. (2017a and 2017b) and Faraz et al. (2017). These efforts are described next.

Chakraborti (2006) and Jardim et al. (2017b) [another resulting paper of this work, see Annex B] derived formulas for $L^*,$ using the unconditional and the conditional perspectives, respectively and the exact distributions formulas for the in-control marginal run length and conditional average run length in each case. Although these distributions and the resulting equations are not in a closed form, they can be easily solved numerically, using many available software packages, for example, such as the R language. Since these methods are based on an exact distribution and yields very accurate results easily, using numerical methods to solve the integrals involved, these are henceforth called "the Exact Methods".

On the other hand, Goedhart et al. (2016 and 2017a), derived formulas for the adjusted limit factor under the unconditional and the conditional perspective, respectively, using sophisticated approximations for the *CFAR* c.d.f. Furthermore, realizing the complexity of the approximations, Goedhart et al. (2018) presented an alternative and simpler approximate formula for the conditional perspective solution based on the theory of tolerance intervals [see Krishnamoorthy and Mathew (2009)]. However, this simpler formula requires the quantile of a non-central chi-square distribution, which is not tabulated in many textbooks and is not provided in a popular software like MSExcel, so its calculation may still require relatively advanced statistical skills. Given this, in the present work, like mentioned in Chapter 1, as an aside, it is derived an even simpler approximate formula in which the quantile of the non-central chi-square distribution is replaced by a central chi-square quantile [using a result given by Cox and Reid (1987)], which is more readily available. All these methods are called “the Approximate Methods” to emphasize the fact that they are derived using some approximations (to the distribution of the $CFAR$ or $CARL_0$).

In addition to the exact and the approximate methods, there are adjustments to the $\bar{X}$ chart control limits considered by Saleh et al. (2015b) under the conditional perspective and the *EPC* using the bootstrap approach proposed by Gandy and Kvaloy (2013). Finally, Faraz et al. (2017) also proposed a method to adjust the $\bar{X}$ chart, however, their adjustment was not based in the *EPC*, instead, it was based on the *equal-tailed tolerance interval* together with the *Bonferroni Inequality* [see Krishnamoorthy and Mathew (2009, p. 4 and p.10)], which generates wider adjusted control limits if compared to the adjusted limits derived under the *EPC*. Faraz et al. (2017) defined their method as “*exact*”, however it is not actually *exact*, since it is based in an inequality. Also, their results are extremely different from all the other methods under the conditional perspective, but this is still included in this work for comparison (and it is called here as “Exact Method”, since this was the definition used by them).

Figure 2 shows a flowchart for the current state of the art regarding the adjustment of Phase II control limits to achieve some desired nominal in-control performance for the $\bar{X}$ chart in the face of parameter estimation with Phase I data.

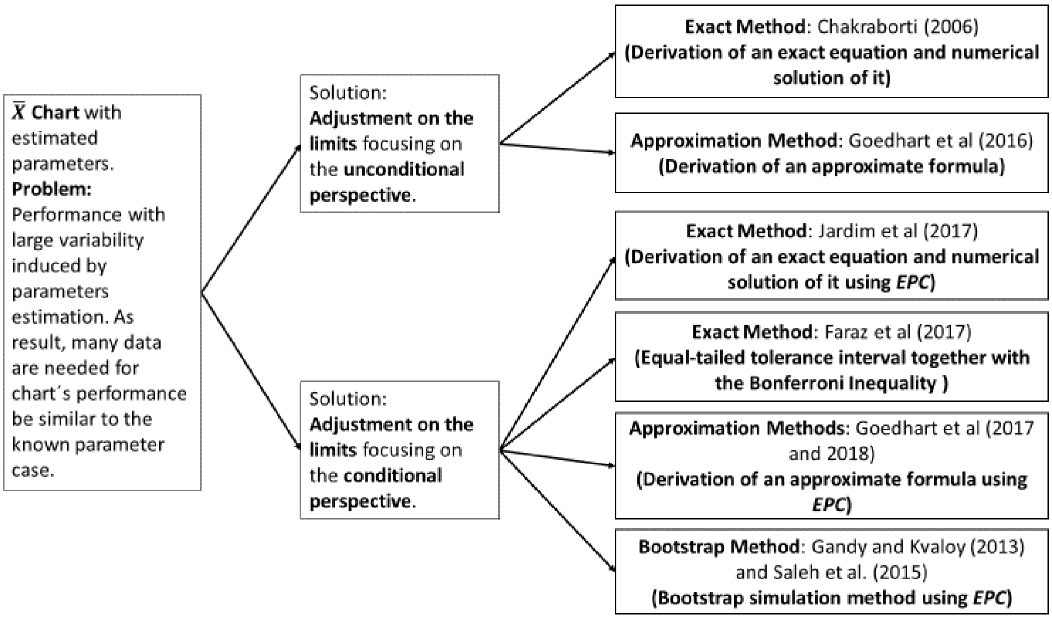


Figure 2. Adjusting the $\overline{X}$ chart control limits for a guaranteed in-control performance

2.2. The Control Limits of the $\bar{X}$ Chart

In the present work, the observations of the process quality variable (X) are considered i.i.d. and normally distributed, like is traditionally done for the $\bar{X}$ Control Chart. When the process is in control, $X\sim N(\mu_0, \sigma_0^2)$; when the process is out of control, $X\sim N(\mu_1, \sigma_0^2)$, with $\mu_1 \neq \mu_0$. Thus, the process standard deviation is assumed to remain at the in-control value σ_0, consistently with the purpose of detecting a shift in the mean. In the ideal case (KK), the in-control process mean (μ_0) and standard deviation (σ_0) are both known or specified. In this situation, the upper and lower control limits (UCL and LCL) of the L-sigma $\bar{X}$ Control Chart with subgroups of size n are given, respectively by

$$UCL = \mu_0 + L\frac{\sigma_0}{\sqrt{n}} \tag{1}$$

and

$$LCL = \mu_0 - L\frac{\sigma_0}{\sqrt{n}}, \tag{2}$$

where the control limit factor L is either a value such as 3 (the widely used “3 sigma limits”) or is chosen so as to provide a nominal in-control average run length such as 370.4 or a false-alarm rate α. In the latter case, we have $L = z_{\alpha/2} = \Phi^{-1}(1-\alpha/2)$, where $\Phi(\cdot)$ denotes the standard normal c.d.f. Thus, the usual 3-sigma limits correspond to a nominal false alarm rate of $\alpha = 0.0027$. However, as noted in the beginning of the Chapter and in Chapter 1, in practice μ_0 or σ_0 are usually unknown and need to be estimated from a Phase I data, consisting of m subgroups of size n, taken from the process when it is in control.

In Cases UU and UK, the most common estimator for the mean μ_0 is the $\bar{\bar{X}}$, the grand mean of the m Phase I samples:

$$\bar{\bar{X}} = \frac{1}{m}\sum_{i=1}^{m} \bar{X}_i, \tag{3}$$

where $\bar{X}_i = \frac{1}{n}\sum_{j=1}^{n} X_{ij}$, $i = 1,2,\dots,m$, $j = 1,2,\dots,n$ and X_{ij} denotes the j-th observation of the i-th Phase I sample. In cases UU and KU, a highly recommended estimator for the standard deviation is the pooled sample standard deviation (S_p), which is given by the square root of the average of the sample variances of the Phase I samples. Thus,

$$S_p = \sqrt{\frac{1}{m}\sum_{i=1}^{m} S_i^2}, \text{ where } S_i^2 = \frac{1}{n-1}\sum_{j=1}^{n}\left(X_{i,j} - \bar{X}_i\right)^2 \tag{4}$$

Mahmoud et al. (2010) showed that, among multi-sample estimators of the standard deviation, S_p is preferable to a more traditional estimator, like $\bar{S}/c_{4,b}$, where $\bar{S} = \sum_{i=1}^{m} S_i/n$ and $c_{4,b}$ is the unbiasing constant defined as [see, Montgomery (2013)]:

$$c_{4,b} = \frac{[\Gamma(b/2)\sqrt{2}]}{\Gamma((b-1)/2)\sqrt{b-1}}. \tag{5}$$

Where $b = m(n-1)+1$ and Γ is the gamma function. So, in the present work, the $\bar{\bar{X}}$ and S_p are considered (instead of $\bar{S}/c_{4,b}$,). In the literature, two other pooled estimators of the standard deviation have been also considered, the unbiased $S_p/c_{4,b}$ and the biased, but minimum mean squared error, estimator $c_{4,b}S_p$ (see Mahmoud et al., 2010 and Saleh et al., 2015a,b). Since these three estimators provide similar results as $c_4 \approx 1$ for relatively small values of m and n [such as $m = 25$ and $n = 4$ — again, see Mahmoud et al. (2010) for a quantitative comparison], in the present work, we consider just the S_p estimator. Note that it is not considered here the range-based estimators since some authors have recommended against their use because of lack of robustness (see again Mahmoud et al., 2010). Anyway, all the formulas and results presented here can be easily modified for other estimators of standard deviation.

When μ_0 or σ_0 are estimated, most authors recommended replacing the constant L in equations (1) and (2) by a constant L^* called the adjusted control limit factor that needs to be found so that a desired nominal in-control chart performance, defined in terms of a suitable performance criterion, is achieved, for a chosen set of estimators and given the available amount of Phase I data. This means that, different from L, L^* may vary depending on the amount of data (*m* and *n*) available to estimate μ_0 and σ_0 and the type (unbiased, biased, minimum variance, etc.) of estimators ($\hat{\mu}_0$ and $\hat{\sigma}_0$) one uses to estimate the parameters. When the amount of Phase I data is sufficiently large (i.e., when *m* and or *n* tend to infinity), L^* is expected to be equal to (converge to) L.

In order to study the effects of the estimation of the process parameter(s) on the performance of a control chart in general, it is convenient to begin with a study of the Phase II probability of a signal given the estimator(s), the so-called conditional probability of a signal (*CPS*), as noted in Chapter 1. Note again that when the process is in-control, a signal represents a false alarm and its probability is called the false alarm rate. As noted earlier, the conditional false alarm rate is denoted *CFAR*. These are discussed in the next section.

2.3. The Conditional Probability of a Signal and the Conditional False Alarm Rate

A signal occurs when, in a given Phase II sample, the average $\bar{X}$ lies outside the control limits. Given the expressions for the control limits (Eq. 1 and Eq. 2) and replacing μ_0 and σ_0 by $\bar{\bar{X}}$ and S_p when appropriate, the conditional probability of a signal (CPS) for any Phase II sample can be written, respectively, for Case UU KU and UK are

$$CPS_{\delta,UU} = P\left(Signal|\bar{\bar{X}}, S_p\right) = 1 - P\left(\bar{\bar{X}} - L\frac{S_p}{\sqrt{n}} \leq \bar{X} \leq \bar{\bar{X}} + L\frac{S_p}{\sqrt{n}}\right), \quad (6)$$

$$CPS_{\delta,KU} = P\left(Signal|S_p\right) = 1 - P\left(\mu_0 - L\frac{S_p}{\sqrt{n}} \le \bar{X} \le \mu_0 + L\frac{S_p}{\sqrt{n}}\right) \quad (7)$$

and

$$CPS_{\delta,UK} = P\left(Signal|\bar{\bar{X}}\right) = 1 - P\left(\bar{\bar{X}} - L\frac{\sigma_0}{\sqrt{n}} \le \bar{X} \le \bar{\bar{X}} + L\frac{\sigma_0}{\sqrt{n}}\right). \quad (8)$$

In the remainder of this work, subscripts UU, KU and KU will be used when necessary to indicate the cases, as above. It is then evident (as denoted in the left-hand member of (6), (7) and (8)) that the probability of a signal in Phase II is conditioned on the value of the estimators $\bar{\bar{X}}$ and S_p in (6), estimator S_p in (7) and estimator $\bar{\bar{X}}$ in (8). Before proceeding, it is convenient to define some notations. Let μ denote the actual process mean in Phase II (being it in control or out of control). Let's also define the scaled shift of the mean as

$$\delta = \frac{\mu - \mu_0}{\sigma_0}. \quad (9)$$

When $\mu = \mu_0$, $\delta = 0$ and the process mean is in control. When $\mu = \mu_1 \ne \mu_0$, $\delta \ne 0$ and the process mean is out of control. Also, it is known that $Y = m(n-1)\, S_p^2/\sigma_0^2$ follows a chi-square distribution with $m(n-1)$ degrees of freedom and $Z = \left(\frac{\bar{\bar{X}} - \mu_0}{\sigma_0}\right)\sqrt{mn}$ follows a standard normal distribution (see Chakraborti, 2000). Note that Y and Z are proportional to the estimation errors defined respectively as S_p^2/σ_0^2 and $\left(\frac{\bar{\bar{X}} - \mu_0}{\sigma_0}\right)$. Recalling that $X \sim N(\mu, \sigma_0^2)$ implies that $\bar{X} \sim N(\mu, \sigma_0^2/n)$ where $\mu = \mu_0 + \delta\sigma_0$ (see Eq. 9), the conditional probability of a signal (CPS) for any Phase II sample, can be written, for Case UU, KU and UK

$$CPS_{\delta,UU} = P(Signal|Z, Y, \delta)$$

$$= 1 - \left[\Phi\left(\frac{Z}{\sqrt{m}} + L\sqrt{\frac{Y}{m(n-1)}} - \delta\sqrt{n}\right) - \Phi\left(\frac{Z}{\sqrt{m}} - L\sqrt{\frac{Y}{m(n-1)}} - \delta\sqrt{n}\right)\right], \quad (10)$$

$$CPS_{\delta,KU} = P(Signal|Y, \delta)$$

$$= 1 - \left[\Phi\left(L\sqrt{\frac{Y}{m(n-1)}} - \delta\sqrt{n}\right) - \Phi\left(-L\sqrt{\frac{Y}{m(n-1)}} - \delta\sqrt{n}\right)\right] \quad (11)$$

and

$$CPS_{\delta,UK} = P(Signal|Z, \delta)$$

$$= 1 - \left[\Phi\left(\frac{Z}{\sqrt{m}} + L - \delta\sqrt{n}\right) - \Phi\left(\frac{Z}{\sqrt{m}} - L - \delta\sqrt{n}\right)\right]. \quad (12)$$

Expressions (10), (11) and (12) are convenient because they express the conditional probability of a signal in terms of two random variables with well-known distributions, namely a standard normal distribution and a chi-square distribution. These general expressions apply to both in-control and out-of-control situations. As noted, in the in-control case $\delta = 0$, whereas in the out-of-control case $\delta \neq 0$. Hence in the in-control situation, the conditional probability of a signal, namely, the Conditional False Alarm Rate, $CFAR$, is expressed for Case KU and Case UU by:

$$CFAR_{UU} = CPS_{0,UU} = P(Signal|Z, Y, \delta = 0)$$

$$= 1 - \left[\Phi\left(\frac{Z}{\sqrt{m}} + L\sqrt{\frac{Y}{m(n-1)}}\right) - \Phi\left(\frac{Z}{\sqrt{m}} - L\sqrt{\frac{Y}{m(n-1)}}\right)\right], \quad (13)$$

$$CFAR_{KU} = CPS_{0,KU} = P(Signal|Y, \delta = 0) = 2\Phi\left(-L\sqrt{\frac{Y}{m(n-1)}}\right) \quad (14)$$

and

$$CFAR_{UK} = CPS_{0,UK} = P(Signal|Z, \delta = 0)$$

$$= 1 - \left[\Phi\left(\frac{Z}{\sqrt{m}} + L\right) - \Phi\left(\frac{Z}{\sqrt{m}} - L\right)\right]. \quad (15)$$

respectively. Given that the in-control conditional run length (RL_0) distribution of the $\bar{X}$ chart is geometric with parameter $CFAR$ (see, for example, Chakraborti (2000)), then its expected value, the conditional in-control average run length $CARL_0$, is:

$$CARL_0 = \frac{1}{CFAR}, \qquad 0 \leq CFAR \leq 1. \quad (16)$$

Also, let $CARL_{0,UU}$, $CARL_{0,KU}$ and $CARL_{0,UK}$ denote the conditional in-control average run lengths in Case UU, Case KU and Case UK respectively. Henceforth, when $CARL_0$, $CFAR$ or CPS_δ in a given equation do not receive a subscript UU, KU or UK, this means that the equation is general for the 3 cases (as in Eq. 16). Also, let's define the $CARL_\delta$ as the notation of the conditional average run length in general, i.e., for the in-control and out-of-control situations. So, note that $CARL_\delta = 1/CPS_\delta$.

Figure 3 shows a 3D contour plot of $CFAR_{UU}$ in function of Z and Y when $m = 30$, $n = 5$ and $\alpha = 0.0027$ (i.e., $L = 3$). Note that the values of $CFAR_{UU}$ can be significantly different from 0.0027 for many combinations of Z and Y. When parameters are estimated, the false alarm rate conditioned on the estimates ($CFAR_{UU}$) can assume very large values such as 0.02 a values more than 7 times larger than the nominal 0.0027.

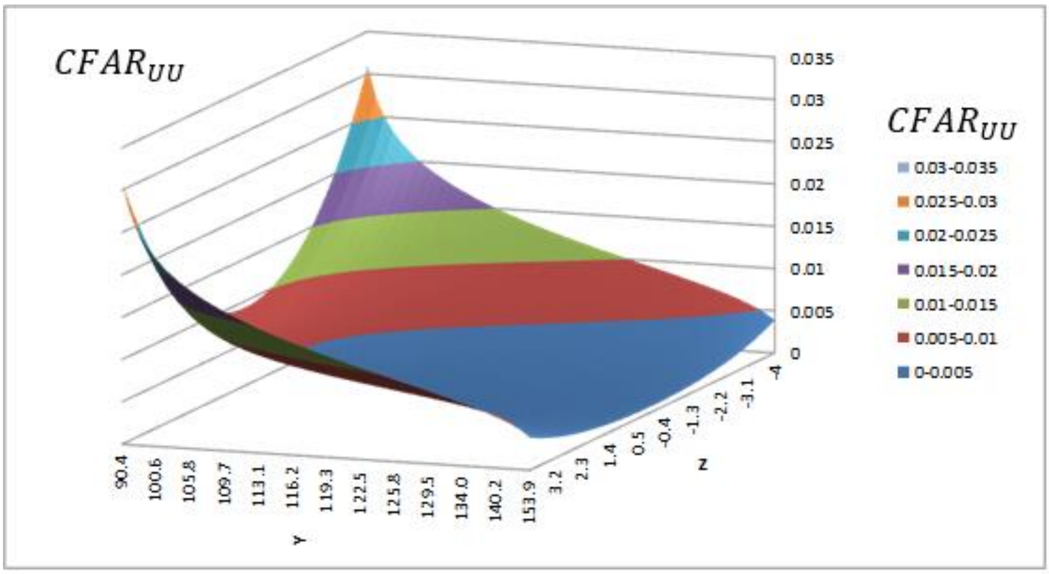


Figure 3. $\boldsymbol{CFAR_{UU}}$ as function of $\boldsymbol{Z}$ and $\mathbf{Y}$ when $\boldsymbol{m = 30}$, $\boldsymbol{n = 5}$ and $\boldsymbol{\alpha = 0.0027}$ (i.e., $\boldsymbol{L = 3}$).

To visualize the effect of the number of Phase I samples, m, Figure 4 next presents the $CFAR_{KU}$ curves parametrized by m by plotting $CFAR_{KU}$ as a function of the order (u) of the quantiles of Y. One way to do this is using the probability integral transformation, which yields the fact that the c.d.f. of Y ($F_{\chi^2_{m(n-1)}}(Y)$) has the same distribution of a random variable U, uniformly distributed between 0 and 1. In fact, it is not correct to construct a graph of $CFAR_{KU}$ directly in function of the random variable Y when the graph is parametrized by m. This is because the values of Y *also* depend on values of m since Y follows a chi-square distribution with $m(n-1)$ degrees of freedom. Figure 4 illustrates the curves of $CFAR_{KU} \times u$ for $n = 5$, $m = 10,\ 20, 50, 100, 500$ and $\alpha = 0.0027$ (i.e., $L = 3$).

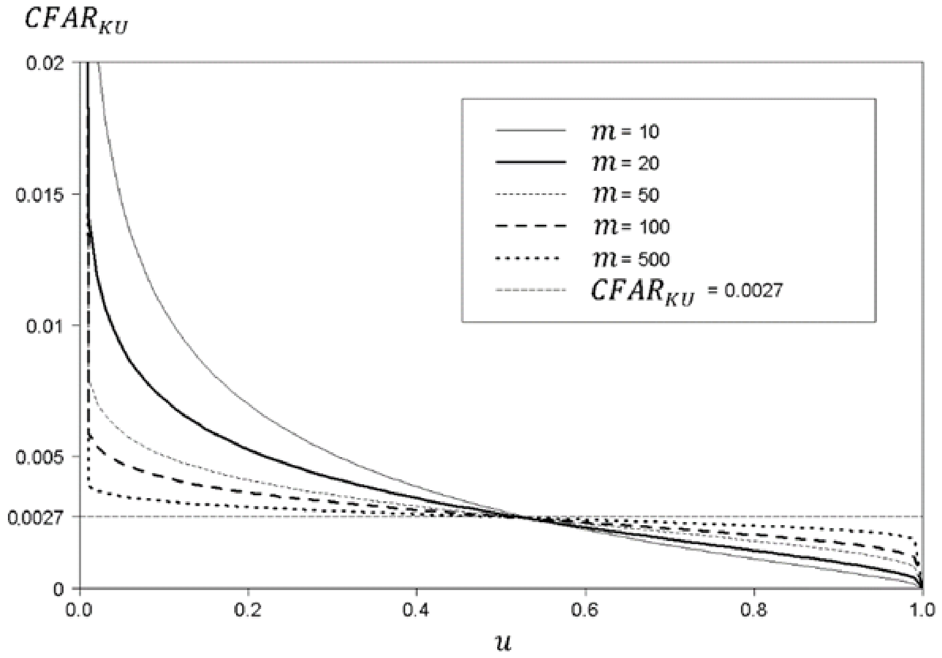


Figure 4. $\boldsymbol{CFAR_{KU}}$ as function of $\boldsymbol{u}$ for $\boldsymbol{n = 5}$, $\boldsymbol{m = 10, 20, 50, 100, 500}$ and $\boldsymbol{\alpha = 0.0027}$ (i.e., $\boldsymbol{L = 3}$).

Figure 4 clearly shows the effect of the number of Phase I samples m on the performance of the $\bar{X}$ control chart when the process standard deviation is estimated. The horizontal line corresponds to the value of the false alarm rate in Case KK and can be considered the target when the 3-Sigma limits are used. So, it can be seen, for $n = 5$, that the curves of $CFAR_{KU}$ are significantly closer to the horizontal line (the target) when *m*, the number of initial reference samples, is larger (compare for example the curves for *m* = 10 and for *m* = 500). This means that the difference between the target and the actual conditional false alarm rate is considerably more likely to be larger when m is small. It is also interesting to note that the effect is different on the two sides of $u = 0.5$ (the 0.5 quantile of Y). This is caused by the skewness of the distribution of $CFAR_{KU}$ which is derived in Chapter 3.

Using Equation (16) it is also possible to plot curves of the conditional in-control average run length in Case KU ($CARL_{0,KU}$), this is showed in Figure A.1 in Appendix A. The conclusions are similar to the one explained above and can been checked in more detail in an resulting paper of this thesis [Jardim et al. (2017b)] in Annex B. Using the same procedure, and the same values for $\boldsymbol{\alpha}$, m and n, Figure 5 shows the $CFAR_{UK}$ curves.

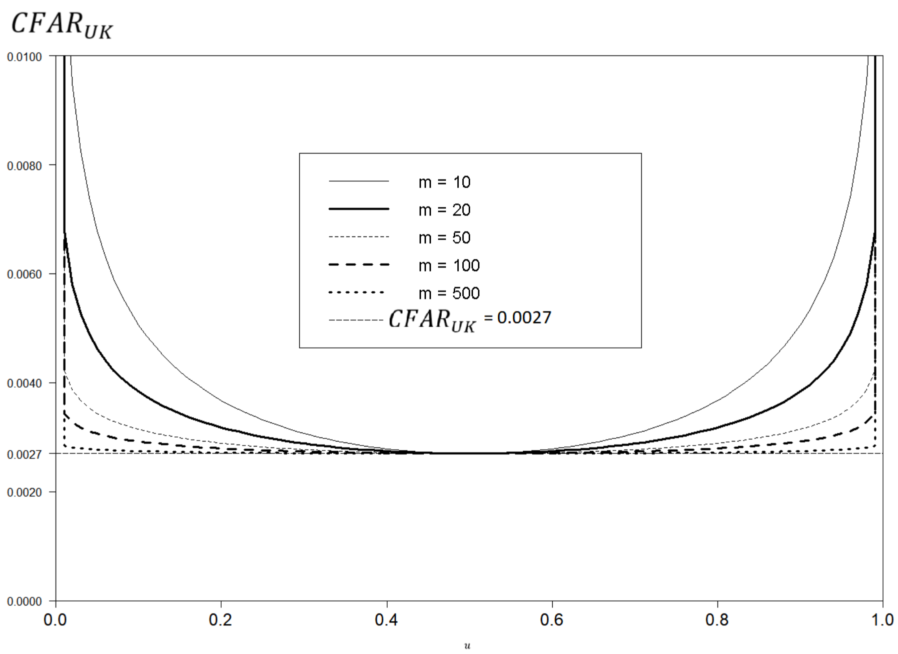


Figure 5. $\boldsymbol{CFAR_{UK}}$ as function of $\boldsymbol{u}$ for $\boldsymbol{n = 5}$, $\boldsymbol{m = 10, 20, 50, 100, 500}$ and $\boldsymbol{\alpha = 0.0027}$ (i.e., $\boldsymbol{L = 3}$).

An interesting behavior of the $CFAR_{UK}$ can be observed in Figure 5: the minimum possible value of $CFAR_{UK}$ is the exact nominal desired value in case KK (i.e., 0.0027, for $L = 3$). The fact that $CFAR_{UK}$ cannot be smaller than the its nominal value is a particular and remarkable property of case UK. Similar conclusion can be made for the $CARL_{0,UK}$ curves which is showed in Figure A.2 in Appendix A.

Figures 3, 4 and 5 provide good insights of the conditional false alarm rate ($CFAR$) behavior when the process mean or the process standard deviation are estimated (such as the minimum possible value of $CFAR$ or "how distant" is it curve from the nominal line). However, they do not clear provide the distributions of this random variable. In the next chapter, the distributions of $CFAR$ and $CARL_0$ are derived. These distributions clearly show the effect of parameters estimation on the $\bar{X}$ chart performance. They are also essential to *calculate exact results* for the Exceedance Probability Criterion (*EPC*), as noted in the Introduction.

3
Derivation of the cumulative distribution functions of the $CARL_0$ and $CFAR$

As explained in the introduction, the control chart´s performance is commonly measure by the probability of the $CARL_0$ (or $CFAR$) exceeding a specified value. This is called the Exceedance Probability Criterion (*EPC*). Thus, it is evident that, to use the *EPC*, it is necessary to know the c.d.f. of the $CARL_0$ and $CFAR$. When it comes to the possibly most well-known control chart of all, the $\bar{X}$ control chart under the normality assumption, the exact c.d.f. expressions of the $CARL_0$ and $CFAR$ are unknown. Given this, in this chapter, we derive the exact c.d.f. expression of the $CARL_0$ (and $CFAR$) of the $\bar{X}$ control chart for the 3 possible parameters case estimation (UU, KU and KU). To do this end, we first derive the c.d.f. expression of the conditional probability of a signal, CPS_δ, which works for the in-control and out-of-control situations (see Equations 10, 11 and 12) and its reciprocal, the conditional average run length $CARL_\delta$. With this c.d.f´s, we also calculate some important properties of the $CARL_0$, such as the mean and standard deviation.

Before proceeding, note that, as shown in Equation (16), $CARL_0$ is a monotonic decreasing function of $CFAR$, so the cumulative distribution function (c.d.f.) of $CFAR$ (F_{CFAR}) is related to the c.d.f of $CARL_0$ (F_{CARL_0}) as shown below:

$$F_{CARL_0}(w) = P(CARL_0 \leq w) = P(1/CFAR \leq w)$$
$$= P(CFAR \geq w^{-1}) = 1 - F_{CFAR}(w^{-1}), \quad w \geq 1. \tag{17}$$

Of course, given that $CARL_\delta = 1/CPS_\delta$ the same works for the general notations CPS_δ and $CARL_\delta$, i.e.:

$$F_{CARL_\delta}(w) = P(CARL_\delta \leq w) = P(1/CPS_\delta \leq w)$$
$$= P(CPS_\delta \geq w^{-1}) = 1 - F_{CPS_\delta}(w^{-1}), \quad w \geq 1. \tag{18}$$

3.1. C.d.f of $CFAR$ (or CPS_δ) and $CARL_0$ (or $CARL_\delta$) in Case UU

In Case UU, the $CPS_{\delta,UU}$ is a function of two random variables (Y and Z), so the derivation of an exact close expression of $F_{CPS_{\delta,UU}}$ requires using the *distribution function technique* and the *conditioning-unconditioning technique* [see Chakraborti (2000)], by first conditioning on Z (see Equation 10) using the following conditional expectation:

$$F_{CPS_{\delta,UU}}(t) = P\left(CPS_{\delta,UU} \le t\right) = E_Z\left(P\left(CPS_{\delta,UU} \le t \middle| Z = z\right)\right)$$

$$= \int_{-\infty}^{\infty} P\left(CPS_{\delta,UU} \le t \middle| Z = z\right) f_Z(z)dz,$$

(19)

where f_Z denotes the probability density function (p.d.f.) of Z.

The next step is to derive an expression of $P\left(CPS_{\delta,UU} \le t \middle| Z = z\right)$. Note that, given z, $P\left(CPS_{\delta,UU} \le t \middle| Z = z\right)$ is a function of only the chi-square random variable Y. So, from Equation (10) one can write:

$$P\left(CPS_{\delta,UU} \le t \middle| Z = z\right)$$

$$= P\left(1 - \left[\Phi\left(\frac{z}{\sqrt{m}} + L\sqrt{\frac{Y}{m(n-1)}} - \delta\sqrt{n}\right) - \Phi\left(\frac{z}{\sqrt{m}} - L\sqrt{\frac{Y}{m(n-1)}} - \delta\sqrt{n}\right)\right] \le t\right)$$

$$= P\left(P\left(\frac{z}{\sqrt{m}} - L\sqrt{\frac{Y}{m(n-1)}} - \delta\sqrt{n} \le Z_1 \le \frac{z}{\sqrt{m}} + L\sqrt{\frac{Y}{m(n-1)}} - \delta\sqrt{n}\right) \ge 1 - t\right),$$

(20)

where Z_1 also follows a standard normal distribution. So,

$$P\left(CPS_{\delta,UU} \le t \middle| Z = z\right)$$

$$= P\left(P\left(-L\sqrt{\frac{Y}{m(n-1)}} \le Z_1 - \frac{z}{\sqrt{m}} + \delta\sqrt{n} \le L\sqrt{\frac{Y}{m(n-1)}}\right) \ge 1 - t\right)$$

$$= P\left(P\left(\left(Z_1 - \frac{z}{\sqrt{m}} + \delta\sqrt{n}\right)^2 \le \left(L\sqrt{\frac{Y}{m(n-1)}}\right)^2\right) \ge 1 - t\right). \quad (21)$$

Given that $\left(Z_1 - \frac{z}{\sqrt{m}} + \delta\sqrt{n}\right)^2$ follows a non-central qui-square distribution with 1 degree of freedom and non-centrality parameter given by $\left(\frac{z}{\sqrt{m}} - \delta\sqrt{n}\right)^2$, one can define $\left(Z_1 - \frac{z}{\sqrt{m}}\right)^2 = \chi^2_{1,\left[\left(\frac{z}{\sqrt{m}}-\delta\sqrt{n}\right)^2\right]}$, and

$$P\left(CPS_{\delta,UU} \leq t \middle| Z = z\right) = P\left(P\left(\chi^2_{1,\left[\left(\frac{z}{\sqrt{m}}-\delta\sqrt{n}\right)^2\right]} \leq L^2 \frac{Y}{m(n-1)}\right) \geq 1 - t\right)$$

$$= 1 - F_{\chi^2_{m(n-1)}}\left(\frac{m(n-1)F^{-1}_{\chi^2_{1,\left[\left(\frac{z}{\sqrt{m}}-\delta\sqrt{n}\right)^2\right]}}(1-t)}{L^2}\right), \qquad (22)$$

where $F^{-1}_{\chi^2_{1,\left[\frac{z^2}{m}\right]}}$ denotes the inverse of the c.d.f. of $\left(Z_1 - \frac{z}{\sqrt{m}} + \delta\sqrt{n}\right)^2 = \chi^2_{1,\left[\left(\frac{z}{\sqrt{m}}-\delta\sqrt{n}\right)^2\right]}$. Using Equation (22) on Equation (19), we have the final exact expression for the c.d.f. of $CPS_{\delta,UU}$:

$$F_{CPS_{\delta,UU}}(t) = 1 - \int_{-\infty}^{\infty} F_{\chi^2_{m(n-1)}}\left(\frac{m(n-1)F^{-1}_{\chi^2_{1,\left[\left(\frac{z}{\sqrt{m}}-\delta\sqrt{n}\right)^2\right]}}(1-t)}{L^2}\right) f_Z(z)dz. \qquad (23)$$

When $\delta = 0$, as noted earlier, $CPS_{0,UU}$ is the conditional false alarm rate $CFAR_{UU}$, so the exact c.d.f. of $CFAR_{UU}$ is expressed as

$$F_{CFAR_{UU}}(t) = 1 - \int_{-\infty}^{\infty} F_{\chi^2_{m(n-1)}}\left(\frac{m(n-1)F^{-1}_{\chi^2_{1,\left[\left(\frac{z}{\sqrt{m}}\right)^2\right]}}(1-t)}{L^2}\right) f_Z(z)dz. \qquad (24)$$

Using Equation (23) on Equation (18), one has the exact c.d.f. expression of the general conditional average run length, $CARL_{\delta,UU}$ (for the in-control and out-of-control situations) as show below:

$$F_{CARL_{\delta,UU}}(w) = \int_{-\infty}^{\infty} F_{\chi^2_{m(n-1)}}\left(\frac{m(n-1)F^{-1}_{\chi^2_{1,\left[\left(\frac{z}{\sqrt{m}}-\delta\sqrt{n}\right)^2\right]}}\left(1-\frac{1}{w}\right)}{L^2}\right) f_Z(z)dz. \tag{25}$$

Using $\delta = 0$, the exact c.d.f. expression of the in-control conditional average run length, $CARL_{\delta,UU}$ is expressed as:

$$F_{CARL_{0,UU}}(w) = \int_{-\infty}^{\infty} F_{\chi^2_{m(n-1)}}\left(\frac{m(n-1)F^{-1}_{\chi^2_{1,\left[\frac{z^2}{m}\right]}}(1-w^{-1})}{L^2}\right) f_Z(z)dz. \tag{26}$$

Note that Expressions (23), (24), (25) and (26) for the c.d.f's in Case UU are exact, however their evaluation involves calculating the integral using some numerical method, since there is no closed-form solution for this integral. This is not difficult since there are plenty of software that precisely calculate integral numerically (such as MATLAB and R). In this work, the R language is used. Indeed, many well-known c.d.f.'s are expressed in terms of integrals, including the one for the celebrated normal distribution.

Figure 6 and 7 show, respectively, the c.d.f. of the $CFAR_{UU}$ and the $CARL_{0,UU}$, calculated using Equation (24) and (26), for $n = 5$, $m = 10, 20, 50, 100, 500$ and $\alpha = 0.0027$ (i.e., $L = 3$). Note that the vertical lines show the nominal false alarm rate 0.0027 (Figure 6) and the in-control average run length 370.4 (Figure 7). The impact of m on the distributions is clear. When m is small (such as $m = 10$), chances are high that the realized false alarm rate is higher than the nominal one. For example, from Figure 6, for $m = 10$, $P(CFAR_{UU} \geq 0.006) \approx 40\%$, so that there is a 40% chance that the conditional false alarm rate is 122% higher than the nominal 0.0027. Also note the significant difference between the vertical line and the c.d.f. curve for smaller values of m. When m gets larger (such as $m = 500$), the c.d.f curves are much "closer" to the vertical line, meaning that in these cases, the $CFAR_{UU}$ is likely to be not much different from 0.0027. Similar conclusions hold for $CARL_{0,UU}$ curves.

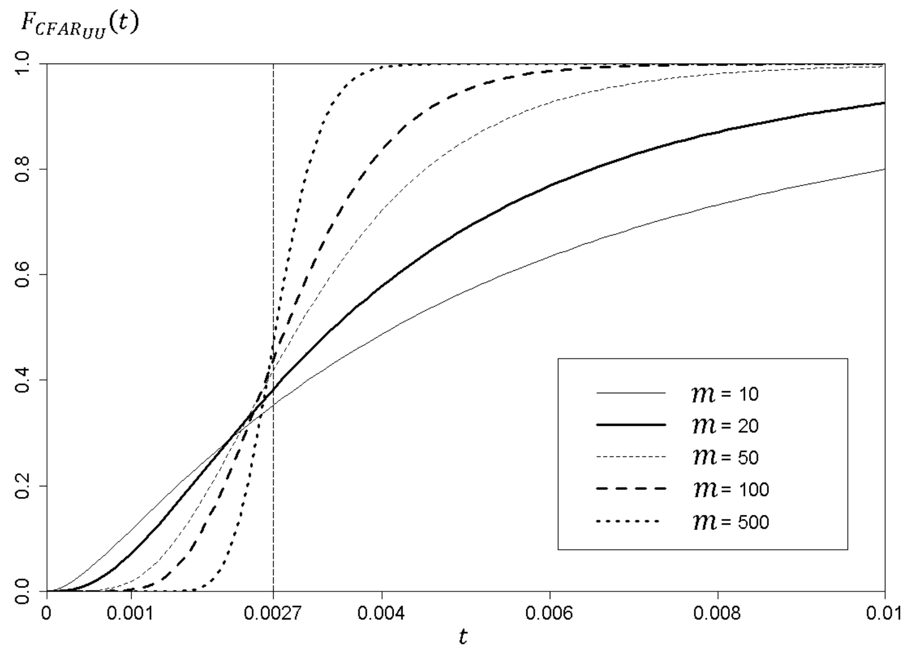


Figure 6. c.d.f. of $\boldsymbol{CFAR_{UU}}$ for $\boldsymbol{n = 5}$, $\boldsymbol{m = 10, 20, 50, 100, 500}$ and $\boldsymbol{\alpha = 0.0027}$ (i.e., $\boldsymbol{L = 3}$).

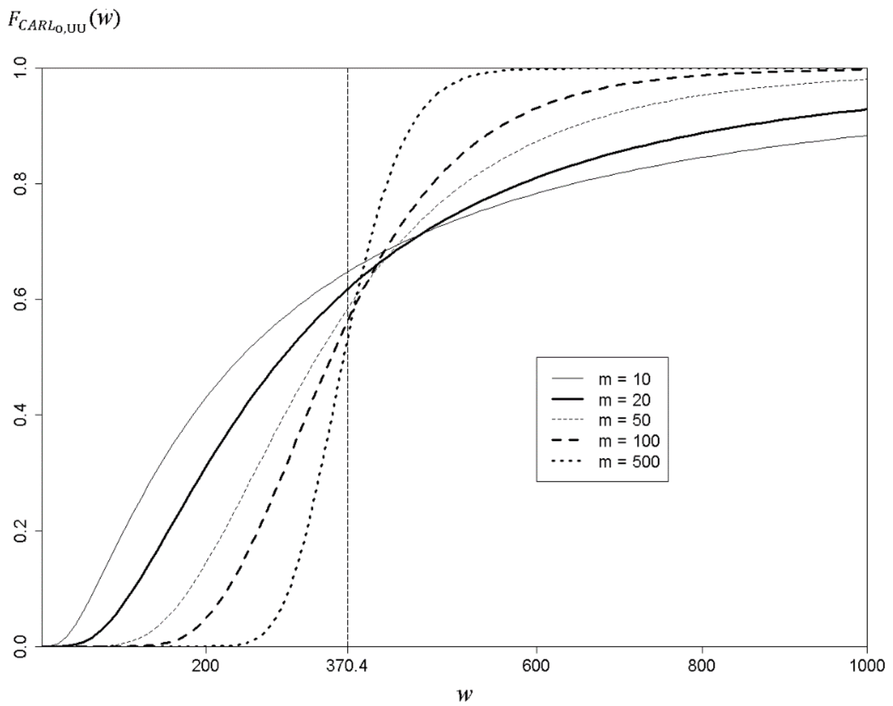


Figure 7. c.d.f. of $\boldsymbol{CARL_{0,UU}}$ for $\boldsymbol{n = 5}$, $\boldsymbol{m = 10, 20, 50, 100, 500}$ and $\boldsymbol{\alpha = 0.0027}$ (i.e., $\boldsymbol{L = 3}$).

To provide further insight, in Figure 8 and 9, it is displayed the p.d.f. of $CARL_{0,UU}$ ($f_{CARL_{0,UU}}$) and $CFAR_{UU}$ ($f_{CFAR_{UU}}$), respectively, calculated by taking the numerical derivatives of the corresponding c.d.f. This was done in the R language using the package “numDeriv” (for details of the codes, see Appendix H). The $f_{CARL_{0,UU}}$ plot shows the large density at values well below 370.4 (including the position of the modes), meaning that when parameters are estimated, in practice, there is a large probability that the $CARL_{0,UU}$ is substantially smaller (and the $CFAR_{UU}$ is substantially larger) than the nominal value, even with a number of Phase I samples quite larger than the usually recommended 25, 30 or 50 Phase I samples. This is reflected in the long right tails of the density functions of $CFAR_{UU}$ and $CARL_{0,UU}$.

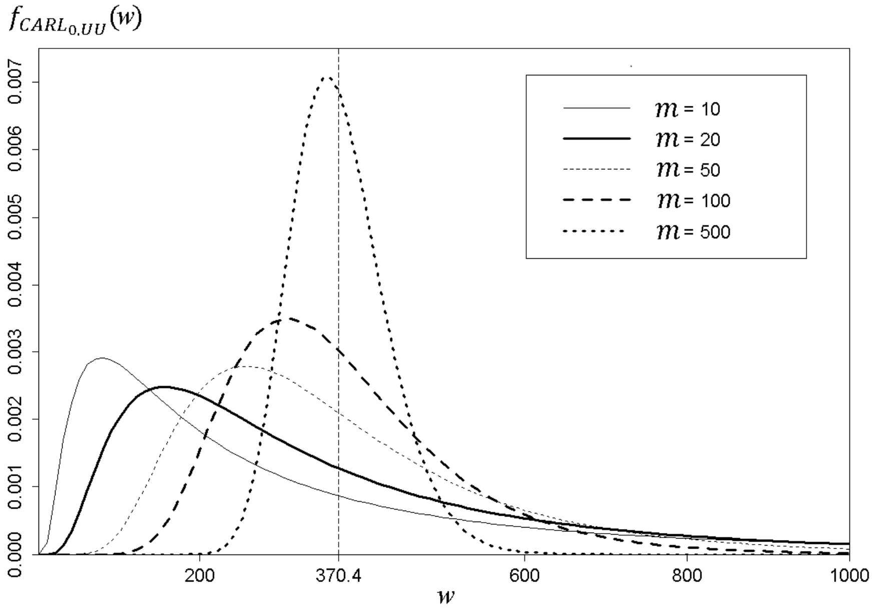


Figure 8. p.d.f. of $\boldsymbol{CARL_{0,UU}}$ for $\boldsymbol{n = 5}$, $\boldsymbol{m = 10, 20, 50, 100, 500}$ and $\boldsymbol{\alpha = 0.0027}$ ($\boldsymbol{L = 3}$).

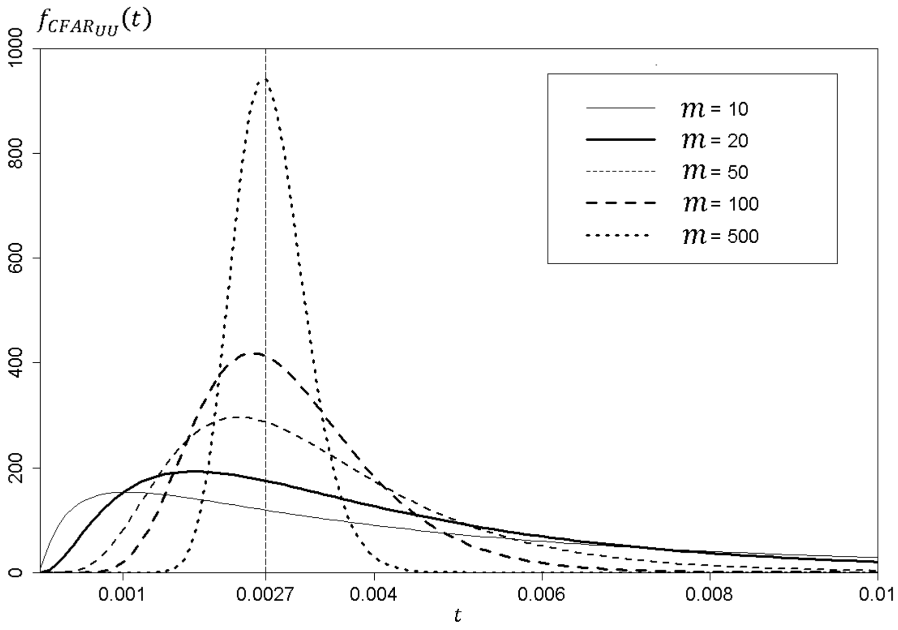


Figure 9. p.d.f. of $\boldsymbol{CFAR_{UU}}$ for $\boldsymbol{n = 5}$, $\boldsymbol{m = 10, 20, 50, 100, 500}$ and $\boldsymbol{\alpha = 0.0027}$ ($\boldsymbol{L = 3}$).

3.2. C.d.f of $CFAR$ (or CPS_δ) and $CARL_0$ (or $CARL_\delta$) in Case KU

From Equation (11), the c.d.f. of the general conditional probability of a signal in case KU ($CPS_{\delta,KU}$) can be obtained similarly to case UU in section 3.1. However, note that in this case, $CPS_{\delta,KU}$ is a function of only one random variable (which is Y), so, the *conditioning-unconditioning method* is not required and the *distribution function* technique is enough in this case. From Eq. (11) one has

$$F_{CPS_{\delta,KU}}(t) = P\left(CPS_{\delta,KU} \le t\right)$$
$$= P\left(1 - \Phi\left(L\sqrt{\frac{Y}{m(n-1)}} - \delta\sqrt{n}\right) - \Phi\left(-L\sqrt{\frac{Y}{m(n-1)}} - \delta\sqrt{n}\right) \le t\right)$$
$$= P\left(P\left(-L\sqrt{\frac{Y}{m(n-1)}} - \delta\sqrt{n} \le Z_1 \le L\sqrt{\frac{Y}{m(n-1)}} - \delta\sqrt{n}\right) \ge 1 - t\right), \quad (27)$$

where Z_1 also follows a standard normal distribution. So

$$F_{CPS_{\delta,KU}}(t) = P\left(CPS_{\delta,KU} \le t\right)$$

$$= P\left(P\left(-L\sqrt{\frac{Y}{m(n-1)}} \leq Z_1 + \delta\sqrt{n} \leq L\sqrt{\frac{Y}{m(n-1)}}\right) \geq 1 - t\right)$$

$$= P\left(P\left(\left(Z_1 + \delta\sqrt{n}\right)^2 \leq \left(L\sqrt{\frac{Y}{m(n-1)}}\right)^2\right) \geq 1 - t\right). \qquad (28)$$

Given that $\left(Z_1 + \delta\sqrt{n}\right)^2$ follows a non-central qui-square distribution with 1 degree of freedom and non-centrality parameter given by $\left(\delta\sqrt{n}\right)^2$, one can define $\left(Z_1 - \frac{z}{\sqrt{m}}\right)^2 = \chi^2_{1,\left[\left(\delta\sqrt{n}\right)^2\right]}$, and

$$F_{CPS\,\delta,KU}(t) = P\left(P\left(\chi^2_{1,\left[\left(\delta\sqrt{n}\right)^2\right]} \leq \left(L\sqrt{\frac{Y}{m(n-1)}}\right)^2\right) \geq 1 - t\right)$$

$$= P\left(F_{\chi^2_{1,\left[\left(\delta\sqrt{n}\right)^2\right]}}\left(\left(L\sqrt{\frac{Y}{m(n-1)}}\right)^2\right) \geq 1 - t\right). \qquad (29)$$

Rearranging the terms, the final exact expression of the c.d.f. of $CPS_{\,\delta,KU}$ is

$$F_{CPS\,\delta,KU}(t) = P\left(Y \geq m(n-1)\frac{F^{-1}_{\chi^2_{1,\left[\left(\delta\sqrt{n}\right)^2\right]}}(1-t)}{L^2}\right)$$

$$= 1 - F_{\chi^2_{m(n-1)}}\left(m(n-1)\frac{F^{-1}_{\chi^2_{1,\left[\left(\delta\sqrt{n}\right)^2\right]}}(1-t)}{L^2}\right). \qquad (30)$$

According to Equation (18), the exact c.d.f. of the general conditional average run length ($CARL_{\delta,KU}$) is:

$$F_{CARL_{\delta,KU}}(w) = F_{\chi^2_{m(n-1)}}\left(m(n-1)\frac{F^{-1}_{\chi^2_{1,\left[\left(\delta\sqrt{n}\right)^2\right]}}\left(1-\frac{1}{w}\right)}{L^2}\right). \qquad (31)$$

When $\delta = 0$, the exact c.d.f. of the conditional false alarm rate ($CFAR_{KU} = CPS_{\,0,KU}$) is expressed as

$$F_{CFAR_{KU}}(t) = 1 - F_{\chi^2_{m(n-1)}}\left(m(n-1)\frac{F^{-1}_{\chi^2_1}(1-t)}{L^2}\right) \tag{32}$$

Given Equation (14) or the fact that $F^{-1}_{\chi^2_1}(1-t) = \left(\Phi^{-1}\left(\frac{t}{2}\right)\right)^2$, it is also possible to derive an alternative exact formula for the c.d.f. of the conditional false alarm rate $CFAR_{KU}$ as shown below.

$$F_{CFAR_{KU}}(t) = P(CFAR_{KU} \leq t) = P\left(2\Phi\left(-L\sqrt{\frac{Y}{m(n-1)}}\right) \leq t\right)$$

$$= P\left(-L\sqrt{\frac{Y}{m(n-1)}} \leq \Phi^{-1}\left(\frac{t}{2}\right)\right) = 1 - F_{\chi^2_{m(n-1)}}\left(m(n-1)\frac{\left(\Phi^{-1}\left(\frac{t}{2}\right)\right)^2}{L^2}\right) \tag{33}$$

Figure 10 shows curves of the c.d.f. $CFAR_{KU}$ ($F_{CFAR_{KU}}$) calculated by Equations (32) and (33) for values of $n = 5$, $m = 10, 20,\ 50, 100,\ 500$ and $\alpha = 0.0027$ (i.e., $L = 3$). Figure 11 clearly show the effect of the number of samples m on $CFAR$ distribution. It is interesting that the curves of $F_{CFAR_{KU}}$ are very similar to the curves of $F_{CFAR_{UU}}$ (compare Figure 10 with Figure 6).

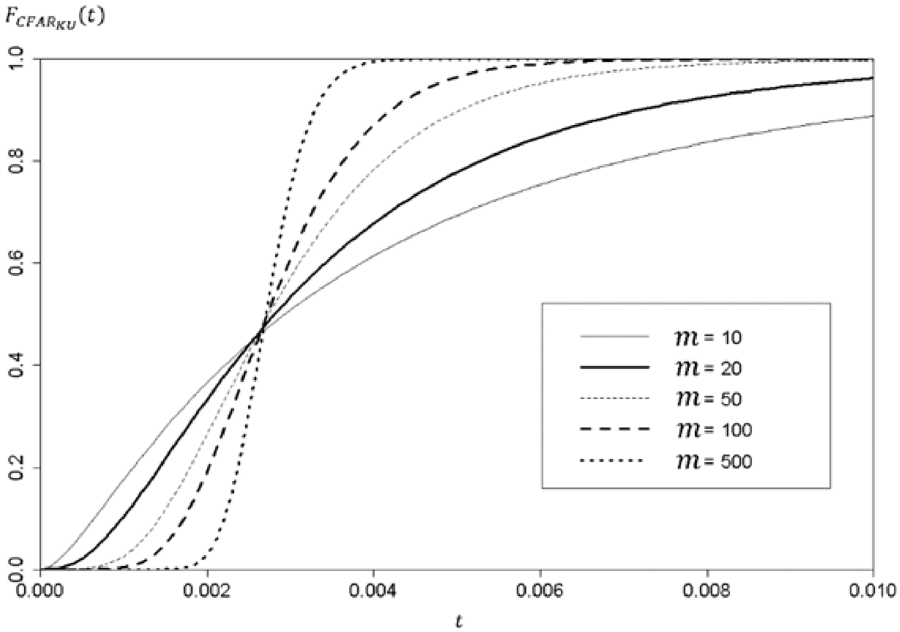


Figure 10. c.d.f. of $\boldsymbol{CFAR_{KU}}$ for $\boldsymbol{n = 5}$, $\boldsymbol{m = 10, 20, 50, 100, 500}$ and $\boldsymbol{\alpha = 0.0027}$ (i.e., $\boldsymbol{L = 3}$).

Using the relationship presented in Equation (17), the exact c.d.f. of the in-control (i.e., $\delta = 0$) conditional average run length ($CARL_{0,KU}$) is expressed as

$$F_{CARL_{0,KU}}(w) = F_{\chi^2_{m(n-1)}}\left(m(n-1)\frac{F^{-1}_{\chi^2_{1,}}\left(1-\frac{1}{w}\right)}{L^2}\right), \quad (34)$$

or, given that $F^{-1}_{\chi^2_1}(1-t) = \left(\Phi^{-1}\left(\frac{t}{2}\right)\right)^2$, the $F_{CARL_{0,KU}}$ can also be expressed as:

$$F_{CARL_{0,KU}}(w) = F_{\chi^2_{m(n-1)}}\left(m(n-1)\frac{\left(\Phi^{-1}\left(\frac{1}{2\mathrm{w}}\right)\right)^2}{L^2}\right). \quad (35)$$

Figure 11 shows the curves of the c.d.f of $CARL_{0,\mathrm{KU}}$ ($F_{CARL_{0,KU}}(t)$) calculated using Equations (34) and (35) for the same values of m and L used in Figure 10. Again, the conclusion is similar to Case UU (compare Figure 12 with Figure 8).

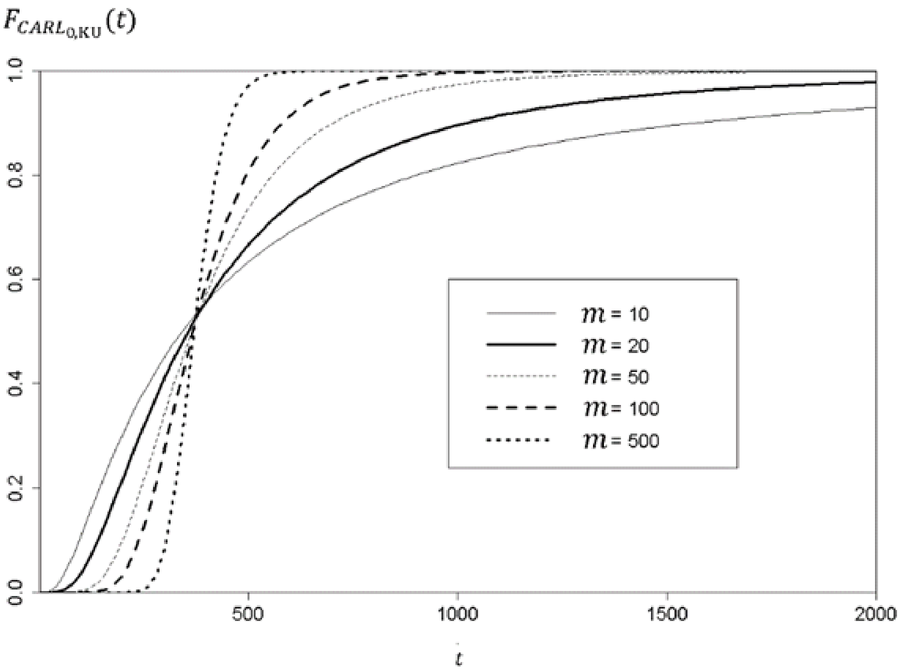


Figure 11. c.d.f. of $\boldsymbol{CARL_{0,KU}}$ for $\boldsymbol{n=5, m=10,20,50,100,500}$ and $\boldsymbol{\alpha=0.0027}$ (i.e., $\boldsymbol{L=3}$).

Figure 12 and 13 display the p.d.f. of $CFAR_{KU}$ ($f_{CFAR_{KU}}(t)$) and $CARL_{0,KU}$ ($f_{CARL_{0,KU}}(t)$) calculated by taking the numerical derivative of (32) and (34) respectively. Like in case UU, these probability density functions have a long right tail for small values of m, meaning that when parameters are estimated in practice there is a large probability of $CFAR$ being substantially larger than 0.0027 (and

$CARL_0$ substantially smaller than 370.4), even with numbers of Phase I samples already quite larger than the usually recommended 25, 30 or 50 Phase I samples. This is a concern in terms of practical consequences, as noted in Chapter 1.

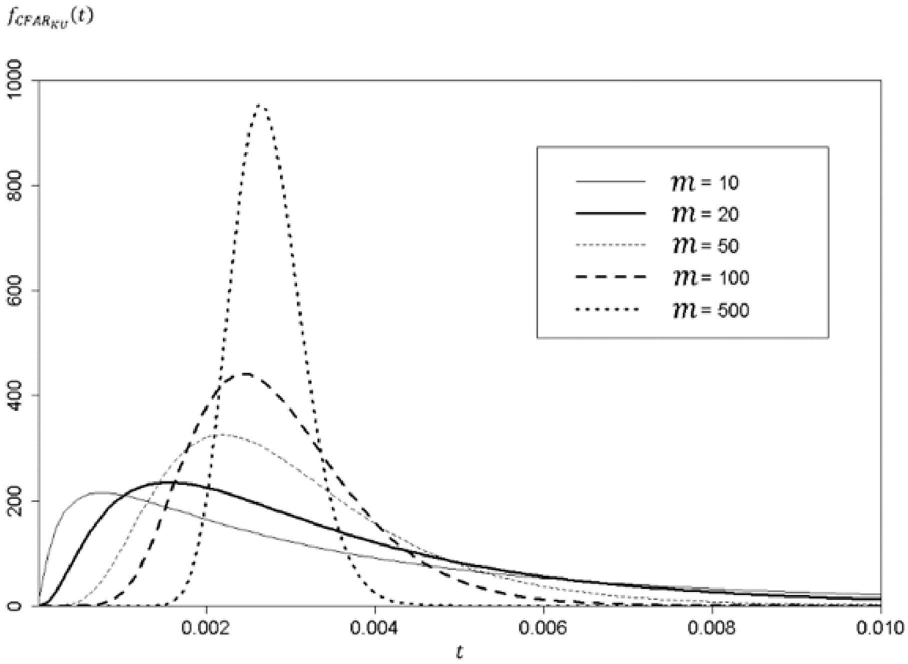


Figure 12. p.d.f of of $\boldsymbol{CFAR_{KU}}$ for $\boldsymbol{n = 5}$, $\boldsymbol{m = 10, 20, 50, 100, 500}$ and $\boldsymbol{\alpha = 0.0027}$ (i.e., $\boldsymbol{L = 3}$).

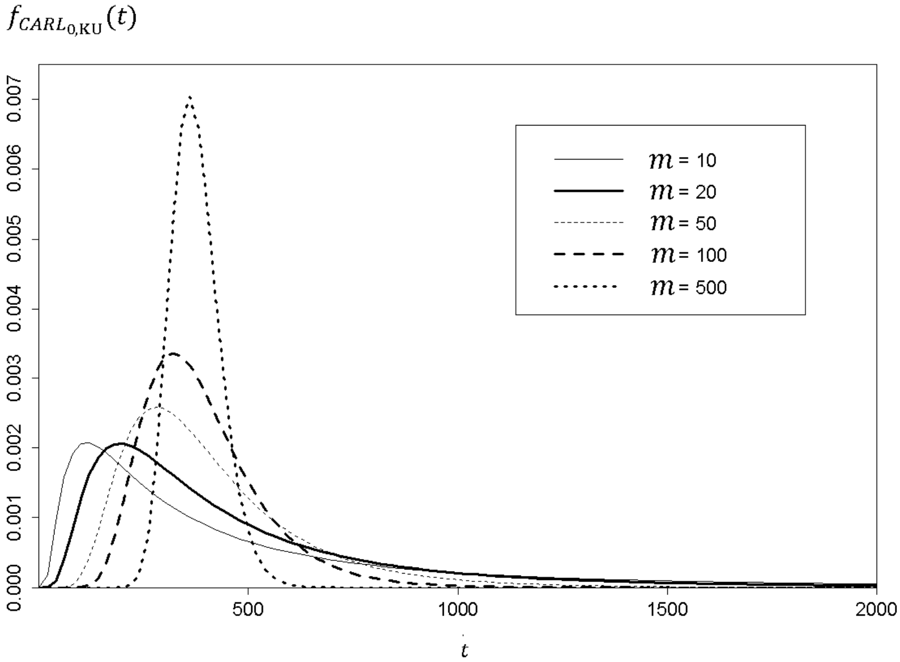


Figure 13. p.d.f. *of* $\boldsymbol{CARL_{0,KU}}$ for $\boldsymbol{n = 5, m = 10, 20, 50, 100, 500}$ and $\boldsymbol{\alpha = 0.0027}$ (i.e., $\boldsymbol{L = 3}$).

3.3. C.d.f of $CFAR$ (or CPS_δ) and $CARL_0$ (or $CARL_\delta$) in Case UK

In case UK, according to Equation (12), the general conditional probability of a signal ($CPS_{\delta,UK}$) is a function of only the random variable Z. Because of this, it is not possible to derive an exact expression for the c.d.f. of $CPS_{\delta,UK}$ ($F_{CPS_{\delta,UU}}$) using the *distribution function method* or the *conditioning-unconditioning method* like was done for cases UU and KU. Given this, in this section, it is presented equations to calculate exact values of $F_{CPS_{\delta,UK}}$ with a search algorithm and also a formula to calculate approximate values of $F_{CPS_{\delta,UK}}$.

As noted in the introduction, the search algorithm provides extremely accurate result, since the user can specify the desired accuracy in order to find a result that is exact up to a specified number of significant digits. In Appendix B, the search algorithm is explained in more detail. For now, the expression of the c.d.f. of $CPS_{\delta,UK}$ ($F_{CPS_{\delta,UK}}$) can be expressed as:

$$F_{CPS_{\delta,UK}}(t) = P\big(CPS_{\delta,UK} \le t\big) = \begin{cases} 0, & t \le 0.0027 \\ \Phi(z_2) - \Phi(z_1), & 0.0027 < t < 1 \\ 1, & t \ge 1 \end{cases},$$

(36)

where z_1 and z_2 are the only two solutions, according to Equation (12), of

$$1 - \left[\Phi\left(\frac{Z}{\sqrt{m}} + L - \delta\sqrt{n}\right) - \Phi\left(\frac{Z}{\sqrt{m}} - L - \delta\sqrt{n}\right)\right] = t,$$

for Z, being $z_1 < \delta\sqrt{mn} < z_2$. Note again that z_1 and z_2 can be precisely found using the search algorithm called the secant method (see Appendix B).

Equation (36) can be explained as follows: there are only two solutions for Z of $CPS_{\delta,UK} = t$ (from Equation (12), see also figure 5 for $CPS_{0,UK}$, i.e., $CFAR_{UK}$) because $CPS_{\delta,UK}$ as a function of Z is decreasing on $(-\infty,\ \delta\sqrt{mn}]$ and increasing on $[\delta\sqrt{mn}, \infty)$ so that $CPS_{\delta,UK}$ varies in the interval $(min\big(CPS_{\delta,UK}\big) = 0.0027, 1\big]$. The value of $CPS_{\delta,UK}$ tends to one when Z tends to $-\infty$ or ∞. Thus, the probability that $CPS_{\delta,UK} < t$ (i.e., the c.d.f. of $CPS_{\delta,UK}$) is the probability that Z belongs to the interval between z_1 and z_2.

The approximate formula for $F_{CPS_{\delta,UK}}$ can be derived starting from Equation (12) also. The details of this derivation are in Appendix C. The final expression is:

$$F_{CPS_{\delta,UK}}(t) = P\left(CPS_{\delta,UK} \le t\right)$$
$$\approx \Phi\left(\delta\sqrt{mn} + \sqrt{\frac{mL^2}{F_{\chi_1^2}^{-1}(1-t)} - 1}\right) - \Phi\left(\delta\sqrt{mn} - \sqrt{\frac{mL^2}{F_{\chi_1^2}^{-1}(1-t)} - 1}\right). \quad (37)$$

When $\delta = 0$, i.e., for the conditional false alarm rate ($CPS_{0,UK} = CFAR_{UK}$), an even simpler approximate formula for $F_{CFAR_{UK}}$ can be derived. The final result is below (for more details on this derivation, see Appendix C).

$$F_{CFAR_{UK}}(t) = P(CFAR_{UK} \le t) \approx F_{\chi_1^2}\left(m\left(\frac{L^2}{F_{\chi_1^2}^{-1}(1-t)} - 1\right)\right). \quad (38)$$

Using equation (18), the approximate formula for the c.d.f. of the general conditional average run length ($CARL_{\delta,UK}$) and the approximate formula for the c.d.f of the in-control average run length ($CARL_{0,UK}$) can be derived, respectively, as:

$$F_{CARL_{\delta,UK}}(w) = P\left(CARL_{\delta,UK} \le w\right)$$
$$\approx \Phi\left(\delta\sqrt{mn} + \sqrt{\frac{mL^2}{F_{\chi_1^2}^{-1}\left(1-\frac{1}{w}\right)} - 1}\right) - \Phi\left(\delta\sqrt{mn} - \sqrt{\frac{mL^2}{F_{\chi_1^2}^{-1}\left(1-\frac{1}{w}\right)} - 1}\right) \quad (39)$$

and

$$F_{CARL_{0,UK}}(w) = P\left(CARL_{0,UK} \le w\right) \approx F_{\chi_1^2}\left(m\left(\frac{L^2}{F_{\chi_1^2}^{-1}\left(1-\frac{1}{w}\right)} - 1\right)\right) \quad (40)$$

The c.d.f.'s and p.d.f.'s of the $CFAR_{UK}$ and $CARL_{0,UK}$ for $m = \{10,20,50,200,500\}$, $n = 5$ and $\alpha = 0.0027$ are shown in Figures 14, 15, 16 and 17. Exact values are calculated according to Equation (36), in grey, and the approximate values are calculated according to Equations (37), (38), (39) and (40) in black. Note that the approximate formulas provides very accurate results.

From these figures, it is evident the difference between the shape of the distributions (either way cdf´s and pdf´s) in case UK and the other cases (UU and KU). However, the impact of m on the $CFAR_{UK}$ and $CARL_{0,UK}$ distributions is also clear. Similarly, to the other cases, when m gets larger (such as $m = 500$), the c.d.f. curves are much “closer” to the vertical line of the nominal value of 0.0027 or 370.4. Note that the p.d.f´s in case UK have a cut (“break”) in the nominal values, differently from the p.d.f.´s in cases UU and KU.

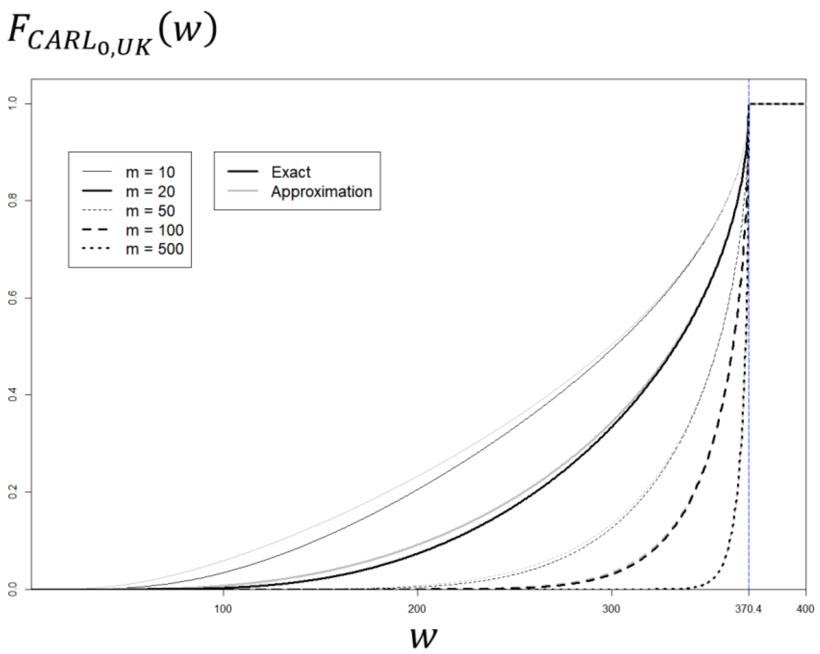


Figure 14. c.d.f. *of* $\boldsymbol{CARL_{0,UK}}$ for $\boldsymbol{n = 5}$, $\boldsymbol{m = 10, 20, 50, 100, 500}$ and $\boldsymbol{\alpha = 0.0027}$ (i.e., $\boldsymbol{L = 3}$).

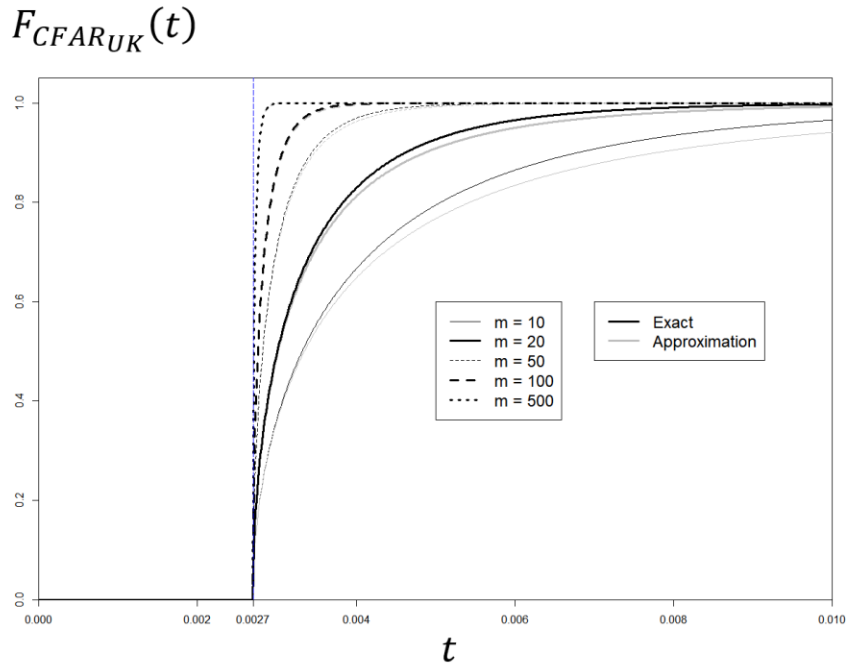


Figure 15. c.d.f. *of* $\boldsymbol{CFAR_{UK}}$ for $\boldsymbol{n = 5}$, $\boldsymbol{m = 10, 20, 50, 100, 500}$ and $\boldsymbol{\alpha = 0.0027}$ (i.e., $\boldsymbol{L = 3}$).

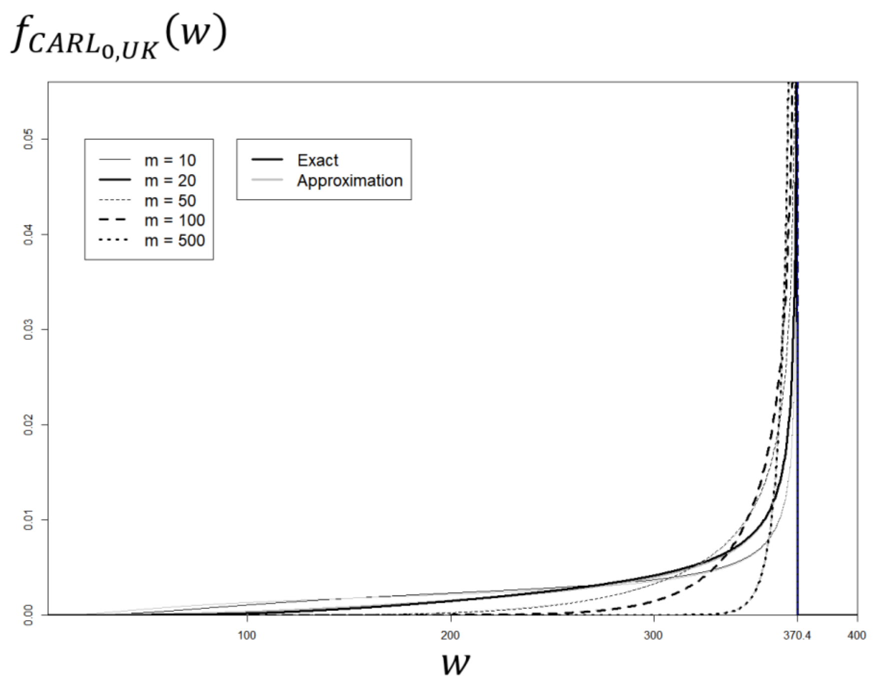


Figure 16. p.d.f. *of* $\boldsymbol{CARL_{0,UK}}$ for $\boldsymbol{n = 5, m = 10, 20, 50, 100, 500}$ and $\boldsymbol{\alpha = 0.0027}$ (i.e., $\boldsymbol{L = 3}$).

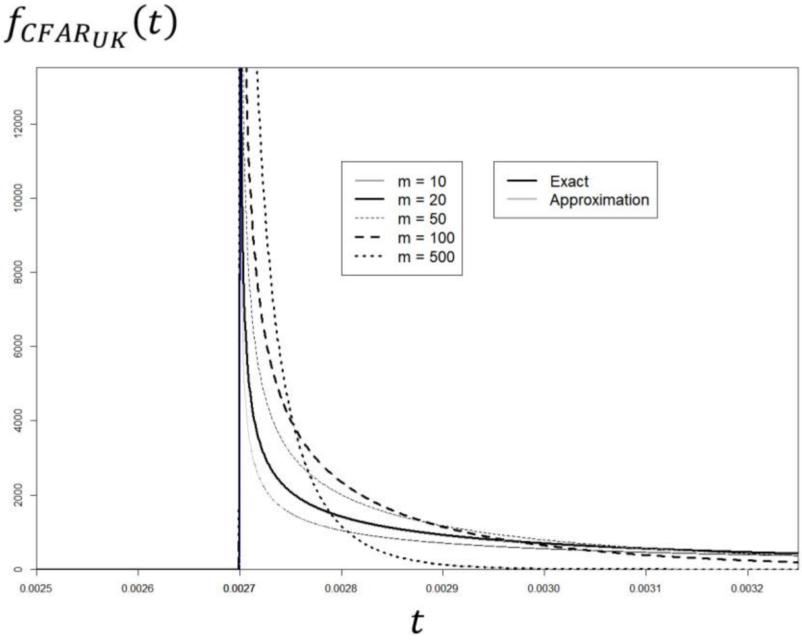


Figure 17. p.d.f. *of* $\boldsymbol{CFAR_{UK}}$ for $\boldsymbol{n=5}$, $\boldsymbol{m=10,20,50,100,500}$ and $\boldsymbol{\alpha=0.0027}$ (i.e., $\boldsymbol{L=3}$).

3.4. The mean and standard deviation of $CFAR$ and $CARL_0$

As explained in Chapters 1 and 2, the mean of the in-control conditional average run length ($CARL_0$) is denoted as the unconditional in-control average run length (ARL_0) and it was one of the most used in-control performance measure of a control chart when parameters are estimated. Given that $\boldsymbol{CARL_0}$ is a non-negative random variable, its mean and standard deviation can be easily calculated using its c.d.f. derived in the previous sections of this chapter. The ARL_0 (i.e., the $E[CARL_0]$) can be expressed as:

$$E[CARL_0] = ARL_0 = \int_0^1 \left(1 - F_{CARL_0}(w)\right) dw. \tag{41}$$

Moreover, the standard deviation of $CARL_0$ (denoted here as $SDARL_0$) can be expressed as

$$SDARL_0 = \sqrt{V[CARL_0]} = \sqrt{E[CARL_0^2] - (E[CARL_0])^2} \tag{42}$$

where,

$$E[CARL_0^2] = 2\int_0^1 w\left(1 - F_{CARL_0}(w)\right)dw \quad (43)$$

Equation (43) is a special case of the expression $k\int_0^\infty w^{k-1}\big(1 - F(w)\big)dw$ which is the k-th central moment of a nonnegative continuous random variable in terms of its c.d.f. This formula is given in Feller (1966), Hong (2012, 2015) and Nadarajah and Mitov (2003), the latter also derived the multivariate analogue. However, an analogue of Formula (43) for a *discrete non-negative integer-valued* random variable was not formally proved in the literature. Thus, as an extension of the present work, we derived this analogue formula and published it in one of the most recognized journals in statistics: The American Statistician (see Annex C).

For the 3 cases (UU, KU and UK), the unconditional in-control average run length (i.e., $E[CARL_0] = ARL_0$) and the standard deviation of the $CARL_0$ [i.e., the $SDARL_0 = SD(CARL_0)$] are presented in Table 1 for some values of m and n and $\alpha = 0.0027$ (i.e., $L = 3$). The difference in the performance (in terms of $SDARL_0$ and ARL_0) between these cases are significant: in the cases UU and KU, the ARL_0 for $n = 5$ are always larger than its nominal value (370.4), and in case UK (and also for cases UU when $n = 9$), the ARL_0 values are always smaller than 370.4. For case UK, ARL_0 and $SDARL_0$ results are invariant in respect to the values of for n [what is expected given Eq. (15)]. Between cases UU and KU, surprisingly, the ARL_0 and $SDARL_0$ are larger in case KU than in Case UU. This is an interesting behavior because only one parameter is estimated in case KU (contrasting with the two parameters estimations in case UU, which would make one to think that the variability in case UU would be larger, but it is not).

As the amount of Phase I samples (m) increases, the ARL_0 in cases UU and KU decreases and converges to 370.4 (for $n = 5$), while, in case UK (or in cases UU for $n = 9$), the ARL_0 increases and converges to 370.4. Note that the convergence of the $ARL_{0,UK}$ is faster than the convergence of $ARL_{0,UU}$ and $ARL_{0,KU}$, i.e., the $ARL_{0,UK}$ reaches a value close to the nominal 370.4 with much less Phase I data than the $ARL_{0,UU}$ and $ARL_{0,KU}$. Furthermore, according to the values of the $SDARL_0$, the variability of the $CARL_{0,UU}$ and $CARL_{0,KU}$ are much larger than that for the $CARL_{0,UK}$ (i.e., $SDARL_{0,KU} > SDARL_{0,UU} \gg SDARL_{0,UK}$). The values of ARL_0 and $SDARL_0$ in Table 1 are exact (calculated numerically using Equations (41) and (42)).

Table 1 - $\boldsymbol{SDARL_0}$ and $\boldsymbol{ARL_0}$ values for cases UU, KU and UU for several values of values of *m* and *n* and $\boldsymbol{\alpha = 0.0027}$ (i.e., a nominal $\boldsymbol{ARL_0}$ of 370.4)

		$ARL_0 = E(CARL_0)$			$SDARL_0 = SD(CARL_0)$		
m	***n***	**Case UU**	**Case KU**	**Case UK**	**Case UU**	**Case KU**	**Case UK**
	3	605.6	748.0	311.0	1565.1	1975.0	61.7
20	***5***	422.4	511.4	311.0	460.3	550.9	61.7
	9	360.3	432.5	311.0	240.0	275.1	61.7
	3	536.9	637.3	319.7	964.3	1159.2	54.6
25	***5***	407.5	477.5	319.7	367.9	425.8	54.6
	9	359.6	418.9	319.7	207.1	231.0	54.6
	3	436.3	477.5	340.9	388.1	425.8	35.1
50	***5***	384.2	418.9	340.9	214.1	231.0	35.1
	9	361.6	393.5	340.9	137.3	144.7	35.1
	3	399.8	418.9	354.2	220.3	231.0	20.7
100	***5***	375.9	393.5	354.2	139.2	144.7	20.7
	9	364.8	381.7	354.2	94.2	96.5	20.7
	3	379.4	385.6	364.6	111.8	113.6	7.9
300	***5***	371.9	377.9	364.6	76.3	77.3	7.9
	9	368.2	374.1	364.6	53.3	53.7	7.9
	3	373.0	374.9	368.6	58.7	59.0	2.5
1000	***5***	370.8	372.6	368.6	41.1	41.2	2.5
	9	369.7	371.5	368.6	28.9	29.0	2.5

3.5. Prediction Bounds for $CFAR$ and $CARL_0$

Since $CFAR$ and $CARL_0$ are both random variables that depend on the parameter estimates, it will be of interest to the practitioner to know how far they can be from the nominal desired values. For example, it is of interest to know, in a given Phase II application, what value (named α_p) will be an upper bound to the $CFAR$, with a certain (high) probability $(1 - p)$. This upper prediction bound to the $CFAR$ will provide a lower bound to the $CARL_0$, both of which can be useful to the practitioner in understanding the in-control chart performance under estimated parameters. Put another way, for a given m and n, it is of practical interest to find the value of $CFAR, \alpha_p$, such that

$$P\left(CFAR > \alpha_p\right) = p \tag{44}$$

This means that the required α_p is the $(1-p)$-quantile of the distribution of $CFAR$. According to Equation (17), this problem is equivalent to finding the value $1/\alpha_p$ such that $P\left(CARL_0 \leq 1/\alpha_p\right) = p$, that is, to finding the p-quantile of $CARL_0$.

In Case UU, to find α_p, one just must, in equation (26), replace t by α_p and make it equal to $1-p$ and solve for α_p. I.e., $\alpha_{p,UU}$ is the solution to the equation

$$\int_{-\infty}^{\infty} F_{\chi^2_{m(n-1)}}\left(\frac{m(n-1)F^{-1}_{\chi^2_{1,\left[\frac{z^2}{m}\right]}}\left(1-\alpha_{p,UU}\right)}{L^2}\right) f_Z(z)dz = p, \tag{45}$$

for given values of L, m, n and p. It is not possible to obtain a closed-form solution to $\alpha_{p,UU}$, because Equation (45) involves an integral over the distribution of Z which can´t be solved analytically [see also Equation (26)]. However, it can be solved numerically in a straightforward way, using a simple search method (like the Secant Method) since $F_{CFAR_{UU}}(t)$ is a monotonic increasing function of t. Moreover, a simple approximate expression for $\alpha_{p,UU}$, shown in equation (46) next, can be obtained from Equations (26) using the one-step Taylor approximation and an approximation for the c.d.f. of a non-central chi-square distribution derived by Cox and Reid (1987). Details of the derivation of this approximation can be found in Appendix D.

$$\alpha_{p,UU} \approx 1 - F_{\chi^2_1}\left(L^2 \frac{F^{-1}_{\chi^2_{m(n-1)}}(p)}{(m+1)(n-1)}\right). \tag{46}$$

In Case KU, using Equation (33), an exact expression of $\alpha_{p,KU}$ can be obtained by solving the following equation for $\alpha_{p,KU}$:

$$F_{\chi^2_{m(n-1)}}\left(m(n-1)\left(-\frac{\Phi^{-1}\left(\frac{\alpha_{p,KU}}{2}\right)}{L}\right)^2\right)=p \tag{47}$$

Rearranging the terms in Equation (47), $\alpha_{p,KU}$ can be expressed as

$$\alpha_{p,KU}=2\Phi\left(-L\sqrt{\frac{F^{-1}_{\chi^2_{m(n-1)}}(p)}{m(n-1)}}\right). \tag{48}$$

In Case UK, one can use Equation (36) for $\delta=0$ and find $\alpha_{p,UK}$ by numerically solving the following system of equations:

$$\begin{cases}\Phi(z_2)-\Phi(z_1)=1-p\\ \Phi\left(\frac{z_i}{\sqrt{m}}+L\right)-\Phi\left(\frac{z_i}{\sqrt{m}}-L\right)=1-\alpha_{p,UK}, \quad i=1,2\end{cases} \tag{49}$$

for, z_2, z_1 and $\alpha_{p,UK}$. Note again that z_2, z_1 and $\alpha_{p,UK}$ can be found via a search algorithm, such as the secant method. From Equation (38) is also possible to derive an approximate formula for $\alpha_{p,KU}$ as show below (for more details see Appendix D):

$$\alpha_{p,UK}\approx 1-F_{\chi^2_1}\left(\frac{L^2}{\frac{F^{-1}_{\chi^2_1}(1-p)}{m}+1}\right) \tag{50}$$

Table 2 shows the values of α_p and $1/\alpha_p$ for $p=0.05$ (i.e., the 0.95 quantile of $CFAR$ and the 0.05 quantile of $CARL_0$) and $p=0.1$ (the 0.9 quantile of $CFAR$ and the 0.1 quantile of $CARL_0$) for some values of m and n in Cases UU, KU and UK. For Cases UU and UK, the exact values were calculated numerically using Equations (45) and (49) and a search method (respectively) and — in bold — the values obtained using the simple approximations given by Equations (46) and (50). We considered $\alpha=0.0027$ ($L=3$). Table 1 shows that when m and/or n are small, the values of $CFAR$ that are exceeded only with a probability of 5% or 10% are much higher than the desired α. For example, in Case UU, for $m=25$ and $n=5$ (values suggested in many manuals and textbooks, see Montgomery, 2009), $\alpha_{0.05}=$

0.0098 — more than 3 times the nominal false alarm rate of 0.0027. This means that, for a small amount of Phase I data, such as $m = 25$ and $n = 5$, if the 3-Sigma limits are used, the spreads of $CFAR$ and $CARL_0$ are large, meaning that the realization of these random variables may be very different from the nominal values of the false alarm rate or average run length (0.0027 and 370.4, respectively). Also note that the approximation works well for $m \geq 50$ in Cases UU and UK.

In cases UU and KU, the values of α_p and $1/\alpha_p$ are not so close to the nominal ones, even for large values of m. For example, for $m = 300$ and $n = 5$, in both cases (UU and KU), $\alpha_{0.05} \geq 0.0037$ (a value more than 37% larger than 0.0027) and $1/\alpha_{0.05} < 267$ (a value 30% smaller than the desired nominal value of 370.4). It is also interesting to note that the higher quantiles of $CFAR$ (or lower quantiles of $CARL_0$) in Case KU are smaller (or larger, for $CARL_0$) compared to those in Case UU. Also, for Case UK, the quantiles of $CFAR$ and $CARL_0$ are invariant with respect to the values of n (note that for the $\bar{X}$ chart n must be greater than 1) and, compared to cases UU and KU, are much closer to the nominal desired values (0.0027 and 370.4 respectively). These results open the question as to the minimum number of Phase I samples (n) required to guarantee a desired quantile for $CFAR$ and $CARL_0$. In the next chapter this problem is addressed.

Table 2. The 0.95 quantile of $\boldsymbol{CFAR}$, the 0.05 quantile of $\boldsymbol{CARL_0}$ ($\boldsymbol{p = 0.5}$), the 0.9 quantile of $\boldsymbol{CFAR}$ and the 0.1 quantile of $\boldsymbol{CARL_0}$ ($\boldsymbol{p = 0.1}$) for $\boldsymbol{\alpha = 0.0027}$ ($\boldsymbol{L = 3}$).

Case	p	n	α_p (m = 25)		$1/\alpha_p$ (m = 25)		α_p (m = 50)		$1/\alpha_p$ (m = 50)	
Case UU	$p = 0.05$	5	0.0098	**0.0094**	102.4	**106.3**	0.0066	**0.0064**	152.5	**155.4**
		10	0.0071	**0.0067**	140.1	**149.7**	0.0052	**0.0050**	193.6	**199.9**
		20	0.0060	**0.0054**	167.8	**186.3**	0.0045	**0.0043**	223.1	**234.6**
		25	0.0057	**0.0051**	174.5	**196.6**	0.0043	**0.0041**	230.5	**244.0**
	$p = 0.10$	5	0.0078	**0.0076**	128.8	**131.7**	0.0055	**0.0055**	180.2	**182.4**
		10	0.0060	**0.0058**	167.2	**173.8**	0.0046	**0.0045**	218.8	**223.3**
		20	0.0051	**0.0048**	195.3	**207.1**	0.0041	**0.0039**	246.1	**253.5**
		25	0.0049	**0.0046**	202.4	**216.2**	0.0040	**0.0038**	252.9	**261.4**
		$m \rightarrow$	**100**				**300**			
	p	n	α_p		$1/\alpha_p$		α_p		$1/\alpha_p$	
	$p = 0.05$	5	0.0050	**0.0049**	200.7	**202.3**	0.0038	**0.0038**	262.5	**263.1**
		10	0.0042	**0.0041**	239.5	**242.9**	0.0034	**0.0034**	292.0	**293.0**
		20	0.0037	**0.0037**	266.7	**272.5**	0.0032	**0.0032**	311.9	**313.5**
		25	0.0037	**0.0036**	273.6	**280.3**	0.0032	**0.0031**	316.9	**318.7**
	$p = 0.10$	5	0.0044	**0.0044**	226.3	**227.6**	0.0035	**0.0035**	281.8	**282.2**
		10	0.0038	**0.0038**	260.7	**263.2**	0.0033	**0.0033**	306.5	**307.2**
		20	0.0035	**0.0035**	284.3	**288.2**	0.0031	**0.0031**	322.9	**324.0**
		25	0.0034	**0.0034**	290.3	**294.7**	0.0031	**0.0030**	327.0	**328.2**
Case KU		$m \rightarrow$	**25**				**50**			
	p	n	α_p		$1/\alpha_p$		α_p		$1/\alpha_p$	
	$p = 0.05$	5	0.0081		123.6		0.0059		168.7	
		10	0.0057		176.3		0.0046		218.1	
		20	0.0045		221.1		0.0039		256.7	
		25	0.0043		233.8		0.0037		267.1	
	$p = 0.10$	5	0.0065		154.4		0.0050		198.7	
		10	0.0049		205.8		0.0041		244.0	
		20	0.0041		246.7		0.0036		277.7	
		25	0.0039		258.0		0.0035		286.6	
		$m \rightarrow$	**100**				**300**			
	p	n	α_p		$1/\alpha_p$		α_p		$1/\alpha_p$	
	$p = 0.05$	5	0.0047		211.3		0.0037		267.1	
		10	0.0039		254.1		0.0034		297.6	
		20	0.0035		285.5		0.0031		318.5	
		25	0.0034		293.7		0.0031		323.8	
	$p = 0.10$	5	0.0042		238.0		0.0035		286.6	
		10	0.0036		275.5		0.0032		312.1	
		20	0.0033		302.0		0.0030		329.2	
		25	0.0032		308.9		0.0030		333.5	
Case UK		$m \rightarrow$	**25**				**50**			
	p	n	α_p		$1/\alpha_p$		α_p		$1/\alpha_p$	
	0.05	$n \geq 1$	0.0049	**0.0052**	204.1	**191.5**	0.0038	**0.0038**	265.9	**260.4**
	0.10		0.0042	**0.0044**	237.1	**228.6**	0.0034	**0.0035**	290.8	**287.5**
		$m \rightarrow$	**100**				**300**			
	p	n	α_p		$1/\alpha_p$		α_p		$1/\alpha_p$	
	0.05	$n \geq 1$	0.0032	**0.0032**	310.5	**308.6**	0.0029	**0.0029**	348.3	348.0
	0.10		0.0031	**0.0031**	326.3	**325.3**	0.0028	**0.0028**	354.6	354.4

Observation: For Case UU and UK, the values in bold were obtained using the approximate formulas, the other values are exact

4
Number of Phase I samples required for a guaranteed in-control performance

In the context of parameter estimation and control charting, another relevant question for the practitioner is the amount of the Phase I data, without adjusting the control limits, that can ensure a "satisfactory" performance of the chart both in terms of in-control robustness. Epprecht et al. (2015), formulated this problem and derived the minimum number m of Phase I reference samples that guarantees with a specified high probability $1-p$ (say, 0.9), that the $CFAR$ does not exceed a nominal α by more than a given percentage ε (e.g. 20%) for the S and S^2 charts. As noted in previous chapters, this is called the Exceedance Probability Criterion (*EPC*). The quantity ε provides some flexibility, as an allowance, for the user to guarantee the minimum performance (lower bound) for the *CFAR*. Indeed, for all cases, it would be impossible to guarantee a high probability that $CFAR$ does not exceed the exact nominal values α (i.e., for $\varepsilon = 0$), which is its median (the 0.5-quantile) when *m* is infinite and thus cannot be a different quantile of it (other than the median), no matter the number of Phase I samples for cases UU and KU. This behavior can be verified in the *CFAR* c.d.f.´s figures for Cases UU and KU in Chapter 2. For case UK, since the minimum possible value of *CFAR* is α, $CFAR$ will be equal or exceed the nominal α with 100% of probability no matter the number of Phase I samples.

Thus, for $\varepsilon > 0$, the following formulation to the $\bar{X}$ chart, in cases UU, KU and UK is considered: Given the values of n, α, ε and p, find the minimum number of in-control Phase I samples, m, such that

$$P(CFAR \leq (1+\varepsilon)\alpha) = 1-p. \quad (51)$$

This problem is like the one in section 2.5, with the difference that now α_p is given and is equal to a tolerated upper bound to the false alarm rate (that is, $\alpha_p =$

$\alpha_{TOL} = (1+\varepsilon)\alpha$) and m is the unknown that needs to be found. Note that, since m is an integer, a perfect match is generally not possible, so, re-stating the problem, m should be the smallest integer such that $P(CFAR \leq (1+\varepsilon)\alpha) \geq 1-p$ is true. Also, note that this problem is equivalent to finding the smallest m such that $P(CARL_0 \leq 1/((1+\varepsilon)\alpha) \leq p$ is true.

In order to find the value of m in all cases (UU, KU and UK), an exact formula is not available. For cases UU and KU, this is because the c.d.f's of $CFAR$ involve a quantile of a chi-square variable whose number of degrees of freedom is also a function of the unknown m. For case UK, this is because the c.d.f. of *CFAR* can´t be expressed in a closed form expression (as shown on Chapter 3). However, for all cases, m can be found using a simple search method (as the Secant Method, for example) since $F_{CFAR}(t)$ is a monotonic increasing function of m. Basically, this means that for Cases UU, KU and UU, we need to solve, respectively,

$$\int_{-\infty}^{\infty} F_{\chi^2_{m(n-1)}}\left(\frac{m(n-1)F^{-1}_{\chi^2_{1,\left[\frac{z^2}{m}\right]}}(1-(1+\varepsilon)\alpha)}{L^2}\right) f_Z(z)dz = p, \tag{51}$$

$$F_{\chi^2_{m(n-1)}}\left(m(n-1)\left(-\frac{\Phi^{-1}\left(\frac{(1+\varepsilon)\alpha}{2}\right)}{L}\right)^2\right) = p \tag{52}$$

and

$$\begin{cases} \Phi(z_2) - \Phi(z_1) = 1 - p \\ \Phi\left(\frac{z_i}{\sqrt{m}} + L\right) - \Phi\left(\frac{z_i}{\sqrt{m}} - L\right) = 1 - (1+\varepsilon)\alpha, \quad i = 1,2 \end{cases} \tag{53}$$

for m, z_2 and z_1. For case UK, using Equations (38) and (51), rearranging the terms, it is also possible to derive an approximate formula for the minimum m, which is

$$m \approx \left\lceil \frac{F^{-1}_{\chi^2_1}(1-p)}{\frac{L^2}{F^{-1}_{\chi^2_1}(1-(1+\varepsilon)\alpha)} - 1} \right\rceil, \tag{54}$$

where $\lceil a \rceil$ denotes the smallest integer greater or equal to a.

Table 3, for cases UU, KU and UK shows the exact minimum number of in-control Phase I samples, m, for $\varepsilon = 0.1, 0.2, 0.3, 0.4, 0.5$, $\alpha = 0.0027$, $p = 5\%, 10\%, 15\%$, and n = 5, 10, 20 and 25. For case UK, Table 3 also shows the results for the approximate formula (54). As it can be seen, for small values of n, one needs a large number of reference samples (m) to guarantee such conditional performance for most of the cases. Also note that, in case KU, m is invariant with respect of n and the approximate formula (54) works well.

Table 3. Minimum number of in control Phase I samples, m, required for $P(CFAR \leq (1+\varepsilon)\alpha) = 1 - p$ or $P(CARL_0 \leq 1/(1+\varepsilon)\alpha) = p$ with $\alpha = 0.0027$ ($L = 3$).

			$\varepsilon = 10\%$			$\varepsilon = 20\%$			$\varepsilon = 30\%$			$\varepsilon = 40\%$			$\varepsilon = 50\%$		
Case	n	$p \rightarrow$	0.05	0.10	0.15	0.05	0.10	0.15	0.05	0.10	0.15	0.05	0.10	0.15	0.05	0.10	0.15
UU	5		3687	2285	1536	1029	649	446	507	324	226	314	203	144	219	144	103
	10		1701	1077	742	492	321	229	250	167	122	159	108	80	114	78	59
	20		871	571	409	270	185	138	145	102	77	97	68	52	72	51	39
	25		717	477	346	230	160	120	126	89	69	85	61	47	64	46	36
KU	5		3588	2185	1435	975	595	393	468	287	190	283	174	116	194	120	80
	10		1595	971	638	433	265	175	208	128	85	126	78	52	87	53	36
	20		756	460	303	206	126	83	99	61	40	60	37	25	41	26	17
	25		598	365	240	163	100	66	78	48	32	48	29	20	33	20	14
UK	$n \geq 2$	Exact	191	135	103	97	68	53	65	46	36	50	35	27	40	28	22
		Approx.	**195**	**138**	**105**	**101**	**71**	**54**	**69**	**49**	**37**	**53**	**37**	**29**	**43**	**31**	**24**

The results from Table 3 are quite interesting. We see that in some cases (for example, when $\varepsilon = 0.1$ and n = 5), the minimum numbers of reference samples required are much larger than the 25 or 30 subgroups, which are the numbers usually proposed in most manuals and textbooks (see Montgomery, 2009); they can also be larger than the 200 or 300 samples proposed by authors who focused on the unconditional ARL_0 (see Quesenberry, 1993 and others) and even larger than the recent number proposed by Saleh et al. (2015), who focused on the standard deviation of $CARL_0$ as an additional performance metric (they recommended using $m = 1200$ when $\alpha = 0.0027$ is used). One can see, as might be expected, that in Case UU, more Phase I samples are needed than in Case KU.

It is also interesting to note that the minimum total amount of data ($m \times n$) required has a different behavior in each case (KU and UU) as n increases. For Case UU, when n increase, for many of the situations, the total amount of data needed increases (although doesn´t increase much), while in the Case KU, for the majority of the situations, the total amount of data decreases with n (again, although doesn´t decrease much). This means that, for Case UU, in most of the situations, increasing n, also increases the cost to design the control chart, since a few more Phase I observations will need to be collected. Note that, in case UU, for $\varepsilon = 20\%$ and $p = 0.05$ the total amount of Phase I data required for $n = 5$ is 5145 and for $n = 25$ is 5750 (both large values). For case KU, the total amount of Phase 1 data is much smaller compared to the other cases.

Table 4. Minimum total amount of Phase I observations, $(m \times n)$, required for $P(CFAR \leq (1 + \varepsilon)\alpha) = 1 - p$ or $P(CARL_0 \leq 1/(1 + \varepsilon)\alpha) = p$ with $\alpha = 0.0027$ $(L = 3)$.

			$\varepsilon = 10\%$			$\varepsilon = 20\%$			$\varepsilon = 30\%$			$\varepsilon = 40\%$			$\varepsilon = 50\%$		
Case	n	$p \rightarrow$	0.05	0.10	0.15	0.05	0.10	0.15	0.05	0.10	0.15	0.05	0.10	0.15	0.05	0.10	0.15
UU	5		18435	11425	7680	5145	3245	2230	2535	1620	1130	1570	1015	720	1095	720	515
	10		17010	10770	7420	4920	3210	2290	2500	1670	1220	1590	1080	800	1140	780	590
	20		17420	11420	8180	5400	3700	2760	2900	2040	1540	1940	1360	1040	1440	1020	780
	25		17925	11925	8650	5750	4000	3000	3150	2225	1725	2125	1525	1175	1600	1150	900
KU	5		17940	10925	7175	4875	2975	1965	2340	1435	950	1415	870	580	970	600	400
	10		15950	9710	6380	4330	2650	1750	2080	1280	850	1260	780	520	870	530	360
	20		15120	9200	6060	4120	2520	1660	1980	1220	800	1200	740	500	820	520	340
	25		14950	9125	6000	4075	2500	1650	1950	1200	800	1200	725	500	825	500	350
UK	$n \geq 2$	Exact	382	270	206	194	136	106	130	92	72	100	70	54	80	56	44
		Approx.	**390**	**276**	**210**	**202**	**142**	**108**	**138**	**98**	**74**	**106**	**74**	**58**	**86**	**62**	**48**

In case the required m for the tolerated α_{TOL} value and the specified p is not feasible and relaxing the value of either ε or p or both is unacceptable on practical grounds, a possible solution is to change the value of the control limit factor L (instead of using $L = 3$ – the most common 3-sigma limits), given a fixed value of m and n, at hand, in order to satisfy the exceedance probability criterion in the in-control situation. This is discussed in the next chapter.

5
Adjustment of the limits for a guaranteed conditional in-control performance

In the previous chapter, we saw that the minimum numbers of reference samples required to guarantee some conditional in-control performances can be very large and may be infeasible in many practical situations. Given this practical hurdle, in this chapter, for all cases (UU, KU and UK), it is presented the exact adjusted control limits for the $\bar{X}$ chart (for any values of m and n) that limit to a low value, p, the probability that the conditional false alarm rate ($CFAR$) exceeds a tolerated value (α_{tol}) in the spirit of the *EPC*. Remember that α_{tol} is greater than α by a percentage ε. So basically, the idea is to replace the limit factor L in Equations (1) and (2) by L^*, where L^* represents the value of the control limit factor that guarantees that

$$P(CFAR \geq \alpha_{tol}) = p \text{ or } P(CARL_0 \geq 1/\alpha_{tol}) = 1 - p, \tag{55}$$

for a given value of $\alpha_{tol} = (1 + \varepsilon)\alpha$, m and n. This means that by using L^*, instead of L (the unadjusted limit factor), the user can guarantee that the probability of the conditional false alarm rate being greater than a specified tolerated bound (α_{tol}), is small (p). Note that L is constant and L^* can change according to the values of α, p, ε, m and n.

Adjustment for the case UU is presented in section 5.1. Also in this section, there is a comparison between the adjustment proposed here and other adjustments methods presented in the literature. Adjustments for cases KU and UU are respectively in sections 5.2 and 5.1.

5.1
Adjustment in Case UU

In the literature three main group of methods to adjust the $\bar{X}$ chart control limits in terms of *EPC* are proposed: One is called the “exact methods”, another group is called “approximate methods” and the third, composed of only one method, is called “bootstrap method”. In next sections these adjustments methods will be explained in more details.

5.1.1. Exact Methods

From Equation (55), it is clear that the c.d.f. of $CARL_0$ is needed to apply the *EPC*. So, using the c.d.f. of $CARL_0$ derived in Chapter 2 [Equation 24] and replacing $[(1+\varepsilon)\alpha]^{-1}$ for t, the exact adjusted control limit factor (L^*) can be obtained by solving the following equation for given values of α, m, n, ε and p.

$$\int_{-\infty}^{\infty} F_{\chi^2_{m(n-1)}}\left(\frac{m(n-1)F^{-1}_{\chi^2_{1,}\left[\frac{z^2}{m}\right]}(1-(1+\varepsilon)\alpha)}{L^{*2}}\right)\phi(z)dz = p. \qquad (56)$$

This solution is denoted L^*_{CE} (CE stands for Conditional Exact) and can be obtained with a software like R. Equation (56) is exact, since the formula for the c.d.f. is exact, but the solution L^* must be found using numerically, since there is no closed form solution for the integral in (56). This type of an analysis goes back to Chakraborti (2006), who adjusted the control limits of the $\bar{X}$ chart under the *unconditional perspective*. This is also similar to the adjustment methods proposed by Diko et. al. (2015) in the context of using the $\bar{X}$ and S charts jointly to monitor the mean and Diko et al. (2017) for various spread charts under the *unconditional perspective*. In some of these papers, this method has been referred to as the “numerical method” but the fact is that the method is exact since the expression for the c.d.f. is exact and numerical refers to the solution that is obtained by solving the equation that involves the c.d.f. which involves calculating the integral using some numerical methods. This is indeed the case for many c.d.f.’s of distributions including the one for the celebrated normal distribution. Also, Equation (56) relates to the theory of Tolerance Intervals. Krishnamoorthy and Mathew (2009, p. 30) give an equation which is equivalent to Eq. (56) where the single-sample estimators S and $\bar{X}$ are used instead of S_p and $\bar{\bar{X}}$ respectively. But, they did not make the relationship with the $\bar{X}$ control chart.

Faraz et al. (2017) also proposed an, what they called, “exact method” to adjust the $\bar{X}$ chart, however, their adjustment was not based in the *EPC* and, instead, it was based on the on the *equal-tailed tolerance interval* together with the *Bonferroni Inequality* [see Krishnamoorthy and Mathew (2009, p. 4 and p.10)], which generates wider adjusted control limits if compared to the adjusted limits derived under the *EPC*. The final formula for L^* is denoted by L^*_{CE2} ($CE2$ stands for Conditional Exact 2) and is given by

$$L^*_{CE2} = \frac{t_{\left(1-\frac{1}{p}, n(m-1), z_{1-\frac{(1+\varepsilon)\alpha}{2}}\right)}}{\sqrt{m}}, \tag{57}$$

Where, $t_{\left(1-\frac{1}{p}, n(m-1), z_{1-\frac{(1+\varepsilon)\alpha}{2}}\right)}$ is the $\left(1-\frac{1}{p}\right)$-quantile of a non-central t-student distribution with $n(m-1)$ degrees of freedom and non-centrality parameter $z_{1-\frac{(1+\varepsilon)\alpha}{2}}$, which is the $\left(1-\frac{(1+\varepsilon)\alpha}{2}\right)$-quantile of a standard normal distribution. For more details of the derivation of L^*_{CE2}, see Faraz et al. (2017).

5.1.2. Approximate Methods

Goedhart et al. (2017) derived an approximate formula for L^* by finding an approximate distribution of $CFAR$. This was accomplished by expressing the $CFAR$, using a two-step Taylor approximation, as approximately a linear combination of a scaled chi and a chi-square random variable. Considering the fact that the chi-square part was more dominant, they approximated the distribution of $CFAR$ by the distribution of a $a\frac{\chi_b^2}{b}$ random variable. After this, they applied the Wilson-Hilferty approximation to a chi-square to achieve normality. Finally, to solve the resulting equation for L^*, they applied a one-step Taylor approximation. The final approximate formula for L^* is denoted by L^*_{CA1} ($CA1$ stands for Conditional Approximation 1) and it is given by

$$L^*_{CA1} \approx L + \frac{\Phi^{-1}(1-p) - g(L)}{g'(L)}. \tag{58}$$

Here $g(L)$ and $g'(L)$ are functions of the expectation and the variance of $CFAR$ and their derivatives, respectively. The complete expressions of $g(L)$ and $g'(L)$ are presented in Appendix E. From the expression of *CFAR* in Equation (13), it is possible to derive an alternative and simpler approximate formula for L^*, denoted here by L^*_{CA2}, which is given by

$$L^*_{CA2} \approx \sqrt{m(n-1)\frac{F^{-1}_{\chi^2_{1,\left[\frac{1}{m}\right]}}(1-(1+\varepsilon)\alpha)}{F^{-1}_{\chi^2_{m(n-1)}}(p)}}, \tag{59}$$

where $F^{-1}_{\chi^2_{m(n-1)}}(p)$ denotes the p-quantile of a central chi-square distribution with $m(n-1)$ degrees of freedom and $F^{-1}_{\chi^2_{1,\left[\frac{1}{m}\right]}}(1-(1+\varepsilon)\alpha)$ denotes the $(1-(1+\varepsilon)\alpha)$-quantile of a non-central chi-square distribution with 1 degree of freedom and non-centrality parameter $\frac{1}{m}$. Formula (59) is given by Goedhart et al. (2018), but they started its proof from an already existing result given by Krishnamoorthy and Mathew (2009). Given this, in the Appendix D, we provide a detailed derivation of (59) starting from Equation (13). Note that L^*_{CA2} requires a non-central chi-square quantile, which (as noted in the Introduction) is not tabulated in most textbooks in Statistics and not available in popular software such as Excel, so its calculation will require relatively advanced statistical skills. Given this, and using (15), in this work, we proposed the following even simpler approximate formula for L^* (here denoted by L^*_{CA3}).

$$L^*_{CA3} \approx \sqrt{(n-1)(m+1)\frac{F^{-1}_{\chi^2_1}(1-(1+\varepsilon)\alpha)}{F^{-1}_{\chi^2_{m(n-1)}}(p)}} \tag{60}$$

Note that there is no non-centrality parameter in (60), since $F^{-1}_{\chi^2_1}(1-(1+\varepsilon)\alpha)$ is the $(1-(1+\varepsilon)\alpha)$-quantile of a central chi-square distribution with 1 degree of freedom. Derivation of (60) is also provided in Appendix D.

5.1.3. Bootstrap Method

Saleh et al. (2015) suggested finding the adjusted limit factor L^* under the conditional perspective, using the EPC and the bootstrap approach of Gandy and Kvaløy (2013). In order to do this, the users, with the help of software (like SAS, R, etc.), should generate B bootstrap estimates of the in-control process mean and the standard deviation (μ_k^*, σ_k^*), $k = 1,2,\ldots,B$, with $\mu_k^* \sim N\left(\bar{\bar{X}}, S_p^2/nm\right)$, $\sigma_k^* \sim \sqrt{S_p^2 \frac{\chi_v^2}{v}}$ and $v = m(n-1)$. Note that, with this, the idea is to consider that $\bar{\bar{X}}$ and S_p are respectively the real in-control process mean and standard deviation, which are estimated respectively by μ_k^* and σ_k^* for each k. By considering a very large value for B, let say $B = 1000$, we have "access to the (bootstrap) population of μ_k^* $(\mu_1^*, \mu_2^*, \ldots, \mu_B^*)$ and σ_k^* $(\sigma_1^*, \sigma_2^*, \ldots, \sigma_B^*)$".

Recalling that $Y = m(n-1)\, S_p^2/\sigma_0^2$ and $Z = \left(\frac{\bar{\bar{X}}-\mu_0}{\sigma_0}\right)\sqrt{mn}$, using Equation (13), the $CFAR$ can be written as

$$CFAR\left(\bar{\bar{X}}, S_p\right) = 1 - \Phi\left(\frac{\bar{\bar{X}}-\mu_0}{\sigma_0}\sqrt{n} + L^* \frac{S_p}{\sigma_0}\right) + \Phi\left(\frac{\bar{\bar{X}}-\mu_0}{\sigma_0}\sqrt{n} - L^* \frac{S_p}{\sigma_0}\right). \qquad (61)$$

Considering that $\bar{\bar{X}}$ and $S_p/c_{4,b}$ are respectively the true in-control process mean and standard deviation and μ_k^* and σ_k^* are respectively the estimators $\bar{\bar{X}}$ and S_p (according to the bootstrap method), for each μ_k^* and σ_k^* $(k = 1,2,\ldots,b)$, the user must find the value of L_k^* that satisfies the following equation:

$$1 - \Phi\left(\frac{\mu_k^* - \bar{\bar{X}}}{S_p}\sqrt{n} + L_k^* \frac{\sigma_k^*}{S_p}\right) + \Phi\left(\frac{\mu_k^* - \bar{\bar{X}}}{S_p}\sqrt{n} - L_k^* \frac{\sigma_k^*}{S_p}\right) = (1+\varepsilon)\alpha. \qquad (62)$$

The solution to Equation (62) is given by

$$L_k^* = \frac{S_p\sqrt{F^{-1}_{\chi^2_{1,}\left[\frac{\mu_k^*-\bar{\bar{X}}}{S_p}\sqrt{n}\right]^2}(1-(1+\varepsilon)\alpha)}}{\sigma_k^*}, \quad k = 1,2,\dots,B \tag{63}$$

where $F^{-1}_{\chi^2_{1,}\left[\frac{\mu_k^*-\bar{\bar{X}}}{S_p}\sqrt{n}\right]^2}(1-(1+\varepsilon)\alpha)$ denotes the $(1-(1+\varepsilon)\alpha)$-quantile of the distribution of a non-central qui-square random variable with 1 degree of freedom and non-centrality parameter given by $\left(\frac{\mu_k^*-\bar{\bar{X}}}{S_p}\sqrt{n}\right)^2$. This formula is derived in Appendix F. Note that Saleh et al. (2015b) provided a rather complicated approximate method to find the values of L_k^*. However, we argue that no approximation is needed since one can derive the exact formula for L_k^* shown in Equation (63). Finally, the required L^*, here denoted by L_{boot}^*, is found as the $(1-p)$-quantile of the collection of bootstrap estimators $(L_1^*, L_2^*, \dots, L_B^*)$.

The method described in this section is often named as the *parametric* bootstrap, since the underline distribution is known. This method can also be considered a Monte Carlo simulation.

5.1.4. Adjustment Results and Discussion in Case UU

Table 5 presents the adjusted control limit factors (L^*) obtained under the conditional perspective for $\varepsilon = 0\%$ and $p = 5\%$, i.e., the values of L_{CE}^* , L_{CE2}^*, L_{CA1}^*, L_{CA2}^*, L_{CA3}^* and L_{boot}^* that make $P(CARL_0 \geq 370.4)$, equal or close to 95%. The L_{CE}^*, L_{CE2}^* L_{CA1}^*, L_{CA2}^*, L_{CA3}^* and L_{boot}^* values are obtained using the methods presented in Sections 5.1.1, 5.1.2 and 5.1.3. i.e., the exact adjusted limit factors L_{CE}^* and L_{CE2}^* are calculated according to Equation (56) and Equations (57). The approximate adjusted limit factors L_{CA1}^*, L_{CA2}^* and L_{CA3}^* are calculated according to Formulas (58), (59) and (60). The adjusted limit factor obtained from the bootstrap method, L_{boot}^*, is also calculated considering $B = 1{,}000$ bootstrap simulations implemented in R. Also, for comparison purposes, the first four columns in gray show the results for the unadjusted limit factor ($L^* = L = 3$) and for each L_{CE}^*, L_{CE2}^*, L_{CA1}^*, L_{CA2}^*, L_{CA3}^* and L_{boot}^*. Table 5 shows the exact unconditional ARL_0 value calculated according to Equation (41), the $SDARL_0$, calculated according to

Equation (42), and the exact $P(CARL_0 \geq 370.4)$ calculated according to Equation (25).

From Table 5, excepted for L^*_{CE2}, all methods yield very similar $P(CARL_0 \geq 370.4)$ values, close to the target. I.e., for all values of L^*_{CE} , L^*_{CA1}, L^*_{CA2}, L^*_{CA3} and L^*_{boot}, the probability $P(CARL_0 \geq 370.4)$ is very close to 95%, being the formula for L^*_{CE} (the one proposed in this work) the most precise one, giving a $P(CARL_0 \geq 370.4)$ of exactly 95%. As noted in the introduction, since L^*_{CE2} is not based on the *EPC*, it results are quite different compared to the other methods: this adjustment factor (L^*_{CE2}) is always much larger than the other, generating very larger values of of $P(CARL_0 \geq 370.4)$, not even closer to the goal which is 95% for $p = 0.05$.

Also from Table 5, for all cases and values of L^*_{CE} , L^*_{CE2}, L^*_{CA1}, L^*_{CA2}, L^*_{CA3} and L^*_{boot}, the values of the unconditional ARL_0 are seen to be much larger than 370.4, often more than 3 times larger. This is also true for the $SDARL_0$ values. For example, for $m = 25$, $n = 5$ and $L^*_J = 3.47$, one has $SDARL_0 = 3630.2$ and $ARL_0 = 2552.5$. The large variability is compensated by the large expectation resulting in getting the desired exceedance probability (equal or close to 95%). This means that, despite taking into account the mean and the variability of $CARL_0$, the conditional perspective with the exceedance probability criterion (*EPC*) does not control these popular aspects of the conditional run length distribution.

Since the adjusted limits are wider than the unadjusted limits (note that $L^* > 3$ for all cases in Table 2), this may give the impression that the out-of-control performance may be deteriorated after the adjustments. However, as will be shown in Chapter 6, this is true just for small values of m and n (like $m = 25$ and $n = 5$) in Case UU, but for most of the other cases, the out-of-control performance will be similar to the one with unadjusted limits (especially for $m \geq 50$, $n \geq 5$, $p \geq 0.1$). If the practitioner is still not satisfied with the very large values of ARL_0 and $SDARL_0$ [such concern is evident in Saleh et al. (2015a,b) who focused mainly on the $SDARL_0$ as the performance measure] in the latter case, he/she can increase the value of ε or p (accepting a smaller lowest tolerated bound for $CARL_0$ or a smaller $P(CARL_0 \geq 370.4)$). This will decrease the value of ARL_0 and $SDARL_0$ while the amount of data remains the same. The possibility of this allowance or practical trade-off is a remarkable "feature" of the conditional perspective. To visualize the trade-off, Tables 5 and 6 show the adjusted control limit factors (L^*) under the conditional perspective, respectively, for the pair $\varepsilon = 20\%$ and $p = 5\%$, and the pair $\varepsilon = 20\%$ and $p = 20\%$, i.e., the values of L^* that make, respectively, $P(CARL_0 \geq 308.6) = P(CFAR \leq 0.0031) \cong 95\%$ and $P(CARL_0 \geq 308.6) = P(CFAR \leq 0.0031) \cong 80\%$. Note that in these cases, the values of ARL_0 and $SDARL_0$ are much smaller compared with the values in Table 5 for the same amount of data. For example, for $m = 50$ and $n = 5$, considering the exact method proposed here (L^*_{CE}) in Table 5 (i.e., for $\varepsilon = 0\%$ and $p = 5\%$), the $ARL_0 = 1157.1$ and the $SD(CARL_0) = 807.6$, now, considering the same amount of data ($m = 50$ and $n = 5$), from Table 6 (i.e., for $\varepsilon = 20\%$ and $p = 20\%$), one has $ARL_0 = 561.0$ and $SD(CARL_0) = 338.7$: a reduction of 51.52% in the expectation and 58.06% in the standard deviation. Note that the unconditional ARL_0 is still much larger than the nominal (370.4). Under the ECP, it is unlike that the unconditional ARL_0 will be close to the nominal value (unless ε or p are extremely large, such as $\varepsilon = 50\%$ or $p = 40\%$). This can be seen as a negative point of the *ECP*, since most practitioners are used to work with the nominal values when parameters are known. But keep in mind that the unconditional ARL_0 being larger than the nominal 370.4 is the inevitable counterpart of guaranteeing with a high probability that the $CARL_0$ will not be below the minimum tolerated value. So, given that the OOC performance of the chart is not significantly affected, the large values of ARL_0 is not something necessarily bad.

Table 5. Values of $L^*$when $L^* = L$, $L^* = L^*_{CE}$, $L^* = L^*_{CA1}$, $L^* = L^*_{CA2}$, $L^* = L^*_{CA3}$, $L^* = L^*_{boot}$ and their corresponding ARL_0, $SDARL_0$ and $P\big(CFAR \le \alpha(1+\varepsilon)\big)$ for $L = 3$, $\alpha = 0.0027$, $\varepsilon = 0\%$ and $p = 5\%$ for Case UU

		Unadjusted Limits				Exact Method [Our Proposal]				Aproximate Method 1 [Goedhart et al. (2017)]				Aproximate Method 2 [Goedhart et al. (2018)]			
m	n	$L^* = L$	$ARL_0 = E(CARL_0)$	$SDARL_0 = SD(CARL_0)$	$\varepsilon = 0\%$ $P(CFAR \le 0.0027) = P(CARL_0 \ge 370.4)$	$L^* = L^*_{CE}$	$ARL_0 = E(CARL_0)$	$SDARL_0 = SD(CARL_0)$	$\varepsilon = 0\%$ $P(CFAR \le 0.0027) = P(CARL_0 \ge 370.4)$	$L^* = L^*_{CA1}$	$ARL_0 = E(CARL_0)$	$SDARL_0 = SD(CARL_0)$	$\varepsilon = 0\%$ $P(CFAR \le 0.0027) = P(CARL_0 \ge 370.4)$	$L^* = L^*_{CA2}$	$ARL_0 = E(CARL_0)$	$SDARL_0 = SD(CARL_0)$	$\varepsilon = 0\%$ $P(CFAR \le 0.0027) = P(CARL_0 \ge 370.4)$
	3	3.00	569.5	1045.9	42.70%	3.66	11547.4	86932.1	95.00%	3.64	10670.9	76617.4	94.62%	3.65	11010.6	80579.9	94.77%
25	*5*	3.00	418.5	380.3	40.50%	3.47	2552.5	3630.2	95.00%	3.47	2596.6	3708.6	95.15%	3.45	2394.2	3351.9	94.41%
	9	3.00	364.2	210.3	37.70%	3.35	1278.9	951.7	95.00%	3.36	1309.6	979.0	95.31%	3.33	1168.7	854.9	93.66%
	3	3.00	448.2	401.3	44.46%	3.43	2327.9	3126.7	95.00%	3.43	2356.0	3173.8	95.11%	3.42	2283.7	3053.1	94.82%
50	*5*	3.00	389.1	217.4	42.69%	3.31	1157.1	807.6	95.00%	3.31	1165.5	814.6	95.11%	3.30	1125.9	781.6	94.57%
	9	3.00	363.8	138.3	40.35%	3.23	790.5	348.2	95.00%	3.23	792.7	349.3	95.06%	3.22	760.1	332.4	94.06%
	3	3.00	418.3	278.6	45.35%	3.34	1433.2	1244.6	95.00%	3.34	1446.1	1258.1	95.11%	3.33	1417.4	1228.0	94.86%
75	*5*	3.00	381.6	166.8	43.82%	3.24	879.5	452.5	95.00%	3.24	881.0	453.4	95.03%	3.24	865.5	444.0	94.67%
	9	3.00	365.0	110.4	41.76%	3.18	662.9	224.3	95.00%	3.18	662.3	224.0	94.97%	3.17	647.7	218.2	94.28%
	3	3.00	404.9	223.8	45.91%	3.28	1119.9	761.9	95.00%	3.29	1125.5	766.4	95.08%	3.28	1111.6	755.1	94.88%
100	*5*	3.00	378.3	140.2	44.54%	3.20	759.9	322.0	95.00%	3.20	759.6	321.8	94.99%	3.20	751.7	317.9	94.73%
	9	3.00	365.9	94.6	42.67%	3.15	602.6	170.9	95.00%	3.15	601.4	170.5	94.93%	3.14	593.2	167.8	94.42%
	3	3.00	392.5	170.1	46.61%	3.23	864.1	436.8	95.00%	3.23	864.7	437.2	95.02%	3.23	860.4	434.7	94.91%
150	*5*	3.00	375.3	111.3	45.45%	3.16	648.3	213.1	95.00%	3.16	647.1	212.6	94.94%	3.16	644.3	211.5	94.81%
	9	3.00	367.1	76.3	43.84%	3.12	542.6	117.8	95.00%	3.12	541.4	117.5	94.90%	3.11	537.7	116.5	94.59%
	3	3.00	386.6	142.3	47.04%	3.19	751.3	314.1	95.00%	3.19	750.5	313.7	94.97%	3.19	749.2	313.0	94.93%
200	*5*	3.00	373.9	95.0	46.02%	3.14	593.8	164.5	95.00%	3.14	592.5	164.1	94.92%	3.14	591.4	163.7	94.85%
	9	3.00	367.8	65.7	44.59%	3.10	511.7	97.2	95.00%	3.10	510.7	97.0	94.90%	3.10	508.7	96.5	94.68%
	3	3.00	383.2	124.7	47.34%	3.17	686.7	249.7	95.00%	3.17	685.5	249.2	94.95%	3.17	685.3	249.1	94.94%
250	*5*	3.00	373.2	84.3	46.41%	3.12	560.7	136.7	95.00%	3.12	559.5	136.3	94.91%	3.12	559.1	136.2	94.88%
	9	3.00	368.3	58.6	45.11%	3.09	492.4	82.7	95.00%	3.09	491.5	82.5	94.90%	3.09	490.3	82.2	94.74%

		Unadjusted Limits				Aproximate Method 3 [Our Proposal]				Bootstrap Method [Saleh et al. (2015)]				Exact Method 2 [Faraz et al. (2017)]			
m	n	$L^* = L$	$ARL_0 = E(CARL_0)$	$SDARL_0 = SD(CARL_0)$	$\varepsilon = 0\%$ $P(CFAR \le 0.0027) = P(CARL_0 \ge 370.4)$	$L^* = L^*_{CA3}$	$ARL_0 = E(CARL_0)$	$SDARL_0 = SD(CARL_0)$	$\varepsilon = 0\%$ $P(CFAR \le 0.0027) = P(CARL_0 \ge 370.4)$	$L^* = L^*_{boot}$	$ARL_0 = E(CARL_0)$	$SDARL_0 = SD(CARL_0)$	$\varepsilon = 0\%$ $P(CFAR \le 0.0027) = P(CARL_0 \ge 370.4)$	$L^* = L^*_{CE@}$	$ARL_0 = E(CARL_0)$	$SDARL_0 = SD(CARL_0)$	$\varepsilon = 0\%$ $P(CFAR \le 0.0027) = P(CARL_0 \ge 370.4)$
	3	3.00	569.5	1045.9	42.70%	3.65	11166.2	82383.7	94.84%	3.70	14172.0	121042.0	95.87%	3.85	27414.8	>10000	97.76%
25	*5*	3.00	418.5	380.3	40.50%	3.46	2419.4	3395.9	94.51%	3.49	2746.6	3977.4	95.60%	3.64	4933.0	8261.1	98.49%
	9	3.00	364.2	210.3	37.70%	3.33	1179.4	864.2	93.81%	3.35	1263.8	938.3	94.84%	3.52	2397.5	2006.2	99.15%
	3	3.00	448.2	401.3	44.46%	3.42	2289.9	3063.4	94.84%	3.40	2076.8	2713.0	93.83%	3.57	3921.3	5983.3	98.18%
50	*5*	3.00	389.1	217.4	42.69%	3.30	1128.6	783.8	94.61%	3.32	1203.0	845.9	95.56%	3.44	1852.2	1413.8	98.93%
	9	3.00	363.8	138.3	40.35%	3.22	761.7	333.3	94.12%	3.22	783.3	344.5	94.79%	3.36	1277.5	612.1	99.54%
	3	3.00	418.3	278.6	45.35%	3.33	1419.0	1229.7	94.87%	3.32	1333.9	1141.3	94.00%	3.45	2165.6	2046.1	98.37%
75	*5*	3.00	381.6	166.8	43.82%	3.24	866.4	444.5	94.69%	3.24	878.8	452.0	94.98%	3.35	1299.1	717.4	99.11%
	9	3.00	365.0	110.4	41.76%	3.17	648.4	218.5	94.31%	3.18	673.8	228.6	95.46%	3.29	995.5	361.6	99.67%
	3	3.00	404.9	223.8	45.91%	3.28	1112.3	755.7	94.89%	3.31	1231.2	853.1	96.31%	3.39	1594.8	1161.0	98.48%
100	*5*	3.00	378.3	140.2	44.54%	3.20	752.1	318.1	94.75%	3.20	749.5	316.8	94.66%	3.30	1069.2	481.6	99.21%
	9	3.00	365.9	94.6	42.67%	3.14	593.5	167.9	94.44%	3.15	608.8	173.0	95.35%	3.25	863.5	260.8	99.74%
	3	3.00	392.5	170.1	46.61%	3.23	860.6	434.8	94.92%	3.23	865.3	437.6	95.03%	3.31	1150.4	612.9	98.60%
150	*5*	3.00	375.3	111.3	45.45%	3.16	644.5	211.6	94.81%	3.16	642.9	211.0	94.74%	3.24	860.1	297.3	99.32%
	9	3.00	367.1	76.3	43.84%	3.11	537.9	116.6	94.60%	3.12	543.1	117.9	95.05%	3.21	733.1	172.9	99.81%
	3	3.00	386.6	142.3	47.04%	3.19	749.3	313.1	94.93%	3.19	745.1	311.0	94.79%	3.27	967.4	423.3	98.72%
200	*5*	3.00	373.9	95.0	46.02%	3.14	591.4	163.8	94.85%	3.13	582.4	160.8	94.25%	3.21	761.6	220.6	99.39%
	9	3.00	367.8	65.7	44.59%	3.10	508.7	96.6	94.69%	3.11	525.4	100.3	96.25%	3.18	666.9	132.8	99.84%
	3	3.00	383.2	124.7	47.34%	3.17	685.4	249.1	94.95%	3.17	674.1	244.3	94.45%	3.24	860.1	325.6	98.76%
250	*5*	3.00	373.2	84.3	46.41%	3.12	559.1	136.2	94.88%	3.12	566.6	138.4	95.42%	3.19	701.2	177.8	99.42%
	9	3.00	368.3	58.6	45.11%	3.09	490.3	82.2	94.74%	3.09	489.3	82.1	94.61%	3.16	625.4	109.5	99.86%

Table 6. Values of $\boldsymbol{L^*}$when $\boldsymbol{L^* = L}$, $\boldsymbol{L^* = L^*_{CE}}$, $\boldsymbol{L^* = L^*_{CA1}}$, $\boldsymbol{L^* = L^*_{CA2}}$, $\boldsymbol{L^* = L^*_{CA3}}$, $\boldsymbol{L^* = L^*_{boot}}$ and their corresponding $\boldsymbol{ARL_0}$, $\boldsymbol{SDARL_0}$ and $\boldsymbol{P\big(CFAR \le \alpha(1+\varepsilon)\big)}$ for $\boldsymbol{L = 3}$, $\boldsymbol{\alpha = 0.0027}$, $\boldsymbol{\varepsilon = 20}$% and $\boldsymbol{p = 5}$% for Case UU

		Unadjusted Limits				Exact Method [Our Proposal]				Aproximate Method 1 [Goedhart et al. (2017)]				Aproximate Method 2 [Goedhart et al. (2018)]			
m	**n**	$L^* = L$	$ARL_0 = E(CARL_0)$	$SDARL_0 = SD(CARL_0)$	$\varepsilon = 20\%$ $P(CFAR \le 0.0032) = P(CARL_0 \ge 308.6)$	$L^* = L^*_{CE}$	$ARL_0 = E(CARL_0)$	$SDARL_0 = SD(CARL_0)$	$\varepsilon = 20\%$ $P(CFAR \le 0.0032) = P(CARL_0 \ge 308.6)$	$L^* = L^*_{CA1}$	$ARL_0 = E(CARL_0)$	$SDARL_0 = SD(CARL_0)$	$\varepsilon = 20\%$ $P(CFAR \le 0.0032) = P(CARL_0 \ge 308.6)$	$L^* = L^*_{CA2}$	$ARL_0 = E(CARL_0)$	$SDARL_0 = SD(CARL_0)$	$\varepsilon = 20\%$ $P(CFAR \le 0.0032) = P(CARL_0 \ge 308.6)$
	3	3.00	569.5	1045.9	49.99%	3.59	8120.5	49788.3	95.00%	3.58	7836.0	47093.1	94.82%	3.58	7747.0	46262.7	94.77%
25	**5**	3.00	418.5	380.3	50.61%	3.41	1959.4	2611.3	95.00%	3.41	2004.6	2686.6	95.20%	3.39	1839.3	2413.6	94.40%
	9	3.00	364.2	210.3	51.49%	3.29	1015.8	723.1	95.00%	3.30	1036.8	741.1	95.28%	3.27	929.5	650.4	93.63%
	3	3.00	448.2	401.3	54.87%	3.36	1792.5	2259.1	95.00%	3.37	1820.0	2302.2	95.15%	3.36	1759.1	2206.8	94.81%
50	**5**	3.00	389.1	217.4	57.21%	3.24	921.6	615.2	95.00%	3.25	925.2	618.1	95.06%	3.24	897.2	595.7	94.56%
	9	3.00	363.8	138.3	60.33%	3.17	640.2	271.4	95.00%	3.17	639.4	271.0	94.97%	3.16	616.1	259.3	94.05%
	3	3.00	418.3	278.6	58.11%	3.27	1129.9	933.9	95.00%	3.28	1137.5	941.5	95.09%	3.27	1117.7	921.7	94.85%
75	**5**	3.00	381.6	166.8	61.60%	3.18	708.8	350.2	95.00%	3.18	707.3	349.3	94.95%	3.18	697.8	343.7	94.66%
	9	3.00	365.0	110.4	66.20%	3.12	540.7	176.3	95.00%	3.12	538.3	175.3	94.87%	3.11	528.6	171.6	94.28%
	3	3.00	404.9	223.8	60.61%	3.22	893.0	581.1	95.00%	3.22	894.2	582.0	95.02%	3.22	886.5	576.0	94.88%
100	**5**	3.00	378.3	140.2	64.98%	3.14	616.2	251.1	95.00%	3.14	613.7	249.9	94.90%	3.14	609.7	248.0	94.73%
	9	3.00	365.9	94.6	70.63%	3.09	493.3	134.9	95.00%	3.09	490.9	134.1	94.82%	3.08	485.8	132.5	94.42%
	3	3.00	392.5	170.1	64.48%	3.17	696.8	338.2	95.00%	3.17	694.6	337.0	94.93%	3.16	693.9	336.6	94.91%
150	**5**	3.00	375.3	111.3	70.11%	3.10	529.1	167.5	95.00%	3.10	526.5	166.5	94.84%	3.10	525.9	166.3	94.81%
	9	3.00	367.1	76.3	77.11%	3.06	445.9	96.2	95.00%	3.06	444.0	95.7	94.80%	3.06	442.1	95.2	94.58%
	3	3.00	386.6	142.3	67.48%	3.13	609.4	245.1	95.00%	3.13	606.7	243.8	94.88%	3.13	607.8	244.3	94.93%
200	**5**	3.00	373.9	95.0	73.98%	3.08	486.3	129.9	95.00%	3.08	484.0	129.2	94.82%	3.08	484.4	129.3	94.85%
	9	3.00	367.8	65.7	81.71%	3.04	421.5	77.3	95.00%	3.04	419.9	76.9	94.80%	3.04	419.0	76.7	94.68%
	3	3.00	383.2	124.7	69.97%	3.11	559.1	195.7	95.00%	3.11	556.3	194.6	94.85%	3.11	558.0	195.3	94.94%
250	**5**	3.00	373.2	84.3	77.08%	3.06	460.2	108.2	95.00%	3.06	458.2	107.6	94.81%	3.06	458.9	107.8	94.88%
	9	3.00	368.3	58.6	85.17%	3.03	406.1	65.8	95.00%	3.03	404.9	65.6	94.80%	3.03	404.5	65.5	94.74%

		Unadjusted Limits				Aproximate Method 3 [Our Proposal]				Bootstrap Method [Saleh et al. (2015)]				Exact Method 2 [Faraz et al. (2017)]			
m	**n**	$L^* = L$	$ARL_0 = E(CARL_0)$	$SDARL_0 = SD(CARL_0)$	$\varepsilon = 20\%$ $P(CFAR \le 0.0032) = P(CARL_0 \ge 308.6)$	$L^* = L^*_{CA3}$	$ARL_0 = E(CARL_0)$	$SDARL_0 = SD(CARL_0)$	$\varepsilon = 20\%$ $P(CFAR \le 0.0032) = P(CARL_0 \ge 308.6)$	$L^* = L^*_{boot}$	$ARL_0 = E(CARL_0)$	$SDARL_0 = SD(CARL_0)$	$\varepsilon = 20\%$ $P(CFAR \le 0.0032) = P(CARL_0 \ge 308.6)$	$L^* = L^*_{CE2}$	$ARL_0 = E(CARL_0)$	$SDARL_0 = SD(CARL_0)$	$\varepsilon = 20\%$ $P(CFAR \le 0.0027) = P(CARL_0 \ge 370.4)$
	3	3.00	569.5	1045.9	49.99%	3.58	7845.8	47184.6	94.83%	3.57	7372.9	42829.5	94.51%	3.79	18866.1	>40000	97.80%
25	**5**	3.00	418.5	380.3	50.61%	3.39	1857.0	2442.4	94.49%	3.42	2063.2	2784.8	95.45%	3.58	3761.3	5887.3	98.54%
	9	3.00	364.2	210.3	51.49%	3.27	937.3	656.8	93.78%	3.32	1134.2	824.9	96.32%	3.47	1898.8	1522.1	99.20%
	3	3.00	448.2	401.3	54.87%	3.36	1763.4	2213.6	94.84%	3.36	1754.2	2199.2	94.79%	3.51	2992.8	4273.9	98.22%
50	**5**	3.00	389.1	217.4	57.21%	3.24	899.1	597.2	94.60%	3.24	905.3	602.2	94.71%	3.38	1469.0	1073.3	98.98%
	9	3.00	363.8	138.3	60.33%	3.16	617.3	259.9	94.10%	3.16	630.4	266.5	94.64%	3.30	1031.9	476.6	99.57%
	3	3.00	418.3	278.6	58.11%	3.27	1118.8	922.8	94.87%	3.31	1321.0	1128.0	96.73%	3.39	1696.4	1525.1	98.41%
75	**5**	3.00	381.6	166.8	61.60%	3.18	698.4	344.1	94.69%	3.18	710.3	351.0	95.04%	3.29	1043.3	553.9	99.15%
	9	3.00	365.0	110.4	66.20%	3.11	529.0	171.8	94.30%	3.11	528.5	171.6	94.27%	3.24	809.8	283.9	99.70%
	3	3.00	404.9	223.8	60.61%	3.22	887.0	576.4	94.89%	3.22	897.7	584.7	95.09%	3.33	1264.8	881.0	98.52%
100	**5**	3.00	378.3	140.2	64.98%	3.14	610.0	248.1	94.74%	3.13	590.0	238.5	93.81%	3.24	864.3	374.8	99.25%
	9	3.00	365.9	94.6	70.63%	3.09	486.0	132.5	94.44%	3.09	501.4	137.5	95.57%	3.20	705.1	205.6	99.77%
	3	3.00	392.5	170.1	64.48%	3.16	694.1	336.7	94.92%	3.18	725.1	354.6	95.79%	3.25	923.8	472.9	98.65%
150	**5**	3.00	375.3	111.3	70.11%	3.10	526.1	166.4	94.81%	3.09	501.5	157.2	93.02%	3.19	700.0	233.3	99.36%
	9	3.00	367.1	76.3	77.11%	3.06	442.1	95.2	94.59%	3.06	447.1	96.5	95.12%	3.15	601.2	136.9	99.83%
	3	3.00	386.6	142.3	67.48%	3.13	607.8	244.3	94.93%	3.14	623.8	252.0	95.57%	3.21	781.7	329.2	98.76%
200	**5**	3.00	373.9	95.0	73.98%	3.08	484.4	129.3	94.85%	3.07	477.4	127.1	94.25%	3.15	622.1	173.8	99.42%
	9	3.00	367.8	65.7	81.71%	3.04	419.1	76.8	94.68%	3.05	429.4	79.0	95.94%	3.12	548.2	105.4	99.86%
	3	3.00	383.2	124.7	69.97%	3.11	558.0	195.3	94.94%	3.12	568.0	199.4	95.45%	3.18	697.9	254.5	98.80%
250	**5**	3.00	373.2	84.3	77.08%	3.06	458.9	107.8	94.88%	3.06	459.3	107.9	94.91%	3.13	574.2	140.6	99.46%
	9	3.00	368.3	58.6	85.17%	3.03	404.5	65.5	94.74%	3.03	412.6	67.0	95.91%	3.10	514.8	87.1	99.88%

Table 7. Values of $\boldsymbol{L}^*$when $\boldsymbol{L}^* = \boldsymbol{L}$, $\boldsymbol{L}^* = \boldsymbol{L}^*_{\boldsymbol{CE}}$, $\boldsymbol{L}^* = \boldsymbol{L}^*_{\boldsymbol{CA1}}$, $\boldsymbol{L}^* = \boldsymbol{L}^*_{\boldsymbol{CA2}}$, $\boldsymbol{L}^* = \boldsymbol{L}^*_{\boldsymbol{CA3}}$, $\boldsymbol{L}^* = \boldsymbol{L}^*_{\boldsymbol{boot}}$ and their corresponding $\boldsymbol{ARL_0}$, $\boldsymbol{SDARL_0}$ and $\boldsymbol{P\big(CFAR \le \alpha(1+\varepsilon)\big)}$ for $\boldsymbol{L = 3}$, $\boldsymbol{\alpha = 0.0027}$, $\boldsymbol{\varepsilon = 20}$% and $\boldsymbol{p = 20}$% for Case UU

		Unadjusted Limits				Exact Method [Our Proposal]				Aproximate Method 1 [Goedhart et al. (2017)]				Aproximate Method 2 [Goedhart et al. (2018)]			
m	n	$L^* = L$	$ARL_0 = E(CARL_0)$	$SDARL_0 = SD(CARL_0)$	$\varepsilon = 20\%$ $P(CFAR \le 0.0032) = P(CARL_0 \ge 308.6)$	$L^* = L^*_{CE}$	$ARL_0 = E(CARL_0)$	$SDARL_0 = SD(CARL_0)$	$\varepsilon = 20\%$ $P(CFAR \le 0.0032) = P(CARL_0 \ge 308.6)$	$L^* = L^*_{CA1}$	$ARL_0 = E(CARL_0)$	$SDARL_0 = SD(CARL_0)$	$\varepsilon = 20\%$ $P(CFAR \le 0.0032) = P(CARL_0 \ge 308.6)$	$L^* = L^*_{CA2}$	$ARL_0 = E(CARL_0)$	$SDARL_0 = SD(CARL_0)$	$\varepsilon = 20\%$ $P(CFAR \le 0.0032) = P(CARL_0 \ge 308.6)$
	3	3.00	569.5	1045.9	49.99%	3.28	1847.4	5453.7	80.00%	3.30	2015.6	6182.0	81.48%	3.28	1838.4	5415.7	79.91%
25	***5***	3.00	418.5	380.3	50.61%	3.19	852.9	926.5	80.00%	3.21	914.9	1011.2	82.00%	3.19	843.5	913.7	79.67%
	9	3.00	364.2	210.3	51.49%	3.14	584.4	372.6	80.00%	3.15	616.2	397.2	82.24%	3.13	574.6	365.1	79.24%
	3	3.00	448.2	401.3	54.87%	3.16	817.1	849.9	80.00%	3.18	867.6	915.7	81.80%	3.16	813.5	845.2	79.86%
50	***5***	3.00	389.1	217.4	57.21%	3.11	561.0	338.7	80.00%	3.11	579.0	351.8	81.47%	3.10	557.1	335.8	79.66%
	9	3.00	363.8	138.3	60.33%	3.07	455.5	181.0	80.00%	3.07	464.4	185.2	81.32%	3.06	451.0	178.9	79.31%
	3	3.00	418.3	278.6	58.11%	3.12	631.0	460.4	80.00%	3.13	655.7	482.5	81.54%	3.12	629.0	458.7	79.87%
75	***5***	3.00	381.6	166.8	61.60%	3.07	484.6	222.4	80.00%	3.08	493.5	227.3	81.11%	3.07	482.4	221.1	79.71%
	9	3.00	365.0	110.4	66.20%	3.04	415.8	129.0	80.00%	3.04	420.1	130.6	80.91%	3.04	413.1	128.0	79.42%
	3	3.00	404.9	223.8	60.61%	3.09	552.7	326.2	80.00%	3.10	567.8	337.1	81.31%	3.09	551.4	325.3	79.89%
100	***5***	3.00	378.3	140.2	64.98%	3.05	448.3	171.9	80.00%	3.05	453.8	174.4	80.89%	3.05	446.8	171.2	79.75%
	9	3.00	365.9	94.6	70.63%	3.02	395.7	103.8	80.00%	3.03	398.3	104.6	80.69%	3.02	393.9	103.2	79.50%
	3	3.00	392.5	170.1	64.48%	3.06	480.7	217.1	80.00%	3.06	488.5	221.3	81.00%	3.06	480.0	216.7	79.91%
150	***5***	3.00	375.3	111.3	70.11%	3.03	412.1	124.4	80.00%	3.03	414.9	125.4	80.63%	3.03	411.2	124.1	79.80%
	9	3.00	367.1	76.3	77.11%	3.01	374.8	78.3	80.00%	3.01	376.2	78.6	80.46%	3.01	373.7	78.0	79.61%
	3	3.00	386.6	142.3	67.48%	3.04	446.0	168.9	80.00%	3.05	450.9	171.1	80.81%	3.04	445.5	168.7	79.92%
200	***5***	3.00	373.9	95.0	73.98%	3.02	393.4	100.9	80.00%	3.02	395.2	101.5	80.50%	3.01	392.8	100.8	79.84%
	9	3.00	367.8	65.7	81.71%	3.00	363.7	64.8	80.00%	3.00	364.5	65.0	80.36%	3.00	362.9	64.7	79.68%
	3	3.00	383.2	124.7	69.97%	3.03	425.1	141.2	80.00%	3.03	428.5	142.6	80.68%	3.03	424.8	141.1	79.94%
250	***5***	3.00	373.2	84.3	77.08%	3.01	381.7	86.6	80.00%	3.01	383.0	86.9	80.41%	3.01	381.3	86.5	79.86%
	9	3.00	368.3	58.6	85.17%	2.99	356.5	56.4	80.00%	2.99	357.1	56.5	80.30%	2.99	356.0	56.3	79.73%

		Unadjusted Limits				Aproximate Method 3 [Our Proposal]				Bootstrap Method [Saleh et al. (2015)]				Exact Method 2 [Faraz et al. (2017)]			
m	n	$L^* = L$	$ARL_0 = E(CARL_0)$	$SDARL_0 = SD(CARL_0)$	$\varepsilon = 20\%$ $P(CFAR \le 0.0032) = P(CARL_0 \ge 308.6)$	$L^* = L^*_{CA3}$	$ARL_0 = E(CARL_0)$	$SDARL_0 = SD(CARL_0)$	$\varepsilon = 20\%$ $P(CFAR \le 0.0032) = P(CARL_0 \ge 308.6)$	$L^* = L^*_{boot}$	$ARL_0 = E(CARL_0)$	$SDARL_0 = SD(CARL_0)$	$\varepsilon = 20\%$ $P(CFAR \le 0.0032) = P(CARL_0 \ge 308.6)$	$L^* = L^*_{CE2}$	$ARL_0 = E(CARL_0)$	$SDARL_0 = SD(CARL_0)$	$\varepsilon = 20\%$ $P(CFAR \le 0.0027) = P(CARL_0 \ge 370.4)$
	3	3.00	569.5	1045.9	49.99%	3.28	1857.2	5495.3	80.09%	3.27	1782.4	5180.9	79.36%	3.47	3810.6	>5000	89.67%
25	***5***	3.00	418.5	380.3	50.61%	3.19	850.6	923.3	79.92%	3.21	906.3	999.4	81.74%	3.35	1453.7	1800.6	91.52%
	9	3.00	364.2	210.3	51.49%	3.13	579.0	368.4	79.58%	3.14	598.1	383.2	81.00%	3.28	943.9	662.4	93.89%
	3	3.00	448.2	401.3	54.87%	3.16	815.3	847.5	79.93%	3.17	827.1	862.8	80.37%	3.30	1302.6	1518.9	90.90%
50	***5***	3.00	389.1	217.4	57.21%	3.10	558.2	336.6	79.76%	3.11	568.2	343.9	80.60%	3.22	826.0	539.5	93.00%
	9	3.00	363.8	138.3	60.33%	3.06	451.9	179.3	79.45%	3.07	460.2	183.2	80.71%	3.18	656.2	279.4	95.54%
	3	3.00	418.3	278.6	58.11%	3.12	629.6	459.2	79.91%	3.13	658.1	484.7	81.68%	3.23	918.2	726.4	91.48%
75	***5***	3.00	381.6	166.8	61.60%	3.07	482.8	221.3	79.76%	3.07	486.0	223.1	80.17%	3.17	669.1	326.9	93.67%
	9	3.00	365.0	110.4	66.20%	3.04	413.4	128.1	79.49%	3.05	428.8	133.8	82.63%	3.13	566.5	186.3	96.25%
	3	3.00	404.9	223.8	60.61%	3.09	551.7	325.5	79.91%	3.10	565.7	335.5	81.13%	3.19	764.0	481.9	91.83%
100	***5***	3.00	378.3	140.2	64.98%	3.05	447.0	171.3	79.78%	3.04	438.5	167.4	78.32%	3.14	595.2	241.0	94.07%
	9	3.00	365.9	94.6	70.63%	3.02	394.1	103.3	79.55%	3.02	395.0	103.6	79.80%	3.11	520.4	143.7	96.65%
	3	3.00	392.5	170.1	64.48%	3.06	480.1	216.8	79.92%	3.06	488.1	221.1	80.95%	3.14	626.3	297.9	92.24%
150	***5***	3.00	375.3	111.3	70.11%	3.03	411.3	124.1	79.82%	3.03	412.7	124.6	80.14%	3.10	521.6	164.1	94.52%
	9	3.00	367.1	76.3	77.11%	3.01	373.8	78.0	79.64%	3.01	375.0	78.3	80.05%	3.08	471.7	102.8	97.10%
	3	3.00	386.6	142.3	67.48%	3.04	445.6	168.7	79.93%	3.05	455.4	173.1	81.52%	3.11	561.7	222.4	92.52%
200	***5***	3.00	373.9	95.0	73.98%	3.01	392.9	100.8	79.85%	3.02	396.9	102.0	80.94%	3.08	483.8	129.1	94.80%
	9	3.00	367.8	65.7	81.71%	3.00	363.0	64.7	79.70%	2.99	360.2	64.1	78.45%	3.06	445.2	82.5	97.35%
	3	3.00	383.2	124.7	69.97%	3.03	424.8	141.1	79.94%	3.04	431.7	143.8	81.28%	3.09	522.7	180.7	92.68%
250	***5***	3.00	373.2	84.3	77.08%	3.01	381.3	86.5	79.87%	3.01	383.2	87.0	80.47%	3.06	459.9	108.1	94.97%
	9	3.00	368.3	58.6	85.17%	2.99	356.0	56.3	79.74%	2.99	358.6	56.7	80.99%	3.05	428.1	70.0	97.51%

5.2. Adjustment in Case KU

In Case KU, since the expression for the c.d.f. of $CARL_0$ is in a simple closed form given by Eq. (35), one can derive a closed-form expression for L^*_{KU}. Using Eq. (35), replacing L by L^*_{KU} and rearranging the terms, one has:

$$L_{KU}^* = \frac{\Phi^{-1}\left(\frac{(1+\varepsilon)\alpha}{2}\right)}{\sqrt{\frac{F_{\chi^2_{m(n-1)}}^{-1}(p)}{m(n-1)}}} \quad (64)$$

Table 8 shows the exact values of L_{KU}^* in case KU for some values of m and n, for $p = 5\%, 10\%, 15\%$ and 20%, $\varepsilon = 0\%, 10\%$ and 20%, and $\alpha = 0.0027$. As noted earlier, this means that the values of L_{KU}^* used in the control limits, provide $P(CFAR \geq (1+\varepsilon)\alpha) = 1-p$ or $P(CARL_0 \geq 1/(1+\varepsilon)\alpha) = 1-p$. So, for example, if the user has 25 reference samples each one with size 9 from a Phase I analysis, then to guarantee $P(CARL_0 \geq 370.4) = 90\%$, he/she should replace L by $L^* = 3.21$ in case KU in Equations (1) and (2).

Since there is no other method in the literature to find L_{KU}^* to be compared with Equation (63) and this equation already gives exact results with a very simple formula, Table 8 does not show other properties of $CARL_0$ as in Tables 5, 6 and 7 [such as the ARL_0 and the $SD(CARL_0)$]. As in case UU, for most of the combinations of parameters, L_{KU}^* is larger than 3. This makes the control limits wider, so the out-of-control performance may be severely affected (since wider control limits causes true alarms to take longer to signal). This will be seen in more detail in Section 5.4.

Table 8. Values of L^*_{KU} to achieve $P(CFAR \geq (1+\varepsilon)\alpha) = 1 - p$ or $P(CARL_0 \geq 1/(1+\varepsilon)\alpha) = 1 - p$ for $n = 3, 5, 9, 15$, $\varepsilon = 0, 0.10, 0.20$, $p = 0.05, 0.1, 0, 15, 0.2$ and $m = 25, 50, 250, 500, 1000$ in Case KU

		m																	
		25			**50**			**100**			**250**			**500**			**1000**		
p	***n***	***ε* → 0**	**0.1**	**0.2**	**0**	**0.1**	**0.2**	**0**	**0.1**	**0.2**	**0**	**0.1**	**0.2**	**0**	**0.1**	**0.2**	**0**	**0.1**	**0.2**
p = 0.05	**3**	3.60	3.56	3.53	3.40	3.37	3.33	3.27	3.24	3.21	3.17	3.13	3.11	3.11	3.08	3.06	3.08	3.05	3.02
	5	3.40	3.37	3.33	3.27	3.24	3.21	3.19	3.16	3.13	3.11	3.08	3.06	3.08	3.05	3.02	3.06	3.03	3.00
	9	3.27	3.24	3.21	3.19	3.16	3.13	3.13	3.10	3.07	3.08	3.05	3.02	3.06	3.03	3.00	3.04	3.01	2.98
	15	3.20	3.17	3.14	3.14	3.11	3.08	3.10	3.07	3.04	3.06	3.03	3.00	3.04	3.01	2.99	3.03	3.00	2.97
p = 0.1	**3**	3.46	3.42	3.39	3.31	3.27	3.24	3.21	3.18	3.15	3.13	3.10	3.07	3.09	3.06	3.03	3.06	3.03	3.01
	5	3.31	3.27	3.24	3.21	3.18	3.15	3.14	3.11	3.09	3.09	3.06	3.03	3.06	3.03	3.01	3.04	3.01	2.99
	9	3.21	3.18	3.15	3.14	3.11	3.09	3.10	3.07	3.04	3.06	3.03	3.01	3.04	3.01	2.99	3.03	3.00	2.97
	15	3.15	3.12	3.10	3.11	3.08	3.05	3.07	3.04	3.02	3.05	3.02	2.99	3.03	3.00	2.98	3.02	2.99	2.97
p = 0.15	**3**	3.36	3.33	3.30	3.25	3.21	3.18	3.17	3.14	3.11	3.10	3.07	3.05	3.07	3.04	3.01	3.05	3.02	2.99
	5	3.25	3.21	3.18	3.17	3.14	3.11	3.12	3.09	3.06	3.07	3.04	3.01	3.05	3.02	2.99	3.04	3.01	2.98
	9	3.17	3.14	3.11	3.12	3.09	3.06	3.08	3.05	3.02	3.05	3.02	2.99	3.04	3.01	2.98	3.02	3.00	2.97
	15	3.12	3.09	3.07	3.09	3.06	3.03	3.06	3.03	3.00	3.04	3.01	2.98	3.03	3.00	2.97	3.02	2.99	2.96
p = 0.2	**3**	3.29	3.26	3.23	3.20	3.17	3.14	3.14	3.11	3.08	3.08	3.05	3.03	3.06	3.03	3.00	3.04	3.01	2.98
	5	3.20	3.17	3.14	3.14	3.11	3.08	3.09	3.06	3.04	3.06	3.03	3.00	3.04	3.01	2.98	3.03	3.00	2.97
	9	3.14	3.11	3.08	3.09	3.06	3.04	3.07	3.04	3.01	3.04	3.01	2.98	3.03	3.00	2.97	3.02	2.99	2.96
	15	3.10	3.07	3.04	3.07	3.04	3.01	3.05	3.02	2.99	3.03	3.00	2.97	3.02	2.99	2.97	3.02	2.99	2.96

5.3. Adjustment in Case UK

In case UK, since there is no closed-form solution for the expression of the c.d.f. of $CARL_0$ and $CFAR$, one can find L^* by solving the following system of equations:

$$\begin{cases} \Phi(z_2) - \Phi(z_1) = 1 - p \\ \left[\Phi\left(\frac{z_i}{\sqrt{m}} + L^*_{UK}\right) - \Phi\left(\frac{z_i}{\sqrt{m}} - L^*_{UK}\right)\right] = 1 - (1+\varepsilon)\alpha, \ i = 1,2 \end{cases} \tag{65}$$

for L^*_{UK}, z_2 and z_1. This can be done numerically with a search algorithm. Using the approximate formula for the c.d.f. of $CARL_0$ and $CFAR$ given by Equation (40), one can also derive the following approximate formula for L^*_{UK}:

$$L^*_{UK} \approx \sqrt{\left(\frac{F^{-1}_{\chi^2_1}(1-p)}{m} + 1\right) F^{-1}_{\chi^2_1}(1-(1+\varepsilon)\alpha)} \tag{66}$$

Table 9 shows the exact [using Equation (65)] and approximate [in bold, using equation (66)] values of L^*_{UK} in case UK for some values of m (note that L^*_{UK}

is invariant with respect to n), for $\varepsilon = 0, 0.05, 0.1, 0.15, 0.2$, $p = 0.05, 0.1, 0, 15, 0.2$ and $\alpha = 0.0027$. As can be seen in Table 9, the approximate formula (66) provides accurate results.

Since the adjustments based on the *EPC* of the $\bar{X}$ control chart for the 3 cases (UU, KU and UK) in most of the parameters combinations make the limits wider than the 3-sigma limits (note that often $L^* > 3$), arises the question of the impact this will have on the out-of-control performance of the chart when these adjusted limits are used. In the next section, we analyze this question.

Table 9. Values of $\boldsymbol{L^*}$ to achieve $\boldsymbol{P(CFAR \geq (1+\varepsilon)\alpha) = 1 - p}$ or $\boldsymbol{P(CARL_0 \geq 1/(1+\varepsilon)\alpha) = 1 - p}$ for $\boldsymbol{\varepsilon = 0, 0.05, 0.1, 0.15, 0.2}$, $\boldsymbol{p = 0.05, 0.1, 0, 15, 0.2}$ and some values of $\boldsymbol{m}$ in CASE UK

		ε									
		0		**0.05**		**0.1**		**0.15**		**0.2**	
m		**Exact**	**Approx**	**Exact**	**Approx**	**Exact**	**Approx**	**Exact**	**Approx**	**Exact**	**Approx**
***m* = 25**	**0.05**	3.19	**3.22**	3.18	**3.21**	3.16	**3.19**	3.15	**3.18**	3.14	**3.16**
	0.1	3.14	**3.16**	3.13	**3.14**	3.11	**3.13**	3.10	**3.11**	3.09	**3.10**
	0.15	3.11	**3.12**	3.10	**3.11**	3.08	**3.09**	3.07	**3.08**	3.06	**3.06**
	0.2	3.09	**3.10**	3.08	**3.08**	3.06	**3.07**	3.05	**3.05**	3.03	**3.04**
***m* = 50**	**0.05**	3.11	**3.11**	3.09	**3.10**	3.08	**3.08**	3.06	**3.07**	3.05	**3.06**
	0.1	3.08	**3.08**	3.06	**3.06**	3.05	**3.05**	3.03	**3.04**	3.02	**3.02**
	0.15	3.06	**3.06**	3.04	**3.05**	3.03	**3.03**	3.02	**3.02**	3.00	**3.00**
	0.2	3.05	**3.05**	3.03	**3.03**	3.02	**3.02**	3.00	**3.01**	2.99	**2.99**
***m* = 75**	**0.05**	3.07	**3.08**	3.06	**3.06**	3.04	**3.05**	3.03	**3.03**	3.02	**3.02**
	0.1	3.05	**3.05**	3.04	**3.04**	3.02	**3.02**	3.01	**3.01**	2.99	**3.00**
	0.15	3.04	**3.04**	3.02	**3.03**	3.01	**3.01**	3.00	**3.00**	2.98	**2.98**
	0.2	3.03	**3.03**	3.02	**3.02**	3.00	**3.00**	2.99	**2.99**	2.98	**2.98**
***m* = 100**	**0.05**	3.05	**3.06**	3.04	**3.04**	3.03	**3.03**	3.01	**3.01**	3.00	**3.00**
	0.1	3.04	**3.04**	3.02	**3.03**	3.01	**3.01**	3.00	**3.00**	2.98	**2.98**
	0.15	3.03	**3.03**	3.02	**3.02**	3.00	**3.00**	2.99	**2.99**	2.97	**2.97**
	0.2	3.02	**3.02**	3.01	**3.01**	2.99	**3.00**	2.98	**2.98**	2.97	**2.97**
***m* = 150**	**0.05**	3.04	**3.04**	3.02	**3.02**	3.01	**3.01**	2.99	**2.99**	2.98	**2.98**
	0.1	3.03	**3.03**	3.01	**3.01**	3.00	**3.00**	2.98	**2.98**	2.97	**2.97**
	0.15	3.02	**3.02**	3.01	**3.01**	2.99	**2.99**	2.98	**2.98**	2.96	**2.96**
	0.2	3.02	**3.02**	3.00	**3.00**	2.99	**2.99**	2.97	**2.97**	2.96	**2.96**
***m* = 200**	**0.05**	3.03	**3.03**	3.01	**3.01**	3.00	**3.00**	2.98	**2.99**	2.97	**2.97**
	0.1	3.02	**3.02**	3.00	**3.01**	2.99	**2.99**	2.98	**2.98**	2.96	**2.96**
	0.15	3.02	**3.02**	3.00	**3.00**	2.99	**2.99**	2.97	**2.97**	2.96	**2.96**
	0.2	3.01	**3.01**	3.00	**3.00**	2.98	**2.98**	2.97	**2.97**	2.96	**2.96**
***m* = 250**	**0.05**	3.02	**3.02**	3.01	**3.01**	2.99	**2.99**	2.98	**2.98**	2.97	**2.97**
	0.1	3.02	**3.02**	3.00	**3.00**	2.99	**2.99**	2.97	**2.97**	2.96	**2.96**
	0.15	3.01	**3.01**	3.00	**3.00**	2.98	**2.98**	2.97	**2.97**	2.96	**2.96**
	0.2	3.01	**3.01**	2.99	**2.99**	2.98	**2.98**	2.97	**2.97**	2.95	**2.95**
***m* = 500**	**0.05**	3.01	**3.01**	3.00	**3.00**	2.98	**2.98**	2.97	**2.97**	2.96	**2.96**
	0.1	3.01	**3.01**	2.99	**2.99**	2.98	**2.98**	2.97	**2.97**	2.95	**2.95**
	0.15	3.01	**3.01**	2.99	**2.99**	2.98	**2.98**	2.96	**2.96**	2.95	**2.95**
	0.2	3.00	**3.00**	2.99	**2.99**	2.98	**2.98**	2.96	**2.96**	2.95	**2.95**
***m* = 1000**	**0.05**	3.01	**3.01**	2.99	**2.99**	2.98	**2.98**	2.96	**2.96**	2.95	**2.95**
	0.1	3.00	**3.00**	2.99	**2.99**	2.97	**2.97**	2.96	**2.96**	2.95	**2.95**
	0.15	3.00	**3.00**	2.99	**2.99**	2.97	**2.97**	2.96	**2.96**	2.95	**2.95**
	0.2	3.00	**3.00**	2.99	**2.99**	2.97	**2.97**	2.96	**2.96**	2.95	**2.95**

5.4
Out–of-control Performance Analysis after the Adjustments

In this section, we analyze the impact of the adjustments proposed in this work on the out-of-control performance of the $\bar{X}$ chart for the three cases (UU, KU and UK). As we noted is the last sections, in most situations the adjustment leads to widening the interval between the control limits (see Table 5,6,7,8 and 9 where often $L^* > 3$). In these cases, the out-of-control conditional *ARL* (i.e., the $CARL_{\delta,UU}$ with $\delta \neq 0$) will be larger with the adjusted limits than with the unadjusted limits. This is the price to pay for guaranteeing a desired in-control performance. So, it is important to assess the deterioration in the $CARL_{\delta,UU}$ due to the adjustment. This assessment will enable the user to choose an appropriate compromise, in terms of $m, n, \varepsilon,$ and $p,$ since the out-of-control performance deterioration is lesser with larger *m* and *n*, and also with larger values of ε and p. For example, for *m* = 25 and *n* = 5 (a typical amount of reference data in practice and according to traditional recommendations), the adjustments proposed in the last sections enable achieving the desired conditional in-control performance in terms of the *EPC* (for example, $P(CARL_0 > 370.4) = 90\%$), however they may produce the undesirable effect of deteriorating the out-of-control performance. Still considering this typical amount of data (i.e., *m* = 25 and *n* = 5), to detect a shift in the process mean of the size of one process standard deviation (i.e., $|\delta| = 1$), with no adjustment, the chart **in case UU**, for example, will have $P\left(CARL_{1,UU} > 7.25\right) = 10\%$, which means that the average number of samples until a true alarm will be most likely below 10 samples. However, with the adjustment (in order to achieve $P\left(CARL_{0,UU} > 370.4\right) = 90\%$), the chart will have $P\left(CARL_{1,UU} > 15.98\right) = 10\%$: a difference of 8.23 (more than 100%) on the 0.9-quantile of the $CARL_{1,UU}$. Note that an out-of-control *ARL* of 15.98 may be unacceptable for the practitioner. However, with *m* = 50 and *n* = 5, with no adjustment $P\left(CARL_{1,UU} > 6.55\right) = 10\%$ and with adjustment, $P\left(CARL_{1,UU} > 9.99\right) = 10\%$. So, with *m* = 50 and *n* = 5, either with or without the adjustment, the $CARL_{1,UU}$ will most likely be below 10 samples, however, only with the adjustment one can guarantee that $P\left(CARL_{0,UU} > 370.4\right) = 90\%$. Also, the difference between the 0.9-quantiles of the $CARL_{1,UU}$ with and without the adjustment, is 3.44 (about 50%). So, a particular user may consider adjusting the limits with m = 50 and n = 5 a good compromise solution between the number and size of subgroups to collect in Phase I, a desired nominal in-control performance and a reasonable out-of-control performance of the $\bar{X}$ chart.

It becomes evident from the above example that knowing the prediction bound for the $CARL_\delta$, with adjusted and with unadjusted limits, is useful for assessing the deterioration (increase) in the $CARL_\delta$ due to the adjustments. The lower prediction bound for $CARL_\delta$ can be calculated similarly as presented in Section 3.5 for bounds for $CARL_0$ in the in-control situation. That is, for given δ, m and n, we can use the distributions of $CARL_{\delta,}$ for all the three cases (UU, KU and UK) derived in Chapter 3 and use them to find a lower bound (denoted $Q_{p_{OOC}}$) that has only a low (specified) probability p_{OOC} (e.g. 0.10) of being exceeded. Formally: for a given p_{OOC} and $\delta \neq 0$, one must find $Q_{p_{OOC}}$ for

$$P\left(CARL_{\delta,UU} > Q_{p_{OOC}}\right) = p_{OOC} \tag{67}$$

Thus, $Q_{p_{OOC}}$ is the $(1 - p_{OOC})$-quantile of the $CARL_{\delta,}$ distribution. Since the $CARL_{\delta,}$ is the realized average number of samples until a true alarm, the smaller the $Q_{p_{OOC}}$ the better the chart's OOC performance. Table 10, 11 and 12, respectively for cases UU, KU and UU presents the values of $Q_{p_{OOC}}$ with the adjusted limits (proposed in the last sections) for $\varepsilon = 0$ and $p = 0.1$ (in grey) and with unadjusted limits, $L = 3$ (in white), for the same values of m and n, for mean shifts $|\delta| = 0.5$, $|\delta| = 1$ and $|\delta| = 1.5$ and for $p_{OOC} = 0.05$ and $p_{OOC} = 0.1$. Also, these tables show the differences (in bold) between the $Q_{p_{OOC}}$ values with the adjusted and the unadjusted limits, respectively, to enable a direct performance comparison.

An examination of Table 10, for Case UU, shows that for $|\delta| = 1$ (a shift in the mean of one standard deviation) and $p_{OOC} = 0.05$, $m = 25$ and $n = 5$, the difference between the $Q_{p_{OOC}}$ values, with and without the adjustment, is of 10.87 samples on average. This is a difference of about 100%, but note that we are considering a 0.95 quantile, a small sample size and a very small number of initial samples. If one increases just the number of samples making $m = 50$ (and maintaining $n = 5$, $p_{OOC} = 0.05, |\delta| = 1)$ the difference between the $Q_{p_{OOC}}$ values, with and without the adjustment, reduces to just 4.12 samples in average (a good improvement). Moreover, if one just increases the sample size by making $n = 10$ (and maintaining $m = 25$, $p_{OOC} = 0.05, |\delta| = 1)$ the difference between the $Q_{p_{OOC}}$ values, with and without the adjustment, reduces to just 0.86 samples in average. This means that for "$n = 10$ and $m = 25$" or "$n = 5$ and $m = 50$" (i.e., a total amount pf Phase I data of 250 observations) the impact of the adjustments (for $\varepsilon = 0$ and $p = 0.1$) in the OOC performance is not so large when $|\delta| = 1$ (compared to when $n = 5$ and $m = 25$).

For a larger shift, say $|\delta| = 1.5$, the difference between the values of $Q_{p_{OOC}}$ with and without adjustment is only 1.14 samples on average making $n = 5$, $m = 25$ and $p_{OOC} = 0.05$. So, for shifts of this magnitude or larger (i.e., $|\delta| \geq 1.5$), the impact of the adjustment on the out-of-control performance is small for any value of n and m. However, for a smaller shift, say $|\delta| = 0.5$, the $Q_{p_{OOC}}$ is large in most cases. For example, for $p_{OOC} = 0.05$, $m = 25$ and $n = 5$, $Q_{p_{OOC}}$ is 107.85 and 351.98 with unadjusted and adjusted limits, respectively. That is an increase of 244.13 samples on average after the adjustments. This shows that, for smaller shifts (such as $|\delta| = 0.5$), the impact of the adjustment on the OOC performance is significantly negative. However, this is not a surprise since the $\bar{X}$ chart is usually not recommended for signaling mean shifts smaller than $|\delta| = 1$ standard deviation (even in Case KK).

As can be seen in Tables 11 and 12, for case KU and UK the situations are slightly different when $|\delta| = 1$. For these shift size, the maximum difference between the $Q_{p_{OOC}}$ values, with and without the adjustment, is of 6.12 samples on average (this is for $p_{OOC} = 0.05$, $m = 25$ and $n = 5$, a small number and size of samples). If the user considers that an increase of less than 10 samples on average on the $Q_{p_{OOC}}$, a satisfactory impact on the OOC performance, in cases KU and UK, the impact when $|\delta| = 1$ is satisfactory for any value of n and m. When $|\delta| = 0.5$ and $|\delta| = 1.5$ the conclusion for cases KU and UU is similar for cases UU.

Table 10. The 0.95 and 0.9 quantiles of $\boldsymbol{CARL_{\delta,UU}}$ with adjusted limits ($\boldsymbol{\alpha = 0.0027}$, $\boldsymbol{p = 0.1}$ and $\boldsymbol{\varepsilon = 0}$) in grey and unadjusted limits ($\boldsymbol{L = 3}$) in white for different values of $\boldsymbol{m}$, $\boldsymbol{n}$ and $\boldsymbol{\delta}$ (Case UU)

			25			**50**			**100**			**300**			**1000**		
δ	p_{OOC}	n	unadj.	adj. $p = 0.1$ $\varepsilon = 0$	difference	unadj.	adj. $p = 0.1$ $\varepsilon = 0$	difference	unadj.	adj. $p = 0.1$ $\varepsilon = 0$	difference	unadj.	adj. $p = 0.1$ $\varepsilon = 0$	difference	unadj.	adj. $p = 0.1$ $\varepsilon = 0$	difference
$\delta = \pm 1.5$	$p_{OOC} = 0.05$	**5**	2.21	3.36	**1.14**	1.97	2.47	**0.50**	1.83	2.09	**0.26**	1.71	1.82	**0.11**	1.64	1.69	**0.05**
		10	1.10	1.17	**0.07**	1.08	1.11	**0.03**	1.07	1.08	**0.02**	1.05	1.06	**0.01**	1.05	1.05	**0.00**
		15	1.01	1.01	**0.01**	1.01	1.01	**0.00**	1.00	1.01	**0.00**	1.00	1.00	**0.00**	1.00	1.00	**0.00**
		20	1.00	1.00	**0.00**	1.00	1.00	**0.00**	1.00	1.00	**0.00**	1.00	1.00	**0.00**	1.00	1.00	**0.00**
	$p_{OOC} = 0.1$	**5**	2.02	2.95	**0.93**	1.86	2.30	**0.44**	1.76	2.00	**0.23**	1.67	1.78	**0.10**	1.62	1.67	**0.05**
		10	1.08	1.14	**0.06**	1.07	1.10	**0.03**	1.06	1.08	**0.02**	1.05	1.06	**0.01**	1.05	1.05	**0.00**
		15	1.01	1.01	**0.01**	1.00	1.01	**0.00**	1.00	1.01	**0.00**	1.00	1.00	**0.00**	1.00	1.00	**0.00**
		20	1.00	1.00	**0.00**	1.00	1.00	**0.00**	1.00	1.00	**0.00**	1.00	1.00	**0.00**	1.00	1.00	**0.00**
$\delta = \pm 1$	$p_{OOC} = 0.05$	**5**	9.27	20.14	**10.87**	7.37	11.50	**4.12**	6.33	8.27	**1.95**	5.45	6.21	**0.77**	4.99	5.32	**0.34**
		10	2.46	3.32	**0.86**	2.21	2.62	**0.40**	2.06	2.28	**0.21**	1.93	2.02	**0.09**	1.85	1.90	**0.04**
		15	1.45	1.66	**0.22**	1.37	1.48	**0.11**	1.33	1.38	**0.06**	1.29	1.31	**0.03**	1.26	1.27	**0.01**
		20	1.15	1.23	**0.08**	1.13	1.16	**0.04**	1.11	1.13	**0.02**	1.09	1.10	**0.01**	1.09	1.09	**0.00**
	$p_{OOC} = 0.1$	**5**	7.75	15.98	**8.23**	6.55	9.99	**3.44**	5.84	7.56	**1.72**	5.21	5.93	**0.72**	4.87	5.19	**0.32**
		10	2.27	3.00	**0.73**	2.10	2.46	**0.36**	1.99	2.19	**0.20**	1.89	1.98	**0.09**	1.84	1.88	**0.04**
		15	1.39	1.58	**0.19**	1.34	1.43	**0.10**	1.31	1.36	**0.05**	1.27	1.30	**0.02**	1.26	1.27	**0.01**
		20	1.13	1.20	**0.07**	1.11	1.15	**0.03**	1.10	1.12	**0.02**	1.09	1.10	**0.01**	1.08	1.09	**0.00**
$\delta = \pm 0.5$	$p_{OOC} = 0.05$	**5**	107.85	351.98	**244.13**	75.24	151.07	**75.83**	58.80	90.45	**31.65**	46.05	57.15	**11.10**	39.75	44.30	**4.55**
		10	29.02	56.41	**27.39**	22.55	33.35	**10.80**	18.99	24.16	**5.17**	16.03	18.06	**2.04**	14.47	15.36	**0.89**
		15	13.34	21.55	**8.21**	10.90	14.40	**3.50**	9.50	11.25	**1.75**	8.30	9.01	**0.72**	7.65	7.97	**0.32**
		20	7.72	11.23	**3.51**	6.52	8.08	**1.56**	5.82	6.61	**0.80**	5.20	5.53	**0.33**	4.86	5.01	**0.15**
	$p_{OOC} = 0.1$	**5**	81.29	249.12	**167.82**	62.14	121.44	**59.30**	51.59	78.40	**26.82**	42.82	52.94	**10.13**	38.23	42.56	**4.33**
		10	23.89	45.11	**21.22**	19.77	28.88	**9.11**	17.35	21.95	**4.60**	15.24	17.15	**1.91**	14.08	14.94	**0.86**
		15	11.42	18.08	**6.65**	9.81	12.85	**3.04**	8.84	10.43	**1.59**	7.97	8.65	**0.68**	7.49	7.80	**0.31**
		20	6.78	9.71	**2.92**	5.98	7.35	**1.38**	5.48	6.21	**0.73**	5.03	5.35	**0.32**	4.78	4.92	**0.14**

Table 11. The 0.95 and 0.9 quantiles of $CARL_{\delta,KU}$ with adjusted limits ($\alpha = 0.0027$, $p = 0.1$ and $\varepsilon = 0$) in grey and unadjusted limits ($L = 3$) in white for different values of m, n and δ (Case KU)

			m														
			25			**50**			**100**			**300**			**1000**		
δ		n	unadj.	adj. $p = 0.1$ $\varepsilon = 0$	difference	unadj.	adj. $p = 0.1$ $\varepsilon = 0$	difference	unadj.	adj. $p = 0.1$ $\varepsilon = 0$	difference	unadj.	adj. $p = 0.1$ $\varepsilon = 0$	difference	unadj.	adj. $p = 0.1$ $\varepsilon = 0$	difference
$\delta = \pm 1.5$	$p_{OOC} = 0.05$	**5**	1.99	2.70	**0.72**	1.84	2.20	**0.36**	1.75	1.96	**0.21**	1.67	1.76	**0.10**	1.62	1.67	**0.05**
		10	1.07	1.11	**0.04**	1.06	1.08	**0.02**	1.05	1.07	**0.01**	1.05	1.06	**0.01**	1.05	1.05	**0.00**
		15	1.00	1.01	**0.00**	1.00	1.01	**0.00**	1.00	1.00	**0.00**	1.00	1.00	**0.00**	1.00	1.00	**0.00**
		20	1.00	1.00	**0.00**	1.00	1.00	**0.00**	1.00	1.00	**0.00**	1.00	1.00	**0.00**	1.00	1.00	**0.00**
	$p_{OOC} = 0.1$	**5**	1.87	2.48	**0.61**	1.77	2.10	**0.33**	1.70	1.90	**0.19**	1.64	1.73	**0.09**	1.61	1.65	**0.05**
		10	1.06	1.10	**0.03**	1.06	1.08	**0.02**	1.05	1.06	**0.01**	1.05	1.05	**0.01**	1.05	1.05	**0.00**
		15	1.00	1.01	**0.00**	1.00	1.00	**0.00**	1.00	1.00	**0.00**	1.00	1.00	**0.00**	1.00	1.00	**0.00**
		20	1.00	1.00	**0.00**	1.00	1.00	**0.00**	1.00	1.00	**0.00**	1.00	1.00	**0.00**	1.00	1.00	**0.00**
$\delta = \pm 1$	$p_{OOC} = 0.05$	**5**	7.48	13.60	**6.12**	6.39	9.22	**2.83**	5.74	7.25	**1.51**	5.16	5.83	**0.67**	4.85	5.16	**0.31**
		10	2.12	2.56	**0.45**	2.00	2.26	**0.26**	1.93	2.09	**0.16**	1.86	1.94	**0.08**	1.82	1.86	**0.04**
		15	1.33	1.43	**0.10**	1.30	1.36	**0.06**	1.28	1.32	**0.04**	1.26	1.28	**0.02**	1.25	1.26	**0.01**
		20	1.10	1.14	**0.03**	1.10	1.11	**0.02**	1.09	1.10	**0.01**	1.08	1.09	**0.01**	1.08	1.08	**0.00**
	$p_{OOC} = 0.1$	**5**	6.60	11.56	**4.96**	5.87	8.33	**2.46**	5.42	6.80	**1.38**	5.00	5.64	**0.63**	4.76	5.07	**0.30**
		10	2.03	2.43	**0.40**	1.94	2.18	**0.24**	1.89	2.04	**0.15**	1.84	1.91	**0.08**	1.81	1.85	**0.04**
		15	1.30	1.40	**0.10**	1.28	1.34	**0.06**	1.27	1.31	**0.04**	1.25	1.27	**0.02**	1.25	1.26	**0.01**
		20	1.10	1.13	**0.03**	1.09	1.11	**0.02**	1.09	1.10	**0.01**	1.08	1.09	**0.01**	1.08	1.08	**0.00**
$\delta = \pm 0.5$	$p_{OOC} = 0.05$	**5**	77.10	195.57	**118.47**	59.81	107.25	**47.44**	50.20	73.35	**23.16**	42.14	51.51	**9.37**	37.90	42.06	**4.16**
		10	20.21	31.86	**11.65**	17.62	23.67	**6.05**	16.02	19.50	**3.48**	14.57	16.20	**1.64**	13.75	14.54	**0.79**
		15	9.45	12.69	**3.24**	8.62	10.46	**1.84**	8.08	9.20	**1.12**	7.58	8.14	**0.56**	7.29	7.57	**0.28**
		20	5.62	6.94	**1.32**	5.25	6.04	**0.78**	5.02	5.51	**0.49**	4.79	5.04	**0.25**	4.65	4.78	**0.13**
	$p_{OOC} = 0.1$	**5**	62.99	152.39	**89.41**	52.14	91.50	**39.36**	45.68	66.09	**20.41**	39.98	48.71	**8.73**	36.84	40.85	**4.01**
		10	18.15	28.13	**9.97**	16.37	21.82	**5.45**	15.23	18.46	**3.24**	14.15	15.73	**1.57**	13.53	14.31	**0.78**
		15	8.80	11.70	**2.90**	8.20	9.91	**1.71**	7.81	8.87	**1.06**	7.43	7.98	**0.54**	7.21	7.49	**0.28**
		20	5.34	6.54	**1.21**	5.07	5.81	**0.74**	4.89	5.37	**0.47**	4.72	4.97	**0.25**	4.62	4.74	**0.13**

Table 12. The 0.95 and 0.9 quantiles of $CARL_{\delta,UK}$ with adjusted limits ($\alpha = 0.0027$, $p = 0.1$ and $\varepsilon = 0$) in grey and unadjusted limits ($L = 3$) in white for different values of m, n and δ (Case UK)

			m														
			25			50			100			300			1000		
δ		n	unadj.	adj. $p = 0.1$ $\varepsilon = 0$	difference	unadj.	adj. $p = 0.1$ $\varepsilon = 0$	difference	unadj.	adj. $p = 0.1$ $\varepsilon = 0$	difference	unadj.	adj. $p = 0.1$ $\varepsilon = 0$	difference	unadj.	adj. $p = 0.1$ $\varepsilon = 0$	difference
$\delta = \pm 1.5$	$p_{OOC} = 0.05$	5	1.96	2.21	**0.25**	1.82	1.93	**0.11**	1.74	1.79	**0.05**	1.66	1.67	**0.01**	1.62	1.62	**0.00**
		10	1.09	1.11	**0.03**	1.07	1.08	**0.01**	1.06	1.07	**0.01**	1.05	1.05	**0.00**	1.05	1.05	**0.00**
		15	1.01	1.01	**0.00**	1.01	1.01	**0.00**	1.00	1.00	**0.00**	1.00	1.00	**0.00**	1.00	1.00	**0.00**
		20	1.00	1.00	**0.00**	1.00	1.00	**0.00**	1.00	1.00	**0.00**	1.00	1.00	**0.00**	1.00	1.00	**0.00**
	$p_{OOC} = 0.1$	5	1.86	2.08	**0.22**	1.76	1.86	**0.10**	1.70	1.74	**0.05**	1.64	1.65	**0.01**	1.61	1.61	**0.00**
		10	1.07	1.10	**0.02**	1.06	1.07	**0.01**	1.06	1.06	**0.00**	1.05	1.05	**0.00**	1.05	1.05	**0.00**
		15	1.01	1.01	**0.00**	1.00	1.01	**0.00**	1.00	1.00	**0.00**	1.00	1.00	**0.00**	1.00	1.00	**0.00**
		20	1.00	1.00	**0.00**	1.00	1.00	**0.00**	1.00	1.00	**0.00**	1.00	1.00	**0.00**	1.00	1.00	**0.00**
$\delta = \pm 1$	$p_{OOC} = 0.05$	5	7.29	9.25	**1.96**	6.27	7.06	**0.79**	5.66	6.00	**0.34**	5.12	5.22	**0.10**	4.82	4.85	**0.03**
		10	2.31	2.64	**0.34**	2.12	2.26	**0.14**	2.00	2.07	**0.06**	1.90	1.92	**0.02**	1.84	1.84	**0.01**
		15	1.42	1.53	**0.11**	1.35	1.40	**0.05**	1.31	1.34	**0.02**	1.28	1.29	**0.01**	1.26	1.26	**0.00**
		20	1.14	1.19	**0.04**	1.12	1.14	**0.02**	1.11	1.11	**0.01**	1.09	1.09	**0.00**	1.08	1.09	**0.00**
	$p_{OOC} = 0.1$	5	6.50	8.18	**1.68**	5.80	6.51	**0.71**	5.37	5.69	**0.32**	4.97	5.07	**0.09**	4.75	4.78	**0.03**
		10	2.16	2.46	**0.30**	2.03	2.16	**0.13**	1.95	2.01	**0.06**	1.87	1.89	**0.02**	1.82	1.83	**0.01**
		15	1.37	1.47	**0.10**	1.32	1.37	**0.04**	1.30	1.32	**0.02**	1.27	1.28	**0.01**	1.25	1.26	**0.00**
		20	1.13	1.17	**0.04**	1.11	1.13	**0.02**	1.10	1.11	**0.01**	1.09	1.09	**0.00**	1.08	1.08	**0.00**
$\delta = \pm 0.5$	$p_{OOC} = 0.05$	5	73.59	107.39	**33.80**	57.88	70.07	**12.20**	49.04	53.96	**4.92**	41.58	42.91	**1.33**	37.62	37.97	**0.35**
		10	24.85	34.15	**9.30**	20.28	23.79	**3.52**	17.65	19.11	**1.46**	15.38	15.78	**0.41**	14.15	14.26	**0.11**
		15	12.21	16.07	**3.85**	10.26	11.77	**1.50**	9.11	9.75	**0.64**	8.11	8.29	**0.18**	7.56	7.61	**0.05**
		20	7.29	9.25	**1.96**	6.27	7.06	**0.79**	5.66	6.00	**0.34**	5.12	5.22	**0.10**	4.82	4.85	**0.03**
	$p_{OOC} = 0.1$	5	61.35	88.67	**27.32**	51.06	61.61	**10.55**	44.96	49.41	**4.45**	39.59	40.84	**1.26**	36.64	36.98	**0.34**
		10	21.30	29.01	**7.71**	18.25	21.35	**3.09**	16.41	17.75	**1.34**	14.76	15.15	**0.39**	13.84	13.95	**0.11**
		15	10.70	13.95	**3.25**	9.38	10.72	**1.34**	8.57	9.16	**0.59**	7.83	8.01	**0.17**	7.42	7.47	**0.05**
		20	6.50	8.18	**1.68**	5.80	6.51	**0.71**	5.37	5.69	**0.32**	4.97	5.07	**0.09**	4.75	4.78	**0.03**

Finally, note that a small difference in $Q_{p_{OOC}}$ values means that the adjustment guarantees the in-control performance as specified and does not significantly deteriorate the OOC performance of the chart. If we consider $Q_{p_{OOC}} \approx 10$ to be an acceptable OOC performance, Table 10, for case UU, shows that both the unadjusted and the adjusted limits do not work well for $|\delta| = 1$ when $n = 5$ and $m = 25$. But in all other cases, for example, when $|\delta| \geq 1$ and $n \geq 10$ or when $|\delta| \geq 1$ and $m \geq 50$, the $Q_{p_{OOC}}$ values are either less than or close to 10 with the adjusted limits, which means the adjustment works well. For cases KU and UK, still considering $Q_{p_{OOC}} \approx 10$ a good performance, from Tables 10 and 11, one can see that the adjustment works well for any values of n and m for $|\delta| \geq 1$. The analysis can be easily replicated for other values of α, p, p_{OOC} and ε. In Appendix G, it is shown tables for adjustments with the combination $\varepsilon = 0.20$ and $p = 0.1$ and the combination $\varepsilon = 0.20$ and $p = 0.2$ for the three cases (UU, KU and UK). The conclusions are similar,

Hence, the adjusted limits are recommended for "$n \geq 10$ and $m \geq 25$" or for "$n \geq 5$ and $m \geq 50$" in case UU and for "$n \geq 5$ and $m \geq 25$" in cases KU and UU in order to guarantee a high probability (such as 0.9) that the conditional in-control average run length is greater than a nominal in-control average run length value (such as 370.4) and to guarantee that a $Q_{pooc} \approx 10$.

6
Conclusions and recommendations

Recently in the literature, the performance of the $\bar{X}$ control chart under normality when parameters are estimated (i.e., when the in-control process mean or the in-control standard deviation are estimated) has been measured based on the conditional in-control average run length ($CARL_0$) or the conditional false alarm rate ($CFAR$). This is because $CARL_0$ and $CFAR$ take into account what is called *practitioner-to-practitioner variability*. Since $CFAR$ and $CARL_0$ are random variables (conditioned on the parameter estimates, i.e., changing from practitioners to practitioner), some authors suggested measuring the performance of the control charts according to the probability of the $CARL_0$ (or the $CFAR$) being at least (or at most) equal to some specified nominal value [this is the Exceedance Probability Criterion (*EPC*)].

In this work, the effects of parameters estimation on the two-sided Phase II $\bar{X}$ control chart was analyzed by deriving the exact $CARL_0$ and $CFAR$ distributions (c.d.f.) in 3 cases: when both the process mean and standard deviation of the process are unknown and need to be estimated (case UU), when only the process standard deviation is estimated (case KU; IC mean specified/known) and when only the process mean is estimated (case UK; IC standard deviation specified/known). For these three cases, previous authors did not provide the exact distributions of these random variables, but instead (just for case UU), they relied on simulations and approximations. The $\bar{X}$ chart in cases KU and UK has not been analyzed so far in this context of the effect of parameter estimation on the conditional performance.

Using the expressions of the c.d.f. of $CARL_0$ and $CFAR$ derived here, the exact upper quantiles of $CFAR$ (and lower quantiles of $CARL_0$) of the $\bar{X}$ chart, which constitute prediction bounds for the $CFAR$ and $CARL_0$ respectively, were calculated and tabulated for the three cases of parameter estimations (UU, KU and UK). These results show that when m or n are small, the values of $CFAR$ that are exceeded (or the values of $CARL_0$ that are not attained) with a small probability of 5% or 10% are much higher (lower) than the desired (nominal) false alarm rate (or than the nominal ARL_0), meaning that the realized false alarm rate may be much larger than expected (or the $CARL_0$ much smaller than expected) when the $\bar{X}$ chart is designed with estimated parameters.

In order to avoid unacceptably low (high) $CARL_0$ ($CFAR$) values, exact expressions were derived, based on the c.d.f.´s derived here, in order to calculate the value of m (the number of Phase I samples) required to guarantee a desired in-control performance in terms of the *EPC* for some Phase I sample sizes (n) for the three cases: UU, KU and UK. These results showed that depending on the practitioner's tolerances and on the subgroups size (n), m can be very large, such as 2,000 samples of size 5 (i.e., a total of 10,000 Phase I data points). This number is larger than the ones recommended in most textbooks and manuals of Statistical Process Control (SPC), and even larger than the numbers recommended by some recent authors who focused on the mean and standard deviation of $CARL_0$ and $CFAR$.

Given the unpractically large numbers (m) of Phase I samples required, in this work, we derived corrections to control limits so that some desired in-control performance in terms of the *EPC* is achieved. Unlike other authors, who use approximations or bootstrapping to propose new adjustment factors for Case UU only, our results were based on the exact c.d.f. of the $CARL_0$ and $CFAR$ for all 3 cases: UU, KU and UK. Moreover, for case UU a detailed comparison between the existing adjustment methods in the literature (approximate formulas and bootstrapping) and the methods proposed in this work was presented. The conclusion is that all of the adjusted methods generate very similar results. Therefore, the recommendation is to use the easiest one. The approximate method derived in this work is the simplest one because, different than the others, it just depends on the quantiles of central chi-square distributions which is tabulated in all text books in statistics. All other adjustment methods will require more advanced statistical skills, like the calculation of the quantile of a non-central chi-square distribution or bootstrapping. However, the solution provided by the exact equations also derived in this work, may be more attractive to be incorporate in a software to calculate the control limits, since this solution yields exact results without requiring much computational time.

The impact of these corrections on the chart's conditional out-of-control performance was also analyzed in this work. Previous authors have tackled this issue only very briefly and focusing mainly only on the unconditional out-of-control run length. Here, the out-of-control performance with and without adjustments were based on some quantiles of the conditional out-of-control average run length. As expected, the deterioration is more severe for smaller sample sizes (n) and smaller number of Phase I samples (m), for example $n = 5$ and $m = 25$. It is also substantial for smaller shifts in the mean (such as $\delta = 0.5$ stnadard deviations). Note, however, that the $\bar{X}$ chart should not be used to detect such small shifts and that even in the ideal (and most often unrealistic) "standards known" case (case KK), its out-of-control average run length is unacceptably large. On the other hand, if one considers a 1 or more standard deviation shift in the mean ($\delta = 1$ or more), the impact on the out-of-control performance is not that substantial in most situations for the 3 cases (UU, KU and UK). This impact is reduced with larger sample sizes and larger numbers of Phase I samples. For case UU, the results presented here leads to a recommendation of using the adjusted limits for at least "$n = 10$ and $m = 25$" or "$n = 5$ and $m = 50$", that is, for at least 250 reference data points (note that this required minimum total number of data points is much smaller than in the case of unadjusted limits). For cases KU and UU, the recommendation is using at least $n = 5$ and $m = 25$ to estimate the parameters (125 reference data points). With these recommended amounts of data and the adjusted limits, the user can strike a balance between a desired nominal in-control conditional performance and a reasonable out-of-control shift detection capability.

Finally, based on the results presented in this work, it seems that, when constructing control charts with estimated control limits, the Exceedance Probability Criterion (EPC) has some imperfections which the practitioners should be aware of. The EPC, controls the probability that the $CARL_0$ is greater than some tolerated value. This approach implicitly considers the variability of $CARL_0$ but it neither controls this variability nor the expected value of $CARL_0$ (the ARL_0), which can also assume extremely large values. Adjusting the limits under the EPC, the large variability of $CARL_0$ is compensated by its large expectation resulting in the desired large probability that the lowest tolerated $CARL_0$ value is exceeded. In conclusion, there is still room for improvement when it comes to designing the Shewhart control charts with unknown parameters. For example, finding a method that controls the *EPC* together with the variability and the expectation of $CARL_0$. One can most likely say the same thing for other types of control charts.

7 References

Albers, W. and Kallenberg, W. C. M. (2004a). “Are Estimated Control Charts in Control?” *Statistics* 38(1), pp. 67-79.

Albers, W. and Kallenberg, W. C. M. (2004b). “Estimation in Shewhart control charts: effects and corrections”. *Metrika* 59(3), pp. 207–234.

Albers, W. and Kallenberg, W. C. M. (2005). “New Corrections for Old Control Charts”. *Quality Engineering* 17, pp. 457–473.

Albers, W., Kallenberg, W. C. M. and Nurdiati S. (2005) “Exceedance probabilities for parametric control charts”. *Statistics* Vol. 39 Issue 5, pp. 429-443.

Aly, A. A., Saleh, N. A., Mahmoud, M. A. and Woodall, W. H. (2015) “A Reevaluation of the Adaptive Exponentially Weighted Moving Average Control Chart When Parameters are Estimated.” *Quality and Reliability Engineering International* 31, pp. 1611–1622

Bischak, D. P. and Trietsch, D. (2007). “The Rate of False Signals in X Charts with Estimated Limits”. *Journal of Quality Technology* 39(1), pp. 54 - 65.

Capizzi, G. and Masarotto, G. (2003). “An adaptive exponentially weighted moving average control chart.” *Technometrics*, 45(3), pp.199-207.

Chakraborti, S. (2000). “Run Length, Average Run Length, and False Alarm Rate of Shewhart X-bar Chart: Exact Derivations by Conditioning”. Communications in Statistics - Simulation and Computation, 29(1), pp. 61 – 81.

Chakraborti, S. (2006). “Parameter Estimation and Design Considerations in Prospective Applications of the $\bar{X}$ Chart”. *Journal of Applied Statistics* 33(4), pp. 439-459.

Chakraborti, S. (2007). “Run Length Distribution and Percentiles: The Shewhart $\bar{X}$ Chart with Unknown Parameters”. *Quality Engineering* 19(2), pp. 119–127.

Chakraborti, S.; Graham, M. A.; and Human, S. W. (2009): “Phase I Statistical Process Control Charts: an Overview and Some Results”. *Quality Engineering*, 21, pp. 52-62.

Chen, G. (1997). “The mean and Standard Deviation of the Run Length Distributions of $\bar{X}$ Charts When Control Limits are Estimated”. *Statistica Sinica* 7, pp. 789-798.

Cox, D., & Reid, N. (1987). “Approximations to Noncentral Distributions”. *The Canadian Journal of Statistics / La Revue Canadienne De Statistique*, 15(2), pp. 105-114.

Diko, M. D., Chakraborti, S. and Graham, M. A. (2015). “Monitoring the process mean when standards are unknown: A classic problem revisited”. *Quality Reliability Engineering International* 32 (2). pp. 609–622.

Diko, M. D., Goedhart, R., Chakraborti, S., Does R. J. M. M. & Epprecht, E. K. (2017) “Phase II control charts for monitoring dispersion when parameters are estimated”. *Quality Engineering*. Published online.

Epprecht, E. K.; Loureiro, L. D.; Chakraborti, S. (2015). “Effect of the Amount of Phase I Data on the Phase II Performance of S^2 and S Control Charts”. *Journal of Quality Technology* 47(2). pp. 139-155.

Faraz A.; William H. Woodall W. H.; and Heuchenne C. (2015). “Guaranteed conditional performance of the S2 control chart with estimated parameters”, *International Journal of Production Research*, 53:14, pp. 4405-4413.

Faraz A.; Heuchenne C. H.; and Saniga E. (2017). “An exact method for designing Shewhart $\bar{X}$ and S^2 control charts to guarantee in-control performance”, *International Journal of Production Research,* doi: 10.1080/00207543.2017.1384580

Feller, W. (1966). “An Introduction to Probability Theory and Its Applications.”, Vol. 2. 2nd ed. New York: Wiley.

Gandy, A. and Kvaløy, J. T. (2013). “Guaranteed Conditional Performance of Control Charts via Bootstrap Methods”. *Scandinavian Journal of Statistics* 40, pp. 647-668.

Ghosh, B. K.; Reynolds, M. R., Jr.; and Hui, Y. V. (1981). “Shewhart $\bar{X}$-Chart with Estimated Process Variance”. *Communications in Statistics – Theory and Methods* 10(18), pp. 1797-1822.

Goedhart, R.; Schoonhoven, M. and Does, R.J.M.M. (2016). Correction factors for Shewhart X and $\bar{X}$ control charts to achieve desired unconditional ARL. *International Journal of Production Research* 54:24, pp: 7464-7479.

Goedhart, R.; Schoonhoven, M. and Does, R.J.M.M. (2017a). “Guaranteed In-Control Performance for the X and $\bar{X}$ Control Charts*”. Journal of Quality Technology* 49(2). pp. 155-171.

Goedhart, R.; da Silva, M. M.; Schoonhoven, M.; Epprecht, E. K.; Chakraborti, S.; Does, R. J.; & Veiga, Á. (2017b). “Shewhart control charts for dispersion adjusted for parameter estimation”. *IISE Transactions*, 1-11.

Goedhart, R.; Schoonhoven, M. and Does, R.J.M.M. (2018). “On guaranteed in-control performance for the Shewhart $\bar{X}$ and X control charts*”. Journal of Quality Technology* 50(1). pp. 130-132.

Guo, B.; and Wang, B. X. (2017). “The Design of the S^2 Control Charts Based on Conditional Performance via Exact Methods”. *Quality and Reliability Engineering International*.

Hong, L. (2012), “A Remark on the Alternative Expectation Formula”, The American Statistician, 66, 232–233.

Hong, L (2015), “Another Remark on the Alternative Expectation Formula”, The American Statistician, 69:3, 157-159.

Jardim, F. S.; Chakraborti, S; and Epprecht, E. K. (2017a) “Two perspectives for designing a Phase II control chart with estimated parameters: The case of the

Shewhart $\bar{X}$ Chart". Technical Memorandum 06/2017, Department of Industrial Engineering, PUC-Rio. Submitted Paper.

Jardim, F. S.; Chakraborti, S; and Epprecht, E. K. (2017b) "New Insights On the $\bar{X}$ Chart with Estimated Parameters: Number of Samples and Limit Adjustments". *Technical Memorandum*, Department of Industrial Engineering, PUC-Rio. Submitted Paper.

Loureiro, L. D.; Epprecht, E.K.; Chakraborti S. and Jardim F. S. (2017) "In-control performance of the joint Phase II $\bar{X}-$ S control charts when parameters are estimated", *Quality Engineering*, 30(2), pp. 253-267.

Jensen, W. A.; Jones-Farmer, L. A.; Champ, C. W.; and Woodall, W. H. (2006) "Effects of Parameter Estimation on Control Chart Properties: A Literature Review". *Journal of Quality Technology*, 38(4), pp. 349-364.

Jones-Farmer, L. A.; Woodall, W. H.; Steiner, S. H.; and Champ, C. W. (2014). "An Overview of Phase I Analysis for Process Improvement and Monitoring". *Journal of Quality Technology* 46(3), pp. 265 280.

Krishnamoorthy, K. and Mathew, T. (2009). *Statistical Tolerance Regions: Theory, Applications, and Computation*. Wiley, New York.

Kumar, N. and Chakraborti, S. (2014) "Phase-II Shewhart-type control charts for monitoring time between events and effects of parameter estimation". *Quality and Reliability Engineering International*. DOI: 10.1002/qre.1752

Mahmoud, M. A.; Henderson, G. R.; Epprecht, E. K.; and Woodall, W. H. (2010) "Estimating the Standard Deviation in Quality Control Applications". *Journal of Quality Technology*, 42(4), pp. 348-357.

Montgomery, D.C. (2009). "*Introduction to Statistical Quality Control*", 7th ed. John Wiley & Sons, Hoboken, NJ.

Moskowitz, H.; Plante, R. D. and Wardell, D. G. (1994). "Using run-length distributions of control charts to detect false alarms". *Production and Operations Management*, 3(3), pp. 217–239.

Nadarajah, S. and Mitov, K.V. (2003). “Product Moments of Multivariate Random Vectors”. Communications in Statistics: Theory and Methods, 32(1), 47-60

Psarakis, S.; Vyniou, A. K.; and Castagliola, P. (2014). “Some Recent Developments on the Effects of Parameter Estimation on Control Charts”. *Quality and Reliability Engineering International* 30. pp. 641-650.

Quesenberry, C. P. (1993). “The Effect of Sample Size on Estimated Limits for $\bar{X}$ and X Control Charts”. *Journal of Quality Technology* 25(4), pp. 237-247.

R Core Team (2014). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

Saleh, N. A.; Mahmoud, M. A.; Keefe, M. J.; and Woodall, W. H. (2015a) “The Difficulty in Designing Shewhart $\bar{X}$ and X Control Charts with Estimated Parameters”. Journal of Quality Technology 47(2). pp. 127-138.

Saleh, N. A.; Mahmoud, M. A.; Jones-Farmer L. A.; Zwetsloot I. and Woodall W.H.; (2015b) “Another look at the EWMA control chart with estimated parameters”. *Journal of Quality Technology* 47(4) pp. 363–382.

Schroeder, R.G.; Linderman, K. and Zhang, D. (2005). Evolution of quality: first fifty issues of production and operations management. Production and Operations Management 14 (4), pp. 468–481.

Shewhart, W. A. (1931). “Economic Control of Quality of Manufactured Product”. Van Nostrand, New York, NY. Reprinted by ASQC Milwaukee, 1980.

Shewhart, W. A. (1939). “Statistical Method from the View-point of Quality Control”. Dover Publications, New York, NY

Trietsch, D. and Bischak, D. (1998). “The Rate of False Signals for Control Charts with Limits Estimated from Small Samples”. In: *Proceedings of the 33rd Annual Conference, The Operational Research Society of New Zealand*, pp. 232-241

Woodall, W. H., and Montgomery, D. C. (1999). “Research Issues and Ideas in Statistical Process Control”. *Journal of Quality Technology*, 31(4), pp. 376-386.

Appendix A - Extras Plots of the $CARL_0$ Curves

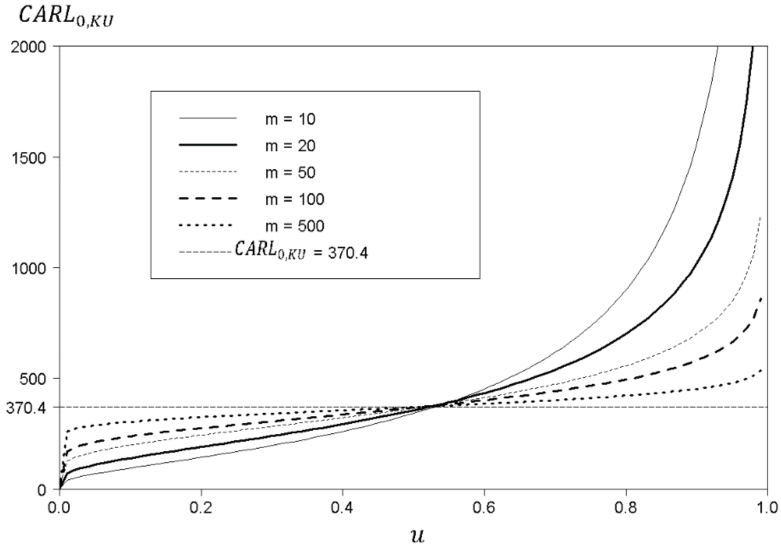


Figure A. 1. $\boldsymbol{CARL_{0,KU}}$ as function of $\boldsymbol{u}$ for $\boldsymbol{n = 5}$, $\boldsymbol{m = 10, 20, 50, 100, 500}$ and $\boldsymbol{\alpha = 0.0027}$ (i.e., $\boldsymbol{L = 3}$).

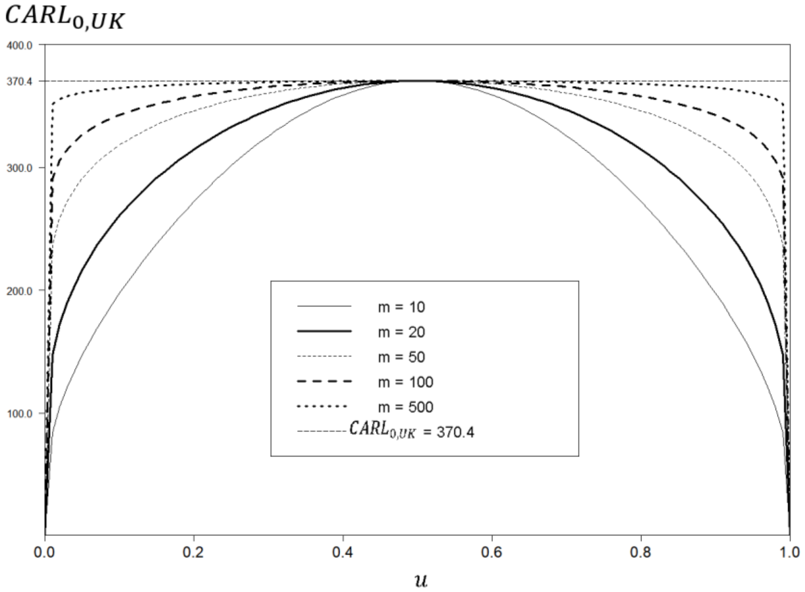


Figure A. 2. $\boldsymbol{CARL_{0,UK}}$ as function of $\boldsymbol{u}$ for $\boldsymbol{n = 5}$, $\boldsymbol{m = 10, 20, 50, 100, 500}$ and $\boldsymbol{\alpha = 0.0027}$ (i.e., $\boldsymbol{L = 3}$).

Appendix B – The Search Algorithm

The search algorithm used to determine the solution of several equations in this work is the secant method, which is used to find the root of a monotonic univariate function. Let $f(x)$ be a monotonic univariate function (note that this is the case of $CARL_{0,UU}$, $CARL_{0,KU}$ and all the c.d.f.´s). Given a value b, one often has to solve an equation such as:

$$f(x^*) = b \tag{B.1}$$

for x^*. Note that the solution of the Eq. (B.1) is equivalent in finding the root of the function $g(x) = f(x) - b$. Because of this, the secant method can be used to find the root of $g(x)$ which is equivalent in finding the $x^*$for

$$g(x^*) = f(x^*) - b = 0 \tag{B.2}$$

So, using Eq. (B.2), the secant method can be applied using the following steps:

1- Starting with initial values x_0^* and x_1^* where, $x_0^* \leq x^* \leq x_1^*$, one construct a line through the points (x_0^*, $g(x_0^*)$) and (x_1^*, $g(x_1^*)$)). In slope-intercept form, this line has the equation:

$$y = \frac{g(x_1^*) - g(x_0^*)}{x_1^* - x_0^*}(x - x_1^*) + g(x_1^*). \tag{B.3}$$

2- One finds the root of this line – the value of x such that $y = 0$ – by solving the following equation for x_2^*:

$$y = \frac{g(x_1^*) - g(x_0^*)}{x_1^* - x_0^*}(x_2^* - x_1^*) + g(x_1^*) = 0, \tag{B.4}$$

which is

$$x_2^* = x_1^* - \ g(x_1^*)\frac{g x_1^* - x_0^*}{g(x_1^*) - g(x_0^*)}. \tag{B.5}$$

3- One then repeat step 1 using x_1^* and x_2^* (if $f(x)$ is a monotonic increasing function) or using x_2^* and x_0^* (if $f(x)$ is a monotonic decreasing function) instead of x_0^* and x_1^*. One continues this process, solving for x_3^*, x_4^*, etc., until one reaches a sufficiently high level of precision (a sufficiently small difference between x_n^* and x_{n-1}^*).

Appendix C – Derivation of the approximate formula for $F_{CPS_{\delta,UK}}(t)$

To derive an approximate formula for $F_{CPS_{\delta,UK}}(t)$, note from Equation (11) that

$$CPS_{\delta,UK} = 1 - \left[\Phi\left(\frac{Z}{\sqrt{m}} + L - \delta\sqrt{n}\right) - \Phi\left(\frac{Z}{\sqrt{m}} - L - \delta\sqrt{n}\right)\right]. \qquad (C.1)$$

Given that $F_{CPS_{\delta,UK}}(t) = P\left(CPS_{\delta,UK} \leq t\right)$, one has

$$F_{CPS_{\delta,UK_{UK,0}}}(t) = P\left(CPS_{\delta,UK} \leq t\right)$$

$$= P\left(1 - \left[\Phi\left(\frac{Z}{\sqrt{m}} + L - \delta\sqrt{n}\right) - \Phi\left(\frac{Z}{\sqrt{m}} - L - \delta\sqrt{n}\right)\right] \leq t\right), \qquad (C.2)$$

where Z_1follows a standard normal distribution, so

$$F_{CPS_{\delta,UK}}(t) = P\left(CPS_{\delta,UK} \leq t\right)$$

$$= P\left(P\left(\frac{Z}{\sqrt{m}} - L - \delta\sqrt{n} \leq Z_1 \leq \frac{Z}{\sqrt{m}} + L - \delta\sqrt{n}\right) \geq 1 - t\right)$$

$$= P\left(P\left(-L \leq Z_1 - \left(\frac{Z}{\sqrt{m}} - \delta\sqrt{n}\right) \leq +L\right) \geq 1 - t\right)$$

$$= P\left(P\left(\left(Z_1 - \left(\frac{Z}{\sqrt{m}} - \delta\sqrt{n}\right)\right)^2 \leq L^2\right) \geq 1 - t\right). \qquad (C.3)$$

Note that $\left(Z_1 - \left(\frac{Z}{\sqrt{m}} - \delta\sqrt{n}\right)\right)^2$ is a random variable in which the distribution is a non central chi-square distribution with 1 degree of freedom (d.f.) and non-centrality parameter $\left(\frac{Z}{\sqrt{m}} - \delta\sqrt{n}\right)^2$. Let's define

$$\left(Z_1 - \left(\frac{Z}{\sqrt{m}} - \delta\sqrt{n}\right)\right)^2 = \chi^2_{1,\left[\left(\frac{Z}{\sqrt{m}} - \delta\sqrt{n}\right)^2\right]}. \qquad (C.4)$$

So, continuing the derivation, one has:

$$F_{CPS_{\delta,UK}}(t) = P\left(CPS_{\delta,UK} \leq t\right) = P\left(P\left(\chi^2_{1,\left[\left(\frac{Z}{\sqrt{m}} - \delta\sqrt{n}\right)^2\right]} \leq L^2\right) \geq 1 - t\right) \qquad (C.5)$$

Cox and Reid (1987) derived the following simple approximation for a non-central chi square distribution.

$$P\left(\chi^2_{1,\left[\left(\frac{Z}{\sqrt{m}}-\delta\sqrt{n}\right)^2\right]} \le L^2\right) \approx P\left(\chi^2_1 \le \frac{L^2}{1+\left(\frac{Z}{\sqrt{m}}-\delta\sqrt{n}\right)^2}\right), \qquad (C.6)$$

where χ^2_1 follows a central chi-square distribution with 1 d.f. So, one has

$$F_{CPS_{\delta,UK}}(t) = P\left(CPS_{\delta,UK} \le t\right) \approx P\left(P\left(\chi^2_1 \le \frac{L^2}{1+\left(\frac{Z}{\sqrt{m}}-\delta\sqrt{n}\right)^2}\right) \ge 1-t\right).$$

$(C.7)$

Rearranging the terms in C.7:

$$F_{CPS_{\delta,UK}}(t) = P\left(CPS_{\delta,UK} \le t\right) \approx P\left(F_{\chi^2_1}\left(\frac{L^2}{1+\left(\frac{Z}{\sqrt{m}}-\delta\sqrt{n}\right)^2}\right) \ge 1-t\right)$$

$$= P\left(\left(\frac{Z}{\sqrt{m}}-\delta\sqrt{n}\right)^2 \le \frac{L^2}{F^{-1}_{\chi^2_1}(1-t)}-1\right)$$

$$= P\left(-\sqrt{\frac{L^2}{F^{-1}_{\chi^2_1}(1-t)}-1} \le \frac{Z}{\sqrt{m}}-\delta\sqrt{n} \le \sqrt{\frac{L^2}{F^{-1}_{\chi^2_1}(1-t)}-1}\right)$$

$$= P\left(\sqrt{m}\left(\delta\sqrt{n}-\sqrt{\frac{L^2}{F^{-1}_{\chi^2_1}(1-t)}-1}\right) \le Z \le \sqrt{m}\left(\delta\sqrt{n}+\sqrt{\frac{L^2}{F^{-1}_{\chi^2_1}(1-t)}-1}\right)\right)$$

$$= \Phi\left(\sqrt{m}\left(\delta\sqrt{n}+\sqrt{\frac{L^2}{F^{-1}_{\chi^2_1}(1-t)}-1}\right)\right) - \Phi\left(\sqrt{m}\left(\delta\sqrt{n}-\sqrt{\frac{L^2}{F^{-1}_{\chi^2_1}(1-t)}-1}\right)\right) \quad (C.8)$$

Rearranging the terms in C.8 again, finally:

$$F_{CPS_{\delta,UK}}(t) = P\left(CPS_{\delta,UK} \le t\right)$$

$$\approx \Phi\left(\delta\sqrt{mn}+\sqrt{m\frac{mL^2}{F^{-1}_{\chi^2_1}(1-t)}-1}\right) - \Phi\left(\delta\sqrt{mn}-\sqrt{\frac{mL^2}{F^{-1}_{\chi^2_1}(1-t)}-1}\right). \quad (C.9)$$

Note that if $\delta = 0$, $CPS_{0,UK} = CFAR_{UK}$. Using C.7, an alternatively approximate formula for the c.d.f. of $CFAR_{UK}$ can be derived as follows

$$F_{CFAR_{UK}}(t) = P(CFAR_{UK} \leq t) \approx P\left(P\left(\chi_1^2 \leq \frac{L^2}{1+\left(\frac{Z}{\sqrt{m}}\right)^2}\right) \geq 1-t\right)$$

$$\approx P\left(F_{\chi_1^2}\left(\frac{L^2}{1+\frac{Z^2}{m}}\right) \geq 1-t\right) = P\left(\frac{L^2}{1+\frac{Z^2}{m}} \geq F_{\chi_1^2}^{-1}(1-t)\right)$$

$$= P\left(Z^2 \leq m\left(\frac{L^2}{F_{\chi_1^2}^{-1}(1-t)} - 1\right)\right)$$

Finally, since Z^2 also follows a chi-square distribution with 1 d.f., one has

$$F_{CFAR_{UK}}(t) = P(CFAR_{UK} \leq t) \approx P\left(\chi_1^2 \leq m\left(\frac{L^2}{F_{\chi_1^2}^{-1}(1-t)} - 1\right)\right)$$

$$= F_{\chi_1^2}\left(m\left(\frac{L^2}{F_{\chi_1^2}^{-1}\left(1-\frac{1}{t}\right)} - 1\right)\right) \qquad (C.10)$$

Appendix D - Derivation of the Approximate Equations of $\alpha_{p,UU}$ L^*_{CA2} and L^*_{CA3} for case UU

Here we derive $\alpha_{p,UU}$, L^*_{CA2} and L^*_{CA3} for case UU. Note that, from Equation (13), $CFAR$, given z, is a function of the chi-square random variable Y. So, from Equation (13) one can write the c.d.f of $CFAR$ conditioned on $Z = z$ as:

$P(CFAR \leq t|Z = z)$

$$= P\left(1 - \left[\Phi\left(\frac{z}{\sqrt{m}} + L^*\sqrt{\frac{Y}{m(n-1)}}\right) - \Phi\left(\frac{z}{\sqrt{m}} - L^*\sqrt{\frac{Y}{m(n-1)}}\right)\right] \leq t\right)$$

$$= P\left(P\left(\frac{z}{\sqrt{m}} - L^*\sqrt{\frac{Y}{m(n-1)}} \leq Z_1 \leq \frac{z}{\sqrt{m}} + L^*\sqrt{\frac{Y}{m(n-1)}}\right) \geq 1 - t\right). \quad (D.1)$$

where Z_1 also follows a standard normal distribution. So

$$P(CFAR \leq t|Z = z)$$

$$= P\left(P\left(-L^*\sqrt{\frac{Y}{m(n-1)}} \leq Z_1 - \frac{z}{\sqrt{m}} \leq L^*\sqrt{\frac{Y}{m(n-1)}}\right) \geq 1 - t\right)$$

$$= P\left(P\left(\left(Z_1 - \frac{z}{\sqrt{m}}\right)^2 \leq \left(L^*\sqrt{\frac{Y}{m(n-1)}}\right)^2\right) \geq 1 - t\right). \quad (D.2)$$

Given that $\left(Z_1 - \frac{z}{\sqrt{m}}\right)^2$ follows a non-central chi-square distribution with 1 degree of freedom and non-centrality parameter $\frac{z^2}{m}$, one can define $\left(Z_1 - \frac{z}{\sqrt{m}}\right)^2 = \chi^2_{1,\left[\frac{z^2}{m}\right]}$, so

$$P(CFAR \leq t|Z = z) = P\left(P\left(\chi^2_{1,\left[\frac{z^2}{m}\right]} \leq L^{*2}\frac{Y}{m(n-1)}\right) \geq 1 - t\right) \quad (D.3)$$

This leads to

$$P(CFAR \leq t|Z = z) = 1 - F_{\chi^2_{m(n-1)}}\left(\frac{m(n-1)F^{-1}_{\chi^2_{1,\left[\frac{z^2}{m}\right]}}(1-t)}{L^{*2}}\right), \quad (D.4)$$

where Z is a standard normal random variable, m is the number of Phase I samples, n is the size of each sample, $F_{\chi^2_{m(n-1)}}$ is the cumulative distribution function of a chi-square distribution with $m(n-1)$ degrees of freedom, and $F^{-1}_{\chi^2_{1,\left[\frac{z^2}{m}\right]}}(1-t)$ is the $(1-t)$-quantile of the distribution of a non-central chi-square distribution with 1 degree of freedom and non-centrality parameter $\frac{z^2}{m}$.

Let $w = z^2/m$, so one has

$$P(CFAR \leq t|Z=z) = 1 - F_{\chi^2_{m(n-1)}}\left(\frac{m(n-1)F^{-1}_{\chi^2_{1,[\mathrm{w}]}}(1-t)}{\left(\frac{L^*}{C_{4,b}}\right)^2}\right) = g(w). \quad (\mathrm{D}.5)$$

Thus, using the one-step Taylor approximation for $g(w)$ around the point $\frac{1}{m}$ gives

$$P(CFAR \leq t|Z=z) = g(w) \approx g\left(\frac{1}{m}\right) + \left(\mathrm{w} - \frac{1}{m}\right)\frac{dg(w)}{dx}\bigg|_{w=\frac{1}{m}}. \qquad (D.6)$$

Taking the expectation in respect to Z on both sides, one has the cumulative distribution function of $CFAR$ as shown below and also an approximation for it.

$$\begin{aligned} P(CFAR \leq t) &= E_Z\big(P(CFAR \leq t|Z=z)\big) \\ &\approx E_Z\left(g\left(\frac{1}{m}\right) + \left(W - \frac{1}{m}\right)\frac{dg(z)}{dx}\bigg|_{z=\frac{1}{m}}\right) \\ &= g\left(\frac{1}{m}\right) + E_Z\left(W - \frac{1}{m}\right)\frac{dg(Z)}{dx}\bigg|_{Z=\frac{1}{m}} \end{aligned} \qquad (D.7)$$

Note that $E(W) = E\left(\frac{Z^2}{m}\right) = \frac{1}{m}E(Z^2) = \frac{1}{m}\big(V(Z) + E^2(Z)\big) = \frac{1}{m}(1+0) = \frac{1}{m}$. This leads to $P(CFAR \leq t) \approx g\left(\frac{1}{m}\right)$. So finally,

$$P(CFAR \leq t) \approx g\left(\frac{1}{m}\right) = 1 - F_{\chi^2_{m(n-1)}}\left(\frac{m(n-1)F^{-1}_{\chi^2_{1,\left[\frac{1}{m}\right]}}(1-t)}{L^{*2}}\right). \qquad (D.8)$$

To find an approximation for L^*, since the goal is $P(CFAR_{UU} \leq (1+\varepsilon)\alpha) = 1-p$, we just need to solve the following approximation for L^*

$$1 - F_{\chi^2_{m(n-1)}}\left(\frac{m(n-1)F^{-1}_{\chi^2_{1,\left[\frac{1}{m}\right]}}(1-(1+\varepsilon)\alpha)}{\left(\frac{L^*}{C_{4,b}}\right)^2}\right) \approx 1-p. \qquad (D.9)$$

Solving for L^*, we get $L^* = L^*_{CA2} \approx C_{4,b}\sqrt{m(n-1)\frac{F^{-1}_{\chi^2_{1,}\left[\frac{1}{m}\right]}(1-(1+\varepsilon)\alpha)}{F^{-1}_{\chi^2_{m(n-1)}}(p)}}$.

Cox and Reid (1987) obtained the following approximation for the c.d.f. of a non-central chi-square distribution:

$$F_{\chi^2_{1,}\left[\frac{1}{m}\right]}(a) \approx F_{\chi^2_1}\left(\frac{a}{1+\frac{1}{m}}\right). \qquad (D.10)$$

Given this approximation, one can write

$$F^{-1}_{\chi^2_{1,}\left[\frac{1}{m}\right]}(b) \approx \left(1+\frac{1}{m}\right)F^{-1}_{\chi^2_1}(b). \qquad (D.11)$$

Replacing b by $1-(1+\varepsilon)\alpha$, and $F^{-1}_{\chi^2_{1,}\left[\frac{1}{m}\right]}(1-(1+\varepsilon)\alpha)$ by $\left(1+\frac{1}{m}\right)F^{-1}_{\chi^2_1}(b)$ in L^*_{CA2}, the final simpler approximation formula is given by:

$$L^*_{CA3} \approx \sqrt{(n-1)(m+1)\frac{F^{-1}_{\chi^2_1}(1-(1+\varepsilon)\alpha)}{F^{-1}_{\chi^2_{m(n-1)}}(p)}}. \qquad (D.12)$$

Replacing L^*_{CA3} by L, $(1+\varepsilon)\alpha$ by $\alpha_{p,UU}$ and rearranging the terms, one has the final expression for $\alpha_{p,UU}$:

$$\alpha_{p,UU} \approx 1 - F_{\chi^2_1}\left(L^2\frac{F^{-1}_{\chi^2_{m(n-1)}}(p)}{(m+1)(n-1)}\right) \qquad (D.13)$$

Appendix E - Expressions of $g(L)$ and $g'(L)$

Here the expressions of $g(L)$ and $g'(L)$ to calculate the approximation shown in (57) are presented. They were derived by Goedhart et al. (2017).

$g(L) = 3\sqrt[3]{(1+\varepsilon)\alpha}\,\frac{f_E(L)^{2/3}}{f_V(L)^{1/2}} - 3\,\frac{f_E(L)}{f_V(L)^{1/2}} + \frac{1}{3}\frac{f_V(L)^{1/2}}{f_E(L)}$ and

$g'(L) = 3\sqrt[3]{(1+\varepsilon)\alpha}B - 3C + \frac{1}{3}D$ with $C = \frac{f_E'(L)f_V(L)^{1/2} - f_E(L)\frac{1}{2}f_V(L)^{-1/2}f_V'(L)}{f_V(L)}$,

$B = \frac{\frac{2}{3}f_E(L)^{-1/3}f_E'(L)f_V(L)^{1/2} - f_E(L)^{2/3}\frac{1}{2}f_V(L)^{-1/2}f_V'(L)}{f_V(L)}$,

$D = \frac{\frac{1}{2}f_V(L)^{-1/2}f_V'(L)f_E(L) - f_V(L)^{1/2}f_E'(L)}{f_E(L)^2}$, where, $f_E(L) = E(CFAR)$, $f_V(L) = V(CFAR)$, $f_E'(L) = \frac{dE(CFAR)}{dL}$ and $f_V'(L) = \frac{dV(CFAR)}{dL}$.

Below, we present the expressions of $E(CFAR)$, $V(CFAR)$, $f_E'(L)$ and $f_V'(L)$. Considering the expression of $CFAR$ given by Equation (13), $E(CFAR)$ can be calculated by

$$E(CFAR) = \int_{-\infty}^{\infty}\int_{0}^{\infty}(CFAR)\,\phi(z)f_Y(y)\,dy\,dz. \qquad (E.1)$$

$V(CFAR)$ is given by $V(CFAR) = E(CFAR^2) - E^2(CFAR)$, where

$$E(CFAR^2) = \int_{-\infty}^{\infty}\int_{0}^{\infty}(CFAR)^2\,\phi(z)f_Y(y)\,dy\,dz. \qquad (E.2)$$

Since $f_E'(L) = \frac{dE(CFAR)}{dL}$, one has: $f_E'(L) = \int_{-\infty}^{\infty}\int_{0}^{\infty} -\frac{1}{C_{4,b}}\sqrt{\frac{Y}{m(n-1)}}\,G\,\phi(z)f_Y(y)\,dy\,dz$,

where $G = \phi\left(\frac{Z}{\sqrt{m}} + \frac{L}{C_{4,b}}\sqrt{\frac{Y}{m(n-1)}}\right) + \phi\left(\frac{Z}{\sqrt{m}} - \frac{L}{C_{4,b}}\sqrt{\frac{Y}{m(n-1)}}\right)$.

Finally, $f_V'(L) = \frac{dV(CFAR)}{dL} = \frac{dE(CFAR^2)}{dL} - 2\,E(CFAR)f_E'(L)$,

where $\frac{dE(CFAR^2)}{dL} = \int_{-\infty}^{\infty}\int_{0}^{\infty} 2CFAR\left(-\frac{1}{C_{4,b}}\sqrt{\frac{Y}{m(n-1)}}\,G\right)\phi(z)f_Y(y)\,dy\,dz$.

Appendix F - Derivation of Formula (63) for the Bootstrap Method

Here it is derived the exact Formula (62) for L_k^*, which is the solution of Equation (61) of the bootstrap method. Rearranging the left hand side of Equation (61), we get

$$P\left(\left(\frac{\mu_k^* - \bar{\bar{X}}}{S_p/c_{4,b}}\right)\sqrt{n} - L_k^* \frac{\sigma_k^*}{S_p} \leq Z_1 \leq \left(\frac{\mu_k^* - \bar{\bar{X}}}{S_p/c_{4,b}}\right)\sqrt{n} + L_k^* \frac{\sigma_k^*}{S_p}\right)$$

$$= 2\big(1 - \Phi(L)\big)(1 + \varepsilon). \qquad \text{(F1)}$$

Where Z_1 follows a standard normal distribution. So

$$P\left(-L_k^* \frac{\sigma_k^* \, c_{4,b}}{S_p} \leq Z_1 - \left(\frac{\mu_k^* - \bar{\bar{X}}}{S_p}\right)\sqrt{n} \leq L_k^* \frac{\sigma_k^*}{S_p}\right)$$

$$= P\left(\left(Z_1 - \left(\frac{\mu_k^* - \bar{\bar{X}}}{S_p}\right)\sqrt{n}\right)^2 \leq \left(L_k^* \frac{\sigma_k^* \, 1}{S_p}\right)^2\right) = 2\big(1 - \Phi(L)\big)(1 + \varepsilon). \quad \text{(F.2)}$$

Defining $W = \left(Z_1 - \left(\frac{\mu_k^* - \bar{\bar{X}}}{S_p}\right)\sqrt{n}\right)^2$ and recognizing that W follows a non-central qui-square distribution with 1 degree of freedom and non-centrality parameter $n\left(\frac{\mu_k^* - \bar{\bar{X}}}{S_p}\right)^2$, we can write

$$P\left(W \leq \left(L_k^* \frac{\sigma_k^*}{S_p}\right)^2\right) = F_W\left(\left(L_k^* \frac{\sigma_k^*}{S_p}\right)^2\right) = 2\big(1 - \Phi(L)\big)(1 + \varepsilon). \quad \text{(F.3)}$$

where F_W denotes the c.d.f. of W. Thus $1 - F_W\left(\left(L_k^* \frac{\sigma_k^*}{S_p}\right)^2\right) = 2\big(1 - \Phi(L)\big)(1 + \varepsilon)$.

Solving this last equation for L_k^*, we get

$$L_k^* = \frac{S_p\sqrt{F_W^{-1}\left(1 - 2\big(1 - \Phi(L)\big)(1 + \varepsilon)\right)}}{\sigma_k^*} \qquad \text{(F.5)}$$

Appendix G – Extra Tables of the Out-of-Control

Table G. 1. The 0.95 and 0.9 quantiles of $CARL_{\delta,UU}$ with adjusted limits ($\alpha = 0.0027$, $p = 0.1$ and $\varepsilon = 0.2$) in grey and unadjusted limits ($L = 3$) in white for different values of m, n and δ (Case UU)

			m = 25			50			100			300			1000		
δ	p_{out}	n	unadj.	adj. $p = 0.1$ $\varepsilon = 0.2$	difference	unadj.	adj. $p = 0.1$ $\varepsilon = 0.2$	difference	unadj.	adj. $p = 0.1$ $\varepsilon = 0.2.$	difference	unadj.	adj. $p = 0.1$ $\varepsilon = 0.$	difference	unadj.	adj. $p = 0.1$ $\varepsilon = 0.2$	difference
$\delta = \pm 1.5$	$p_{OOC} = 0.05$	5	2.21	3.11	**0.89**	1.97	2.33	**0.36**	1.83	1.99	**0.15**	1.71	1.74	**0.03**	1.64	1.63	**-0.01**
		10	1.10	1.15	**0.05**	1.08	1.10	**0.02**	1.07	1.07	**0.01**	1.05	1.05	**0.00**	1.05	1.05	**0.00**
		15	1.01	1.01	**0.00**	1.01	1.01	**0.00**	1.00	1.00	**0.00**	1.00	1.00	**0.00**	1.00	1.00	**0.00**
		20	1.00	1.00	**0.00**	1.00	1.00	**0.00**	1.00	1.00	**0.00**	1.00	1.00	**0.00**	1.00	1.00	**0.00**
	$p_{OOC} = 0.1$	5	2.02	2.75	**0.73**	1.86	2.17	**0.31**	1.76	1.90	**0.14**	1.67	1.71	**0.03**	1.62	1.61	**-0.01**
		10	1.08	1.13	**0.04**	1.07	1.09	**0.02**	1.06	1.07	**0.01**	1.05	1.05	**0.00**	1.05	1.04	**0.00**
		15	1.01	1.01	**0.00**	1.00	1.01	**0.00**	1.00	1.00	**0.00**	1.00	1.00	**0.00**	1.00	1.00	**0.00**
		20	1.00	1.00	**0.00**	1.00	1.00	**0.00**	1.00	1.00	**0.00**	1.00	1.00	**0.00**	1.00	1.00	**0.00**
$\delta = \pm 1$	$p_{OOC} = 0.05$	5	9.27	17.53	**8.26**	7.37	10.25	**2.88**	6.33	7.48	**1.16**	5.45	5.69	**0.24**	4.99	4.91	**-0.08**
		10	2.46	3.09	**0.63**	2.21	2.47	**0.25**	2.06	2.16	**0.10**	1.93	1.93	**0.00**	1.85	1.82	**-0.03**
		15	1.45	1.60	**0.15**	1.37	1.43	**0.06**	1.33	1.35	**0.02**	1.29	1.28	**0.00**	1.26	1.25	**-0.01**
		20	1.15	1.21	**0.05**	1.13	1.15	**0.02**	1.11	1.12	**0.01**	1.09	1.09	**0.00**	1.09	1.08	**-0.01**
	$p_{OOC} = 0.1$	5	7.75	14.04	**6.29**	6.55	8.95	**2.40**	5.84	6.86	**1.02**	5.21	5.44	**0.23**	4.87	4.79	**-0.08**
		10	2.27	2.80	**0.54**	2.10	2.32	**0.23**	1.99	2.08	**0.09**	1.89	1.89	**0.00**	1.84	1.80	**-0.03**
		15	1.39	1.52	**0.13**	1.34	1.39	**0.05**	1.31	1.32	**0.02**	1.27	1.27	**0.00**	1.26	1.24	**-0.01**
		20	1.13	1.18	**0.04**	1.11	1.13	**0.02**	1.10	1.11	**0.01**	1.09	1.09	**0.00**	1.08	1.08	**-0.01**
$\delta = \pm 0.5$	$p_{OOC} = 0.05$	5	107.85	286.20	**178.35**	75.24	126.44	**51.21**	58.80	77.06	**18.26**	46.05	49.47	**3.42**	39.75	38.69	**-1.06**
		10	29.02	48.33	**19.32**	22.55	29.09	**6.54**	18.99	21.31	**2.32**	16.03	16.09	**0.07**	14.47	13.77	**-0.71**
		15	13.34	18.96	**5.63**	10.90	12.85	**1.95**	9.50	10.13	**0.63**	8.30	8.18	**-0.11**	7.65	7.27	**-0.38**
		20	7.72	10.07	**2.35**	6.52	7.33	**0.82**	5.82	6.05	**0.23**	5.20	5.10	**-0.10**	4.86	4.64	**-0.23**
	$p_{OOC} = 0.1$	5	81.29	204.79	**123.50**	62.14	102.35	**40.21**	51.59	67.10	**15.51**	42.82	45.94	**3.12**	38.23	37.22	**-1.01**
		10	23.89	38.91	**15.02**	19.77	25.30	**5.53**	17.35	19.42	**2.06**	15.24	15.30	**0.06**	14.08	13.40	**-0.68**
		15	11.42	16.00	**4.57**	9.81	11.51	**1.70**	8.84	9.42	**0.57**	7.97	7.87	**-0.11**	7.49	7.12	**-0.37**
		20	6.78	8.75	**1.97**	5.98	6.70	**0.72**	5.48	5.70	**0.22**	5.03	4.93	**-0.10**	4.78	4.56	**-0.22**

Table G. 2. The 0.95 and 0.9 quantiles of $CARL_{\delta,UU}$ with adjusted limits ($\alpha = 0.0027$, $p = 0.2$ and $\varepsilon = 0.2$) in grey and unadjusted limits ($L = 3$) in white for different values of m, n and δ (Case UU)

			m = 25			50			100			300			1000		
δ	p_{OOC}	n	unadj.	adj. $p = 0.2$ $\varepsilon = 0.2$	difference	unadj.	adj. $p = 0.2$ $\varepsilon = 0.2$	difference	unadj.	adj. $p = 0.2$ $\varepsilon = 0.2$	difference	unadj.	adj. $p = 0.2$ $\varepsilon = 0.2$	difference	unadj.	adj. $p = 0.1$ $\varepsilon = 0.2$	difference
$\delta = \pm 1.5$	$p_{OOC} = 0.05$	5	2.21	2.72	**0.51**	1.97	2.17	**0.20**	1.83	1.91	**0.08**	1.71	1.71	**0.00**	1.64	1.61	**-0.03**
		10	1.10	1.13	**0.03**	1.08	1.09	**0.01**	1.07	1.07	**0.00**	1.05	1.05	**0.00**	1.05	1.05	**0.00**
		15	1.01	1.01	**0.00**	1.01	1.01	**0.00**	1.00	1.00	**0.00**	1.00	1.00	**0.00**	1.00	1.00	**0.00**
		20	1.00	1.00	**0.00**	1.00	1.00	**0.00**	1.00	1.00	**0.00**	1.00	1.00	**0.00**	1.00	1.00	**0.00**
	$p_{OOC} = 0.1$	5	2.02	2.44	**0.42**	1.86	2.03	**0.17**	1.76	1.83	**0.07**	1.67	1.68	**0.00**	1.62	1.60	**-0.03**
		10	1.08	1.11	**0.03**	1.07	1.08	**0.01**	1.06	1.06	**0.00**	1.05	1.05	**0.00**	1.05	1.04	**0.00**
		15	1.01	1.01	**0.00**	1.00	1.01	**0.00**	1.00	1.00	**0.00**	1.00	1.00	**0.00**	1.00	1.00	**0.00**
		20	1.00	1.00	**0.00**	1.00	1.00	**0.00**	1.00	1.00	**0.00**	1.00	1.00	**0.00**	1.00	1.00	**0.00**
$\delta = \pm 1$	$p_{OOC} = 0.05$	5	9.27	13.79	**4.52**	7.37	8.94	**1.56**	6.33	6.88	**0.56**	5.45	5.46	**0.01**	4.99	4.81	**-0.18**
		10	2.46	2.83	**0.37**	2.21	2.35	**0.13**	2.06	2.10	**0.04**	1.93	1.90	**-0.02**	1.85	1.81	**-0.05**
		15	1.45	1.54	**0.09**	1.37	1.40	**0.03**	1.33	1.33	**0.00**	1.29	1.27	**-0.01**	1.26	1.25	**-0.02**
		20	1.15	1.18	**0.03**	1.13	1.14	**0.01**	1.11	1.11	**0.00**	1.09	1.09	**0.00**	1.09	1.08	**-0.01**
	$p_{OOC} = 0.1$	5	7.75	11.22	**3.47**	6.55	7.86	**1.31**	5.84	6.33	**0.50**	5.21	5.22	**0.01**	4.87	4.70	**-0.17**
		10	2.27	2.58	**0.31**	2.10	2.22	**0.12**	1.99	2.02	**0.03**	1.89	1.87	**-0.02**	1.84	1.79	**-0.05**
		15	1.39	1.47	**0.08**	1.34	1.37	**0.03**	1.31	1.31	**0.00**	1.27	1.26	**-0.01**	1.26	1.24	**-0.02**
		20	1.13	1.16	**0.03**	1.11	1.12	**0.01**	1.10	1.10	**0.00**	1.09	1.08	**0.00**	1.08	1.08	**-0.01**
$\delta = \pm 0.5$	$p_{OOC} = 0.05$	5	107.85	199.14	**91.29**	75.24	102.07	**26.84**	58.80	67.43	**8.63**	46.05	46.21	**0.16**	39.75	37.37	**-2.39**
		10	29.02	39.83	**10.82**	22.55	25.94	**3.40**	18.99	19.84	**0.85**	16.03	15.52	**-0.51**	14.47	13.51	**-0.96**
		15	13.34	16.53	**3.19**	10.90	11.87	**0.97**	9.50	9.65	**0.15**	8.30	7.98	**-0.31**	7.65	7.18	**-0.47**
		20	7.72	9.05	**1.33**	6.52	6.90	**0.38**	5.82	5.83	**0.02**	5.20	5.00	**-0.20**	4.86	4.59	**-0.27**
	$p_{OOC} = 0.1$	5	81.29	145.26	**63.97**	62.14	83.31	**21.17**	51.59	58.93	**7.35**	42.82	42.96	**0.14**	38.23	35.96	**-2.28**
		10	23.89	32.34	**8.45**	19.77	22.64	**2.87**	17.35	18.11	**0.76**	15.24	14.76	**-0.48**	14.08	13.16	**-0.93**
		15	11.42	14.03	**2.60**	9.81	10.66	**0.84**	8.84	8.98	**0.13**	7.97	7.68	**-0.30**	7.49	7.03	**-0.46**
		20	6.78	7.90	**1.11**	5.98	6.32	**0.34**	5.48	5.50	**0.02**	5.03	4.84	**-0.19**	4.78	4.52	**-0.26**

Table G. 3. The 0.95 and 0.9 quantiles of $CARL_{\delta,KU}$ with adjusted limits ($\alpha = 0.0027$, $p = 0.1$ and $\varepsilon = 0.2$) in grey and unadjusted limits ($L = 3$) in white for different values of m, n and δ (Case KU)

			m														
			25			50			100			300			1000		
δ		n	unadj.	adj. $p = 0.1$ $\varepsilon = 0.2$	difference	unadj.	adj. $p = 0.1$ $\varepsilon = 0.2$	difference	unadj.	adj. $p = 0.1$ $\varepsilon = 0.2$	difference	unadj.	adj. $p = 0.1$ $\varepsilon = 0.2$	difference	unadj.	adj. $p = 0.1$ $\varepsilon = 0.2$	difference
$\delta = \pm 1.5$	$p_{OOC} = 0.05$	**5**	1.99	2.52	**0.54**	1.84	2.09	**0.25**	1.75	1.87	**0.12**	1.67	1.69	**0.03**	1.62	1.61	**-0.01**
		10	1.07	1.09	**0.02**	1.06	1.07	**0.01**	1.05	1.06	**0.00**	1.05	1.05	**0.00**	1.05	1.04	**0.00**
		15	1.00	1.01	**0.00**	1.00	1.00	**0.00**	1.00	1.00	**0.00**	1.00	1.00	**0.00**	1.00	1.00	**0.00**
		20	1.00	1.00	**0.00**	1.00	1.00	**0.00**	1.00	1.00	**0.00**	1.00	1.00	**0.00**	1.00	1.00	**0.00**
	$p_{OOC} = 0.1$	**5**	1.87	2.33	**0.46**	1.77	1.99	**0.22**	1.70	1.81	**0.11**	1.64	1.67	**0.03**	1.61	1.59	**-0.01**
		10	1.06	1.08	**0.02**	1.06	1.07	**0.01**	1.05	1.06	**0.00**	1.05	1.05	**0.00**	1.05	1.04	**0.00**
		15	1.00	1.01	**0.00**	1.00	1.00	**0.00**	1.00	1.00	**0.00**	1.00	1.00	**0.00**	1.00	1.00	**0.00**
		20	1.00	1.00	**0.00**	1.00	1.00	**0.00**	1.00	1.00	**0.00**	1.00	1.00	**0.00**	1.00	1.00	**0.00**
$\delta = \pm 1$	$p_{OOC} = 0.05$	**5**	7.48	11.96	**4.48**	6.39	8.26	**1.87**	5.74	6.58	**0.84**	5.16	5.35	**0.18**	4.85	4.76	**-0.09**
		10	2.12	2.41	**0.30**	2.00	2.14	**0.14**	1.93	1.99	**0.06**	1.86	1.85	**0.00**	1.82	1.78	**-0.04**
		15	1.33	1.39	**0.06**	1.30	1.32	**0.03**	1.28	1.29	**0.01**	1.26	1.25	**-0.01**	1.25	1.23	**-0.01**
		20	1.10	1.12	**0.02**	1.10	1.10	**0.01**	1.09	1.09	**0.00**	1.08	1.08	**0.00**	1.08	1.07	**-0.01**
	$p_{OOC} = 0.1$	**5**	6.60	10.24	**3.65**	5.87	7.51	**1.63**	5.42	6.18	**0.76**	5.00	5.18	**0.17**	4.76	4.68	**-0.09**
		10	2.03	2.29	**0.27**	1.94	2.07	**0.13**	1.89	1.95	**0.05**	1.84	1.83	**0.00**	1.81	1.77	**-0.04**
		15	1.30	1.36	**0.06**	1.28	1.31	**0.03**	1.27	1.28	**0.01**	1.25	1.25	**-0.01**	1.25	1.23	**-0.01**
		20	1.10	1.11	**0.02**	1.09	1.10	**0.01**	1.09	1.09	**0.00**	1.08	1.08	**0.00**	1.08	1.07	**-0.01**
$\delta = \pm 0.5$	$p_{OOC} = 0.05$	**5**	77.10	160.71	**83.61**	59.81	90.28	**30.47**	50.20	62.69	**12.49**	42.14	44.65	**2.51**	37.90	36.75	**-1.15**
		10	20.21	27.62	**7.41**	17.62	20.79	**3.17**	16.02	17.27	**1.25**	14.57	14.46	**-0.10**	13.75	13.04	**-0.71**
		15	9.45	11.31	**1.86**	8.62	9.41	**0.79**	8.08	8.33	**0.25**	7.58	7.41	**-0.17**	7.29	6.91	**-0.38**
		20	5.62	6.30	**0.68**	5.25	5.53	**0.27**	5.02	5.06	**0.05**	4.79	4.65	**-0.13**	4.65	4.43	**-0.22**
	$p_{OOC} = 0.1$	**5**	62.99	126.49	**63.50**	52.14	77.51	**25.37**	45.68	56.72	**11.03**	39.98	42.32	**2.34**	36.84	35.74	**-1.11**
		10	18.15	24.52	**6.36**	16.37	19.23	**2.86**	15.23	16.39	**1.16**	14.15	14.06	**-0.10**	13.53	12.84	**-0.69**
		15	8.80	10.47	**1.67**	8.20	8.93	**0.73**	7.81	8.04	**0.23**	7.43	7.27	**-0.17**	7.21	6.84	**-0.37**
		20	5.34	5.96	**0.63**	5.07	5.33	**0.26**	4.89	4.94	**0.05**	4.72	4.59	**-0.13**	4.62	4.39	**-0.22**

Table G. 4. The 0.95 and 0.9 quantiles of $CARL_{\delta,KU}$ with adjusted limits ($\alpha = 0.0027$, $p = 0.2$ and $\varepsilon = 0.2$) in grey and unadjusted limits ($L = 3$) in white for different values of m, n and δ (Case KU)

			m														
			25			50			100			300			1000		
δ		n	unadj.	adj. $p = 0.2$ $\varepsilon = 0.2$	difference	unadj.	adj. $p = 0.2$ $\varepsilon = 0.2$	difference	unadj.	adj. $p = 0.2$ $\varepsilon = 0.2$	difference	unadj.	adj. $p = 0.2$ $\varepsilon = 0.2$	difference	unadj.	adj. $p = 0.2$ $\varepsilon = 0.2$	difference
$\delta = \pm 1.5$	$p_{OOC} = 0.05$	**5**	1.99	2.26	**0.28**	1.84	1.96	**0.12**	1.75	1.80	**0.05**	1.67	1.66	**0.00**	1.62	1.59	**-0.03**
		10	1.07	1.08	**0.01**	1.06	1.06	**0.00**	1.05	1.06	**0.00**	1.05	1.05	**0.00**	1.05	1.04	**0.00**
		15	1.00	1.00	**0.00**	1.00	1.00	**0.00**	1.00	1.00	**0.00**	1.00	1.00	**0.00**	1.00	1.00	**0.00**
		20	1.00	1.00	**0.00**	1.00	1.00	**0.00**	1.00	1.00	**0.00**	1.00	1.00	**0.00**	1.00	1.00	**0.00**
	$p_{OOC} = 0.1$	**5**	1.87	2.11	**0.24**	1.77	1.88	**0.11**	1.70	1.75	**0.04**	1.64	1.64	**0.00**	1.61	1.58	**-0.03**
		10	1.06	1.07	**0.01**	1.06	1.06	**0.00**	1.05	1.05	**0.00**	1.05	1.05	**0.00**	1.05	1.04	**0.00**
		15	1.00	1.00	**0.00**	1.00	1.00	**0.00**	1.00	1.00	**0.00**	1.00	1.00	**0.00**	1.00	1.00	**0.00**
		20	1.00	1.00	**0.00**	1.00	1.00	**0.00**	1.00	1.00	**0.00**	1.00	1.00	**0.00**	1.00	1.00	**0.00**
$\delta = \pm 1$	$p_{OOC} = 0.05$	**5**	7.48	9.71	**2.23**	6.39	7.29	**0.90**	5.74	6.08	**0.34**	5.16	5.13	**-0.03**	4.85	4.66	**-0.18**
		10	2.12	2.26	**0.14**	2.00	2.06	**0.05**	1.93	1.94	**0.01**	1.86	1.83	**-0.03**	1.82	1.77	**-0.05**
		15	1.33	1.35	**0.03**	1.30	1.30	**0.01**	1.28	1.27	**0.00**	1.26	1.25	**-0.01**	1.25	1.23	**-0.02**
		20	1.10	1.11	**0.01**	1.10	1.10	**0.00**	1.09	1.09	**0.00**	1.08	1.08	**-0.01**	1.08	1.07	**-0.01**
	$p_{OOC} = 0.1$	**5**	6.60	8.42	**1.82**	5.87	6.66	**0.79**	5.42	5.73	**0.31**	5.00	4.97	**-0.03**	4.76	4.58	**-0.18**
		10	2.03	2.15	**0.13**	1.94	1.99	**0.05**	1.89	1.90	**0.01**	1.84	1.81	**-0.03**	1.81	1.76	**-0.05**
		15	1.30	1.33	**0.02**	1.28	1.29	**0.01**	1.27	1.26	**0.00**	1.25	1.24	**-0.01**	1.25	1.23	**-0.02**
		20	1.10	1.10	**0.01**	1.09	1.09	**0.00**	1.09	1.08	**0.00**	1.08	1.08	**-0.01**	1.08	1.07	**-0.01**
$\delta = \pm 0.5$	$p_{OOC} = 0.05$	**5**	77.10	116.30	**39.20**	59.81	73.94	**14.13**	50.20	55.15	**4.95**	42.14	41.75	**-0.39**	37.90	35.50	**-2.40**
		10	20.21	23.65	**3.44**	17.62	18.82	**1.20**	16.02	16.17	**0.15**	14.57	13.96	**-0.60**	13.75	12.80	**-0.94**
		15	9.45	10.24	**0.79**	8.62	8.81	**0.20**	8.08	7.98	**-0.11**	7.58	7.23	**-0.35**	7.29	6.83	**-0.46**
		20	5.62	5.87	**0.25**	5.25	5.28	**0.02**	5.02	4.91	**-0.11**	4.79	4.58	**-0.21**	4.65	4.39	**-0.26**
	$p_{OOC} = 0.1$	**5**	62.99	93.05	**30.06**	52.14	63.96	**11.82**	45.68	50.06	**4.38**	39.98	39.61	**-0.37**	36.84	34.53	**-2.31**
		10	18.15	21.11	**2.96**	16.37	17.45	**1.08**	15.23	15.37	**0.14**	14.15	13.57	**-0.58**	13.53	12.61	**-0.92**
		15	8.80	9.50	**0.71**	8.20	8.39	**0.18**	7.81	7.71	**-0.10**	7.43	7.10	**-0.34**	7.21	6.76	**-0.46**
		20	5.34	5.57	**0.23**	5.07	5.09	**0.02**	4.89	4.79	**-0.10**	4.72	4.51	**-0.21**	4.62	4.36	**-0.26**

Table G. 5. The 0.95 and 0.9 quantiles of $CARL_{\delta,UK}$ with adjusted limits ($\alpha = 0.0027$, $p = 0.1$ and $\varepsilon = 0.2$) in grey and unadjusted limits ($L = 3$) in white for different values of m, n and δ (Case UK)

			m														
			25			**50**			**100**			**300**			**1000**		
δ		n	**unadj.**	**adj.** $p=0.1$ $\varepsilon=0.2$	**difference**	**unadj.**	**adj.** $p=0.1$ $\varepsilon=0.2$	**difference**	**unadj.**	**adj.** $p=0.1$ $\varepsilon=0.2$	**difference**	**unadj.**	**adj.** $p=0.1$ $\varepsilon=0.2$	**difference**	**unadj.**	**adj.** $p=0.1$ $\varepsilon=0.2$	**difference**
$\delta=\pm1.5$	$p_{ooc}=0.05$	**5**	1.96	2.10	**0.14**	1.82	1.85	**0.03**	1.74	1.72	**-0.02**	1.66	1.62	**-0.04**	1.62	1.57	**-0.05**
		10	1.09	1.10	**0.02**	1.07	1.07	**0.00**	1.06	1.06	**0.00**	1.05	1.05	**0.00**	1.05	1.04	**-0.01**
		15	1.01	1.01	**0.00**	1.01	1.01	**0.00**	1.00	1.00	**0.00**	1.00	1.00	**0.00**	1.00	1.00	**0.00**
		20	1.00	1.00	**0.00**	1.00	1.00	**0.00**	1.00	1.00	**0.00**	1.00	1.00	**0.00**	1.00	1.00	**0.00**
	$p_{ooc}=0.1$	**5**	1.86	1.98	**0.13**	1.76	1.78	**0.02**	1.70	1.68	**-0.02**	1.64	1.60	**-0.04**	1.61	1.56	**-0.05**
		10	1.07	1.09	**0.01**	1.06	1.07	**0.00**	1.06	1.05	**0.00**	1.05	1.05	**0.00**	1.05	1.04	**-0.01**
		15	1.01	1.01	**0.00**	1.00	1.00	**0.00**	1.00	1.00	**0.00**	1.00	1.00	**0.00**	1.00	1.00	**0.00**
		20	1.00	1.00	**0.00**	1.00	1.00	**0.00**	1.00	1.00	**0.00**	1.00	1.00	**0.00**	1.00	1.00	**0.00**
$\delta=\pm1$	$p_{ooc}=0.05$	**5**	7.29	8.38	**1.09**	6.27	6.45	**0.18**	5.66	5.52	**-0.14**	5.12	4.83	**-0.30**	4.82	4.50	**-0.33**
		10	2.31	2.50	**0.19**	2.12	2.15	**0.03**	2.00	1.98	**-0.03**	1.90	1.84	**-0.06**	1.84	1.77	**-0.07**
		15	1.42	1.48	**0.06**	1.35	1.36	**0.01**	1.31	1.31	**-0.01**	1.28	1.26	**-0.02**	1.26	1.24	**-0.02**
		20	1.14	1.17	**0.02**	1.12	1.13	**0.00**	1.11	1.10	**0.00**	1.09	1.08	**-0.01**	1.08	1.08	**-0.01**
	$p_{ooc}=0.1$	**5**	6.50	7.44	**0.94**	5.80	5.97	**0.17**	5.37	5.24	**-0.13**	4.97	4.69	**-0.29**	4.75	4.43	**-0.32**
		10	2.16	2.33	**0.17**	2.03	2.06	**0.03**	1.95	1.92	**-0.03**	1.87	1.81	**-0.06**	1.82	1.76	**-0.07**
		15	1.37	1.42	**0.06**	1.32	1.33	**0.01**	1.30	1.29	**-0.01**	1.27	1.25	**-0.02**	1.25	1.23	**-0.02**
		20	1.13	1.15	**0.02**	1.11	1.11	**0.00**	1.10	1.10	**0.00**	1.09	1.08	**-0.01**	1.10	1.07	**-0.01**
$\delta=\pm0.5$	$p_{ooc}=0.05$	**5**	73.59	91.98	**18.39**	57.88	60.66	**2.78**	49.04	47.03	**-2.02**	41.58	37.61	**-3.96**	37.62	33.39	**-4.23**
		10	24.85	29.97	**5.12**	20.28	21.09	**0.81**	17.65	17.04	**-0.61**	15.38	14.15	**-1.22**	14.15	12.82	**-1.33**
		15	12.21	14.35	**2.14**	10.26	10.61	**0.35**	9.11	8.85	**-0.27**	8.11	7.56	**-0.55**	7.56	6.96	**-0.60**
		20	7.29	8.38	**1.09**	6.27	6.45	**0.18**	5.66	5.52	**-0.14**	5.12	4.83	**-0.30**	4.82	4.50	**-0.33**
	$p_{ooc}=0.1$	**5**	61.35	76.24	**14.89**	51.06	53.47	**2.41**	44.96	43.14	**-1.82**	39.59	35.84	**-3.74**	36.64	32.53	**-4.11**
		10	21.30	25.55	**4.25**	18.25	18.97	**0.71**	16.41	15.86	**-0.55**	14.76	13.60	**-1.16**	13.84	12.55	**-1.29**
		15	10.70	12.51	**1.81**	9.38	9.69	**0.31**	8.57	8.32	**-0.25**	7.83	7.31	**-0.52**	7.42	6.83	**-0.59**
		20	6.50	7.44	**0.94**	5.80	5.97	**0.17**	5.37	5.24	**-0.13**	4.97	4.69	**-0.29**	4.75	4.43	**-0.32**

Table G. 6. The 0.95 and 0.9 quantiles of $CARL_{\delta,UK}$ with adjusted limits ($\alpha = 0.0027$, $p = 0.2$ and $\varepsilon = 0.2$) in grey and unadjusted limits ($L = 3$) in white for different values of m, n and δ (Case UK)

			m														
			25			**50**			**100**			**300**			**1000**		
δ		n	**unadj.**	**adj.** $p=0.2$ $\varepsilon=0.2$	**difference**	**unadj.**	**adj.** $p=0.2$ $\varepsilon=0.2$	**difference**	**unadj.**	**adj.** $p=0.2$ $\varepsilon=0.2$	**difference**	**unadj.**	**adj.** $p=0.2$ $\varepsilon=0.2$	**difference**	**unadj.**	**adj.** $p=0.2$ $\varepsilon=0.2$	**difference**
$\delta=\pm1.5$	$p_{ooc}=0.05$	**5**	1.96	2.01	**0.05**	1.82	1.81	**-0.01**	1.74	1.70	**-0.04**	1.66	1.61	**-0.05**	1.62	1.57	**-0.05**
		10	1.09	1.09	**0.01**	1.07	1.07	**0.00**	1.06	1.06	**0.00**	1.05	1.05	**-0.01**	1.05	1.04	**-0.01**
		15	1.01	1.01	**0.00**	1.01	1.00	**0.00**	1.00	1.00	**0.00**	1.00	1.00	**0.00**	1.00	1.00	**0.00**
		20	1.00	1.00	**0.00**	1.00	1.00	**0.00**	1.00	1.00	**0.00**	1.00	1.00	**0.00**	1.00	1.00	**0.00**
	$p_{ooc}=0.1$	**5**	1.86	1.90	**0.05**	1.76	1.75	**-0.01**	1.70	1.66	**-0.04**	1.64	1.59	**-0.05**	1.61	1.55	**-0.05**
		10	1.07	1.08	**0.01**	1.06	1.06	**0.00**	1.06	1.05	**0.00**	1.05	1.04	**-0.01**	1.05	1.04	**-0.01**
		15	1.01	1.01	**0.00**	1.00	1.00	**0.00**	1.00	1.00	**0.00**	1.00	1.00	**0.00**	1.00	1.00	**0.00**
		20	1.00	1.00	**0.00**	1.00	1.00	**0.00**	1.00	1.00	**0.00**	1.00	1.00	**0.00**	1.00	1.00	**0.00**
$\delta=\pm1$	$p_{ooc}=0.05$	**5**	7.29	7.70	**0.41**	6.27	6.18	**-0.09**	5.66	5.40	**-0.26**	5.12	4.79	**-0.33**	4.82	4.49	**-0.34**
		10	2.31	2.38	**0.07**	2.12	2.10	**-0.02**	2.00	1.95	**-0.05**	1.90	1.83	**-0.07**	1.84	1.77	**-0.07**
		15	1.42	1.44	**0.02**	1.35	1.35	**-0.01**	1.31	1.30	**-0.02**	1.28	1.26	**-0.02**	1.26	1.24	**-0.02**
		20	1.14	1.15	**0.01**	1.12	1.12	**0.00**	1.11	1.10	**-0.01**	1.09	1.08	**-0.01**	1.08	1.08	**-0.01**
	$p_{ooc}=0.1$	**5**	6.50	6.85	**0.35**	5.80	5.72	**-0.08**	5.37	5.13	**-0.24**	4.97	4.66	**-0.32**	4.75	4.42	**-0.33**
		10	2.16	2.23	**0.06**	2.03	2.02	**-0.02**	1.95	1.90	**-0.05**	1.87	1.80	**-0.06**	1.82	1.76	**-0.07**
		15	1.37	1.39	**0.02**	1.32	1.32	**-0.01**	1.30	1.28	**-0.02**	1.27	1.25	**-0.02**	1.25	1.23	**-0.02**
		20	1.13	1.13	**0.01**	1.11	1.11	**0.00**	1.10	1.09	**-0.01**	1.09	1.08	**-0.01**	1.08	1.07	**-0.01**
$\delta=\pm0.5$	$p_{ooc}=0.05$	**5**	73.59	80.30	**6.71**	57.88	56.53	**-1.34**	49.04	45.38	**-3.67**	41.58	37.17	**-4.41**	37.62	33.27	**-4.35**
		10	24.85	26.74	**1.89**	20.28	19.88	**-0.39**	17.65	16.55	**-1.10**	15.38	14.02	**-1.36**	14.15	12.79	**-1.36**
		15	12.21	13.01	**0.79**	10.26	10.09	**-0.17**	9.11	8.63	**-0.49**	8.11	7.50	**-0.61**	7.56	6.94	**-0.62**
		20	7.29	7.70	**0.41**	6.27	6.18	**-0.09**	5.66	5.40	**-0.26**	5.12	4.79	**-0.33**	4.82	4.49	**-0.34**
	$p_{ooc}=0.1$	**5**	61.35	66.79	**5.44**	51.06	49.90	**-1.15**	44.96	41.65	**-3.31**	39.59	35.42	**-4.16**	36.64	32.42	**-4.22**
		10	21.30	22.87	**1.57**	18.25	17.91	**-0.35**	16.41	15.40	**-1.01**	14.76	13.47	**-1.29**	13.84	12.52	**-1.33**
		15	10.70	11.37	**0.67**	9.38	9.23	**-0.15**	8.57	8.12	**-0.45**	7.83	7.25	**-0.58**	7.42	6.81	**-0.61**
		20	6.50	6.85	**0.35**	5.80	5.72	**-0.08**	5.37	5.13	**-0.24**	4.97	4.66	**-0.32**	4.75	4.42	**-0.33**

Appendix H - Codes in R Language

R CODES FOR CASE UU

before running the codes below, please, download the packages cubature and numDeriv

```
library(cubature) # this package helps compute double integrals
library(numDeriv) # this package helps compute numerical derivations
```

the function secantc below was created to find the root of a monotonic increasing function. Note that the precision is up to 10 decimal places.

```
secantc <- function(fun, x0, x1, tol=1e-10, niter=100000){
  for ( i in 1:niter ) {
    funx1 <- fun(x1)
    funx0 <- fun(x0)
    x2 <- ( (x0*funx1) - (x1*funx0) )/( funx1 - funx0 )
    funx2 <- fun(x2)
    if (abs(funx2) < tol) {
      return(x2)
    }
    if (funx2 < 0)
      x0 <- x2
    else
      x1 <- x2
  }
  stop("exceeded allowed number of iteractions")
}
```

the functions CPS and CARL below computes the Conditional Average Run Length (CARL) and the Conditional Probability of a Signal of the Xbar chart in Case UU for a given values of Z (standard normal random variable), Y (chi-square random variable with n*(m-1) d.f.), limit factor (L), scaled shift in the process mean (delta), number (m) and size (n) of Phase I samples. Note that when delta = 0, CARL will return the in-control CARL (i.e., the CARL0).

```
CPS <- function (Z,Y,delta,L,m,n) {
  a <- 1 - pnorm((-delta*sqrt(n))+((1/sqrt(m))*Z)+(L*sqrt(Y/(m*(n-1)))),0,1) + pnorm((-
delta*sqrt(n))+((1/sqrt(m))+Z)-(L*sqrt(Y/(m*(n-1)))),0,1)
  return(a)
}

CARL <- function (Z,Y,delta,L,m,n) {
```

```
  a <- 1/(1 - pnorm((-delta*sqrt(n))+((1/sqrt(m))*Z)+(L*sqrt(Y/(m*(n-1)))),0,1) + pnorm((-
delta*sqrt(n))+((1/sqrt(m))+Z)-(L*sqrt(Y/(m*(n-1)))),0,1))
  return(a)
}

# the functions CDFCPS and CDFCARL (below) computes, respectively, the c.d.f. (for any
value t) of the Conditional Probability of a Signal (CPS) and CARL of the Xbar chart for a
given limit factor (L), scaled shift in the process mean (delta), number (m) and size (n) of
Phase I samples. Note that when delta = 0, CDFCPS and CDFCARL will return the in-
control c.d.f. of the CPS and CARL (i.e., cdf of the CFAR, the conditional False Alarm Rate
and CARL0).

CDFCPS <- function (t,delta,L,m,n) {
  CFAR <- function (U) {
    a<-1- pchisq((m*(n-1)*qchisq(1-t, df=1, ncp = ((qnorm(U)/sqrt(m))-
(delta*sqrt(n)))^2))/((L)^2),m*(n-1))
    return(a)
  }
  d <- integrate(CFAR,0,1)$val
  return(d)
}

CDFCARL <- function (t,delta,L,m,n) {
  CARL <- function (U) {
    a<-pchisq((m*(n-1)*qchisq(1-(1/t), df=1, ncp = ((qnorm(U)/sqrt(m))-
(delta*sqrt(n)))^2))/(L^2),m*(n-1))
    return(a)
  }
  d <- integrate(CARL,0,1)$val
  return(d)
}

# the functions ARL, ARL2, VARL, SDARL2 and quantileCARL (below) compute,
respectively, the mean, the central second moment, the variance, the standard
deviation and the p-quantile of the CARL0 of the Xbar chart in case UU for a given limit
factor (L), scaled shift in the process mean (delta), number (m) and size (n) of Phase I
samples. Note that if delta = 0, the function returns the in-control values.

ARL <- function (delta,L,m,n) {

  CARL <- function (U) {
    a <- 1/(1 - pnorm((-
delta*sqrt(n))+((1/sqrt(m))*qnorm(U[1],0,1))+(L*sqrt(qchisq(U[2],m*(n-1))/(m*(n-
1)))),0,1) + pnorm((-delta*sqrt(n))+((1/sqrt(m))*qnorm(U[1],0,1))-
(L*sqrt(qchisq(U[2],m*(n-1))/(m*(n-1)))),0,1))
    return(a)
  }
  a <- adaptIntegrate(CARL, lowerLimit = c(0, 0), upperLimit = c(1, 1))$integral
  return (a)

}
```

```
ARL2 <- function (delta,L,m,n) {

  CARL <- function (U) {
    a <- (1/(1 - pnorm((-
delta*sqrt(n))+((1/sqrt(m))*qnorm(U[1],0,1))+(L*sqrt(qchisq(U[2],m*(n-1))/(m*(n-
1)))),0,1) + pnorm((-delta*sqrt(n))+((1/sqrt(m))*qnorm(U[1],0,1))-
(L*sqrt(qchisq(U[2],m*(n-1))/(m*(n-1)))),0,1)))^2
    return(a)
  }
  a <- adaptIntegrate(CARL, lowerLimit = c(0, 0), upperLimit = c(1, 1))$integral
  return (a)
}

VARL <- function (delta,L,m,n) {
  a <- ARL2(delta,L,m,n) - (ARL(delta,L,m,n))^2
  return (a)
}

SDARL <- function (delta,L,m,n) {
  a <- sqrt( ARL2(delta,L,m,n) - (ARL(delta,L,m,n))^2)
  return (a)
}

quantileCARL <-function (p,delta,L,m,n) {
  CDFm <- function (a) {
    a <- CDFCPS(a,delta,L,m,n) - (1-p)
    return(a)
  }
  g<-1/secantc(CDFm,0.002,0.01)
  return(g)
}

# the functions plotCDFCARL (below) plots the c.d.f of the CARL0 of the Xbar chart in
case UU for a given limit factor (L), scaled shift in the process mean (delta), number (m)
and size (n) of Phase I samples. Note that if delta = 0, the function returns the in-control
results.

plotCDFCARL <- function (delta,L,m,n) {


  if (delta == 0) {
    CDFCARL12 <- Vectorize(CDFCARL)
    curve(CDFCARL12(x,delta,L,m,n),1,2000
,ylim=c(0,1),xlab="t",ylab="",cex.axis=1.5,type="l",lty=1,lwd=3,yaxs="i",xaxs="i",xaxt="n
",yaxt="n")
    title(main=paste("P(IC CARL <= t)","for", "L=",L, "m=",m, "n=",n,"delta=", delta ),
line=+2.5)
    xvalues<-c(0,200,400,600,800,1000,1200,1400,1600,1800,2000)
    yvalues<-c(0,0.2,0.4,0.6,0.8,1)
    axis(1,at=xvalues,cex.axis=1.5,las=1)
    axis(2,at=yvalues,cex.axis=1.5,las=1)
```

```
   ARLr <- round(ARL(delta,L,m,n),2)
   CDFmeanr <-round(CDFCARL(ARL(delta,L,m,n),delta,L,m,n),2)
   axis(3,ARLr,cex.axis=1,las=1)
   axis(4,CDFCARL(ARL(delta,L,m,n),delta,L,m,n),cex.axis=1,las=1)
   abline(v=ARL(delta,L,m,n),lty=5.5,col="blue")
   abline(h=CDFCARL(ARL(delta,L,m,n),delta,L,m,n),lty=5.5,col="blue")

   Median <- round(quantileCARL(0.5,delta,L,m,n),2)
   CDFmedianr <-round(CDFCARL(quantileCARL(0.5,delta,L,m,n),delta,L,m,n),2)
   axis(1,Median,cex.axis=1,las=1, line=1)
   axis(4,CDFCARL(quantileCARL(0.5,delta,L,m,n),delta,L,m,n),cex.axis=1,las=1)
   abline(v=quantileCARL(0.5,delta,L,m,n) ,lty=5.5,col="red")
   abline(h=CDFCARL(quantileCARL(0.5,delta,L,m,n),delta,L,m,n),lty=5.5,col="red")
   legend(1250, 0.4, c( paste("MCARL =", Median) , paste("ARL =", ARLr)), cex=1,
lty=c(5.5,5.5),lwd=c(1,1),col=c("red","blue"))
  }

  else {

   CDFCARL12 <- Vectorize(CDFCARL)
   curve(CDFCARL12(x,delta,L,m,n),1,100
,ylim=c(0,1),xlab="t",ylab="",cex.axis=1.5,type="l",lty=1,lwd=3,yaxs="i",xaxs="i",xaxt="n
",yaxt="n")
   title(main=paste("P( OOC CARL <= t)","for", "L=",L, "m=",m, "n=",n,"delta=", delta ),
line=+2.5)
   xvalues<-c(0,20,40,60,80,100)
   yvalues<-c(0,0.2,0.4,0.6,0.8,1)
   axis(1,at=xvalues,cex.axis=1.5,las=1)
   axis(2,at=yvalues,cex.axis=1.5,las=1)

   ARLr <- round(ARL(delta,L,m,n),2)
   CDFmeanr <-round(CDFCARL(ARL(delta,L,m,n),delta,L,m,n),2)
   axis(3,ARLr,cex.axis=1,las=1)
   axis(4,CDFCARL(ARL(delta,L,m,n),delta,L,m,n),cex.axis=1,las=1)
   abline(v=ARL(delta,L,m,n),lty=5.5,col="blue")
   abline(h=CDFCARL(ARL(delta,L,m,n),delta,L,m,n),lty=5.5,col="blue")

   Median <- round(quantileCARL(0.5,delta,L,m,n),2)
   CDFmedianr <-round(CDFCARL(quantileCARL(0.5,delta,L,m,n),delta,L,m,n),2)
   axis(1,Median,cex.axis=1,las=1, line=1)
   axis(4,CDFCARL(quantileCARL(0.5,delta,L,m,n),delta,L,m,n),cex.axis=1,las=1)
   abline(v=quantileCARL(0.5,delta,L,m,n) ,lty=5.5,col="red")
   abline(h=CDFCARL(quantileCARL(0.5,delta,L,m,n),delta,L,m,n),lty=5.5,col="red")
   legend(60, 0.4, c( paste("MCARL =", Median) , paste("ARL =", ARLr)), cex=1,
lty=c(5.5,5.5),lwd=c(1,1),col=c("red","blue"))
  }
}

dev.new()
plotCDFCARL(0,3,25,5)
```

```
dev.new()
plotCDFCARL(0.5,3,25,5)
```

the function plotPDFCARL (below) plots the p.d.f of the CARL0 of the Xbar chart in case UU for a given limit factor (L), scaled shift in the process mean (delta), number (m) and size (n) of Phase I samples. Note that if delta = 0, the function returns the in-control results.

```
plotPDFCARL <- function (delta,L,m,n) {

  if (delta == 0) {
    CDF <- function (h) {
      g <- CDFCARL(h,delta,L,m,n)
      return(g)
    }
    PDF <- function (x) {
      f <- grad(CDF, x)
      return(f)
    }
    PDF2 <- Vectorize(PDF)

curve(PDF2,1.01,2000,xlab="t",ylab="",n=100,cex.axis=1.5,type="l",lty=1,lwd=3,yaxs="i",
xaxs="i",xaxt="n")
    title(main=paste("pdf of the IC CARL","for", "L=",L, "m=",m, "n=",n  ), line=+2.5)
    xvalues<-c(0,200,400,600,800,1000,1200,1400,1600,1800,2000)
    axis(1,at=xvalues,cex.axis=1.5,las=1)

    ARLr <- round(ARL(delta,L,m,n),2)
    axis(3,ARLr,cex.axis=1,las=1)
    abline(v=ARL(delta,L,m,n),lty=5.5,col="blue")

    Median <- round(quantileCARL(0.5,delta,L,m,n),2)
    axis(1,Median,cex.axis=1,las=1,line=1)
    abline(v=quantileCARL(0.5,delta,L,m,n) ,lty=5.5,col="red")
  }

  else {

    CDF <- function (h) {
      g <- CDFCARL(h,delta,L,m,n)
      return(g)
    }
    CDF2 <- Vectorize(CDF )
    PDF <- function (x) {
      f <- grad(CDF2, x)
      return(f)
    }
    PDF2 <- Vectorize(PDF)

curve(PDF2,1.01,100,n=100,xlab="t",ylab="",cex.axis=1.5,type="l",lty=1,lwd=3,yaxs="i",x
axs="i",xaxt="n")
```

```
    title(main=paste("pdf of the OOC CARL","for", "L=",L, "m=",m, "n=",n,"delta=", delta
), line=+2.5)
    xvalues<-c(0,20,40,60,80,100)
    axis(1,at=xvalues,cex.axis=1.5,las=1)

    ARLr <- round(ARL(delta,L,m,n),2)
    axis(3,ARLr,cex.axis=1,las=1)
    abline(v=ARL(delta,L,m,n),lty=5.5,col="blue")

    Median <- round(quantileCARL(0.5,delta,L,m,n),2)
    axis(1,Median,cex.axis=1,las=1,line=1)
    abline(v=quantileCARL(0.5,delta,L,m,n) ,lty=5.5,col="red")

  }
}

dev.new()
plotPDFCARL(0,3,25,5)
dev.new()
plotPDFCARL(0.5,3,25,5)

# the codes below generate a table with the unconditional ARL values for a set of values
of n (row) and m (column), given a value of the scaled shift (delta) and limit factor (L)

m<-c(25,50,75,100,150,200,250)
n<-c(3,5,9)
delta <- 0
L <- 3
ARLtable<-matrix(,nrow = length(n), ncol = length(m))

for (i in 1:length(n)){
  for (j in 1:length(m)){
    ARLtable[i,j] <- ARL(delta,L,m[j],n[i])
    cat(ARLtable[i,j]," ")
  }
}
ARLtable

# the codes below generate a table with the SDARL values for a set of values of n (row)
and m (column), given a value of the scaled shift (delta) and limit factor (L)

m<-c(25,50,75,100,150,200,250)
n<-c(3,5,9)
delta <- 0
L <- 3
SDARLtable<-matrix(,nrow = length(n), ncol = length(m))

for (i in 1:length(n)){
  for (j in 1:length(m)){
    SDARLtable[i,j] <- SDARL(delta,L,m[j],n[i])
    cat(SDARLtable[i,j]," ")
```

```
  }
}
SDARLtable

# the codes below generate a table with the p-quantile values of the CARL0 for a set of values of n (row) and m (column), given a value of the scaled shift (delta) and limit factor (L)

m<-c(25,50,100,300,1000)
n<-c(5,10,20,25)
delta <- 0
L <- 3
p <-0.1
quantileTABLE<-matrix(,nrow = length(n), ncol = length(m))

for (i in 1:length(n)){
  for (j in 1:length(m)){
    quantileTABLE[i,j]<-quantileCARL(p,delta,L,m[j],n[i])
    cat(quantileTABLE[i,j], "  ")
  }
}
quantileTABLE

# the codes below generate a table with tht minimum values of m, which generates P(CFAR<(1+e)*alpha)=1-p for a set of values of n (row) and m (e), given a value of the scaled shift (delta), p, nominal alpha and limit factor (L)

e<-c(0.1,0.2,0.3,0.4,0.5)
n<-c(5,10,20,25)
L <- 3
p <- 0.15
alpha<-0.0027
delta <- 0
MINIMUMmTABLE<-matrix(,nrow = length(n), ncol = length(e))

for (i in 1:length(e)){
  for (j in 1:length(n)){
    CDFm <- function (m) {
      a <- CDFCPS((1+e[i])*alpha,delta,L,m,n[j]) - (1-p)
      return(a)
    }
    MINIMUMmTABLE[j,i]<-ceiling(secantc(CDFm,30,4000))
    cat(MINIMUMmTABLE[j,i],"  ")
  }
}
MINIMUMmTABLE

# the codes below generate a table with the adjusted value of L, which generates P(CFAR<(1+e)alpha)=1-p for a set of values of n (row) and m (column), given a value of the scaled shift (delta), p, nominal alpha and e
```

```
m<-c(25,50,100,300,1000)
n<-c(5,10,20,25)
alpha <- 0.0027
p<-0.05
e<-0.2
delta<-0
adjLtable<-matrix(,nrow = length(n), ncol = length(m))

for (i in 1:length(n)){
  for (j in 1:length(m)){
    CDFaux <- function (s) {
      a <- CDFCPS((1+e)*alpha,delta,s,m[j],n[i])-(1-p)
      return (a)
    }
    adjLtable[i,j]<-secantc(CDFaux,2.1,3.9)
    cat(adjLtable[i,j], " ")
  }
}
adjLtable
```

R CODES FOR CASE KU


```
# before running the codes below, please, download the packages cubature and
numDeriv

library(cubature) # this package helps compute double integrals
library(numDeriv) # this package helps compute numerical derivations

# the function secantc below was created to find the root of a monotonic increasing
function. Note that the precision is up to 10 decimal places.

secantc <- function(fun, x0, x1, tol=1e-10, niter=100000){
  for ( i in 1:niter ) {
    funx1 <- fun(x1)
    funx0 <- fun(x0)
    x2 <- ( (x0*funx1) - (x1*funx0) )/( funx1 - funx0 )
    funx2 <- fun(x2)
    if (abs(funx2) < tol) {
      return(x2)
    }
    if (funx2 < 0)
      x0 <- x2
    else
      x1 <- x2
  }
  stop("exceeded allowed number of iteractions")
}

# the function CARL and CPS below computes the Conditional Average Run Length
(CARL) and the Conditional Probability of a signal of the Xbar chart in Case KU for a given
of Y (chi-square random variable with n*(m-1) d.f., but note that Y = qchisq(U,m*(n-1)),
```

so the function below is actually in fucntion of qchisq(U,m*(n-1)) what is in function of U), limit factor (L), scaled shift in the process mean (delta), number (m) and size (n) of Phase I samples. Note that when delta = 0, CARL will return the in-control CARL (i.e., the CARL0).

```
CPS <- function (U,delta,L,m,n) {
  a <- 1-pnorm((-delta*sqrt(n))+(L*sqrt(qchisq(U,m*(n-1))/(m*(n-1)))),0,1)+pnorm((-
delta*sqrt(n))-(L*sqrt(qchisq(U,m*(n-1))/(m*(n-1)))),0,1)
  return(a)
}

CARL <- function (U,delta,L,m,n) {
  a <- 1/(1-pnorm((-delta*sqrt(n))+(L*sqrt(qchisq(U,m*(n-1))/(m*(n-1)))),0,1)+pnorm((-
delta*sqrt(n))-(L*sqrt(qchisq(U,m*(n-1))/(m*(n-1)))),0,1))
  return(a)
}

# the functions plotCPS and plotCARL below create, respectively, plots of the CPS and
CARL0 curves for various values of m, n=5 and L=3 in function of U. Please, use delta
between 0.5 and 1.5.

plotCPS <- function (delta,L,m,n) {
  if (delta == 0) {

curve(CPS(x,delta,L,10,n),0,1,ylim=c(0,0.02),xlim=c(0,1),cex.axis=1.5,xlab="u",cex.axis=1.
5,type="l",lty=1,col="black",yaxs="i",xaxs="i",ylab="",yaxt="n")
    curve(CPS(x,delta,L,20,n),0,1
,add=TRUE,xlab="u",ylab="CFAR",type="l",lty=1,lwd=3,col="black")
    curve(CPS(x,delta,L,50,n),0,1
,add=TRUE,xlab="u",ylab="CFAR",type="l",lty=2,lwd=3,col="black")
    curve(CPS(x,delta,L,100,n),0,1
,add=TRUE,xlab="u",ylab="CFAR",type="l",lty=2,lwd=3,col="black")
    curve(CPS(x,delta,L,500,n),0,1 ,add=TRUE,xlab="u",ylab="CFAR",type="l",lty=3,
lwd=3,col="black")
    yvalues<-c(0,0.0027,0.005,0.01,0.015,0.02)
    axis(2,at=yvalues,labels=yvalues,cex.axis=1.5,las=1)
    abline(a = 0.0027, b = 0,lty=5.5)
    legend(0.4, 0.0155, c("m = 10","m = 20","m = 50","m = 100","m = 500", "CFAR =
0.0027"), cex=1.5, lty=c(1,1,2,2,3,5.5),lwd=c(3,3,3,3,3,0));
    title(main=paste("CFAR curves","for", "L=",L, "n=",n,"delta=", delta   ))
  }
  else {
    l <- 0.75*delta - 0.225

curve(CPS(x,delta,L,10,n),0,1,ylim=c(0,l),xlim=c(0,1),cex.axis=1.5,xlab="u",cex.axis=1.5,ty
pe="l",lty=1,col="black",yaxs="i",xaxs="i",ylab="")
    curve(CPS(x,delta,L,20,n),0,1
,add=TRUE,xlab="u",ylab="CFAR",type="l",lty=1,lwd=3,col="black")
    curve(CPS(x,delta,L,50,n),0,1
,add=TRUE,xlab="u",ylab="CFAR",type="l",lty=2,lwd=3,col="black")
```

```
   curve(CPS(x,delta,L,100,n),0,1
,add=TRUE,xlab="u",ylab="CFAR",type="l",lty=2,lwd=3,col="black")
   curve(CPS(x,delta,L,500,n),0,1 ,add=TRUE,xlab="u",ylab="CFAR",type="l",lty=3,
lwd=3,col="black")
   legend(0.4, l, c("m = 10","m = 20","m = 50","m = 100","m = 500"), cex=1.5,
lty=c(1,1,2,2,3,5.5),lwd=c(3,3,3,3,3,0));
   title(main=paste("CPS curves","for", "L=",L, "n=",n,"delta=", delta   ))
  }
}

plotCARL <- function (delta,L,m,n) {
  if (delta == 0) {

curve(CARL(x,delta,L,10,n),0,1,ylim=c(0,2000),cex.axis=1.5,yaxt="n",xlab="u",yaxs="i",xa
xs="i",ylab="",type="l",lty=1)
   curve(CARL(x,delta,L,20,n),0,1 ,add=TRUE,xlab="u",ylab="CFAR",type="l",lty=1,lwd=3)
   curve(CARL(x,delta,L,50,n),0,1 ,add=TRUE,xlab="u",ylab="CFAR",type="l",lty=2)
   curve(CARL(x,delta,L,100,n),0,1
,add=TRUE,xlab="u",ylab="CFAR",type="l",lty=2,lwd=3)
   curve(CARL(x,delta,L,500,n),0,1 ,add=TRUE,xlab="u",ylab="CFAR",type="l",lty=3,
lwd=3)
   yvalues<-c(0,370.4,500,1000,1500,2000)
   axis(2,at=yvalues,labels=yvalues,cex.axis=1.5,las=1)
   abline(a = 370.4, b = 0,lty=5.5)
   legend(0.2, 1500, c("m = 10","m = 20","m = 50","m = 100","m = 500", "CFAR = 370.4"),
cex=1.5, lty=c(1,1,2,2,3,5.5),lwd=c(0,3,0,3,3,0));
   title(main=paste("CARL curves","for", "L=",L, "n=",n,"delta=", delta   ))
  }
  else {
   l <- (-76*delta) + 118

curve(CARL(x,delta,L,10,n),0,1,ylim=c(1,l),cex.axis=1.5,xlab="u",yaxs="i",xaxs="i",ylab=""
,type="l",lty=1)
   curve(CARL(x,delta,L,20,n),0,1 ,add=TRUE,xlab="u",ylab="CFAR",type="l",lty=1,lwd=3)
   curve(CARL(x,delta,L,50,n),0,1 ,add=TRUE,xlab="u",ylab="CFAR",type="l",lty=2)
   curve(CARL(x,delta,L,100,n),0,1
,add=TRUE,xlab="u",ylab="CFAR",type="l",lty=2,lwd=3)
   curve(CARL(x,delta,L,500,n),0,1 ,add=TRUE,xlab="u",ylab="CFAR",type="l",lty=3,
lwd=3)
   legend(0.2, l, c("m = 10","m = 20","m = 50","m = 100","m = 500", "CFAR = 370.4"),
cex=1.5, lty=c(1,1,2,2,3,5.5),lwd=c(0,3,0,3,3,0));
   title(main=paste("CARL curves","for", "L=",L, "n=",n,"delta=", delta   ))
  }
}

dev.new()
plotCPS(0,3,25,5)
dev.new()
plotCARL(0,3,25,5)
```

the functions CDFCPS and CDFCARL (below) computes, respectively, the c.d.f. (for any value t) of the Conditional Probability of a Signal (CPS) and CARL of the Xbar chart for a given limit factor (L), scaled shift in the process mean (delta), number (m) and size (n) of Phase I samples. Note that when delta = 0, CDFCPS and CDFCARL will return the in-control c.d.f. of the CPS and CARL (i.e., cdf of the CFAR, the conditional False Alarm Rate and CARL0).

```
CDFCPS <- function (t,delta,L,m,n) {
 a <- 1 - pchisq((m*(n-1)*qchisq(1-t, df=1, ncp = (delta^2)*n))/(L^2),m*(n-1))
 return(a)
}

CDFCARL <- function (t,delta,L,m,n) {
 a <- pchisq((m*(n-1)*qchisq(1-(1/t), df=1, ncp = (delta^2)*n))/(L^2),m*(n-1))
 return(a)
}
```

the functions ARL, ARL2, VARL, SDARL2 and quantileCARL (below) compute, respectively, the mean, the central second moment, the variance, the standard deviation and the p-quantile of the CARL0 of the Xbar chart in case KU for a given limit factor (L), scaled shift in the process mean (delta), number (m) and size (n) of Phase I samples. Note that if delta = 0, the function returns the in-control values.

```
ARL <- function (delta,L,m,n) {
 CARL <- function (U) {
  a <- 1/(1-pnorm((-delta*sqrt(n))+(L*sqrt(qchisq(U,m*(n-1))/(m*(n-1)))),0,1)+pnorm((-
delta*sqrt(n))-(L*sqrt(qchisq(U,m*(n-1))/(m*(n-1)))),0,1))
  return(a)
 }
 a <- integrate(CARL,0,1)$va
 return(a)
}

ARL2 <- function (delta,L,m,n) {
 CARL <- function (U) {
  a <- (1/(1-pnorm((-delta*sqrt(n))+(L*sqrt(qchisq(U,m*(n-1))/(m*(n-1)))),0,1)+pnorm((-
delta*sqrt(n))-(L*sqrt(qchisq(U,m*(n-1))/(m*(n-1)))),0,1)))^2
  return(a)
 }
 a <- integrate(CARL,0,1)$va
 return(a)
}

VARL <- function (delta,L,m,n) {
 a <- ARL2(delta,L,m,n) - (ARL(delta,L,m,n))^2
 return (a)
}

SDARL <- function (delta,L,m,n) {
 a <- sqrt( ARL2(delta,L,m,n) - (ARL(delta,L,m,n))^2)
 return (a)
```

```
}

quantileCARL <-function (p,delta,L,m,n) {
  g <- 1/(1-pchisq((((L^2)*qchisq(p,m*(n-1)))/(m*(n-1))), df=1, ncp = (delta^2)*n))
  return(g)
}
```

the functions plotCDFCARL (below) plots the c.d.f of the CARL0 of the Xbar chart in case KU for a given limit factor (L), scaled shift in the process mean (delta), number (m) and size (n) of Phase I samples. Note that if delta = 0, the function returns the in-control results.

```
plotCDFCARL <- function (delta,L,m,n) {

  if (delta == 0) {
    CDFCARL12 <- Vectorize(CDFCARL)
    curve(CDFCARL12(x,delta,L,m,n),1,2000
,ylim=c(0,1),xlab="t",ylab="",cex.axis=1.5,type="l",lty=1,lwd=3,yaxs="i",xaxs="i",xaxt="n
",yaxt="n")
    title(main=paste("P(IC CARL <= t)","for", "L=",L, "m=",m, "n=",n,"delta=", delta ),
line=+2.5)
    xvalues<-c(0,200,400,600,800,1000,1200,1400,1600,1800,2000)
    yvalues<-c(0,0.2,0.4,0.6,0.8,1)
    axis(1,at=xvalues,cex.axis=1.5,las=1)
    axis(2,at=yvalues,cex.axis=1.5,las=1)

    ARLr <- round(ARL(delta,L,m,n),2)
    CDFmeanr <-round(CDFCARL(ARL(delta,L,m,n),delta,L,m,n),2)
    axis(3,ARLr,cex.axis=1,las=1)
    axis(4,CDFCARL(ARL(delta,L,m,n),delta,L,m,n),cex.axis=1,las=1)
    abline(v=ARL(delta,L,m,n),lty=5.5,col="blue")
    abline(h=CDFCARL(ARL(delta,L,m,n),delta,L,m,n),lty=5.5,col="blue")

    Median <- round(quantileCARL(0.5,delta,L,m,n),2)
    CDFmedianr <-round(CDFCARL(quantileCARL(0.5,delta,L,m,n),delta,L,m,n),2)
    axis(1,Median,cex.axis=1,las=1, line=1)
    axis(4,CDFCARL(quantileCARL(0.5,delta,L,m,n),delta,L,m,n),cex.axis=1,las=1)
    abline(v=quantileCARL(0.5,delta,L,m,n) ,lty=5.5,col="red")
    abline(h=CDFCARL(quantileCARL(0.5,delta,L,m,n),delta,L,m,n),lty=5.5,col="red")
    legend(1250, 0.4, c( paste("MCARL =", Median) , paste("ARL =", ARLr)), cex=1,
lty=c(5.5,5.5),lwd=c(1,1),col=c("red","blue"))
  }

  else {
    CDFCARL12 <- Vectorize(CDFCARL)
    curve(CDFCARL12(x,delta,L,m,n),1,100
,ylim=c(0,1),xlab="t",ylab="",cex.axis=1.5,type="l",lty=1,lwd=3,yaxs="i",xaxs="i",xaxt="n
",yaxt="n")
    title(main=paste("P( OOC CARL <= t)","for", "L=",L, "m=",m, "n=",n,"delta=", delta ),
line=+2.5)
    xvalues<-c(0,20,40,60,80,100)
```

```
   yvalues<-c(0,0.2,0.4,0.6,0.8,1)
   axis(1,at=xvalues,cex.axis=1.5,las=1)
   axis(2,at=yvalues,cex.axis=1.5,las=1)

   ARLr <- round(ARL(delta,L,m,n),2)
   CDFmeanr <-round(CDFCARL(ARL(delta,L,m,n),delta,L,m,n),2)
   axis(3,ARLr,cex.axis=1,las=1)
   axis(4,CDFCARL(ARL(delta,L,m,n),delta,L,m,n),cex.axis=1,las=1)
   abline(v=ARL(delta,L,m,n),lty=5.5,col="blue")
   abline(h=CDFCARL(ARL(delta,L,m,n),delta,L,m,n),lty=5.5,col="blue")

   Median <- round(quantileCARL(0.5,delta,L,m,n),2)
   CDFmedianr <-round(CDFCARL(quantileCARL(0.5,delta,L,m,n),delta,L,m,n),2)
   axis(1,Median,cex.axis=1,las=1, line=1)
   axis(4,CDFCARL(quantileCARL(0.5,delta,L,m,n),delta,L,m,n),cex.axis=1,las=1)
   abline(v=quantileCARL(0.5,delta,L,m,n) ,lty=5.5,col="red")
   abline(h=CDFCARL(quantileCARL(0.5,delta,L,m,n),delta,L,m,n),lty=5.5,col="red")
   legend(60, 0.4, c( paste("MCARL =", Median) , paste("ARL =", ARLr)), cex=1,
lty=c(5.5,5.5),lwd=c(1,1),col=c("red","blue"))
  }
}

dev.new()
plotCDFCARL(0,3,25,3)
dev.new()
plotCDFCARL(0.5,3,25,5)

# the function plotPDFCARL (below) plots the p.d.f of the CARL0 of the Xbar chart in case
KU for a given limit factor (L), scaled shift in the process mean (delta), number (m) and
size (n) of Phase I samples. Note that if delta = 0, the function returns the in-control
results.

plotPDFCARL <- function (delta,L,m,n) {

  if (delta == 0) {
   CDF <- function (h) {
     g <- CDFCARL(h,delta,L,m,n)
     return(g)
   }
   PDF <- function (x) {
     f <- grad(CDF, x)
     return(f)
   }
   PDF2 <- Vectorize(PDF)

curve(PDF2,1.01,2000,xlab="t",ylab="",n=100,cex.axis=1.5,type="l",lty=1,lwd=3,yaxs="i",
xaxs="i",xaxt="n")
   title(main=paste("pdf of the IC CARL","for", "L=",L, "m=",m, "n=",n  ), line=+2.5)
   xvalues<-c(0,200,400,600,800,1000,1200,1400,1600,1800,2000)
   axis(1,at=xvalues,cex.axis=1.5,las=1)
```

```
  ARLr <- round(ARL(delta,L,m,n),2)
  axis(3,ARLr,cex.axis=1,las=1)
  abline(v=ARL(delta,L,m,n),lty=5.5,col="blue")

  Median <- round(quantileCARL(0.5,delta,L,m,n),2)
  axis(1,Median,cex.axis=1,las=1,line=1)
  abline(v=quantileCARL(0.5,delta,L,m,n) ,lty=5.5,col="red")
 }

 else {

  CDF <- function (h) {
   g <- CDFCARL(h,delta,L,m,n)
   return(g)
  }
  CDF2 <- Vectorize(CDF )
  PDF <- function (x) {
   f <- grad(CDF2, x)
   return(f)
  }
  PDF2 <- Vectorize(PDF)

curve(PDF2,1.01,100,n=100,xlab="t",ylab="",cex.axis=1.5,type="l",lty=1,lwd=3,yaxs="i",x
axs="i",xaxt="n")
  title(main=paste("pdf of the OOC CARL","for", "L=",L, "m=",m, "n=",n,"delta=", delta
), line=+2.5)
  xvalues<-c(0,20,40,60,80,100)
  axis(1,at=xvalues,cex.axis=1.5,las=1)

  ARLr <- round(ARL(delta,L,m,n),2)
  axis(3,ARLr,cex.axis=1,las=1)
  abline(v=ARL(delta,L,m,n),lty=5.5,col="blue")

  Median <- round(quantileCARL(0.5,delta,L,m,n),2)
  axis(1,Median,cex.axis=1,las=1,line=1)
  abline(v=quantileCARL(0.5,delta,L,m,n) ,lty=5.5,col="red")

 }
}

dev.new()
plotPDFCARL(0,3,25,5)
dev.new()
plotPDFCARL(0.5,3,25,5)

# the codes below generate a table with the unconditional ARL values for a set of values
of n (row) and m (column), given a value of the scaled shift (delta) and limit factor (L)

m<-c(25,50,75,100,150,200,250)
n<-c(3,5,9)
delta <- 0
```

```
L <- 3
ARLtable<-matrix(,nrow = length(n), ncol = length(m))

for (i in 1:length(n)){
  for (j in 1:length(m)){
    ARLtable[i,j] <- ARL(delta,L,m[j],n[i])
    cat(ARLtable[i,j]," ")
  }
}
ARLtable

# the codes below generate a table with the SDARL values for a set of values of n (row)
and m (column), given a value of the scaled shift (delta) and limit factor (L)

m<-c(25,50,75,100,150,200,250)
n<-c(3,5,9)
delta <- 0
L <- 3
SDARLtable<-matrix(,nrow = length(n), ncol = length(m))

for (i in 1:length(n)){
  for (j in 1:length(m)){
    SDARLtable[i,j] <- SDARL(delta,L,m[j],n[i])
    cat(SDARLtable[i,j]," ")
  }
}
SDARLtable

# the codes below generate a table with the p-quantile values of the CARL0 for a set of
values of n (row) and m (column), given a value of the scaled shift (delta) and limit factor
(L)

m<-c(25,50,100,300,1000)
n<-c(5,10,20,25)
delta <- 0
L <- 3
p <-0.1
quantileTABLE<-matrix(,nrow = length(n), ncol = length(m))


for (i in 1:length(n)){
  for (j in 1:length(m)){
    quantileTABLE[i,j]<-quantileCARL(p,delta,L,m[j],n[i])
    cat(quantileTABLE[i,j], "  ")
  }
}
quantileTABLE

# the codes below generate a table with tht minimum values of m, which generates
P(CFAR<(1+e)*alpha)=1-p for a set of values of n (row) and m (e), given a value of the
scaled shift (delta), p, nominal alpha and limit factor (L)
```

```
e<-c(0.1,0.2,0.3,0.4,0.5)
n<-c(5,10,20,25)
L <- 3
p <- 0.15
alpha<-0.0027
delta <- 0
MINIMUMmTABLE<-matrix(,nrow = length(n), ncol = length(e))

for (i in 1:length(e)){
  for (j in 1:length(n)){
    CDFm <- function (m) {
      a <- CDFCPS((1+e[i])*alpha,delta,L,m,n[j]) - (1-p)
      return(a)
    }
    MINIMUMmTABLE[j,i]<-ceiling(secantc(CDFm,5,4000))
    cat(MINIMUMmTABLE[j,i],"  ")
  }
}
MINIMUMmTABLE

# the codes below generate a table with the adjusted value of L, which generates
P(CFAR<(1+e)alpha)=1-p for a set of values of n (row) and m (column), given a value of
the scaled shift (delta), p, nominal alpha and e

m<-c(25,50,100,300,1000)
n<-c(3,5,9,15)
alpha <- 0.0027
p<-0.05
e<-0
adjLtable<-matrix(,nrow = length(n), ncol = length(m))

adjL <- function (p,m,n,e) {
  alfatol=(1+e)*0.0027
  zalfatol2 <- -1*qnorm(alfatol/2)
  g <-  zalfatol2/sqrt(qchisq(p,m*(n-1))/(m*(n-1)))
  return(g)
}

for (i in 1:length(n)){
  for (j in 1:length(m)){
    adjLtable[i,j]<-adjL(p,m[j],n[i],e)
    cat(adjLtable[i,j], " ")
  }
}
adjLtable
```

R CODES FOR CASE UK

```
# before running the codes below, please, download the packages cubature and
numDeriv
```

```
library(cubature) # this package helps compute double integrals
library(numDeriv) # this package helps compute numerical derivations

# the function secantc below was created to find the root of a monotonic increasing
function. Note that the precision is up to 10 decimal places.

secantc <- function(fun, x0, x1, tol=1e-10, niter=100000){
  for ( i in 1:niter ) {
    funx1 <- fun(x1)
    funx0 <- fun(x0)
    x2 <- ( (x0*funx1) - (x1*funx0) )/( funx1 - funx0 )
    funx2 <- fun(x2)
    if (abs(funx2) < tol) {
      return(x2)
    }
    if (funx2 < 0)
      x0 <- x2
    else
      x1 <- x2
  }
  stop("exceeded allowed number of iteractions")
}

# the function CARL and CPS below computes the Conditional Average Run Length
(CARL) and the Conditional Probability of a signal of the Xbar chart in Case KU for a given
of Z (standard bormal random varianle, but note that Z = qnorm(U,0,1)), so the function
below is actually in fucntion of qchisq(U,m*(n-1)) which is in function of U), limit factor
(L), scaled shift in the process mean (delta), number (m) and size (n) of Phase I samples.
Note that when delta = 0, CARL will return the in-control CARL (i.e., the CARL0).

CPS <- function (U,delta,L,m,n) {
  a <- 1 - pnorm((-delta*sqrt(n))+((1/sqrt(m))*qnorm(U,0,1))+L,0,1) + pnorm((-
delta*sqrt(n))+((1/sqrt(m))*qnorm(U,0,1))-L,0,1)
  return(a)
}

CARL <- function (U,delta,L,m,n) {
  a <- 1/(1 - pnorm((-delta*sqrt(n))+((1/sqrt(m))*qnorm(U,0,1))+L,0,1) + pnorm((-
delta*sqrt(n))+((1/sqrt(m))*qnorm(U,0,1))-L,0,1))
  return(a)
}

# the functions plotCPS and plotCARL below create, respectively, plots of the CPS and
CARL0 curves for various values of m, n=5 and L=3 in function of U. Please, use delta
between 0.5 and 1.5.

plotCPS <- function (delta,L,m,n) {
  if (delta == 0) {
```

```
curve(CPS(x,delta,L,10,n),0,1,ylim=c(0,0.02),xlim=c(0,1),cex.axis=1.5,xlab="u",cex.axis=1.
5,type="l",lty=1,col="black",yaxs="i",xaxs="i",ylab="",yaxt="n")
   curve(CPS(x,delta,L,20,n),0,1
,add=TRUE,xlab="u",ylab="CFAR",type="l",lty=1,lwd=3,col="black")
   curve(CPS(x,delta,L,50,n),0,1
,add=TRUE,xlab="u",ylab="CFAR",type="l",lty=2,lwd=3,col="black")
   curve(CPS(x,delta,L,100,n),0,1
,add=TRUE,xlab="u",ylab="CFAR",type="l",lty=2,lwd=3,col="black")
   curve(CPS(x,delta,L,500,n),0,1 ,add=TRUE,xlab="u",ylab="CFAR",type="l",lty=3,
lwd=3,col="black")
   yvalues<-c(0,0.0027,0.005,0.01,0.015,0.02)
   axis(2,at=yvalues,labels=yvalues,cex.axis=1.5,las=1)
   abline(a = 0.0027, b = 0,lty=5.5)
   legend(0.4, 0.0155, c("m = 10","m = 20","m = 50","m = 100","m = 500", "CFAR =
0.0027"), cex=1.5, lty=c(1,1,2,2,3,5.5),lwd=c(3,3,3,3,3,0));
   title(main=paste("CFAR curves","for", "L=",L, "n=",n,"delta=", delta   ))
  }
  else {
   l <- 0.75*delta - 0.225

curve(CPS(x,delta,L,10,n),0,1,ylim=c(0,l),xlim=c(0,1),cex.axis=1.5,xlab="u",cex.axis=1.5,ty
pe="l",lty=1,col="black",yaxs="i",xaxs="i",ylab="")
   curve(CPS(x,delta,L,20,n),0,1
,add=TRUE,xlab="u",ylab="CFAR",type="l",lty=1,lwd=3,col="black")
   curve(CPS(x,delta,L,50,n),0,1
,add=TRUE,xlab="u",ylab="CFAR",type="l",lty=2,lwd=3,col="black")
   curve(CPS(x,delta,L,100,n),0,1
,add=TRUE,xlab="u",ylab="CFAR",type="l",lty=2,lwd=3,col="black")
   curve(CPS(x,delta,L,500,n),0,1 ,add=TRUE,xlab="u",ylab="CFAR",type="l",lty=3,
lwd=3,col="black")
   legend(0.4, l, c("m = 10","m = 20","m = 50","m = 100","m = 500"), cex=1.5,
lty=c(1,1,2,2,3,5.5),lwd=c(3,3,3,3,3,0));
   title(main=paste("CPS curves","for", "L=",L, "n=",n,"delta=", delta   ))
  }
}

plotCARL <- function (delta,L,m,n) {
  if (delta == 0) {

curve(CARL(x,delta,L,10,n),0,1,ylim=c(0,400),cex.axis=1.5,yaxt="n",xlab="u",yaxs="i",xax
s="i",ylab="",type="l",lty=1)
   curve(CARL(x,delta,L,20,n),0,1 ,add=TRUE,xlab="u",ylab="CFAR",type="l",lty=1,lwd=3)
   curve(CARL(x,delta,L,50,n),0,1 ,add=TRUE,xlab="u",ylab="CFAR",type="l",lty=2)
   curve(CARL(x,delta,L,100,n),0,1
,add=TRUE,xlab="u",ylab="CFAR",type="l",lty=2,lwd=3)
   curve(CARL(x,delta,L,500,n),0,1 ,add=TRUE,xlab="u",ylab="CFAR",type="l",lty=3,
lwd=3)
   yvalues<-c(100,200,300,370.4,400)
   axis(2,at=yvalues,labels=yvalues,cex.axis=1.5,las=1)
   abline(a = 370.4, b = 0,lty=5.5)
```

```
    legend(0.2, 1500, c("m = 10","m = 20","m = 50","m = 100","m = 500", "CFAR = 370.4"),
cex=1.5, lty=c(1,1,2,2,3,5.5),lwd=c(0,3,0,3,3,0));
    title(main=paste("CARL curves","for", "L=",L, "n=",n,"delta=", delta   ))
  }
  else {
    l <- (-76*delta) + 118

curve(CARL(x,delta,L,10,n),0,1,ylim=c(1,l),cex.axis=1.5,xlab="u",yaxs="i",xaxs="i",ylab=""
,type="l",lty=1)
    curve(CARL(x,delta,L,20,n),0,1 ,add=TRUE,xlab="u",ylab="CFAR",type="l",lty=1,lwd=3)
    curve(CARL(x,delta,L,50,n),0,1 ,add=TRUE,xlab="u",ylab="CFAR",type="l",lty=2)
    curve(CARL(x,delta,L,100,n),0,1
,add=TRUE,xlab="u",ylab="CFAR",type="l",lty=2,lwd=3)
    curve(CARL(x,delta,L,500,n),0,1 ,add=TRUE,xlab="u",ylab="CFAR",type="l",lty=3,
lwd=3)
    legend(0.2, l, c("m = 10","m = 20","m = 50","m = 100","m = 500", "CFAR = 370.4"),
cex=1.5, lty=c(1,1,2,2,3,5.5),lwd=c(0,3,0,3,3,0));
    title(main=paste("CARL curves","for", "L=",L, "n=",n,"delta=", delta   ))
  }
}

dev.new()
plotCPS(0,3,25,5)
dev.new()
plotCARL(0,3,25,5)

# the functions CDFCPS and CDFCARL (below) computes, respectively, the c.d.f. (for any
value t) of the Conditional Probability of a Signal (CPS) and CARL of the Xbar chart for a
given limit factor (L), scaled shift in the process mean (delta), number (m) and size (n) of
Phase I samples. Note that when delta = 0, CDFCPS and CDFCARL will return the in-
control c.d.f. of the CPS and CARL (i.e., cdf of the CFAR, the conditional False Alarm Rate
and CARL0).

CDFCARL <- function (x,delta,L,m,n) {
  l <- 1/(1 - pnorm((-delta*sqrt(n))+((1/sqrt(m))*(delta*sqrt(m*n)))+L,0,1) + pnorm((-
delta*sqrt(n))+((1/sqrt(m))*(delta*sqrt(m*n)))-L,0,1))
  CARL <- function (Z) {
    a <- 1/(1 - pnorm((-delta*sqrt(n))+((1/sqrt(m))*Z)+L,0,1) + pnorm((-
delta*sqrt(n))+((1/sqrt(m))*Z)-L,0,1))
    return(a)
  }
  ARL0 <- function (t) {
    k <- CARL(t)-x
    return(k)
  }
  if (x>l) {
    c <- 1
  }
  if ((x<=l) & (x>=1) ) {
    b <- secantc(ARL0,-100,delta*sqrt(m*n))
    c <-  1 -  pnorm((2*(delta*sqrt(m*n)))-b,0,1) + pnorm(b,0,1)
```

```
  }
  if (x<1) {
    c <- 0
  }
  return(c)
}

CDFCPS <- function (t,delta,L,m,n) {
  a <- 1 - CDFCARL(1/t,delta,L,m,n)
  return(a)
}
```

the functions ARL, ARL2, VARL, SDARL2 and quantileCARL (below) compute, respectively, the mean, the central second moment, the variance, the standard deviation and the p-quantile of the CARL0 of the Xbar chart in case KU for a given limit factor (L), scaled shift in the process mean (delta), number (m) and size (n) of Phase I samples. Note that if delta = 0, the function returns the in-control values.

```
ARL <- function (delta,L,m,n) {
  CARL <- function (U) {
    a <- 1/(1 - pnorm((-delta*sqrt(n))+((1/sqrt(m))*qnorm(U,0,1))+L,0,1) + pnorm((-
delta*sqrt(n))+((1/sqrt(m))*qnorm(U,0,1))-L,0,1))
    return(a)
  }
  a <- integrate(CARL,0,1)$va
  return(a)
}

ARL2 <- function (delta,L,m,n) {
  CARL <- function (U) {
    a <- (1/(1 - pnorm((-delta*sqrt(n))+((1/sqrt(m))*qnorm(U,0,1))+L,0,1) + pnorm((-
delta*sqrt(n))+((1/sqrt(m))*qnorm(U,0,1))-L,0,1)))^2
    return(a)
  }
  a <- integrate(CARL,0,1)$va
  return(a)
}

VARL <- function (delta,L,m,n) {
  a <- ARL2(delta,L,m,n) - (ARL(delta,L,m,n))^2
  return (a)
}

SDARL <- function (delta,L,m,n) {
  a <- sqrt( ARL2(delta,L,m,n) - (ARL(delta,L,m,n))^2)
  return (a)
}

quantileCARL <-function (p,delta,L,m,n) {
  l <- 1/(1 - pnorm((-delta*sqrt(n))+((1/sqrt(m))*(delta*sqrt(m*n)))+L,0,1) + pnorm((-
delta*sqrt(n))+((1/sqrt(m))*(delta*sqrt(m*n)))-L,0,1))
```

```
  CDFm <- function (a) {
    a <- CDFCARL(a,delta,L,m,n) - p
    return(a)
  }
  g<-secantc(CDFm,1,l)
  return(g)
}

plotCDFCARL <- function (delta,L,m,n) {

  if (delta == 0) {
    CDFCARL12 <- Vectorize(CDFCARL)
    curve(CDFCARL12(x,delta,L,m,n),1,400
,ylim=c(0,1),xlab="t",ylab="",cex.axis=1.5,type="l",lty=1,lwd=3,yaxs="i",xaxs="i",xaxt="n
",yaxt="n")
    title(main=paste("P(IC CARL <= t)","for", "L=",L, "m=",m, "n=",n,"delta=", delta ),
line=+2.5)
    xvalues<-c(0,100,200,300,400)
    yvalues<-c(0,0.2,0.4,0.6,0.8,1)
    axis(1,at=xvalues,cex.axis=1.5,las=1)
    axis(2,at=yvalues,cex.axis=1.5,las=1)

    ARLr <- round(ARL(delta,L,m,n),2)
    CDFmeanr <-round(CDFCARL(ARL(delta,L,m,n),delta,L,m,n),2)
    axis(3,ARLr,cex.axis=1,las=1)
    axis(4,CDFCARL(ARL(delta,L,m,n),delta,L,m,n),cex.axis=1,las=1)
    abline(v=ARL(delta,L,m,n),lty=5.5,col="blue")
    abline(h=CDFCARL(ARL(delta,L,m,n),delta,L,m,n),lty=5.5,col="blue")

    Median <- round(quantileCARL(0.5,delta,L,m,n),2)
    CDFmedianr <-round(CDFCARL(quantileCARL(0.5,delta,L,m,n),delta,L,m,n),2)
    axis(1,Median,cex.axis=1,las=1, line=1)
    axis(4,CDFCARL(quantileCARL(0.5,delta,L,m,n),delta,L,m,n),cex.axis=1,las=1)
    abline(v=quantileCARL(0.5,delta,L,m,n) ,lty=5.5,col="red")
    abline(h=CDFCARL(quantileCARL(0.5,delta,L,m,n),delta,L,m,n),lty=5.5,col="red")
    legend(1250, 0.4, c( paste("MCARL =", Median) , paste("ARL =", ARLr)), cex=1,
lty=c(5.5,5.5),lwd=c(1,1),col=c("red","blue"))
  }

  else {

    CDFCARL12 <- Vectorize(CDFCARL)
    curve(CDFCARL12(x,delta,L,m,n),1,100
,ylim=c(0,1),xlab="t",ylab="",cex.axis=1.5,type="l",lty=1,lwd=3,yaxs="i",xaxs="i",xaxt="n
",yaxt="n")
    title(main=paste("P( OOC CARL <= t)","for", "L=",L, "m=",m, "n=",n,"delta=", delta ),
line=+2.5)
    xvalues<-c(0,20,40,60,80,100)
    yvalues<-c(0,0.2,0.4,0.6,0.8,1)
    axis(1,at=xvalues,cex.axis=1.5,las=1)
    axis(2,at=yvalues,cex.axis=1.5,las=1)
```

```
   ARLr <- round(ARL(delta,L,m,n),2)
   CDFmeanr <-round(CDFCARL(ARL(delta,L,m,n),delta,L,m,n),2)
   axis(3,ARLr,cex.axis=1,las=1)
   axis(4,CDFCARL(ARL(delta,L,m,n),delta,L,m,n),cex.axis=1,las=1)
   abline(v=ARL(delta,L,m,n),lty=5.5,col="blue")
   abline(h=CDFCARL(ARL(delta,L,m,n),delta,L,m,n),lty=5.5,col="blue")

   Median <- round(quantileCARL(0.5,delta,L,m,n),2)
   CDFmedianr <-round(CDFCARL(quantileCARL(0.5,delta,L,m,n),delta,L,m,n),2)
   axis(1,Median,cex.axis=1,las=1, line=1)
   axis(4,CDFCARL(quantileCARL(0.5,delta,L,m,n),delta,L,m,n),cex.axis=1,las=1)
   abline(v=quantileCARL(0.5,delta,L,m,n) ,lty=5.5,col="red")
   abline(h=CDFCARL(quantileCARL(0.5,delta,L,m,n),delta,L,m,n),lty=5.5,col="red")
   legend(60, 0.4, c( paste("MCARL =", Median) , paste("ARL =", ARLr)), cex=1,
lty=c(5.5,5.5),lwd=c(1,1),col=c("red","blue"))
  }
}

dev.new()
plotCDFCARL(0,3,25,5)
dev.new()
plotCDFCARL(0.5,3,25,5)

# the function plotPDFCARL (below) plots the p.d.f of the CARL0 of the Xbar chart in case
KU for a given limit factor (L), scaled shift in the process mean (delta), number (m) and
size (n) of Phase I samples. Note that if delta = 0, the function returns the in-control
results.

plotPDFCARL <- function (delta,L,m,n) {

  if (delta == 0) {
   CDF <- function (h) {
    g <- CDFCARL(h,delta,L,m,n)
    return(g)
   }
   PDF <- function (x) {
    f <- grad(CDF, x)
    return(f)
   }
   PDF2 <- Vectorize(PDF)

curve(PDF2,1.01,400,xlab="t",ylab="",n=100,cex.axis=1.5,type="l",lty=1,lwd=3,yaxs="i",x
axs="i",xaxt="n")
   title(main=paste("pdf of the IC CARL","for", "L=",L, "m=",m, "n=",n  ), line=+2.5)
   xvalues<-c(0,100,200,300,400)
   axis(1,at=xvalues,cex.axis=1.5,las=1)

   ARLr <- round(ARL(delta,L,m,n),2)
   axis(3,ARLr,cex.axis=1,las=1)
   abline(v=ARL(delta,L,m,n),lty=5.5,col="blue")
```

```
    Median <- round(quantileCARL(0.5,delta,L,m,n),2)
    axis(1,Median,cex.axis=1,las=1,line=1)
    abline(v=quantileCARL(0.5,delta,L,m,n) ,lty=5.5,col="red")
  }

  else {

    CDF <- function (h) {
      g <- CDFCARL(h,delta,L,m,n)
      return(g)
    }
    CDF2 <- Vectorize(CDF )
    PDF <- function (x) {
      f <- grad(CDF2, x)
      return(f)
    }
    PDF2 <- Vectorize(PDF)

curve(PDF2,1.01,100,n=100,xlab="t",ylab="",cex.axis=1.5,type="l",lty=1,lwd=3,yaxs="i",x
axs="i",xaxt="n")
    title(main=paste("pdf of the OOC CARL","for", "L=",L, "m=",m, "n=",n,"delta=", delta
), line=+2.5)
    xvalues<-c(0,20,40,60,80,100)
    axis(1,at=xvalues,cex.axis=1.5,las=1)

    ARLr <- round(ARL(delta,L,m,n),2)
    axis(3,ARLr,cex.axis=1,las=1)
    abline(v=ARL(delta,L,m,n),lty=5.5,col="blue")

    Median <- round(quantileCARL(0.5,delta,L,m,n),2)
    axis(1,Median,cex.axis=1,las=1,line=1)
    abline(v=quantileCARL(0.5,delta,L,m,n) ,lty=5.5,col="red")

  }
}

dev.new()
plotPDFCARL(0,3,25,5)
dev.new()
plotPDFCARL(0.5,3,25,5)

# the codes below generate a table with the unconditional ARL values for a set of values
of n (row) and m (column), given a value of the scaled shift (delta) and limit factor (L)

m<-c(25,50,75,100,150,200,250)
n<-c(3,5,9)
delta <- 0
L <- 3
ARLtable<-matrix(,nrow = length(n), ncol = length(m))
```

```
for (i in 1:length(n)){
  for (j in 1:length(m)){
    ARLtable[i,j] <- ARL(delta,L,m[j],n[i])
    cat(ARLtable[i,j]," ")
  }
}
ARLtable

# the codes below generate a table with the SDARL values for a set of values of n (row)
and m (column), given a value of the scaled shift (delta) and limit factor (L)

m<-c(25,50,75,100,150,200,250)
n<-c(3,5,9)
delta <- 0
L <- 3
SDARLtable<-matrix(,nrow = length(n), ncol = length(m))

for (i in 1:length(n)){
  for (j in 1:length(m)){
    SDARLtable[i,j] <- SDARL(delta,L,m[j],n[i])
    cat(SDARLtable[i,j]," ")
  }
}
SDARLtable

# the codes below generate a table with the p-quantile values of the CARL0 for a set of
values of n (row) and m (column), given a value of the scaled shift (delta) and limit factor
(L). Two axillar functions were created: CDFCARLax and quantileCARLax, for the secant
method work properly

m<-c(25,50,100,300,1000)
n<-c(5,10,20,25)
delta <- 0
L <- 3
p <-0.1
quantileTABLE<-matrix(,nrow = length(n), ncol = length(m))

CDFCARLax <- function (x,delta,L,m,n) {
  l <- 1/(1 - pnorm((-delta*sqrt(n))+((1/sqrt(m))*(delta*sqrt(m*n)))+L,0,1) + pnorm((-
delta*sqrt(n))+((1/sqrt(m))*(delta*sqrt(m*n)))-L,0,1))
  CARL <- function (Z) {
    a <- 1/(1 - pnorm((-delta*sqrt(n))+((1/sqrt(m))*Z)+L,0,1) + pnorm((-
delta*sqrt(n))+((1/sqrt(m))*Z)-L,0,1))
    return(a)
  }
  ARL0 <- function (t) {
    k <- CARL(t)-x
    return(k)
  }
  if (x>l) {
    c <- 1
```

```
  }
  if ((x<=l) & (x>=1) ) {
    b <- secantc(ARL0,-400,delta*sqrt(m*n))
    c <-  1 -  pnorm((2*(delta*sqrt(m*n)))-b,0,1) + pnorm(b,0,1)
  }
  if (x<1) {
    c <- 0
  }
  return(c)
}

quantileCARLax <-function (p,delta,L,m,n) {
  l <- 1/(1 - pnorm((-delta*sqrt(n))+((1/sqrt(m))*(delta*sqrt(m*n)))+L,0,1) + pnorm((-
delta*sqrt(n))+((1/sqrt(m))*(delta*sqrt(m*n)))-L,0,1))
  CDFm <- function (a) {
    a <- CDFCARLax(a,delta,L,m,n) - p
    return(a)
  }
  g<-secantc(CDFm,1,l)
  return(g)
}

for (i in 1:length(n)){
  for (j in 1:length(m)){
    quantileTABLE[i,j]<-quantileCARLax(p,delta,L,m[j],n[i])
    cat(quantileTABLE[i,j], "  ")
  }
}
quantileTABLE

# the codes below generate a table with tht minimum values of m, which generates
P(CFAR<(1+e)*alpha)=1-p for a set of values of n (row) and m (e), given a value of the
scaled shift (delta), p, nominal alpha and limit factor (L)

e<-c(0.1,0.2,0.3,0.4,0.5)
n<-c(5,10,20,25)
L <- 3
p <- 0.15
alpha<-0.0027
delta <- 0
MINIMUMmTABLE<-matrix(,nrow = length(n), ncol = length(e))

for (i in 1:length(e)){
  for (j in 1:length(n)){
    CDFm <- function (m) {
      a <- CDFCPS((1+e[i])*alpha,delta,L,m,n[j]) - (1-p)
      return(a)
    }
    MINIMUMmTABLE[j,i]<-ceiling(secantc(CDFm,5,4000))
    cat(MINIMUMmTABLE[j,i],"  ")
  }
```

```
}
MINIMUMmTABLE

# the codes below generate a table with the adjusted value of L, which generates
P(CFAR<(1+e)alpha)=1-p for a set of values of n (row) and m (column), given a value of
the scaled shift (delta), p, nominal alpha and e

m<-c(25,50,100,300,1000)
n<-c(5,10,20,25)
alpha <- 0.0027
p<-0.05
e<-0.2
delta<-0
adjLtable<-matrix(,nrow = length(n), ncol = length(m))

for (i in 1:length(n)){
  for (j in 1:length(m)){
    CDFaux <- function (s) {
      a <- CDFCPS((1+e)*alpha,delta,s,m[j],n[i])-(1-p)
      return (a)
    }
    adjLtable[i,j]<-secantc(CDFaux,2.1,3.9)
    cat(adjLtable[i,j], " ")
  }
}
adjLtable
```

Annex A – Paper under Review in the Journal of Quality Technology

Two perspectives for designing a Phase II control chart with estimated parameters: The case of the Shewhart $\overline{X}$ Chart

F. Jardim
Department of Industrial Engineering
Pontifical Catholic University of Rio de Janeiro,
Rio de Janeiro, Brazil

S. Chakraborti
Department of Information Systems, Statistics and Management Science,
University of Alabama,
Tuscaloosa, AL, USA

E. Epprecht
Department of Industrial Engineering
Pontifical Catholic University of Rio de Janeiro
Rio de Janeiro, Brazil

Abstract

The impact of parameter estimation on control charts has been studied with great interest in the recent literature. The estimated control limits affect chart performance, often negatively. Guided by the need to design control charts with a specified nominal in-control performance so as to avoid excessive false alarms, two major perspectives are advocated. Under the first, the so-called *unconditional perspective*, control limits are determined so that the in-control *unconditional average run length* equals a specified nominal value. However, the *unconditional perspective* does not account for the *practitioner-to-practitioner variability* inherent in using control charts based on parameter estimates. Thus, more recently, researchers have considered a *second perspective*, called the *conditional perspective*, under which the so-called exceedance probability criterion (EPC) is used to calculate the control limits so that the in-control conditional average run-length is at least equal to a specified nominal value with a high probability. These two perspectives lead to adjusted control limits and various methods have been proposed for calculating these limits. In this paper, we consider the Shewhart $\overline{X}$ chart to illustrate the two perspectives and compare the adjusted control limits resulting from the different adjustment methods under the two perspectives. Summary and recommendations are given.

Key Words: Unconditional and Conditional Run Length and Average Run Length, Control Limit Adjustments, Bootstrap, False Alarm Rate, Guaranteed In-Control Performance, Exceedance probability criterion

1. Introduction

One has to wonder what may still be worth studying when it comes to the most well-known and celebrated control chart of all, namely the Shewhart $\bar{X}$ chart. This chart has been around for more than sixty years, but, as a matter of fact, is that only fairly recently we have started to fully understand and appreciate its performance, particularly when the chart is used with parameters estimated with reference data. For reviews of some of the works on the effect on the performance of control charts when parameters are estimated, see Jensen et al. (2006) and Psarakis et al. (2014). The situation we are concerned with arises in a Phase II monitoring setting where the parameter estimates are obtained from reference data, from a Phase I analysis (with m samples/subgroups each one of size n). For an overview of Phase I analysis the reader is referred to Chakraborti et al. (2009) and Jones-Farmer et al. (2014).

Traditionally, researchers studying the in-control performance of the $\bar{X}$ chart with estimated parameters have focused on the unconditional *in-control run length* distribution, especially on its expected value, the so-called *unconditional* in-control average run length (ARL_0). This is the first perspective, called the unconditional perspective, under which it has been seen that a large amount of Phase I data are required to achieve in-control performance as in the known parameters case when traditional control limits, like the "3-sigma limits", are used [see, for example, Quesenbery (1993) and Chen (1997)]. In this context, some authors have also suggested examining the standard deviation of the in-control run length ($SDRL_0$) distribution [see for example, Chakraborti (2007)]. Note that the unconditional run length distribution is obtained by averaging over the distributions of the estimators, so that the performance under the unconditional perspective is not for a specific control chart with a given set of parameter estimates from a specific Phase I sample, but an "average" performance over an infinite number of possible control charts, each corresponding to one set of parameter estimates from one Phase I sample from the same process (see, Chakraborti (2000)). The average performance idea may be unsettling to some users and thus, recently, an alternative point of view has emerged in the study of performance and design of control charts with estimated parameters. This advocates focusing on the performance of the control chart given (conditional on) the Phase I data from which the parameters are estimated and the control limits are constructed. This is the second perspective, called the *conditional perspective*, that argues that the in-control *conditional* run length distribution (and consequently its various attributes such as the conditional false alarm rate ($CFAR$) or the conditional in-control average run length ($CARL_0$)) are more meaningful in the context of chart design, since they take account of the *practitioner-to-practitioner variability* [see for example, Saleh et al. (2015a) and Epprecht et al. (2015)]. To this end, recognizing the fact that the $CARL_0$ is a random variable, one performance measure under the conditional perspective has been the *Exceedance Probability Criterion* [Albers et al. (2005)], denoted here by *EPC*, under which the probability that the $CFAR$ is smaller than some desired nominal value is set to be high. From a practical standpoint, however, it may be more useful to ensure that the probability that the $CARL_0$ is *greater* than some desired nominal value (such as 370.4) is high. This is how the exceedance probability criterion (*EPC*) is interpreted and used in this paper. Note that the $CFAR$ and the $CARL_0$ formulations are equivalent since the *conditional* run length distribution is geometric so that $CARL_0$ is the reciprocal of $CFAR$.

The performance of a Phase II control chart is related to the amount of Phase I data used in the parameters estimation. Several authors [see, for example, Quesenberry (1993), Chen (1997), Chakraborti (2000, 2006) and Diko et al. (2015)] have noted that for the $\bar{X}$ chart, with estimated parameters, the *unconditional perspective* leads to requiring a very large amount of Phase I data so that some nominal in-control chart performance can be achieved comparable to the known parameters case. In fact, the required amount of data has been shown to be much larger than what has been traditionally recommended, which is m = 25 or 30 Phase I subgroups, each of size $n = 5$. On the other hand, under the conditional perspective, Saleh et al. (2015a,b) and Jardim et al. (2017), for example, have shown that employing the EPC, the required amount of data can be even larger. Thus both approaches, used in conjunction with the traditional control charts, may be infeasible in routine control charting applications.

So, the practitioners face a dilemma while choosing a Phase II control chart with estimated parameters in the control limits. The fact that the unconditional perspective leads to somewhat smaller amount of required Phase I data may give the impression that this perspective is preferable. However, this is not true. In fact, adopting the *unconditional* ARL_0 as a performance criterion does not reduce the chance that the ARL_0 of the Phase II chart, with the estimated control limits, may be unacceptably small relative to a nominal value such as 370.4. In the *unconditional* perspective, one simply does not consider this chance (risk), since one typically focuses on some moments of the in-control unconditional run length distribution, such as the *expected value* (which is also the expected value of the $CARL_0$ distribution), as a chart performance measure to design the chart (that is to calculate the control limits). The expected value measure does not account for the rather large variability in the $CARL_0$ distribution [see for example, Saleh et al. (2015a) and Jardim et al. (2017)].

Given the finding that under both perspectives, large, often impractical, amounts of Phase I data are required to guarantee a traditional nominal in-control performance of the Phase II chart comparable to the known parameter case, some authors have considered using adjustments to the control limits to properly compensate for the effects of parameter estimation and to guarantee a desired in-control performance with the amount of data at hand. Such control limits are called adjusted limits and this adjustment consists of replacing the *limit factor* (L) (usually equal to 3 in the traditional Shewhart $\bar{X}$ chart), by a new (or corrected or adjusted) limit factor (L^*), which yields a specified nominal in-control performance. For example, in the unconditional perspective, the constant 3 in the traditional "3-sigma limits" may be replaced (or adjusted) by a constant ($L^* = 3.15$, say), to guarantee that the ARL_0 has a desired nominal value. On the other hand, under the conditional perspective, one recognizes that the $CARL_0$ is a random variable with a distribution and thus one uses the *EPC* and replaces the traditional *limit factor* (L) by an adjusted limit factor (L^*), to guarantee that with a high probability, the $CARL_0$ is greater than a specified value, say, 370.4. Of course, any adjustment to the control limits also impacts the chart's out-of-control performance and one must carefully balance the gains and losses on both fronts. The conventional wisdom in SPC has been to weigh the chart's in-control performance more heavily, so that too many false alarms relative to what is nominally expected can be avoided, but this must be balanced so that the chart's shift detecting ability is not highly compromised.

Although our discussions are general and apply to all control charts, we use the Shewhart $\bar{X}$ chart here for illustration because of its simplicity and popularity. For this chart, formulas for the adjusted limit factor (L^*) have been derived, methods for finding the solution have been considered, and the results have been tabulated for many cases of interest. To underscore the keen interest in this area of research, note that several articles have appeared in major journals over the last decade on the topic of adjustment of control limits for the $\bar{X}$ chart. These include Chakraborti (2006), Gandy and Kvaløy (2013), Saleh et al. (2015b), Goedhart et al. (2016 and 2017) and Jardim et al. (2017). It is therefore time to examine the various issues and get a better and more comprehensive understanding of the proposals. To this end, we briefly describe these efforts before making a comparison among the various approaches.

Chakraborti (2006) and Jardim et al. (2017) derived formulas for L^*, using the unconditional and the conditional perspective, respectively, using the exact distributions formulas for the in-control marginal run length and conditional average run length in each case. Although these distributions and the resulting equations are not in a closed form, they can be easily solved numerically, using many available software packages. Since these methods are based on an exact distribution and yields very accurate results fairly easily using numerical methods to solve the integrals involved, we call these “the Exact Methods”. On the other hand, Goedhart et al. (2016, 2017) for example, have derived formulas for the adjusted limit factor under the unconditional and the conditional perspective, respectively, using sophisticated approximations. Furthermore, realizing the complexity of the approximations, Goedhart et al. (2018) presented an alternative and simpler approximate formula for the conditional perspective solution, based on some theory of tolerance intervals available in the literature. However, this simpler formula requires the quantile of a *non-central* chi-square distribution, which is not tabulated in many quality control text books and is not provided in a popular software like Excel, so its calculation may still require relatively advanced statistical skills. Given this, in this paper, as an aside, we derive an even simpler approximate formula in terms of the central chi-square percentile. We call all these methods “the Approximate Methods” to emphasize the fact that they are derived using some approximations (to the distribution of the $CARL_0$), and not to imply that they are not of good quality. In fact, the approximations yield good results, particularly for larger values of m, closer to the ones obtained by the exact methods. In addition, adjustments to the $\bar{X}$ chart control limits have been considered by Saleh et al. (2015b) under the conditional perspective and the *EPC*, using the bootstrap approach proposed by Gandy and Kvaloy (2013). Note that since we assume normality, it is possible to find the adjustments (limit factors) *analytically* and the need for bootstrapping may be questionable (as also noted in Goedhart et al. (2017)). Nevertheless, since bootstrap is a powerful method that can be applied more generally, assuming no specific distribution, we include this method and the resulting adjustment factors in our comparisons. Figure 1 shows a flowchart for the current state of the art regarding the adjustment of Phase II control limits to achieve some desired nominal in-control performance for the $\bar{X}$ chart in the face of parameter estimation with Phase I data.

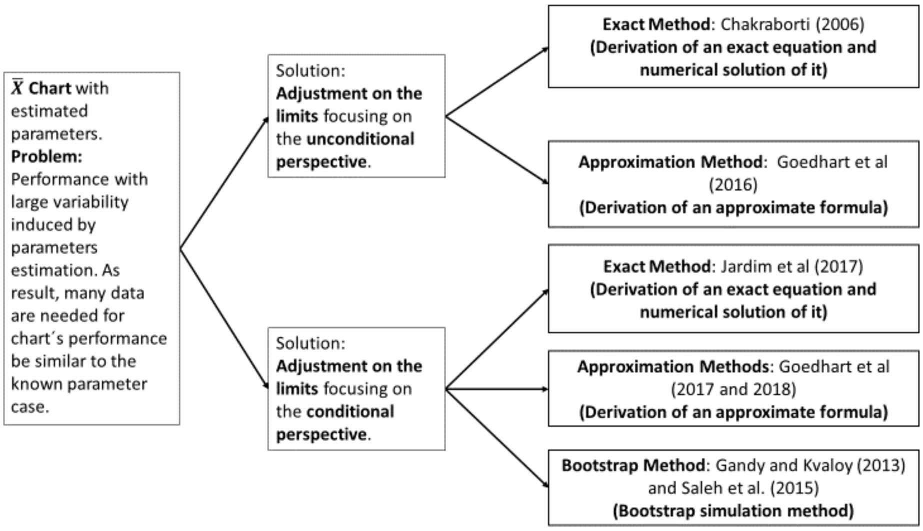


Figure 17. adjusting the $\bar{X}$ chart control limits for a guaranteed in-control performance

The existence of the two perspectives and at least two different methods under each used to determine the adjusted control limit factors may seem perplexing. With this in mind, in this paper we analyze the results from each method under each of the two perspectives. Moreover, for further insight, the performance of the solutions obtained under the unconditional perspective is also examined from the conditional point of view under the *EPC* and vice versa. For example, with the unconditional adjusted limit factor that guarantees a nominal $ARL_0 = 370.4$, we calculate and examine the probability that the conditional ARL_0 (that is $CARL_0$) is at least 370.4. Similarly, having found the adjusted limit that guarantees that the $CARL_0$ is at least 370.4 with a 95% probability (that is using the *EPC*), we calculate the associated value of the ARL_0. This analysis sheds interesting light on the relative performance of the various solutions under the two perspectives. With these results and comparisons, we offer some practical advice and recommendations regarding the design of the $\bar{X}$ control chart when applied with estimated parameters.

Before proceeding, it is important to note that for the Shewhart $\bar{X}$ control chart, when the in-control process mean (μ_0) and the in-control process standard deviation (σ_0) of the underlying normal distribution are estimated by $\hat{\mu}_0$ and $\hat{\sigma}_0$, respectively, the adjusted upper and lower control limits are given by

$$\widehat{UCL} = \hat{\mu}_0 + L^* \frac{\hat{\sigma}_0}{\sqrt{n}}, \tag{1}$$

$$\widehat{LCL} = \hat{\mu}_0 - L^* \frac{\hat{\sigma}_0}{\sqrt{n}}, \tag{2}$$

where L^* is the *adjusted control limit factor* that needs to be found so that a desired nominal in-control chart performance, defined in terms of a suitable performance criterion, is achieved, for a chosen set of estimators and given the available amount of Phase I data. This means that L^* may vary depending on the amount of data (*m* and *n*) available to estimate μ_0 and σ_0 and the type (unbiased, minimum variance, etc.) of estimators ($\hat{\mu}_0$ and $\hat{\sigma}_0$) one uses to estimate the parameters. Note that this formulation takes account of the fact that Phase I data are used to estimate the unknown parameters, which are necessary to establish the control limits. Note that this approach is different from using a traditional control limit factor (L) – also called *uncorrected limit factor* (like the “3-sigma” limits where $L = 3$, which yields a nominal ARL_0 = 370.4). The constant L does not depend on the amount of Phase I data available to set up the Phase II control chart and thus does not take into account the estimation of parameters on the control limits. However, when the amount of data is large (*m* and or *n* tend to infinity), L^* is expected to be equal to (converge to) L.

The remainder of this paper is organized as follows: In Sections 2 and 3, we present the various adjustment methods under the unconditional and conditional perspectives, respectively. Some results and discussion regarding these methods and these two perspectives are presented in Section 4. Finally, some conclusions and recommendations are provided in Section 5.

2. Adjusted Control Limit Factors Under the Unconditional Perspective

Under the unconditional perspective, the adjusted limits are found by first finding the corrected limit factor (L^*) which produces the desired nominal ARL_0. Usually, the desired nominal value of the ARL_0 is taken to be the one in the ideal (albeit typically unrealistic) *parameters known case* (μ_0 and σ_0 are known) such as 370.4 and thus, also assuming normality, $ARL_0 = \left(2\left(1-\Phi(L)\right)\right),^{-1}$ where L is the traditional or the *uncorrected* limit factor, set so as to give the desired nominal ARL_0 through the inverse relation $L = -\Phi^{-1}(1/(2ARL_0))$, where Φ is the cumulative distribution function (c.d.f.) of a standard normal random variable. For example, when the desired nominal $ARL_0 = 370.4$, one has $L = 3$ (the most commonly used 3-sigma limits).

When the in-control process mean (μ_0) and standard deviation (σ_0) are estimated by calculating $\hat{\mu}_0$ and $\hat{\sigma}_0$ from m Phase I samples each of size n, the Phase II control limits and consequently the run length distribution depends on the estimators $\hat{\mu}_0$ and $\hat{\sigma}_0$. Hence for given values of $\hat{\mu}_0$ and $\hat{\sigma}_0$, the run length distribution is referred to as the conditional run length distribution, which is geometric with probability of success equal to the conditional probability of a signal, $CPS(\hat{\mu}_0, \hat{\sigma}_0)$, is given by

$$CPS(\hat{\mu}_0, \hat{\sigma}_0) = 1 - P\left(\hat{\mu}_0 - L^* \frac{\hat{\sigma}_0}{\sqrt{n}} \le \bar{X}_l \le \hat{\mu}_0 + L^* \frac{\hat{\sigma}_0}{\sqrt{n}}\right)$$

Hence the *conditional* average run length $CARL(\hat{\mu}_0, \hat{\sigma}_0)$ is equal to

$$\begin{aligned} CARL(\hat{\mu}_0, \hat{\sigma}_0) &= \left(CPS(\hat{\mu}_0, \hat{\sigma}_0)\right)^{-1} \\ &= \left[P\left(\hat{\mu}_0 - L^* \frac{\hat{\sigma}_0}{\sqrt{n}} \le \bar{X}_l \le \hat{\mu}_0 + L^* \frac{\hat{\sigma}_0}{\sqrt{n}}\right)\right]^{-1}, \quad (3) \end{aligned}$$

where $\bar{X}_l$ denotes the l^{th} Phase II sample of size n. Note that Equation (3) applies both to the in-control and out-of-control states of the process. When the process is in control, the *CPS* is the conditional probability of a false alarm, denoted $CFAR$, and the *CARL* is the conditional in-control average run length, denoted $CARL_0$.

Being a function of $\hat{\mu}_0$ and $\hat{\sigma}_0$, the conditional average run length $CARL(\hat{\mu}_0, \hat{\sigma}_0)$ is a random variable and plays a crucial role in the performance of the control charts under both the unconditional and the conditional perspectives. The distribution of $CARL(\hat{\mu}_0, \hat{\sigma}_0)$ is considered in Jardim et al. (2017) and the reader is referred to that work for interesting insights and more details. For our purposes here, however, the mean and the second moment of the $CARL(\hat{\mu}_0, \hat{\sigma}_0)$ distribution are sufficient and are given by

$$ARL = \mathrm{E}\big(CARL(\hat{\mu}_0, \hat{\sigma}_0)\big) = \int_{-\infty}^{\infty}\int_{0}^{\infty} CARL(\hat{\mu}_0, \hat{\sigma}_0)\, f_{\hat{\sigma}_0}(\hat{\sigma}_0) d\hat{\sigma}_0 f_{\hat{\mu}_0}(\hat{\mu}_0) d\hat{\mu}_{0,} \quad (4)$$

$$\mathrm{E}\big(CARL^2(\hat{\mu}_0, \hat{\sigma}_0)\big) = \int_{-\infty}^{\infty}\int_{0}^{\infty} CARL^2(\hat{\mu}_0, \hat{\sigma}_0)\, f_{\hat{\sigma}_0}(\hat{\sigma}_0) d\hat{\sigma}_0 f_{\hat{\mu}_0}(\hat{\mu}_0) d\hat{\mu}_{0,} \quad (5)$$

where $f_{\hat{\mu}_0}(\hat{\mu}_0)$ and $f_{\hat{\sigma}_0}(\hat{\sigma}_0)$ denote the p.d.f. of $\hat{\mu}_0$ and $\hat{\sigma}_0$, respectively. Note that ARL is also the mean of the unconditional run-length distribution. Some authors have used the notation $AARL$ for ARL but we continue to use the latter to avoid confusion.

Given (4) and (5), the standard deviation of $CARL(\hat{\mu}_0, \hat{\sigma}_0)$, denoted $SDARL$, can be calculated from

$$SDARL = SD\big(CARL(\hat{\mu}_0, \hat{\sigma}_0)\big)$$

$$= \sqrt{\mathrm{E}\big(CARL^2(\hat{\mu}_0, \hat{\sigma}_0)\big) - \mathrm{E}^2\big(CARL(\hat{\mu}_0, \hat{\sigma}_0)\big)}. \quad (6)$$

Again, note that Equations (4) and (6) apply to both in- and out-of-control cases.

Typically, $\hat{\mu}_0$ is taken to be the grand mean of the m Phase I sample means ($\bar{\bar{X}}$). Thus, $\bar{\bar{X}} = \frac{1}{m}\sum_{i=1}^{m}\bar{X}_i$, where $\bar{X}_i = \frac{1}{n}\sum_{j=1}^{n}X_{ij}$, $i = 1,2,\dots,m$, $j = 1,2,\dots,n$, and X_{ij} denotes the j-*th* observation of the i-*th* sample. The observations X_{ij} are assumed to be normally distributed with mean μ_0 and standard deviation σ_0 where both are assumed unknown. For the standard deviation we consider the unbiased pooled estimator $\hat{\sigma}_0 = S_p/c_{4,b}$, where $S_p = \sqrt{\frac{1}{m}\sum_{i=1}^{m}S_i^2}$, $S_i^2 = \frac{1}{n-1}\sum_{j=1}^{n}\big(X_{i,j} - \bar{X}_i\big)^2$ and $c_{4,b}$ is the unbiasing constant for $b = m(n-1)+1$, where $c_{4,b} = \big[\Gamma(b/2)\sqrt{2}\big]/\big[\Gamma((b-1)/2)\sqrt{b-1}\big]$ and Γ is the gamma function. A comment about the standard deviation is in order. The unbiased pooled estimator we use here has been highly recommended in the recent literature [see Mahmoud et al. (2010) and Saleh et al. (2015a, b)]. Other estimators of the standard deviation, based on the average range or the average standard deviation could also be considered, but we leave that for the future.

It is known that (see, for example, Chakraborti, 2000) when the process is in control, (i) $Y = m(n-1)\, S_p^2/\sigma_0^2$ follows a central chi-square distribution with

$m(n-1)$ d.f., (ii) $Z = \left(\frac{\bar{\bar{X}}-\mu_0}{\sigma_0}\right)\sqrt{mn}$ follows a standard normal distribution and (iii) $\bar{X}_l \sim N(\mu_0, \sigma_0^2/n)$. Thus the conditional false alarm rate, $CPS\left(\bar{\bar{X}}, S_p | IC\right)$, can be conveniently expressed (see, for example, Chakraborti (2000)) by:

$$CFAR(Z,Y) = 1 - \left(\Phi\left(\frac{Z}{\sqrt{m}} + \frac{L^*}{c_{4,b}}\sqrt{\frac{Y}{m(n-1)}}\right) - \Phi\left(\frac{Z}{\sqrt{m}} - \frac{L^*}{c_{4,b}}\sqrt{\frac{Y}{m(n-1)}}\right)\right). \quad (7)$$

The random variable $CFAR(Z,Y)$ (or its inverse, the in-control conditional average run-length, given by $CARL_0 = [CFAR(Z,Y)]^{-1}$) plays a key role in the study of the performance of the Phase II Shewhart $\bar{X}$ chart with estimated parameters.

Finally, using (4) and (7), the in-control (IC) *unconditional* average run length (ARL_0) can be expressed as

$$ARL_0 = E(CARL_0) = \mathrm{E}[CARL(\hat{\mu}_0, \hat{\sigma}_0)|IC]$$
$$= E[[CFAR(Z,Y)]^{-1}] = \int_{-\infty}^{\infty}\int_{0}^{\infty} [CFAR(z,y)]^{-1} f_Y(y) dy \phi(z)\, dz, \quad (8)$$

where ϕ denotes the p.d.f. of a standard normal distribution and f_Y denotes the p.d.f. of Y, a central chi-square random variable with $m(n-1)$ degrees of freedom. Given that the desired in-control average run length is given by $\left[2\left(1-\Phi(L)\right)\right]^{-1}$, Chakraborti (2006) proposed to obtain the adjusted limit factor (L^*) that, used in (7), so that

$$\int_{-\infty}^{\infty}\int_{0}^{\infty} [CFAR(z,y)]^{-1} f_Y(y)\phi(z)\, dy\, dz = \left[2\left(1-\Phi(L)\right)\right]^{-1}. \quad (9)$$

The resulting solution is called the exact *unconditional* L^* and is denoted L^*_{UCE} (where the subscript *UCE* stands for "UnConditional Exact"). The solution to (9) can be obtained using a software like R that allows numerical integration. Note that Diko et al. (2015 and 2017) also used this formulation to find the exact adjusted control limit for each of the $\bar{X}$ and the S charts when applied jointly, and for various spread charts, including the S chart, respectively.

On the other hand, Goedhart et al. (2016), using a two-step Taylor expansion for $CFAR(z, y)$, derived the following approximate formula for L^*.

$$L_{UCA}^* \approx L - \frac{\left(\frac{\phi^2(L)}{4\bar{\Phi}^3(L)} - \frac{L\phi(L)}{4\bar{\Phi}^2(L)}\right)\left(A + \frac{1}{m}\right) + \frac{\phi^2(L)}{4\bar{\Phi}^3(L)}\left(A - \frac{1}{m}\right)}{2\frac{\phi(L)}{4\bar{\Phi}^2(L)}}, \tag{10}$$

where $A = \frac{L^2}{2(m(n-1)+1)}$ and $\bar{\Phi}(L) = 1 - \Phi(L)$. The factor L_{UCA}^* is referred to as the unconditional approximate solution (*UCA* standing for "UnConditional Approximate"). The exact and the approximate solutions will be compared in Section 4.

3. Adjusted Control Limit Factors Under the Conditional Perspective

As noted before, the $CARL_0$ is a random variable when the process parameters are estimated. The distribution of $CARL_0$ has a large variance when the amount of data used to estimate the parameters is small to moderate, like $m = 25$ and $n = 5$ [see Saleh et al. (2015) and Jardim el al. (2017)], so any Phase II $\bar{X}$ chart runs a considerable risk of having a very different $CARL_0$ from the advertised nominal value, depending on the parameter estimates obtained from the reference samples used in the control limits. In the conditional perspective, one recognizes the randomness of $CARL_0$ and uses the *EPC*, to ensure that the $CARL_0$ is at least, 370.4 (or perhaps a value slightly smaller), with a high probability (such as 0.95). Formally, this can be formulated as

$$P\left(CARL_0 \geq \frac{1}{(1+\varepsilon)}\left(\frac{1}{\alpha}\right)\right) = 1 - p, \tag{11}$$

for a small value p (such as 0.05), where α is the nominal false alarm rate: ($\alpha = 2\big(1 - \Phi(L)\big)$), which is also the false alarm rate in the *known parameters case* when the uncorrected limit factor (L) is used. Note that we can directly pick an α suitable in a given context and use it in (11) or, as it may be more common, pick an L (as in an L-sigma chart), calculate the corresponding α value and use that in (11). The choice is up to the user. The quantity ε is called the *tolerance factor* (meaning that the user will be willing to tolerate a $CARL_0$ that is at least $100\left(\frac{\varepsilon}{1+\varepsilon}\right)$% smaller than the nominal $1/\alpha$ with a high probability (that is with a small specified p). The *tolerance factor* (ε) allows flexibility for the user in the face of the inherent uncertainty in the random variable $CARL_0$. For example, in Section 4, it will be seen that if the user is willing to tolerate a $CARL_0$ value 10% ($\varepsilon = 1/9$) smaller than the nominal (for example, 333.4 as opposed to 370.4), the required amount of Phase I data for achieving the same performance will be a lot less, compared to the situation when no tolerance is allowed (i.e. when $\varepsilon = 0$).

Exact Method

From Equation (11), it is clear that the c.d.f. of $CARL_0$ is needed to apply the *EPC*. Taking $\hat{\mu}_0 = \bar{\bar{X}}$ and $\hat{\sigma}_0 = S_p/c_{4,b}$, Jardim et al. (2017), showed that the exact c.d.f. of the $CARL_0$ can be expressed as

$$P(CARL_0 \leq t) = \int_{-\infty}^{\infty} F_{\chi^2_{m(n-1)}}\left(\frac{m(n-1)F^{-1}_{\chi^2_{1,}\left[\frac{z^2}{m}\right]}\left(1-\frac{1}{t}\right)}{\left(\frac{L^*}{C_{4,b}}\right)^2}\right)\phi(z)dz\,,\quad t \geq 1 \tag{12}$$

where $F^{-1}_{\chi^2_{1,}\left[\frac{z^2}{m}\right]}(1-\frac{1}{t})$ denotes the $(1-\frac{1}{t})$-quantile of a non-central chi-square distribution with 1 degree of freedom and non-centrality parameter $\frac{z^2}{m}$ and $F_{\chi^2_{m(n-1)}}$ denotes the c.d.f. of a central chi-square random variable with $m(n-1)$ degrees of freedom. So, using (11), (12) and substituting $[(1+\varepsilon)\alpha]^{-1}$ for t, the exact adjusted control limit factor (L^*) can be obtained by solving the following equation, for given values of α, m, n, ε and p.

$$\int_{-\infty}^{\infty} F_{\chi^2_{m(n-1)}}\left(\frac{m(n-1)F^{-1}_{\chi^2_{1,}\left[\frac{z^2}{m}\right]}(1-(1+\varepsilon)\alpha)}{\left(\frac{L^*}{C_{4,b}}\right)^2}\right)\phi(z)dz = p. \tag{13}$$

This solution is denoted L^*_{CE} (CE stands for *Conditional Exact*) and can be obtained with a software like R. We emphasize that Equation (13) is exact, since the formula for the c.d.f. is exact, but the solution L^* must be found using a computer code, since there is no closed form solution for the c.d.f. in (12) or the integral in (13). This type of an analysis goes back to Chakraborti (2006). This is also similar to the adjustment methods given by Diko et. al. (2015) in the context of using the $\bar{X}$ and S charts jointly to monitor the mean and Diko et al. (2017) for various spread charts under the unconditional perspective. In some of these papers, this method has been referred to as the “numerical method” but the fact is that the method is exact since the expression for the c.d.f. is exact and “numerical” refers to the solution that is obtained by solving the equation that involves the c.d.f. which involves calculating the integral using some numerical methods. This is indeed the case for many c.d.f.'s of distributions including the one for the celebrated normal distribution.

Approximate Methods

Goedhart et al. (2017) derived an approximate formula for L^* by finding an approximate distribution of the $CFAR$. Details can be found in their paper. The final approximate formula for L^* is denoted by L^*_{CA1} ($CA1$ stands for *Conditional Approximation* 1) and is given by

$$L^*_{CA1} \approx L + \frac{\Phi^{-1}(1-p) - g(L)}{g'(L)}. \tag{14}$$

Here $g(L)$ and $g'(L)$ are functions of the expectation and the variance of $CFAR$ and their derivatives, respectively.

However, note that starting from (7), it is possible to derive an alternative, simpler approximate formula for L^*, denoted here by L^*_{CA2}, given by

$$L^*_{CA2} \approx C_{4,b} \sqrt{m(n-1)\frac{F^{-1}_{\chi^2_{1,\left[\frac{1}{m}\right]}}(1-(1+\varepsilon)\alpha)}{F^{-1}_{\chi^2_{m(n-1)}}(p)}}, \tag{15}$$

where $F^{-1}_{\chi^2_{m(n-1)}}(p)$ denotes the p-quantile of a central chi-square distribution with $m(n-1)$ degrees of freedom and $F^{-1}_{\chi^2_{1,\left[\frac{1}{m}\right]}}(1-(1+\varepsilon)\alpha)$ denotes the $(1-(1+\varepsilon)\alpha)$-quantile of a *non-central* chi-square distribution with 1 degree of freedom and non-centrality parameter $\frac{1}{m}$. Formula (16) is in fact given by Goedhart et al. (2018), which they found by starting from an existing result in Krishnamoorthy and Mathew (2009). We provide a more detailed derivation of (15) starting from Equation (7) in the supplementary material. Note that L^*_{CA2} requires a *non-central* chi-square quantile, which is not tabulated in many text books in Statistics and not available in popular software such as Excel, so its calculation will require some relatively advanced statistical skills of the practitioner. Given this, using a result from Cox and Reid (1987), we derive the following even simpler approximation formula for L^* (denoted by L^*_{CA3}) in terms of central chi-square percentiles

$$L_{CA3}^{*} \approx C_{4,b}\sqrt{(n-1)(m+1)\frac{F_{\chi_1^2}^{-1}(1-(1+\varepsilon)\alpha)}{F_{\chi_{m(n-1)}^2}^{-1}(p)}} \quad (16)$$

Derivation of (16) is also provided in the supplementary material.

Bootstrap Method

Finally, Saleh et al. (2015) suggested finding the adjusted limit factor L^{*} under the conditional perspective, using the EPC and the bootstrap approach of Gandy and Kvaløy (2013). In order to do this, the users, with the help of software (like SAS, R, etc.), should generate B bootstrap estimates of the in-control process mean and the standard deviation (μ_k^*, σ_k^*), $k = 1,2,\ldots,B$, with $\mu_k^* \sim N\left(\bar{\bar{X}}, S_p^2/nmc_{4,b}^2\right)$, $\sigma_k^* \sim \sqrt{S_p^2\frac{\chi_v^2}{vc_{4,b}^2}}$ and $v = m(n-1)$. Note that, with this, the idea is to consider that $\bar{\bar{X}}$ and $S_p/c_{4,b}$ are respectively the real in-control process mean and standard deviation, which are estimated respectively by μ_k^* and σ_k^* for each k. By considering a very large value for B, let say $B = 1000$, we have "access to the (bootstrap) population of μ_k^* $(\mu_1^*, \mu_2^*, \ldots, \mu_B^*)$ and σ_k^* $(\sigma_1^*, \sigma_2^*, \ldots, \sigma_B^*)$".

Recalling that $Y = m(n-1)\,S_p^2/\sigma_0^2$ and $Z = \left(\frac{\bar{\bar{X}}-\mu_0}{\sigma_0}\right)\sqrt{mn}$, using (7), the $CFAR$ can be written as

$$CFAR\left(\bar{\bar{X}}, S_p\right) = 1 - \Phi\left(\frac{\bar{\bar{X}}-\mu_0}{\sigma_0}\sqrt{n} + L^{*}\frac{S_p}{c_{4,b}\sigma_0}\right) + \Phi\left(\frac{\bar{\bar{X}}-\mu_0}{\sigma_0}\sqrt{n} - L^{*}\frac{S_p}{c_{4,b}\sigma_0}\right). \quad (17)$$

Considering that $\bar{\bar{X}}$ and $S_p/c_{4,b}$ are respectively the true in-control process mean and standard deviation and μ_k^* and σ_k^* are respectively the estimators $\bar{\bar{X}}$ and $S_p/c_{4,b}$ (according to the bootstrap method), for each μ_k^* and σ_k^*, the user must find the value of L_k^* that satisfies the following equation:

$$1 - \Phi\left(\frac{\mu_k^* - \bar{\bar{X}}}{S_p/c_{4,b}}\sqrt{n} + L_k^*\frac{\sigma_k^*\,c_{4,b}}{S_p}\right) + \Phi\left(\frac{\mu_k^* - \bar{\bar{X}}}{S_p/c_{4,b}}\sqrt{n} - L_k^*\frac{\sigma_k^*\,c_{4,b}}{S_p}\right) = (1+\varepsilon)\alpha. \quad (18)$$

Thus, for each pair of values of μ_k^* and σ_k^*, for each bootstrap sample ($k = 1,2,\ldots,b$) the user finds the value of L_k^* that solves (18) where α is the desired false alarm rate and ε is the tolerance factor defined earlier. The solution to Equation (18) is given by

$$L_k^* = \frac{S_p \sqrt{F^{-1}_{\chi^2_{1,\left[\frac{\mu_k^* - \bar{\bar{X}}}{S_p/c_{4,b}}\sqrt{n}\right]^2}}(1-(1+\varepsilon)\alpha)}}{C_{4,b}\sigma_k^*}, \quad k = 1,2,\dots,B \tag{19}$$

Note that the derivation of this formula is also presented in the supplementary material. This formula for L_k^* is much simpler than the rather complicated approximate method given in Saleh et al. (2015). However, we argue that no approximation is needed since one can derive the exact formula for L_k^* shown in Equation (19).

Finally, the required L^*, here denoted by L_{boot}^*, is found as the $(1-p)$-quantile of the collection of bootstrap estimators $(L_1^*, L_2^*, \dots, L_B^*)$.

4. Results and Discussion

In this section, we present the adjusted limit factors under the unconditional perspective for a nominal ARL_0 of 370.4 $(\alpha = 0.0027)$, $m = 13, 15, 20, 25, 50, 75, 100, 150, 200, 250$ and $n = 3, 5, 9$. Under the conditional perspective, for the adjusted limit factors, using the EPC, we use $P(CARL_0 \geq 370.4) = 0.95$ and $P(CARL_0 \geq 308.6) = 0.80$ (using $\alpha = 0.0027, p = 5\%, 20\%$ and $\varepsilon = 0\%$ and 20%, respectively). For the adjusted limit factors along with the unadjusted limit factor $L = 3$, we calculate the corresponding exact unconditional ARL_0 values using Equation (8), the exact standard deviation of $CARL_0$ ($SDARL_0$) using Equation (6), and the exact values of $P(CARL_0 \geq 370.4)$ and $P(CARL_0 \geq 308.6)$ using Equation (13).

The results along with some discussion are presented in subsections 4.1 and 4.2, for the unconditional and the conditional perspective, respectively.

4.1 Results and Discussion Under the Unconditional Perspective

Table 1 presents the adjusted limit factors (L^*) under the unconditional perspective. For comparison, the first five columns in grey show the results when the unadjusted factor $L = 3$ is used in the control limits. Note that in order to solve Equation (9) for L_{UCE}^*, the Secant search method implemented in R was used with $|370.4 - ARL_0| \leq 10^{-10}$ as the stopping rule.

Table 13. unconditional perspective results: values of $L^*$when $L^* = L$, $L^* = L^*_{UCE}$, $L^* = L^*_{UCA}$ and their corresponding ARL_0, $SDARL_0$ and $P\big(CFAR \leq \alpha(1+\varepsilon)\big)$ for $L = 3, \alpha = 0.0027, \varepsilon = 0\%$ and $\varepsilon = 20\%$

		Unadjusted Limits					Exact Method [Chakraborti (2006)]					Aproximate Method [Goedhart et al. (2016)]				
m	n	$L^* = L$	$ARL_0 = E(CARL_0)$	$SDARL_0 = SD(CARL_0)$	$\varepsilon=0\%$ $P(CFAR \leq 0.0027) = P(CARL_0 \geq 370.4)$	$\varepsilon=20\%$ $P(CFAR \leq 0.0032) = P(CARL_0 \geq 308.6)$	$L^* = L^*_{UCE}$	$ARL_0 = E(CARL_0)$	$SDARL_0 = SD(CARL_0)$	$\varepsilon=0\%$ $P(CFAR \leq 0.0027) = P(CARL_0 \geq 370.4)$	$\varepsilon=20\%$ $P(CFAR \leq 0.0032) = P(CARL_0 \geq 308.6)$	$L^* = L^*_{UCA}$	$ARL_0 = E(CARL_0)$	$SDARL_0 = SD(CARL_0)$	$\varepsilon=0\%$ $P(CFAR \leq 0.0027) = P(CARL_0 \geq 370.4)$	$\varepsilon=20\%$ $P(CFAR \leq 0.0032) = P(CARL_0 \geq 308.6)$
13	5	3.00	500.6	887.0	38.23%	45.32%	2.92	370.4	587.9	28.65%	35.27%	2.96	435.9	734.0	33.79%	40.73%
	9	3.00	375.3	343.6	35.03%	44.57%	3.00	370.4	337.9	34.41%	43.92%	3.04	430.4	407.7	41.57%	51.25%
15	5	3.00	473.7	700.7	38.74%	46.40%	2.93	370.4	505.4	30.06%	37.32%	2.97	420.88	598.8	34.53%	42.06%
	9	3.00	371.1	304.3	35.62%	45.98%	3.00	370.4	303.6	35.52%	45.87%	3.03	417.2	351.8	41.74%	52.22%
20	5	3.00	436.9	480.7	39.74%	48.71%	2.95	370.4	389.1	32.65%	41.32%	2.98	400.6	430.2	35.99%	44.84%
	9	3.00	366.1	244.7	36.80%	48.98%	3.00	370.4	248.3	37.53%	49.74%	3.03	398.9	271.8	42.16%	54.42%
25	3	3.00	569.5	1045.9	42.70%	49.99%	2.89	370.4	579.8	28.72%	35.46%	2.90	390.7	623.5	30.42%	37.28%
	5	3.00	418.5	380.3	40.50%	50.61%	2.97	370.4	326.3	34.45%	44.34%	2.98	390.7	348.9	37.08%	47.10%
	9	3.00	364.2	210.3	37.70%	51.49%	3.01	370.4	214.6	38.90%	52.72%	3.02	389.8	228.3	42.55%	56.36%
50	3	3.00	448.2	401.3	44.46%	54.87%	2.95	370.4	315.9	34.67%	44.86%	2.95	376.1	322.1	35.45%	45.68%
	5	3.00	389.1	217.4	42.69%	57.21%	2.99	370.4	204.7	38.99%	53.49%	2.99	376.3	208.7	40.16%	54.68%
	9	3.00	363.8	138.3	40.35%	60.33%	3.01	370.4	141.3	42.25%	62.15%	3.01	376.1	143.9	43.88%	63.69%
75	3	3.00	418.3	278.6	45.35%	58.11%	2.96	370.4	239.9	37.42%	50.13%	2.97	373.1	242.1	37.89%	50.62%
	5	3.00	381.6	166.8	43.82%	61.60%	2.99	370.4	160.9	41.01%	58.88%	2.99	373.2	162.4	41.72%	59.57%
	9	3.00	365.0	110.4	41.76%	66.20%	3.00	370.4	112.4	43.70%	67.95%	3.01	373.1	113.4	44.68%	68.82%
100	3	3.00	404.9	223.8	45.91%	60.61%	2.97	370.4	200.9	39.09%	53.86%	2.98	371.9	201.9	39.40%	54.18%
	5	3.00	378.3	140.2	44.54%	64.98%	2.99	370.4	136.7	42.22%	62.80%	2.99	372.0	137.5	42.71%	63.26%
	9	3.00	365.9	94.6	42.67%	70.63%	3.00	370.4	95.9	44.55%	72.22%	3.00	372.0	96.4	45.22%	72.77%
150	3	3.00	392.5	170.1	46.61%	64.48%	2.98	370.4	158.6	41.06%	59.18%	2.98	371.1	159.0	41.25%	59.36%
	5	3.00	375.3	111.3	45.45%	70.11%	3.00	370.4	109.5	43.66%	68.54%	3.00	371.2	109.8	43.93%	68.78%
	9	3.00	367.1	76.3	43.84%	77.11%	3.00	370.4	77.1	45.56%	78.38%	3.00	371.1	77.3	45.95%	78.66%
200	3	3.00	386.6	142.3	47.04%	67.48%	2.99	370.4	135.2	42.25%	63.07%	2.99	370.8	135.3	42.38%	63.19%
	5	3.00	373.9	95.0	46.02%	73.98%	3.00	370.4	93.9	44.51%	72.75%	3.00	370.8	94.1	44.70%	72.90%
	9	3.00	367.8	65.7	44.59%	81.71%	3.00	370.4	66.3	46.16%	82.73%	3.00	370.8	66.4	46.42%	82.89%
250	3	3.00	383.2	124.7	47.34%	69.97%	2.99	370.4	119.7	43.07%	66.16%	2.99	370.7	119.9	43.16%	66.25%
	5	3.00	373.2	84.3	46.41%	77.08%	3.00	370.4	83.5	45.09%	76.08%	3.00	370.7	83.6	45.23%	76.18%
	9	3.00	368.3	58.6	45.11%	85.17%	3.00	370.4	59.0	46.56%	85.99%	3.00	370.7	59.0	46.75%	86.09%

OBSERVATION: ALL DOUBLE INTEGRALS WERE CALCULATED IN R USING THE FUNCTION 'ADAPTINTEGRATE' FROM THE PACKAGE 'CUBATURE', WITH DEFAULT ENTRIES. ALL SIMPLE INTEGRALS WERE CALCULATED IN R USING THE DEFAULT FUNCTION 'INTEGRATE' WITH DEFAULT ENTRIES.

From Table 1, we see that when the unadjusted limit factor ($L^* = 3$) is used in the Phase II control limits, the attained ARL_0 differs considerably from the nominal ARL_0 value of 370.4 in many cases, especially when m and n are small. In this situation, it is interesting to note that, for a fixed m, the ARL_0 values are *larger* than 370.4 when n is small (for example, $n = 3$) and *smaller* than 370.4 when n is large (for example, $n = 9$). An ARL_0 value larger than 370.4 may give the impression that the chart performance is better compared to the known parameters case, but, as Quesenberry (1993) noted, this is not true. The large unconditional ARL_0 value for $L^* = 3$ is due to a combination of an increased rate of very short runs until a false alarm and just a few extremely long runs until a false alarm, and this is clearly undesirable, since a "quick" false alarm event is obviously unwanted. On the other hand, when the unadjusted limit factor ($L^* = 3$) is used in the control limits, under the conditional perspective and the exceedance probability criterion, the probability that the $CARL_0$ is greater than 370.4 is small (below 50%) for all values of m and n and this may also be a problem for the unadjusted limit factor $L^* = 3$. Even when the tolerance factor ε is increased to 20%, for $L^* = 3$, in most of the cases, this exceedance probability is still small [for example, for $m = 75$ and $n = 9$, one has $P(CARL_0 \geq 308.6) = 66.30\%$]. This means that one can expect the attained ARL_0 values smaller than the nominal and in more than 30% of the cases and that is a problem for the practical implementation. Only when m and n are large, the $P(CARL_0 \geq 308.6)$, for $L^* = 3$,increases considerably [for example, for $m = 250$ and $n = 9$, one has $P(CARL_0 \geq 308.6) = 85.17\%$]. This can be explained looking to the standard deviation of the $CARL_0$ (that is, $SDARL_0$). Note that the $SDARL_0$ is very large for small values of m and n indicating that, in practice, the realized ARL_0 (i.e., the $CARL_0$) will *not* (most likely) be close to the unconditional ARL_0. However, when m and n are large, $SDARL_0$ decreases and the $P(CARL_0 \geq 308.6)$ increases.

Also from Table 1, we see that the unconditional adjusted limit factor proposed by Chakraborti (2006), L^*_{UCE}, achieves a precise ARL_0 equal to 370.4 for all values of m or n. On the other hand, with the approximate limit factor from formula (10), proposed by Goedhart et al. (2016), the adjusted limit factor L^*_{UCA} does not achieve an

ARL_0 close to 370.4 for small values of m. For example, for $m = 20$ and $n = 5$, $L^*_{UCA} =$ 2.98 and $ARL_0 = 400.6$. However, for larger values of m (like $m \geq 100$), the approximate solution does achieve results close to 370.4, indicating a satisfactory in-control performance with respect to the nominal ARL_0 *under the unconditional perspective*. However, it is interesting to note that the $P(CARL_0 \geq 370.4)$, $P(CARL_0 \geq 308.6)$ and the $SDARL_0$ values for both of these adjusted control limits factors (L^*_{UCE} and L^*_{UCA}) are very similar to the respective values obtained when the unadjusted limit factor ($L^* = 3$) is used for all values of m, n and ε. This means that these unconditional limit factors are not satisfactory under the conditional perspective, since even after the adjustment of the control limits (which produces an desired expected value of the $CARL_0$, that is the unconditional ARL_0 gets equal or close to 370,4), the variability of the $CARL_0$ is still large (despite being a little smaller with $L^* = L^*_{UCE}$ and $L^* = L^*_{UCA}$, than with $L^* = 3$). So, even though the $ARL_0 \approx 370.4$, for L^*_{UCE} and L^*_{UCA}, the chances are high that in a given instance, the ARL_0 for a Phase II chart for a given set of estimates from a set of Phase I reference data can be very different from the nominal 370.4. This may not be satisfactory. Next, we present and discuss results for the conditional perspective.

4.2 Results and Discussion Under the Conditional Perspective

Table 2 presents the adjusted control limit factors (L^*) obtained under the conditional perspective for $\varepsilon = 0\%$ and $p = 5\%$, i.e., the values of L^*_{CE} , L^*_{CA1}, L^*_{CA2}, L^*_{CA3} and L^*_{boot} that make $P(CARL_0 \geq 370.4)$, equal to 95%. Also, as in Table 1, for comparison purposes, the first four columns in gray show the results for the unadjusted limit factor ($L^* = L = 3$) and for each L, L^*_{CE} , L^*_{CA1}, L^*_{CA2}, L^*_{CA3} and L^*_{boot}. Table 2 shows the exact unconditional ARL_0 value calculated according to Equation (8), the $SDARL_0$ value calculated according to Equation (6), and the exact $P(CARL_0 \geq 370.4)$ value calculated according to Equation (12).

From Table 2, we see that under the conditional perspective, the five methods presented in Section 3 yield very similar $P(CARL_0 \geq 370.4)$ values close to the target. i.e., for all values of L^*_{CE} , L^*_{CA1}, L^*_{CA2}, L^*_{CA3} and L^*_{boot}, the probability $P(CARL_0 \geq 370.4)$ is very close to the specified 95%, the method proposed by Jardim et al. (2017) being the most precise one, yielding $P(CARL_0 \geq 370.4)$ exactly equql to 95%.

Also from Table 2, for all cases and values of L_{CE}^* , L_{CA1}^*, L_{CA2}^*, L_{CA3}^* and L_{boot}^*, the values of the unconditional ARL_0 values are seen to be much larger than 370.4, often more than 3 times larger. This is also true for the $SDARL_0$ values. For example, for $m = 20$, $n = 5$ and $L_J^* = 3.54$, one has $SDARL_0 = 7804.6$ and $ARL_0 = 3840.8$. The large variability is compensated by the large expectation resulting in getting the desired exceedance probability (equal or close to 95%). This means that, despite taking into account the mean and the variability of $CARL_0$, the conditional perspective and the exceedance probability criterion does not control these popular aspects of the run-length distribution.

Since the adjusted limits are wider than the unadjusted limits (note that $L^* > 3$ for all cases in Table 2), this may give the impression that the out-of-control performance may be deteriorated after the adjustments. However, as shown by Jardim et al. (2017), this is true just for small values of m and n (like $m = 25$ and $n = 5$), but for most of the other cases, the out-of-control performance will be similar to the one with unadjusted limits (especially for $m \geq 50, n \geq 5, p \geq 0.1$ and $\varepsilon \geq 0\%$). However, if the practitioner is still not satisfied with the very large values of ARL_0 and $SDARL_0$ [such concern is evident in Saleh et al. (2015a,b) who focused mainly on the $SDARL_0$ as the performance measure] in the latter case, he/she can increase the value of ε or p (accepting a smaller lowest tolerated bound for $CARL_0$ or a smaller $P(CARL_0 \geq 370.4)$). This will decrease the value of ARL_0 and $SDARL_0$ while the amount of data remains the same. The possibility of this allowance or practical trade-off may be a useful feature of the conditional perspective. To visualize the trade-off, Tables 3 and 4 show the adjusted control limit factors (L^*) under the conditional perspective, respectively, for the pair $\varepsilon = 20\%$ and $p = 5\%$, and the pair $\varepsilon = 20\%$ and $p = 20\%$, i.e., the values of L^* that make, respectively, $P(CARL_0 \geq 308.6) = P(CFAR \leq 0.0031) \cong 95\%$ and $P(CARL_0 \geq 308.6) = P(CFAR \leq 0.0031) \cong 20\%$. Note that in these cases, the values of ARL_0 and $SDARL_0$ are much smaller compared with the values in Table 2 for the same amount of data. For example, for $m = 50$ and $n = 5$, considering the exact method by Jardim et al. (2017) in Table 2 (i.e., for $\varepsilon = 0\%$ and $p = 5\%$), the $ARL_0 = 1157.1$ and the $SD(CARL_0) = 807.6$, now, considering the same amount of data ($m = 50$ and $n = 5$), from Table 4 (i.e., for $\varepsilon = 20\%$ and $p = 20\%$), one has $ARL_0 = 561.0$ and $SD(CARL_0) = 338.7$: a reduction of 48% in the expectation and 42% in the standard deviation. Note that the unconditional ARL_0 is still much larger than the nominal (370.4). Under the *EPC*, it is unlikely that the unconditional ARL_0 will be close to the nominal value (unless ε or p are extremely large, such as $\varepsilon = 50\%$ or $p = 40\%$) which may raise some questions in practice.

Table 14 - conditional perspective results: : values of $L^*$when $L^* = L$, $L^* = L^*_{CE}$, $L^* = L^*_{CA1}$, $L^* = L^*_{CA2}$, $L^* = L^*_{CA3}$, $L^* = L^*_{boot}$ and their corresponding ARL_0, $SDARL_0$ and $P\big(CFAR \leq \ \alpha(1+\varepsilon)\big)$ for $L = 3$, $\alpha = 0.0027$, $\varepsilon = 0\%$ and $p = 5\%$

		Unadjusted Limits				Exact Method [Jardim et al. (2017)]				Aproximate Method 1 [Goedhart et al. (2017)]				Aproximate Method 2 [Goedhart et al. (2018)]				Aproximate Method 3 [Our Proposal]				Bootstrap Method [Saleh et al. (2015)]			
m	n	$L^* = L$	$ARL_0 = E(CARL_0)$	$SDARL_0 = SD(CARL_0)$	$P(CFAR \leq 0.0027)_{\varepsilon=0\%} = P(CARL_0 \geq 370.4)$	$L^* = L^*_{CE}$	$ARL_0 = E(CARL_0)$	$SDARL_0 = SD(CARL_0)$	$P(CFAR \leq 0.0027)_{\varepsilon=0\%} = P(CARL_0 \geq 370.4)$	$L^* = L^*_{CA1}$	$ARL_0 = E(CARL_0)$	$SDARL_0 = SD(CARL_0)$	$P(CFAR \leq 0.0027)_{\varepsilon=0\%} = P(CARL_0 \geq 370.4)$	$L^* = L^*_{CA2}$	$ARL_0 = E(CARL_0)$	$SDARL_0 = SD(CARL_0)$	$P(CFAR \leq 0.0027)_{\varepsilon=0\%} = P(CARL_0 \geq 370.4)$	$L^* = L^*_{CA3}$	$ARL_0 = E(CARL_0)$	$SDARL_0 = SD(CARL_0)$	$P(CFAR \leq 0.0027)_{\varepsilon=0\%} = P(CARL_0 \geq 370.4)$	$L^* = L^*_{boot}$	$ARL_0 = E(CARL_0)$	$SDARL_0 = SD(CARL_0)$	$P(CFAR \leq 0.0027)_{\varepsilon=0\%} = P(CARL_0 \geq 370.4)$
13	5	3.00	500.6	887.0	38.23%	3.72	13080.2	**104498.3**	95.00%	3.70	11891.4	**89745.4**	94.56%	3.69	11403.6	**83970.0**	94.36%	3.70	11969.7	**90696.5**	94.59%	3.73	13589.2	**111093.5**	94.75%
	9	3.00	375.3	343.6	35.03%	3.54	3029.6	4615.3	95.00%	3.56	3200.4	4941.1	95.40%	3.50	2526.2	3682.3	93.42%	3.51	2619.7	3852.5	93.77%	3.55	3172.5	4887.6	95.08%
15	5	3.00	473.7	700.7	38.74%	3.65	7980.4	**33805.0**	95.00%	3.65	7680.9	**31995.0**	94.80%	3.63	7107.6	**28621.7**	94.36%	3.64	7353.5	**30052.3**	94.55%	3.69	9387.3	**42753.8**	95.43%
	9	3.00	371.1	304.3	35.62%	3.49	2377.1	2983.8	95.00%	3.51	2501.3	3175.6	95.42%	3.46	2032.2	2463.0	93.45%	3.46	2087.2	2544.9	93.74%	3.49	2372.2	2976.3	94.72%
20	5	3.00	436.9	480.7	39.74%	3.54	3840.8	7804.6	95.00%	3.54	3875.9	7897.2	95.06%	3.52	3534.8	7009.4	94.38%	3.53	3596.5	7168.1	94.51%	3.56	4091.6	8471.2	95.10%
	9	3.00	366.1	244.7	36.80%	3.41	1613.1	1474.2	95.00%	3.42	1671.6	1538.5	95.38%	3.38	1437.3	1283.4	93.56%	3.38	1458.3	1305.9	93.76%	3.40	1578.6	1436.4	94.51%
25	3	3.00	569.5	1045.9	42.70%	3.66	11547.4	**86932.1**	95.00%	3.64	10670.9	**76617.4**	94.62%	3.65	11010.6	**80579.9**	94.77%	3.65	11166.2	**82383.7**	94.84%	3.70	14172.0	**121042.0**	95.87%
	5	3.00	418.5	380.3	40.50%	3.47	2552.5	3630.2	95.00%	3.47	2596.6	3708.6	95.15%	3.45	2394.2	3351.9	94.41%	3.46	2419.4	3395.9	94.51%	3.49	2746.6	3977.4	95.60%
	9	3.00	364.2	210.3	37.70%	3.35	1278.9	951.7	95.00%	3.36	1309.6	979.0	95.31%	3.33	1168.7	854.9	93.66%	3.33	1179.4	864.2	93.81%	3.35	1263.8	938.3	94.84%
50	3	3.00	448.2	401.3	44.46%	3.43	2327.9	3126.7	95.00%	3.43	2356.0	3173.8	95.11%	3.42	2283.7	3053.1	94.82%	3.42	2289.9	3063.4	94.84%	3.40	2076.8	2713.0	93.83%
	5	3.00	389.1	217.4	42.69%	3.31	1157.1	807.6	95.00%	3.31	1165.5	814.6	95.11%	3.30	1125.9	781.6	94.57%	3.30	1128.6	783.8	94.61%	3.32	1203.0	845.9	95.56%
	9	3.00	363.8	138.3	40.35%	3.23	790.5	348.2	95.00%	3.23	792.7	349.3	95.06%	3.22	760.1	332.4	94.06%	3.22	761.7	333.3	94.12%	3.22	783.3	344.5	94.79%
75	3	3.00	418.3	278.6	45.35%	3.34	1433.2	1244.6	95.00%	3.34	1446.1	1258.1	95.11%	3.33	1417.4	1228.0	94.86%	3.33	1419.0	1229.7	94.87%	3.32	1333.9	1141.3	94.00%
	5	3.00	381.6	166.8	43.82%	3.24	879.5	452.5	95.00%	3.24	881.0	453.4	95.03%	3.24	865.5	444.0	94.67%	3.24	866.4	444.5	94.69%	3.24	878.8	452.0	94.98%
	9	3.00	365.0	110.4	41.76%	3.18	662.9	224.3	95.00%	3.18	662.3	224.0	94.97%	3.17	647.7	218.2	94.28%	3.17	648.4	218.5	94.31%	3.18	673.8	228.6	95.46%
100	3	3.00	404.9	223.8	45.91%	3.28	1119.9	761.9	95.00%	3.29	1125.5	766.4	95.08%	3.28	1111.6	755.1	94.88%	3.28	1112.3	755.7	94.89%	3.31	1231.2	853.1	96.31%
	5	3.00	378.3	140.2	44.54%	3.20	759.9	322.0	95.00%	3.20	759.6	321.8	94.99%	3.20	751.7	317.9	94.73%	3.20	752.1	318.1	94.75%	3.20	749.5	316.8	94.66%
	9	3.00	365.9	94.6	42.67%	3.15	602.6	170.9	95.00%	3.15	601.4	170.5	94.93%	3.14	593.2	167.8	94.42%	3.14	593.5	167.9	94.44%	3.15	608.8	173.0	95.35%
150	3	3.00	392.5	170.1	46.61%	3.23	864.1	436.8	95.00%	3.23	864.7	437.2	95.02%	3.23	860.4	434.7	94.91%	3.23	860.6	434.8	94.92%	3.23	865.3	437.6	95.03%
	5	3.00	375.3	111.3	45.45%	3.16	648.3	213.1	95.00%	3.16	647.1	212.6	94.94%	3.16	644.3	211.5	94.81%	3.16	644.5	211.6	94.81%	3.16	642.9	211.0	94.74%
	9	3.00	367.1	76.3	43.84%	3.12	542.6	117.8	95.00%	3.12	541.4	117.5	94.90%	3.11	537.7	116.5	94.59%	3.11	537.9	116.6	94.60%	3.12	543.1	117.9	95.05%
200	3	3.00	386.6	142.3	47.04%	3.19	751.3	314.1	95.00%	3.19	750.5	313.7	94.97%	3.19	749.2	313.0	94.93%	3.19	749.3	313.1	94.93%	3.19	745.1	311.0	94.79%
	5	3.00	373.9	95.0	46.02%	3.14	593.8	164.5	95.00%	3.14	592.5	164.1	94.92%	3.14	591.4	163.7	94.85%	3.14	591.4	163.8	94.85%	3.13	582.4	160.8	94.25%
	9	3.00	367.8	65.7	44.59%	3.10	511.7	97.2	95.00%	3.10	510.7	97.0	94.90%	3.10	508.7	96.5	94.68%	3.10	508.7	96.6	94.69%	3.11	525.4	100.3	96.25%
250	3	3.00	383.2	124.7	47.34%	3.17	686.7	249.7	95.00%	3.17	685.5	249.2	94.95%	3.17	685.3	249.1	94.94%	3.17	685.4	249.1	94.95%	3.17	674.1	244.3	94.45%
	5	3.00	373.2	84.3	46.41%	3.12	560.7	136.7	95.00%	3.12	559.5	136.3	94.91%	3.12	559.1	136.2	94.88%	3.12	559.1	136.2	94.88%	3.12	566.6	138.4	95.42%
	9	3.00	368.3	58.6	45.11%	3.09	492.4	82.7	95.00%	3.09	491.5	82.5	94.90%	3.09	490.3	82.2	94.74%	3.09	490.3	82.2	94.74%	3.09	489.3	82.1	94.61%

Observation: All double integrals were calculated in R using the function 'adaptIntegrate' from the package 'cubature', with the default maximum tolerance (10^{-5}), unless for the values in bold. In these cases, the maximum tolerance was (10^{-2}). This means that the precision is affect in these cases. All simple integrals were calculated in R using the default function 'integrate' with default entries.

Table 3 - conditional perspective results: : value$L^* = L$ s of $L^*$when, $L^* = L^*_{CE}$, $L^* = L^*_{CA1}$, $L^* = L^*_{CA2}$, $L^* = L^*_{CA3}$, $L^* = L^*_{boot}$ and their corresponding ARL_0, $SDARL_0$ and $P\big(CFAR \leq \alpha(1+\varepsilon)\big)$ for $L = 3$, $\alpha = 0.0027$, $\varepsilon = 20\%$ and $p = 5\%$

		Unadjusted Limits				Exact Method [Jardim et al. (2017)]				Aproximate Method 1 [Goedhart et al. (2017)]				Aproximate Method 2 [Goedhart et al. (2018)]				Aproximate Method 3 [Our Proposal]				Bootstrap Method [Saleh et al. (2015)]			
m	n	$L^* = L$	$ARL_0 = E(CARL_0)$	$SDARL_0 = SD(CARL_0)$	$\varepsilon = 20\%$ $P(CFAR \leq 0.0032) = P(CARL_0 \geq 308.6)$	$L^* = L^*_{CE}$	$ARL_0 = E(CARL_0)$	$SDARL_0 = SD(CARL_0)$	$\varepsilon = 20\%$ $P(CFAR \leq 0.0032) = P(CARL_0 \geq 308.6)$	$L^* = L^*_{CA1}$	$ARL_0 = E(CARL_0)$	$SDARL_0 = SD(CARL_0)$	$\varepsilon = 20\%$ $P(CFAR \leq 0.0032) = P(CARL_0 \geq 308.6)$	$L^* = L^*_{CA2}$	$ARL_0 = E(CARL_0)$	$SDARL_0 = SD(CARL_0)$	$\varepsilon = 20\%$ $P(CFAR \leq 0.0032) = P(CARL_0 \geq 308.6)$	$L^* = L^*_{CA3}$	$ARL_0 = E(CARL_0)$	$SDARL_0 = SD(CARL_0)$	$\varepsilon = 20\%$ $P(CFAR \leq 0.0032) = P(CARL_0 \geq 308.6)$	$L^* = L^*_{boot}$	$ARL_0 = E(CARL_0)$	$SDARL_0 = SD(CARL_0)$	$\varepsilon = 20\%$ $P(CFAR \leq 0.0032) = P(CARL_0 \geq 308.6)$
13	*5*	3.00	500.6	887.0	45.32%	3.65	9181.9	**59709.1**	95.00%	3.64	8794.8	**55831.8**	94.80%	3.62	8013.5	**48319.8**	94.33%	3.63	8372.9	**51723.0**	94.56%	3.65	9006.2	**57933.7**	94.91%
	9	3.00	375.3	343.6	44.57%	3.48	2318.0	3309.2	95.00%	3.50	2472.3	3585.1	95.48%	3.44	1936.0	2646.1	93.37%	3.44	2001.4	2757.5	93.71%	3.49	2427.9	3505.3	95.35%
15	*5*	3.00	473.7	700.7	46.40%	3.59	5766.9	**21241.7**	95.00%	3.59	5748.4	**21145.3**	94.98%	3.56	5141.1	**18051.4**	94.33%	3.57	5302.3	**18856.6**	94.52%	3.60	6043.4	**22703.8**	95.25%
	9	3.00	371.1	304.3	45.98%	3.43	1838.6	2179.0	95.00%	3.44	1943.5	2332.1	95.48%	3.39	1574.4	1802.0	93.41%	3.40	1613.3	1856.7	93.69%	3.41	1668.6	1935.0	94.05%
20	*5*	3.00	436.9	480.7	48.71%	3.48	2888.1	5399.3	95.00%	3.48	2956.4	5564.7	95.17%	3.46	2660.3	4856.7	94.36%	3.46	2702.5	4956.3	94.49%	3.49	3073.2	5850.0	95.44%
	9	3.00	366.1	244.7	48.98%	3.35	1268.8	1104.7	95.00%	3.35	1313.2	1151.4	95.38%	3.32	1132.2	963.2	93.52%	3.32	1147.3	978.7	93.71%	3.35	1286.1	1122.9	95.15%
25	*3*	3.00	569.5	1045.9	49.99%	3.59	8120.5	**49788.3**	95.00%	3.58	7836.0	**47093.1**	94.82%	3.58	7747.0	**46262.7**	94.77%	3.58	7845.8	**47184.6**	94.83%	3.57	7372.9	**42829.5**	94.51%
	5	3.00	418.5	380.3	50.61%	3.41	1959.4	2611.3	95.00%	3.41	2004.6	2686.6	95.20%	3.39	1839.3	2413.6	94.40%	3.39	1857.0	2442.4	94.49%	3.42	2063.2	2784.8	95.45%
	9	3.00	364.2	210.3	51.49%	3.29	1015.8	723.1	95.00%	3.30	1036.8	741.1	95.28%	3.27	929.5	650.4	93.63%	3.27	937.3	656.8	93.78%	3.32	1134.2	824.9	96.32%
50	*3*	3.00	448.2	401.3	54.87%	3.36	1792.5	2259.1	95.00%	3.37	1820.0	2302.2	95.15%	3.36	1759.1	2206.8	94.81%	3.36	1763.4	2213.6	94.84%	3.36	1754.2	2199.2	94.79%
	5	3.00	389.1	217.4	57.21%	3.24	921.6	615.2	95.00%	3.25	925.2	618.1	95.06%	3.24	897.2	595.7	94.56%	3.24	899.1	597.2	94.60%	3.24	905.3	602.2	94.71%
	9	3.00	363.8	138.3	60.33%	3.17	640.2	271.4	95.00%	3.17	639.4	271.0	94.97%	3.16	616.1	259.3	94.05%	3.16	617.3	259.9	94.10%	3.16	630.4	266.5	94.64%
75	*3*	3.00	418.3	278.6	58.11%	3.27	1129.9	933.9	95.00%	3.28	1137.5	941.5	95.09%	3.27	1117.7	921.7	94.85%	3.27	1118.8	922.8	94.87%	3.31	1321.0	1128.0	96.73%
	5	3.00	381.6	166.8	61.60%	3.18	708.8	350.2	95.00%	3.18	707.3	349.3	94.95%	3.18	697.8	343.7	94.66%	3.18	698.4	344.1	94.69%	3.18	710.3	351.0	95.04%
	9	3.00	365.0	110.4	66.20%	3.12	540.7	176.3	95.00%	3.12	538.3	175.3	94.87%	3.11	528.6	171.6	94.28%	3.11	529.0	171.8	94.30%	3.11	528.5	171.6	94.27%
100	*3*	3.00	404.9	223.8	60.61%	3.22	893.0	581.1	95.00%	3.22	894.2	582.0	95.02%	3.22	886.5	576.0	94.88%	3.22	887.0	576.4	94.89%	3.22	897.7	584.7	95.09%
	5	3.00	378.3	140.2	64.98%	3.14	616.2	251.1	95.00%	3.14	613.7	249.9	94.90%	3.14	609.7	248.0	94.73%	3.14	610.0	248.1	94.74%	3.13	590.0	238.5	93.81%
	9	3.00	365.9	94.6	70.63%	3.09	493.3	134.9	95.00%	3.09	490.9	134.1	94.82%	3.08	485.8	132.5	94.42%	3.09	486.0	132.5	94.44%	3.09	501.4	137.5	95.57%
150	*3*	3.00	392.5	170.1	64.48%	3.17	696.8	338.2	95.00%	3.17	694.6	337.0	94.93%	3.16	693.9	336.6	94.91%	3.16	694.1	336.7	94.92%	3.18	725.1	354.6	95.79%
	5	3.00	375.3	111.3	70.11%	3.10	529.1	167.5	95.00%	3.10	526.5	166.5	94.84%	3.10	525.9	166.3	94.81%	3.10	526.1	166.4	94.81%	3.09	501.5	157.2	93.02%
	9	3.00	367.1	76.3	77.11%	3.06	445.9	96.2	95.00%	3.06	444.0	95.7	94.80%	3.06	442.1	95.2	94.58%	3.06	442.1	95.2	94.59%	3.06	447.1	96.5	95.12%
200	*3*	3.00	386.6	142.3	67.48%	3.13	609.4	245.1	95.00%	3.13	606.7	243.8	94.88%	3.13	607.8	244.3	94.93%	3.13	607.8	244.3	94.93%	3.14	623.8	252.0	95.57%
	5	3.00	373.9	95.0	73.98%	3.08	486.3	129.9	95.00%	3.08	484.0	129.2	94.82%	3.08	484.4	129.3	94.85%	3.08	484.4	129.3	94.85%	3.07	477.4	127.1	94.25%
	9	3.00	367.8	65.7	81.71%	3.04	421.5	77.3	95.00%	3.04	419.9	76.9	94.80%	3.04	419.0	76.7	94.68%	3.04	419.1	76.8	94.68%	3.05	429.4	79.0	95.94%
250	*3*	3.00	383.2	124.7	69.97%	3.11	559.1	195.7	95.00%	3.11	556.3	194.6	94.85%	3.11	558.0	195.3	94.94%	3.11	558.0	195.3	94.94%	3.12	568.0	199.4	95.45%
	5	3.00	373.2	84.3	77.08%	3.06	460.2	108.2	95.00%	3.06	458.2	107.6	94.81%	3.06	458.9	107.8	94.88%	3.06	458.9	107.8	94.88%	3.06	459.3	107.9	94.91%
	9	3.00	368.3	58.6	85.17%	3.03	406.1	65.8	95.00%	3.03	404.9	65.6	94.80%	3.03	404.5	65.5	94.74%	3.03	404.5	65.5	94.74%	3.03	412.6	67.0	95.91%

OBSERVATION: ALL DOUBLE INTEGRALS WERE CALCULATED IN R USING THE FUNCTION 'ADAPTINTEGRATE' FROM THE PACKAGE 'CUBATURE', WITH THE DEFAULT MAXIMUM TOLERANCE (10^{-5}), UNLESS FOR THE VALUES IN BOLD. IN THESE CASES, THE MAXIMUM TOLERANCE WAS (10^{-3}). THIS MEANS THAT THE PRECISION IS AFFECT IN THESE CASES. ALL SIMPLE INTEGRALS WERE CALCULATED IN R USING THE DEFAULT FUNCTION 'INTEGRATE' WITH DEFAULT ENTRIES.

PUC-Rio - Certificação Digital Nº 1312436/CA

Table 4 - conditional perspective results: : values of $L^*$when $L^* = L$, $L^* = L^*_{CE}$, $L^* = L^*_{CA1}$, $L^* = L^*_{CA2}$, $L^* = L^*_{CA3}$, $L^* = L^*_{boot}$ and their corresponding ARL_0, $SDARL_0$ and $P\big(CFAR \le \alpha(1+\varepsilon)\big)$ for $L = 3$, $\alpha = 0.0027$, $\varepsilon = 20\%$ and $p = 20\%$

		Unadjusted Limits				Exact Method [Jardim et al. (2017)]			
m	n	$L^* = L$	$ARL_0 = E(CARL_0)$	$SDARL_0 = SD(CARL_0)$	$\varepsilon = 20\%$ $P(CFAR \le 0.0032) = P(CARL_0 \ge 308.6)$	$L^* = L^*_{CE}$	$ARL_0 = E(CARL_0)$	$SDARL_0 = SD(CARL_0)$	$\varepsilon = 20\%$ $P(CFAR \le 0.0032) = P(CARL_0 \ge 308.6)$
13	5	3.00	500.6	887.0	45.32%	3.34	2092.3	**6552.3**	80.00%
	9	3.00	375.3	343.6	44.57%	3.25	939.3	1077.8	80.00%
15	5	3.00	473.7	700.7	46.40%	3.31	1620.1	**3650.0**	80.00%
	9	3.00	371.1	304.3	45.98%	3.22	825.4	816.1	80.00%
20	5	3.00	436.9	480.7	48.71%	3.24	1097.8	1560.3	80.00%
	9	3.00	366.1	244.7	48.98%	3.17	670.9	511.9	80.00%
25	3	3.00	569.5	1045.9	49.99%	3.28	1847.4	**5453.7**	80.00%
	5	3.00	418.5	380.3	50.61%	3.19	852.9	926.5	80.00%
	9	3.00	364.2	210.3	51.49%	3.14	584.4	372.6	80.00%
50	3	3.00	448.2	401.3	54.87%	3.16	817.1	849.9	80.00%
	5	3.00	389.1	217.4	57.21%	3.11	561.0	338.7	80.00%
	9	3.00	363.8	138.3	60.33%	3.07	455.5	181.0	80.00%
75	3	3.00	418.3	278.6	58.11%	3.12	631.0	460.4	80.00%
	5	3.00	381.6	166.8	61.60%	3.07	484.6	222.4	80.00%
	9	3.00	365.0	110.4	66.20%	3.04	415.8	129.0	80.00%
100	3	3.00	404.9	223.8	60.61%	3.09	552.7	326.2	80.00%
	5	3.00	378.3	140.2	64.98%	3.05	448.3	171.9	80.00%
	9	3.00	365.9	94.6	70.63%	3.02	395.7	103.8	80.00%
150	3	3.00	392.5	170.1	64.48%	3.06	480.7	217.1	80.00%
	5	3.00	375.3	111.3	70.11%	3.03	412.1	124.4	80.00%
	9	3.00	367.1	76.3	77.11%	3.01	374.8	78.3	80.00%
200	3	3.00	386.6	142.3	67.48%	3.04	446.0	168.9	80.00%
	5	3.00	373.9	95.0	73.98%	3.02	393.4	100.9	80.00%
	9	3.00	367.8	65.7	81.71%	3.00	363.7	64.8	80.00%
250	3	3.00	383.2	124.7	69.97%	3.03	425.1	141.2	80.00%
	5	3.00	373.2	84.3	77.08%	3.01	381.7	86.6	80.00%
	9	3.00	368.3	58.6	85.17%	2.99	356.5	56.4	80.00%

		Aproximate Method 1 [Goedhart et al. (2017)]				Aproximate Method 2 [Goedhart et al. (2018)]			
m	n	$L^* = L^*_{CA1}$	$ARL_0 = E(CARL_0)$	$SDARL_0 = SD(CARL_0)$	$\varepsilon = 20\%$ $P(CFAR \le 0.0032) = P(CARL_0 \ge 308.6)$	$L^* = L^*_{CA2}$	$ARL_0 = E(CARL_0)$	$SDARL_0 = SD(CARL_0)$	$\varepsilon = 20\%$ $P(CFAR \le 0.0032) = P(CARL_0 \ge 308.6)$
13	5	3.35	2117.8	**6667.3**	80.20%	3.32	1932.4	**5846.9**	78.64%
	9	3.27	1013.6	1184.7	81.97%	3.24	892.2	1011.0	78.57%
15	5	3.31	1673.6	**3814.6**	80.61%	3.29	1512.7	**3325.1**	78.65%
	9	3.24	884.9	889.0	82.02%	3.21	788.0	770.9	78.56%
20	5	3.25	1146.8	1649.9	81.06%	3.23	1042.6	1460.6	78.69%
	9	3.19	706.9	545.3	81.89%	3.16	646.5	489.3	78.57%
25	3	3.30	2015.6	**6182.0**	81.48%	3.28	1838.4	**5415.7**	79.91%
	5	3.21	914.9	1011.2	82.00%	3.19	843.5	913.7	79.67%
	9	3.15	616.2	397.2	82.24%	3.13	574.6	365.1	79.24%
50	3	3.18	867.6	915.7	81.80%	3.16	813.5	845.2	79.86%
	5	3.11	579.0	351.8	81.47%	3.10	557.1	335.8	79.66%
	9	3.07	464.4	185.2	81.32%	3.06	451.0	178.9	79.31%
75	3	3.13	655.7	482.5	81.54%	3.12	629.0	458.7	79.87%
	5	3.08	493.5	227.3	81.11%	3.07	482.4	221.1	79.71%
	9	3.04	420.1	130.6	80.91%	3.04	413.1	128.0	79.42%
100	3	3.10	567.8	337.1	81.31%	3.09	551.4	325.3	79.89%
	5	3.05	453.8	174.4	80.89%	3.05	446.8	171.2	79.75%
	9	3.03	398.3	104.6	80.69%	3.02	393.9	103.2	79.50%
150	3	3.06	488.5	221.3	81.00%	3.06	480.0	216.7	79.91%
	5	3.03	414.9	125.4	80.63%	3.03	411.2	124.1	79.80%
	9	3.01	376.2	78.6	80.46%	3.01	373.7	78.0	79.61%
200	3	3.05	450.9	171.1	80.81%	3.04	445.5	168.7	79.92%
	5	3.02	395.2	101.5	80.50%	3.01	392.8	100.8	79.84%
	9	3.00	364.5	65.0	80.36%	3.00	362.9	64.7	79.68%
250	3	3.03	428.5	142.6	80.68%	3.03	424.8	141.1	79.94%
	5	3.01	383.0	86.9	80.41%	3.01	381.3	86.5	79.86%
	9	2.99	357.1	56.5	80.30%	2.99	356.0	56.3	79.73%

		Aproximate Method 3 [Our Proposal]				Bootstrap Method [Saleh et al. (2015)]			
m	n	$L^* = L^*_{CA3}$	$ARL_0 = E(CARL_0)$	$SDARL_0 = SD(CARL_0)$	$\varepsilon = 20\%$ $P(CFAR \le 0.0032) = P(CARL_0 \ge 308.6)$	$L^* = L^*_{boot}$	$ARL_0 = E(CARL_0)$	$SDARL_0 = SD(CARL_0)$	$\varepsilon = 20\%$ $P(CFAR \le 0.0032) = P(CARL_0 \ge 308.6)$
13	5	3.33	2001.9	**6150.3**	80.46%	3.33	2004.7	**6162.6**	79.28%
	9	3.25	918.7	1048.5	80.20%	3.25	948.0	1090.2	80.25%
15	5	3.30	1551.8	**3442.3**	80.30%	3.32	1729.4	**3988.8**	81.22%
	9	3.22	805.3	791.8	80.01%	3.21	795.2	779.6	78.85%
20	5	3.23	1056.8	1486.1	80.05%	3.23	1062.2	1495.9	79.17%
	9	3.17	654.3	496.5	79.72%	3.17	659.2	501.1	79.33%
25	3	3.28	1857.2	**5495.3**	80.09%	3.27	1782.4	**5180.9**	79.36%
	5	3.19	850.6	923.3	79.92%	3.21	906.3	999.4	81.74%
	9	3.13	579.0	368.4	79.58%	3.14	598.1	383.2	81.00%
50	3	3.16	815.3	847.5	79.93%	3.17	827.1	862.8	80.37%
	5	3.10	558.2	336.6	79.76%	3.11	568.2	343.9	80.60%
	9	3.06	451.9	179.3	79.45%	3.07	460.2	183.2	80.71%
75	3	3.12	629.6	459.2	79.91%	3.13	658.1	484.7	81.68%
	5	3.07	482.8	221.3	79.76%	3.07	486.0	223.1	80.17%
	9	3.04	413.4	128.1	79.49%	3.05	428.8	133.8	82.63%
100	3	3.09	551.7	325.5	79.91%	3.10	565.7	335.5	81.13%
	5	3.05	447.0	171.3	79.78%	3.04	438.5	167.4	78.32%
	9	3.02	394.1	103.3	79.55%	3.02	395.0	103.6	79.80%
150	3	3.06	480.1	216.8	79.92%	3.06	488.1	221.1	80.95%
	5	3.03	411.3	124.1	79.82%	3.03	412.7	124.6	80.14%
	9	3.01	373.8	78.0	79.64%	3.01	375.0	78.3	80.05%
200	3	3.04	445.6	168.7	79.93%	3.05	455.4	173.1	81.52%
	5	3.01	392.9	100.8	79.85%	3.02	396.9	102.0	80.94%
	9	3.00	363.0	64.7	79.70%	2.99	360.2	64.1	78.45%
250	3	3.03	424.8	141.1	79.94%	3.04	431.7	143.8	81.28%
	5	3.01	381.3	86.5	79.87%	3.01	383.2	87.0	80.47%
	9	2.99	356.0	56.3	79.74%	2.99	358.6	56.7	80.99%

OBSERVATION: ALL DOUBLE INTEGRALS WERE CALCULATED IN R USING THE FUNCTION ‘ADAPTINTEGRATE’ FROM THE PACKAGE ‘CUBATURE’, WITH THE DEFAULT MAXIMUM TOLERANCE (10^{-5}), UNLESS FOR THE VALUES IN BOLD. IN THESE CASES, THE MAXIMUM TOLERANCE WAS (10^{-3}). THIS MEANS THAT THE PRECISION IS AFFECT IN THESE CASES. ALL SIMPLE INTEGRALS WERE CALCULATED IN R USING THE DEFAULT FUNCTION ‘INTEGRATE’ WITH DEFAULT ENTRIES.

5. Summary and Recommendations

In this paper, we analyze the recently proposed methods for adjusting the $\bar{X}$ control chart limits to achieve a desired in-control performance when parameters are estimated with a given amount of Phase I data, under two perspectives, called the unconditional and the conditional. We calculate and analyze the performance results for the available methods under these two perspectives. We also propose a new and simple approximate adjusted limit factor under the conditional perspective.

Based on our results, it is seen that, when constructing control charts with estimated control limits, both perspectives have some imperfections which the practitioners should be aware of. While the unconditional perspective does control the expected value of the $CARL_0$, it does not control (nor consider) the variability of the $CARL_0$, which can be very large even for a relatively large amount of reference data (such as 100 samples of size 9). This means that, in these cases, chances may be high that for a specific application the realized $CARL_0$ assumes undesirable small values relative to the specified nominal, which leads to many false alarms. On the other hand, under the conditional perspective, with the EPC, one can control the probability that the realized ARL_0 is greater than some desired nominal value. This approach implicitly considers the variability of $CARL_0$ but it neither controls the expected value nor the this variability of $CARL_0$ (the ARL_0), which can also be extremely large.

In conclusion, there is still room for improvement when it comes to designing the Shewhart control charts with unknown parameters. One can most likely say the same thing for control charts other than the Shewhart $\bar{X}$ chart, and this will be examined in the future.

Acknowledgements

This research was supported by the CNPq (Brazilian Council for Scientific and Technological Development) through project numbers 201172/2016-0 (international split PhD scholarship for the first author), 401523/2014-4 (travel grant for the second author), 308677/2015-3 (research grant for the 3rd author) and CAPES (Brazilian Coordination for the Improvement of Higher Education Personnel) through a national PhD scholarship for the first author. We are thankful to Prof. Ronald Does from The University of Amsterdam for his insightful comments on an earlier version which has improved presentation.

REFERENCES

Chakraborti, S. (2000). “Run Length, Average Run Length and False Alarm Rate of Shewhart X-bar Chart: Exact Derivations by Conditioning”. *Journal of Applied Statistics*, 33(4), pp. 439-459.

Chakraborti, S. (2006). “Parameter Estimation and Design Considerations in Prospective Applications of the X Chart”. Communication in Statistics - Simulation and Computation 29(1), pp. 61-81.

Chakraborti, S. (2007). “Run Length Distribution and Percentiles: The Shewhart Chart with Unknown Parameters”. *Quality Engineering* 19, pp. 119–127.

Chakraborti, S.; Graham, M. A.; and Human, S. W. (2009). “Phase I Statistical Process Control Charts: an Overview and Some Results”. *Quality Engineering* 21, pp. 52-62.

Chen, G. (1997). “The mean and Standard Deviation of the Run Length Distributions of $\bar{X}$ Charts When Control Limits are Estimated”. *Statistica Sinica* 7, pp. 789-798.

Cox, D., & Reid, N. (1987). “Approximations to Noncentral Distributions”. *The Canadian Journal of Statistics / La Revue Canadienne De Statistique*, 15(2), pp. 105-114.

Diko, M. D., Chakraborti, S. and Graham, M. A. (2015). “Monitoring the process mean when standards are unknown: A classic problem revisited”. *Quality Reliability Engineering International* 32 (2). pp. 609–622.

Diko, M. D., Goedhart, R., Chakraborti, S., Does R. J. M. M. & Epprecht, E. K. (2017) “Phase II control charts for monitoring dispersion when parameters are estimated”. *Quality Engineering*. Published online.

Epprecht, E. K.; Loureiro, L. D.; Chakraborti, S. (2015). “Effect of the Amount of Phase I Data on the Phase II Performance of S2 and S Control Charts”. *Journal of Quality Technology* 47(2). pp. 139-155.

Gandy, A. and Kvaløy, J. T. (2013). “Guaranteed Conditional Performance of Control Charts via Bootstrap Methods”. *Scandinavian Journal of Statistics* 40, pp. 647-668.

Goedhart, R.; Schoonhoven, M. and Does, R.J.M.M. (2018). “Recent Developments and Future Directions of Estimating Control Chart Limits*”. Journal of Quality Technology* 50(1). *Paper Accepted.*

Goedhart, R.; Schoonhoven, M. and Does, R.J.M.M. (2017). "Guaranteed In-Control Performance for the X and $\bar{X}$ Control Charts*". Journal of Quality Technology 49(2*). pp. 155-171.

Goedhart, R.; Schoonhoven, M. and Does, R.J.M.M. (2016). Correction factors for Shewhart X and $\bar{X}$ control charts to achieve desired unconditional ARL. *International Journal of Production Research*. DOI: 10.1080/00207543.2016.1193251

Jardim, F. S.; Chakraborti, S; and Epprecht, E. K. (2017) "Effect of the Amount of Phase I Data on the Conditional Performance of the Phase II $\bar{X}$ Chart". *Technical Memorandum 01/2017, Department of Industrial Engineering, PUC-Rio*.

Jensen, W. A.; Jones-Farmer, L. A.; Champ, C. W. and Woodall, W. H. (2006) "Effects of Parameter Estimation on Control Chart Properties: A Literature Review". *Journal of Quality Technology*, 38(4), pp. 349-364.

Jones-Farmer, L. A.; Woodall, W. H.; Steiner, S. H. and Champ, C. W. (2014). "An Overview of Phase I Analysis for Process Improvement and Monitoring". *Journal of Quality Technology 46(3)*, pp. 265 280.

Krishnamoorthy, K. and Mathew, T. (2009). "Statistical Tolerance Regions: Theory, Applications and Computation". Wiley, New York.

Mahmoud, M. A.; Henderson, G. R.; Epprecht, E. K. and Woodall, W. H. (2010) "Estimating the Standard Deviation in Quality Control Applications". Journal of Quality Technology, 42(4), pp. 348-357.

Montgomery, D.C. (2009). "Introduction to Statistical Quality Control", 7th ed. John Wiley & Sons, Hoboken, NJ.

Psarakis, S.; Vyniou, A. K.; and Castagliola, P. (2014). "Some Recent Developments on the Effects of Parameter Estimation on Control Charts". *Quality and Reliability Engineering International* 30. pp. 641-650.

Quesenberry, C. P. (1993). "The Effect of Sample Size on Estimated Limits for X¯and X Control Charts". *Journal of Quality Technology* 25(4), pp. 237-247.

Saleh, N. A.; Mahmoud, M. A.; Keefe, M. J. and Woodall, W. H. (2015a) "The Difficulty in Designing Shewhart X¯and X Control Charts with Estimated Parameters". *Journal of Quality Technology* 47(2). pp. 127-138.

Saleh, N. A.; Mahmoud, M. A.; Jones-Farmer, L. A., Zwetsloot, I. and Woodall W.H.; (2015b) "Another look at the EWMA control chart with estimated parameters". *Journal of Quality Technology* 47 pp. 363–38

Schoonhoven, M. (2016). “On the Use of the Expected ARL Metric for Control Charts Based on Estimated Parameters”. *Quality Reliability Engineering International* 33 pp. 745–749. doi: 10.1002/qre.2052.

Annex B – Paper Under Review in The Production Operation Management

New Insights On the $\bar{X}$ Chart with Estimated Parameters: Number of Samples and Limit Adjustments

F. Jardim
Department of Industrial Engineering
Pontifical Catholic University of Rio de Janeiro,
Rio de Janeiro, Brazil

S. Chakraborti
Department of Information Systems, Statistics and Management Science,
University of Alabama,
Tuscaloosa, AL, USA

E. Epprecht
Department of Industrial Engineering
Pontifical Catholic University of Rio de Janeiro
Rio de Janeiro, Brazil

Abstract

Performance measures of control charts with estimated parameters are random variables and vary from the nominal across reference samples. In this context, a recent idea has been to study the distribution of the realized (or the conditional) in-control average run length ($CARL_0$) [or, equivalently, the false alarm rate ($CFAR$)] for a given set of estimates from a given reference sample and apply the exceedance probability criterion (EPC) in order to design control charts that ensure desirable in-control performance. Under the EPC, te probability that the conditional in-control average run length (or the false alarm rate) is smaller (or not larger) than a specified value is guaranteed with a high probability, and this helps prevent too many low in-control ARL's (or too many high false alarm rates). In order to apply the EPC, the c.d.f. of the conditional in-control average run length (or the false alarm rate) is necessary. For the two-sided Shewhart $\bar{X}$ control chart, under normality, we derive the exact c.d.f. of the $CARL_0$ and the $CFAR$, currently not available in the literature. Using these key results, we calculate the minimum number of Phase I samples required to guarantee a desired nominal in-control performance with high probability in terms of the EPC. Since the required amount of data can be prohibitively large, we also provide exact formulas for adjustments to the control limits for a given amount of Phase I data; some tabulations are provided. Our adjustment formulas give more accurate results compared to some available methods. The impact of these adjustments on the out-of-control performance of the chart is examined in detail. A summary and some recommendations are provided.

Key Words: Average Run Length, False Alarm Rate, Conditional Run Length Distribution, Exceedance Probability Criterion, Guaranteed In-Control Performance, Out-of-Control Performance, Phase I and Phase II

1. Introduction

Control charts are among the indispensable tools for monitoring process quality in various industries. Managers recognize their value and the importance of performance. The performance of a control chart is measured in terms of the number of observations (or samples) until an alarm (the well-known *run length,* denoted RL). In this setting, one important factor is whether or not the underlying parameters are known or whether they need to be estimated before monitoring can start. Traditionally, the analysis didn't distinguish between these two cases and the focus has been the mean of the RL distribution. Reliance on the mean has been known to be problematic and researchers have suggested using the entire RL distribution and associated measures [see, e.g., Moskowitz et al. (1994)]. Many researchers have studied the effects of parameter estimation [see for example, Quesenberry (1993), Chen (1997), Chakraborti (2000 and 2006) and Goedhart et al. (2016)] and recognized the profound impact it can have on chart performance. They focused almost exclusively on the in-control performance. Many of these researchers have, to this end, considered the in-control run length (RL_0) distribution, accounting for the probability distribution of the estimators, but used the mean (denoted ARL_0) and the standard deviation of the "*unconditional*" RL_0 distribution, which is the in-control run length distribution after "averaging out" the effects of the distribution of the estimators. This distribution is not geometric as is the case in the known parameter case. However, as noted by some authors [see for example, Trietsch and Bischak (1998), Albers and Kallenberg (2004a,b, and 2005), Albers et al. (2005), Bischak and Trietsch (2007), Kumar and Chakraborti (2014), Saleh et al (2015a,b), Epprecht et al. (2015), Faraz et al. (2015), Goedhart et al (2017a,b and 2018) and Jardim et al. (2017)], in a practical application, the *unconditional* in-control RL_0 distribution or its mean, the ARL_0, may not represent the actual performance of the chart. The reason is that the RL_0 distribution conditional on the parameter estimates (denoted CRL_0) as well as the mean of this distribution, the *conditional in-control average run length* (denoted $CARL_0$), are random variables. This is because they are function of the parameter estimates obtained from the Phase I data (reference sample) at hand, and thus may vary significantly from dataset to dataset (the so-called practitioner-to-practitioner variability [see Saleh et al (2015a,b)]. Hence, there is little or no assurance that in a given application, the control chart will maintain the nominal ARL_0 value specified in the design of the chart. This is a very important point for the managers to recognize as it can influence the decision-making process significantly.

The reference sample typically consists of m samples, each of size n, collected from a process considered to be in control in what is called the Phase I analysis. For an overview of Phase I Statistical Process Control, the reader is referred to Chakraborti et al. (2009) and Jones-Farmer et al. (2014). For reviews of the literature on the effects of parameter estimation on the performance of control charts, see for example, Jensen et al. (2006) and Psarakis et al. (2014). The *false alarm rate* for a given set of estimates is called the *conditional false alarm rate* (denoted $CFAR$). It is known that the CRL_0 distribution follows a geometric distribution [see Chakraborti (2000)] with a parameter $CFAR$ and mean $CARL_0$. Indeed, the $CFAR$ and the $CARL_0$ are both random variables with a large variability, so their values may be quite different from their nominal ones [see Saleh et al. (2015a)]. As noted already, this is a key performance issue, since even though the control chart may be designed to have a nominal ARL_0 value such as 370, in a given application, depending on the parameter estimates, one can only say, for example, that the realized in-control average run-length, which is the $CARL_0$, may be anywhere from 200 to 500. Since a lower than nominal ARL_0 points to more false alarms, not properly adjusting for parameter estimation may lead to inefficiency and a loss of confidence in the whole charting process. Recognizing this, Albers and Kallenberg (2005) and Albers et al. (2005) proposed to set up the control chart limits so as to guarantee that the $CARL_0$ has a large probability of exceeding a given tolerated value. This is called the *exceedance probability criterion* (denoted *EPC*); the tolerated value provides a lower prediction bound to the random variable $CARL_0$. Thus, it becomes evident that, in order to use and implement the *EPC*, the distribution of $CARL_0$ is needed. However, even when it comes to the most well-known control chart, the two-sided Shewhart $\bar{X}$ control chart, under the assumption of normality, the exact c.d.f. of the $CARL_0$ is unavailable. So, most authors studying the effect of parameter estimation on the performance of the $\bar{X}$ charts, using the *EPC*, have relied on simulations, bootstrapping, or approximations to the distribution of the $CARL_0$. The study of control charts has been common in the literature [see, for example, Schroeder et al. (2005)], however, the interest in the *conditional* performance of control charts has been truly remarkable and deservedly so; we briefly summarize some recent works in this field.

Saleh et al. (2015a,b) focused on the $\bar{X}$, X and the EWMA charts and proposed examining the standard deviation of the $CARL_0$ distribution, in addition to its average (ARL_0). They showed that an impractically large number of Phase I samples are required in order to guarantee that the ARL_0 is close to a nominal ARL_0 value and the standard deviation of the $CARL_0$ is within at most 10% of the nominal ARL_0 value. Thus, they recommended adjusting the control limits, and suggested using the bootstrap approach of Gandy and Kvaloy (2013) for a given set of Phase I data, which guarantees a desired *IC conditional* performance in terms of the *EPC*. However, they did not provide any adjustments to the limits. Albers and Kallenberg (2004a,b and 2005) and Goedhart et al. (2017a and 2018) studied adjustments to the limits of the X and the $\bar{X}$ chart based on the *EPC* and analytical approximations to the distribution of the $CARL_0$ (rather than using simulations or the bootstrap approach). Other types of control charts have also been studied under the *EPC*. Epprecht et al. (2015) derived the exact c.d.f. of the $CFAR$ for the one-sided S and S^2 charts and found that the required numbers of Phase I samples were much larger than the ones found by previous authors who based their analyses only on the unconditional ARL_0 measure. Kumar and Chakraborti (2014) made a similar analysis for Shewhart-type time between events charts, found the exact c.d.f. of the $CARL_0$ and obtained similar conclusions regarding the required amount of Phase I data. Aly et al. (2015) and Faraz et al (2015) suggested using the *EP*C and the bootstrap approach of Gandy and Kvaloy (2013) to calculate the adjusted limits for the adaptive EWMA chart of Capizzi and Masarotto (2003) and for the *one-sided* S^2 control chart, respectively. Goedhart et al. (2017b) and Faraz et al (2017) calculated the adjusted limits according to the *EPC* for the *one-sided S* chart, but, in this case, they based their analysis on the exact *CFAR* distribution derived by Epprecht et al. (2015). For the *two-sided* S^2 control chart, Guo and Wang (2017), provided adjusted control limits under the *EPC* using a numerical approach to calculate the *CFAR* distribution. Finally, Faraz et al. (2017) also proposed an exact method to adjust the $\bar{X}$ chart, however, their adjustment was based on the equal-tailed tolerance interval together with the Bonferroni Inequality [see Krishnamoorthy and Mathew (2009, p. 4 and p.10)], which generates wider adjusted control limits which of course lead to the undesirable side effect of increasing the out-of-control (OOC) average run length (*ARL*) of the chart compared with the adjusted limits derived under the *EPC*.

Having recognized the important role the c.d.f. of $CARL_0$ plays in the performance of charts under estimated parameters, in this paper, we first derive its expression for the $\bar{X}$ chart. To the best of our knowledge, this result is not available in the literature. Using this result, we obtain the *exact* prediction bounds for the $CARL_0$. These prediction bounds show that when the number of reference samples (*m*) and/or the size of each reference sample (*n*) is small, some lower quantiles of the $CARL_0$ (such as the 0.05 and 0.1-quantiles) distribution are much smaller than the typically desired (nominal) ARL_0. As we explain in more detail in Section 5, this implies that the amount of reference data plays a crucial role in assuring that the Phase II chart performs at, or close to, the desired nominal ARL_0. Under this motivation, we determine the *exact* number of Phase I samples needed to guarantee a desired in-control performance in terms of the *EP*C. According to our results, the required numbers of Phase I samples are in some cases larger than the ones given by Saleh et al. (2015a). Since such large amounts of Phase I data may not always be available in practice, we then consider adjusting the control limits and provide *exact* formulas for the adjustment factor, again based on our c.d.f. expressions, for a given number and size of Phase I samples at hand, so that a desired conditional in-control performance is guaranteed in terms of the *EPC*. Note that the difference between our adjustments and those of Goedhart et al. (2017a and 2018) and Jardim et al. (2017) is that while their results were based on some approximations of the c.d.f. of the $CARL_0$ (or $CFAR$) distribution, our results are based on the exact c.d.f. of the $CARL_0$. We believe our adjustment formulas can be incorporated (programmed) readily and effectively in a software which makes them valuable for practical applications.

Finally, while the adjustments to the control limits under the *EPC* criterion guarantee a specified in-control chart performance, in most cases, they correspond to widening the control limits, which of course leads to the undesirable side effect of increasing the out-of-control (OOC) average run length (*ARL*) of the chart, i.e., reducing its shift detection ability. So, we analyze the impact of the adjustment on $CARL_\delta$, the *conditional* OOC ARL of the $\bar{X}$ chart, where δ is the *scaled shift* in the process mean. Note that some authors (Saleh et al., 2015b, and Goedhart et al., 2017a) have also analyzed the effects of the adjustment (based on the *EPC*) on the OOC performance of the $\bar{X}$ chart. However, Saleh et al. (2015b) mainly focused on the *unconditional* OOC ARL (i.e., the $E(CARL_{\delta})$), thus disregarding the practitioner-to-practitioner variation and Goedhart et al. (2017a) conducted a very limited examination of the $CARL_{\delta}$ distribution by simulation, and displayed a boxplot for only one value of the shift δ in the process mean and for just one pair of values of *n* and *m*. This does not give a complete picture of the impact. We examine the impact of the adjustment on the $CARL_{\delta}$ in much more detail, using our exact c.d.f., for several values of δ, *n* and *m*, and calculate some *exact* quantiles of interest of $CARL_{\delta}$ with and without the adjustment, and make a relative comparison. This analysis is important for the user, who needs to balance between controlling, on one hand, the risk of having a false-alarm rate (or the in-control *ARL*) much higher (lower) than the nominal and, on the other hand, allowing a deterioration of the OOC performance.

In addition to the most usual case of the $\bar{X}$ chart with estimated process mean and standard deviation, we also consider the case where only the process standard deviation is estimated. In this case, the process is considered to be in control when its mean coincides with the target or nominal value μ_0. According to Montgomery (2009; p. 243), "*in processes where the mean of the quality characteristic is controlled by adjustments to the machine, standard or target values of the mean are sometimes helpful in achieving management goals with respect to process performance*". This is equivalent to knowing the in-control mean μ_0, so this case is called "Case KU" (*mean **K**nown, standard deviation **U**nknown*) as in Quesenberry (1993). The case where both the mean and the standard deviations are unknown (chart centered on $\bar{\bar{X}}$) is denoted as Case UU. Case KU was also studied by Ghosh et al. (1981), but they focused on the *unconditional ARL* as the main performance measure criteria. To our knowledge, the *conditional* performance of the $\bar{X}$ chart in Case KU has not been analyzed thus far.

The remainder of this paper is organized as follows: In Section 2 we describe the control limits of the $\bar{X}$ chart and the estimators used while introducing some important notation and assumptions. In Section 3, we study the conditional in-control average run-length ($CARL_0$). In Section 4, we derive the c.d.f.'s for the $CARL_0$ and analyze the effects of the number of reference samples on this distribution. In Section 5, we determine prediction bounds for the $CARL_0$. Sections 6 and 7 address, respectively, the problem of finding the minimum number of reference samples, and the adjustment factors for the control limits that guarantee a specified conditional in-control performance under the EPC. The impact of the adjustment of the control limits in the out-of-control performance of the $\bar{X}$ control chart is studied in Section 8. A summary and conclusions are offered in Section 9.

2. The Control Limits of the $\bar{X}$ Chart

We assume that the observations on the process quality variable (X) are i.i.d. and normally distributed. When the process is in control, $X \sim N(\mu_0, \sigma_0^2)$; when the process is out of control, $X \sim N(\mu_1, \sigma_0^2)$, with $\mu_1 \neq \mu_0$. Thus, the process standard deviation is assumed to remain at the in-control value σ_0, consistently with the purpose of detecting a shift in the mean. In the ideal case, the in-control process mean (μ_0) and standard deviation (σ_0) are both known or specified (this is denoted Case KK: "mean Known, standard deviation Known" by Quesenberry, 1993). In Case KK, the upper and lower control limits (UCL and LCL) of the L-sigma $\bar{X}$ Control Chart with subgroups of size *n* are given by

$$UCL = \mu_0 + L\frac{\sigma_0}{\sqrt{n}} \tag{1}$$

and

$$LCL = \mu_0 - L\frac{\sigma_0}{\sqrt{n}}, \tag{2}$$

where the control limit factor L is either a value such as 3 (the widely used "3 sigma limits") or is chosen so as to provide a nominal in-control average run length such as 370.4 or a false-alarm rate α. In the latter case, we have $L = z_{\alpha/2} = \Phi^{-1}(1 - \alpha/2)$, where $\Phi(\cdot)$ denotes the standard normal c.d.f. Thus, the usual 3-sigma limits correspond to a nominal false alarm rate of $\alpha = 0.0027$. However, in practice μ_0 or σ_0 are usually unknown and need to be estimated from a Phase I analysis, consisting of *m* subgroups of size *n*, taken from the process when it is in control.

In Case UU, for the mean μ_0, we use the most common estimator $\bar{\bar{X}}$, the grand mean of the m Phase I samples: $\bar{\bar{X}} = \frac{1}{m}\sum_{i=1}^{m}\bar{X}_i$, where $\bar{X}_i = \frac{1}{n}\sum_{j=1}^{n}X_{ij}$, $i = 1,2,\dots,m$, $j = 1,2,\dots,n$ and X_{ij} denotes the j-*th* observation of the i-*th* Phase I sample. In both cases, KU and UU, we need to estimate the unknown standard deviation σ_0. For this purpose, we choose the pooled sample standard deviation (S_p), which is given by the square root of the average of the sample variances of the Phase I samples. Thus $S_p = \sqrt{\frac{1}{m}\sum_{i=1}^{m}S_i^2}$, where $S_i^2 = \frac{1}{n-1}\sum_{j=1}^{n}\left(X_{i,j} - \bar{X}_i\right)^2$. We base this choice on the recommendation of Mahmoud et al. (2010), who showed that, among multi-sample estimators of the standard deviation, S_p is preferable to a more traditional estimator, like $\bar{S}/c_4$, where $\bar{S} = \sum_{i=1}^{m}S_i/n$ and c_4 is the unbiasing constant [see, Montgomery (2009)]. Note that, in the literature, two other estimators have been considered, the unbiased S_p/c_4 and the biased, but minimum mean squared error, estimator c_4S_p (see Mahmoud et al., 2010 and Saleh et al., 2015a,b). Since these three estimators provide similar results as $c_4 \approx 1$ for relatively small values of m and n [such as $m = 25$ and $n = 4$ — again, see Mahmoud et al. (2010) for a quantitative comparison], in the present work, we consider just the S_p estimator. Note that we do not consider the range based estimators since some authors have recommended against their use because of lack of robustness. Anyway, all the formulas and results presented here can be easily modified for other estimators of standard deviation.

In order to study the effects of the estimation of the process parameter(s) on the performance of a control chart in general, it is convenient to begin with a study of the Phase II probability of a signal given the estimator(s), the so-called *conditional probability of a signal* (*CPS*). A signal occurs when, for any sample, its average $\bar{X}$ lies outside the control limits. When the process is in-control, a signal represents a false alarm and its probability is called the false-alarm rate. As noted earlier, the *conditional false-alarm rate* is denoted *CFAR*. These are discussed in the next section.

3. The Conditional Probability of a Signal and the Conditional False Alarm Rate

Given the control limits [Equations (1) and (2)] and replacing μ_0 and σ_0 by their respective estimators, $\bar{\bar{X}}$ and S_p, the conditional probabilities of a signal (CPS), for Case KU and Case UU, are given by

$$CPS_{KU} = P\left(Signal|S_p\right) = 1 - P\left(\mu_0 - L\frac{S_p}{\sqrt{n}} \leq \bar{X} \leq \mu_0 + L\frac{S_p}{\sqrt{n}}\right) \quad (3a)$$

and

$$CPS_{UU} = P\left(Signal|\bar{\bar{X}}, S_p\right) = 1 - P\left(\bar{\bar{X}} - L\frac{S_p}{\sqrt{n}} \leq \bar{X} \leq \bar{\bar{X}} + L\frac{S_p}{\sqrt{n}}\right), \quad (3b)$$

respectively.

From Equations (3a) and (3b) it is evident that the conditional probability of a signal in Phase II depends on the value of the estimator S_p in Case KU and on $\bar{\bar{X}}$ and S_p in Case UU. Before proceeding further, let μ denote the process mean in Phase II. Define the *scaled shift* of the mean as

$$\delta = \frac{\mu - \mu_0}{\sigma_0}, \quad (4)$$

where μ is the process mean in Phase II. When $\mu = \mu_0$, $\delta = 0$ and the process mean is in control. When $\mu = \mu_1 \neq \mu_0$, $\delta \neq 0$, and the process mean is out of control. It is well-known that $Y = m(n-1)\, S_p^2/\sigma_0^2$ follows a chi-square distribution with $m(n-1)$ degrees of freedom and $Z = \left(\frac{\bar{\bar{X}} - \mu_0}{\sigma_0}\right)\sqrt{mn}$ follows a standard normal distribution. Recalling that $X \sim N(\mu, \sigma_0^2)$ implies that $\bar{X} \sim N(\mu, \sigma_0^2/n)$ where $\mu = \mu_0 + \delta\sigma_0$, the conditional probability of a signal (CPS) for cases KU and UU can be expressed respectively as

$$CPS_{KU} = P(Signal|Y, \delta) = 1 - \left[\Phi\left(L\sqrt{\frac{Y}{m(n-1)}} - \delta\sqrt{n}\right) - \Phi\left(-L\sqrt{\frac{Y}{m(n-1)}} - \delta\sqrt{n}\right)\right] \quad (5a)$$

and

$$CPS_{UU} = P(Signal|Z, Y, \delta) = 1 - \left[\Phi\left(\frac{Z}{\sqrt{m}} + L\sqrt{\frac{Y}{m(n-1)}} - \delta\sqrt{n}\right) - \Phi\left(\frac{Z}{\sqrt{m}} - L\sqrt{\frac{Y}{m(n-1)}}\right) - \delta\sqrt{n}\right]. \qquad (5b)$$

These general expressions apply to both the in-control and the out-of-control cases.

In the in-control case, $\delta = 0$, and Equations (5a) and (5b) give the conditional false-alarm rate, $CFAR$:

$$CFAR_{KU} = P(Signal|Y, \delta = 0) = 2\Phi\left(-L\sqrt{\frac{Y}{m(n-1)}}\right) \qquad (6a)$$

and

$$CFAR_{UU} = P(Signal|Z, Y, \delta = 0) = 1 - \left[\Phi\left(\frac{Z}{\sqrt{m}} + L\sqrt{\frac{Y}{m(n-1)}}\right) - \Phi\left(\frac{Z}{\sqrt{m}} - L\sqrt{\frac{Y}{m(n-1)}}\right)\right], \qquad (6b)$$

in Case KU and UU, respectively.

Given that the in-control conditional run length distribution is geometric with parameter $CFAR$, its expected value, the conditional in-control average run length, $CARL_0$, is given by:

$$CARL_0 = \frac{1}{CFAR}, \qquad 0 \leq CFAR \leq 1. \qquad (7)$$

From now on, when we don't use a subscript such as KU or UU on the $CARL_0$ (or the $CFAR$ or the CPS) as in Equation (7), it means that the expression is true in both case KU and UU. Also, let $CARL_{0,KU}$ and $CARL_{0,UU}$ denote the conditional in-control average run lengths in Case KU and UU, respectively.

It is interesting to visualize the effect of parameter estimation through the number of Phase I subgroups, m, on the $CFAR$ and $CARL_0$. Figure 1 shows the plots of the $CFAR_{KU}$ (in Panel (a)) and $CARL_{0,KU}$ (in Panel (b)) as a function of $u \in (0, 1)$ corresponding to the quantiles of Y, parametrized by m for a fixed $n = 5$. They were obtained by writing Y as $\mathrm{F}^{-1}_{\chi^2_{m(n-1)}}(U)$ in Equation (6a), where U is a uniform (0, 1) random variable, since $U = F_{\chi^2_{m(n-1)}}(Y)$ from the probability integral transformation [see Epprecht et al. (2015)].

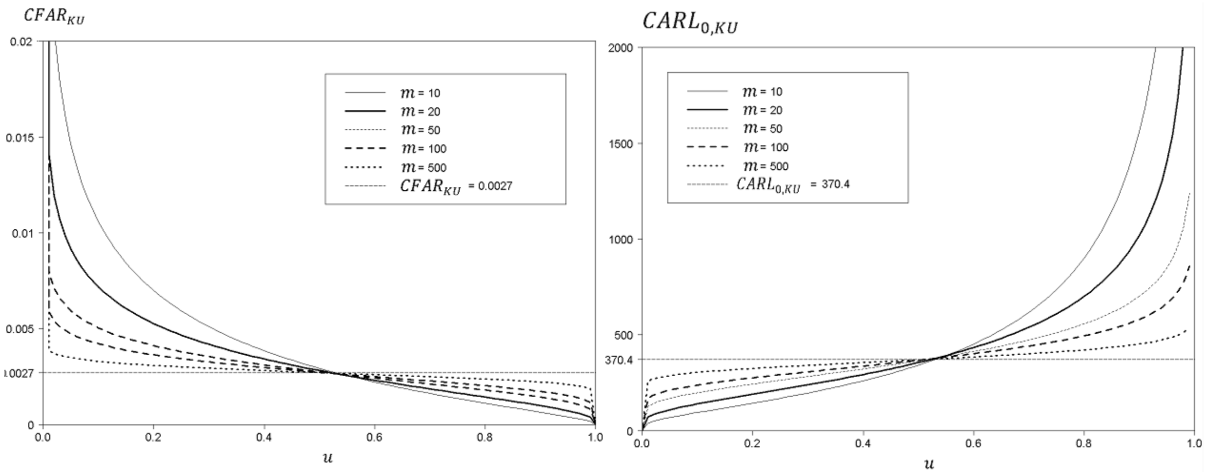


FIGURE 1. $CFAR_{KU}$ (Panel (a)) and $CARL_{0,KU}$ (Panel (b)) as function of u for $n = 5$, $m = 10, 20, 50,100,500$ and $\alpha = 0.0027$ (i.e., $L = 3$).

Figure 1 clearly shows the effect of the number of Phase I samples $\boldsymbol{m}$ on the performance of the $\overline{\boldsymbol{X}}$ control chart when the process standard deviation is estimated. In both panels, the horizontal lines correspond to the value of the nominal false alarm rate 0.0027 and the in-control average run length, 370.4, respectively, in Case KK, when the 3-Sigma limits are used. It can be seen, that for $\boldsymbol{n = 5}$, the $\boldsymbol{CFAR_{KU}}$ and $\boldsymbol{CARL_{0,KU}}$ curves are significantly closer to the horizontal line when *m*, the number of reference samples, is large (compare for example the curves for *m* = 10 and *m* = 500). This means that the difference between the nominal and the realized (conditional) false alarm rate is considerably more likely to be large when $\boldsymbol{m}$ is small. It is also interesting to note that the effect is different on the two sides of $\boldsymbol{u = 0.5}$ (the 0.5 quantile or the median of $\boldsymbol{Y}$). For $\boldsymbol{CFAR_{KU}}$ the situation is worse for the lower quantiles whereas for $\boldsymbol{CARL_{0,KU}}$ the reverse is true. This means that for smaller values of *m*, a lower than average estimator of the standard deviation would produce a higher than nominal conditional false alarm rate while the reverse is true for the conditional in-control average run length. This is caused by the skewness of the distributions of $\boldsymbol{CFAR_{KU}}$ and $\boldsymbol{CARL_{0,KU}}$, which will be seen more clearly in Section 4.

Note that it is possible to construct similar figures in Case UU, which will be a three-dimensional graphic since in this case the $CFAR_{UU}$ (and $CARL_{0,UU}$) is a function of two random variables (not just one). The effect of the number of Phase

I samples (m) in Case UU will be clearly seen in the next section, where the distribution of $CFAR_{UU}$ and $CARL_{0,UU}$ are derived, so it is omitted here.

4. Cumulative Distribution Function of $CFAR$ and $CARL_0$

From Equation (6a), we can easily obtain the c.d.f. of $CFAR_{KU}$

$$F_{CFAR_{KU}}(t) = P(CFAR_{KU} \leq t) = P\left(2\Phi\left(-L\sqrt{\frac{Y}{m(n-1)}}\right) \leq t\right),$$

$$= P\left(-L\sqrt{\frac{Y}{m(n-1)}} \leq \Phi^{-1}\left(\frac{t}{2}\right)\right) = 1 - F_{\chi^2_{m(n-1)}}\left(m(n-1)\left(-\frac{\Phi^{-1}\left(\frac{t}{2}\right)}{L}\right)^2\right),$$

$$0 \leq t \leq 1, \qquad \text{(8a)}$$

where $F_{\chi^2_{m(n-1)}}$ denotes the c.d.f. of a central chi-square distribution with $m(n-1)$ degrees of freedom.

In Case UU, since the $CFAR_{UU}$ is a function of two random variables (Y and Z), the derivation of an exact closed form for expression of $F_{CFAR_{UU}}$ is more involved. To this end, we derive $F_{CFAR_{UU}}$ in Case UU by using the conditioning-unconditioning technique [see Chakraborti (2000)], by first conditioning on Z [see Equation (6b)], calculating the conditional c.d.f., and then unconditioning, by taking the expectation of the conditional c.d.f. over the distribution of Z:

$$F_{CFAR_{UU}}(t) = P(CFAR_{UU} \leq t) = E_Z\big(P(CFAR_{UU} \leq t | Z = z)\big)$$

$$= \int_{-\infty}^{\infty} P(CFAR_{UU} \leq t | Z = z) f_Z(z) dz, \qquad \text{(8b)}$$

where f_Z denotes the probability density function (p.d.f.) of Z.

The next step is to derive the expression for the conditional c.d.f. $P(CFAR_{UU} \leq t | Z = z)$. Note that, given z, $P(CFAR_{UU} \leq t | Z = z)$ is a function of the chi-square random variable Y. So, from Equation (6b) one can write:

$P(CFAR_{UU} \leq t | Z = z)$

$$= P\left(1 - \left[\Phi\left(\frac{z}{\sqrt{m}} + L\sqrt{\frac{Y}{m(n-1)}}\right) - \Phi\left(\frac{z}{\sqrt{m}} - L\sqrt{\frac{Y}{m(n-1)}}\right)\right] \leq t\right)$$

$$= P\left(P\left(\frac{z}{\sqrt{m}} - L\sqrt{\frac{Y}{m(n-1)}} \leq Z_1 \leq \frac{z}{\sqrt{m}} + L\sqrt{\frac{Y}{m(n-1)}}\right) \geq 1-t\right),$$

where Z_1 also follows a standard normal distribution, independent of Z and Y. Therefore

$$P(CFAR_{UU} \leq t | Z = z)$$
$$= P\left(P\left(-L\sqrt{\frac{Y}{m(n-1)}} \leq Z_1 - \frac{z}{\sqrt{m}} \leq L\sqrt{\frac{Y}{m(n-1)}}\right) \geq 1-t\right)$$
$$= P\left(P\left(\left(Z_1 - \frac{z}{\sqrt{m}}\right)^2 \leq \left(L\sqrt{\frac{Y}{m(n-1)}}\right)^2\right) \geq 1-t\right). \tag{8c}$$

Now, since $\left(Z_1 - \frac{z}{\sqrt{m}}\right)^2$ follows a non-central chi-square distribution with 1 degree of freedom and non-centrality parameter $\frac{z^2}{m}$, denoting $\left(Z_1 - \frac{z}{\sqrt{m}}\right)^2 = \chi^2_{1,\left[\frac{z^2}{m}\right]}$, Equation (8c) can be expressed as

$$P(CFAR_{UU} \leq t | Z = z) = P\left(P\left(\chi^2_{1,\left[\frac{z^2}{m}\right]} \leq L^2 \frac{Y}{m(n-1)}\right) \geq 1-t\right) = 1 - F_{\chi^2_{m(n-1)}}\left(\frac{m(n-1)F^{-1}_{\chi^2_{1,\left[\frac{z^2}{m}\right]}}(1-t)}{L^2}\right), \tag{8d}$$

where $F^{-1}_{\chi^2_{1,\left[\frac{z^2}{m}\right]}}(1-t)$ denotes the $(1-t)$-quantile of a non-central chi-square distribution with 1 d.f. and non-centrality parameter $\frac{z^2}{m}$. Applying the result from Equation (8d) in Equation (8b), we have the final expression for the c.d.f. of $CFAR_{UU}$:

$$F_{CFAR_{UU}}(t)$$
$$= 1 - \int_{-\infty}^{\infty} F_{\chi^2_{m(n-1)}}\left(\frac{m(n-1)F^{-1}_{\chi^2_{1,\left[\frac{z^2}{m}\right]}}(1-t)}{L^2}\right) f_Z(z)dz. \tag{8e}$$

For the distribution of the conditional in-control average run-length, note that, in general, as shown in Equation (7), the $CARL_0$ is a monotonically decreasing function of $CFAR$, so that the c.d.f. of $CARL_0$ (F_{CARL_0}) can be obtained from the c.d.f of $CFAR$ (F_{CFAR}). Thus,

$$F_{CARL_0}(w) = P(CARL_0 \leq w) = P(1/CFAR \leq w) = P(CFAR \geq w^{-1}) = 1 - F_{CFAR}(w^{-1}), \quad w \geq 1. \quad (9)$$

Hence, in Case KU,

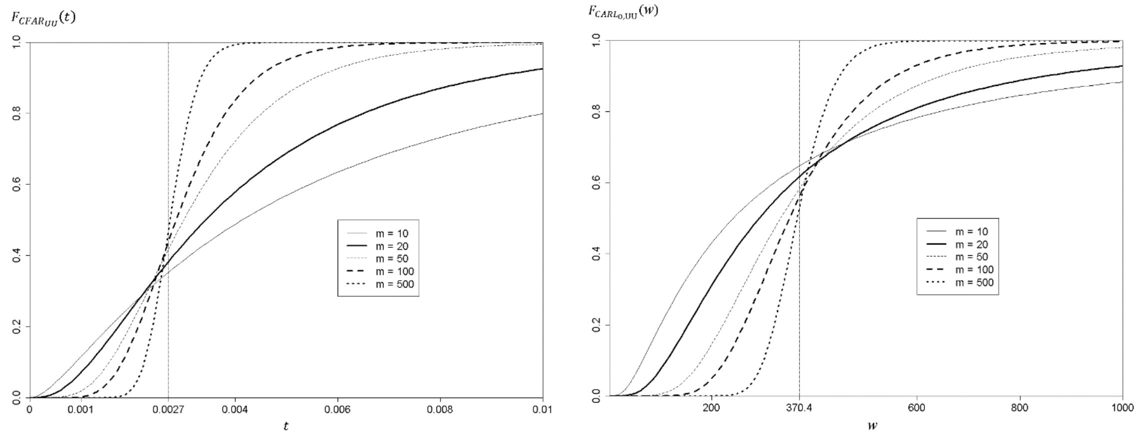
$$F_{CARL_{0,KU}}(w) = F_{\chi^2_{m(n-1)}}\left(m(n-1)\left(-\frac{\Phi^{-1}\left(\frac{w^{-1}}{2}\right)}{L} \right)^2 \right). \quad (10a)$$

In Case UU, the $F_{CARL_{0,UU}}$ can be similarly obtained by using Equations (8e) and (9) and is given by

$$F_{CARL_{0,UU}}(w) = \int_{-\infty}^{\infty} F_{\chi^2_{m(n-1)}}\left(\frac{m(n-1)F^{-1}_{\chi^2_{1,\left[\frac{z^2}{m}\right]}}(1-w^{-1})}{L^2} \right) f_Z(w)dz. \quad (10b)$$

Note that Expressions (8e) and (10b) for the c.d.f's in Case UU are exact, however their evaluation involves calculating the integral using some numerical method, since there is no closed form solution. This is not difficult as will be seen later. Indeed, many well-known c.d.f.'s are expressed in terms of integrals, including the one for the celebrated normal distribution.

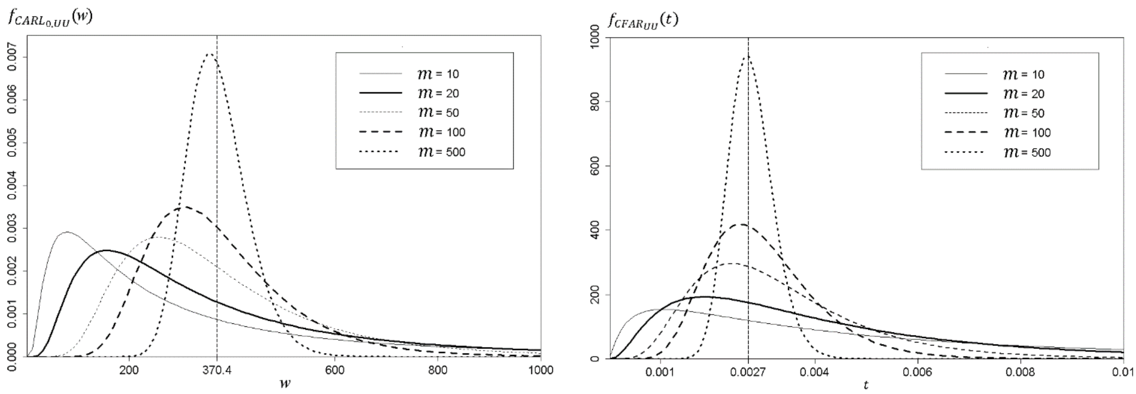
Figure 2 shows the c.d.f. of the $CFAR_{UU}$ and the $CARL_{0,UU}$, in Panel (a) and (b), respectively, calculated using Equation (8e) and (10b), for $n = 5$, $m = 10, 20, 50, 100, 500$ and $\alpha = 0.0027$ (i.e., $L = 3$). Note that the vertical lines show the nominal false alarm rate 0.0027 (in Panel (a)) and the in-control average run length 370.4 (in Panel (b)). The impact of m on the distributions is clear. When m is small (such as $m = 10$), chances are high that the realized false alarm rate is higher than the nominal one. For example, from Figure 2, Panel (a), for $m = 10$, $P(CFAR_{UU} \geq 0.006) \approx 40\%$, so that there is a 40% chance that the conditional false alarm rate is 122% higher than the nominal 0.0027. Also note the significant difference between the vertical line and the c.d.f. curve for smaller values of m. When m gets larger (such as $m = 500$), the c.d.f curves are much "closer" to the vertical line, meaning that in these cases, the $CFAR_{UU}$ is likely to be not much different from 0.0027. Similar explanations hold for $CARL_{0,UU}$. Figures for the c.d.f. of, $CARL_{0,KU}$ and $CFAR_{KU}$ are omitted for space reasons; but the conclusions from them would be similar.



(a) (b)

FIGURE 2. The c.d.f. of $\boldsymbol{CFAR_{UU}}$ (Panel (a)) and $\boldsymbol{CARL_{0,UU}}$ (Panel (b)) for $\boldsymbol{n = 5}$, $\boldsymbol{m = 10, 20, 50, 100,\ 500}$ and $\boldsymbol{\alpha = 0.0027}$ (i.e., $\boldsymbol{L = 3}$).

To provide further insight, in Figure 3 we display the p.d.f. of $CARL_{0,UU}$ ($f_{CARL_{0,UU}}$) and $CFAR_{UU}$ ($f_{CFAR_{UU}}$), in panels (a) and (b), respectively, calculated by taking the numerical derivatives of the corresponding c.d.f. The $f_{CARL_{0,UU}}$ plot shows the large density at values well below 370.4 (including the position of the modes), meaning that when parameters are estimated, in practice, there is a large probability that the $CARL_{0,UU}$ is substantially smaller (and the $CFAR_{UU}$ is substantially larger) than the nominal value, even with a number of Phase I samples quite larger than the usually recommended 25, 30 or 50 Phase I samples. This is reflected in the long right tails of the density functions of $CFAR_{UU}$ and $CARL_{0,UU}$. Note that we also omitted the figures of the p.d.f. of $CARL_{0,KU}$ and $CFAR_{KU}$ because they are similar to Figure 3. In summary, this examination clearly raises concern about the realized $CFAR$ (or the $CARL_0$) being so much different from their nominal values in practical terms and will be discussed further in the next sections through two applications.



(a) (b)

FIGURE 3. p.d.f. of $\boldsymbol{CARL_{0,UU}}$ and $\boldsymbol{CFAR_{UU}}$ for $\boldsymbol{n = 5}$, $\boldsymbol{m = 10, 20, 50, 100, 500}$ and $\boldsymbol{\alpha = 0.0027}$ ($\boldsymbol{L = 3}$).

5. Prediction Bounds for $CFAR$ and $CARL_0$

Since the $CFAR$ (and the $CARL_0$) are both random variables, it is of interest to the practitioner to know how far they can vary from their desired nominal desired values. For example, since the $CFAR$ can take any value between 0 and 1, it may be of interest to know, in a given Phase II application, what an upper bound to the $CFAR$ is, with a certain (high) probability, say $(1 - p)$. In the same spirit as in a confidence bound, this upper bound, denoted α_p, is called an upper prediction bound.

Thus, for a given m and n, it is of interest to find α_p $(0 < \alpha_p < 1)$, for a small p $(0 < p < 1)$ such that

$$P(CFAR \leq \alpha_p) = F_{CFAR}(\alpha_p) = 1 - p \tag{11}$$

which means that α_p is the $(1 - p)$-quantile of the in-control distribution of $CFAR$. Note that Equation (11) can be written as $P(CARL_0 \geq 1/\alpha_p) = 1 - p$, so that $1/\alpha_p$ forms a lower prediction bound to $CARL_0$ and is the p-quantile of the distribution of $CARL_0$. Both of these bounds are useful to the practitioner. In summary, the problem of finding α_p reduces to finding the $(1 - p)$-quantile of the $CFAR$ (or the p-quantile of $CARL_0$) distribution when the process is in-control and as it will be seen, our c.d.f. expressions, derived in Equations (8a) and (8e), are useful to this end.

For example, in Case KU, using Equation (8a), an exact expression for $\alpha_{p,KU}$ is obtained by solving the following equation for $\alpha_{p,KU}$:

$$F_{\chi^2_{m(n-1)}}\left(m(n-1)\left(-\frac{\Phi^{-1}\left(\frac{\alpha_{p,KU}}{2}\right)}{L}\right)^2\right) = p \tag{12}$$

which yields

$$\alpha_{p,KU} = 2\Phi\left(-L\sqrt{\frac{F^{-1}_{\chi^2_{m(n-1)}}(p)}{m(n-1)}}\right). \tag{13}$$

This exact upper bound can be easily calculated for given values of L, m, n and p.
On the other hand, in Case UU, the required $\alpha_{p,UU}$ is the solution to the equation

$$\int_{-\infty}^{\infty} F_{\chi^2_{m(n-1)}}\left(\frac{m(n-1)F^{-1}_{\chi^2_{1,\left[\frac{z^2}{m}\right]}}(1-\alpha_{p,UU})}{L^2}\right) f_Z(z)dz = p, \tag{14}$$

which can be solved, for given values of L, m, n and p, using a simple search method (like the Secant Method). However, a simple approximate expression for $\alpha_{p,UU}$, can be obtained from Equations (8c) and (8d) using a one-step Taylor approximation and an approximation for the c.d.f. of a non-central chi-square distribution in terms of a central chi-square distribution given by Cox and Reid (1987). Details of the derivation of the approximation can be found in Jardim et al. (2017).

$$\alpha_{p,UU} \approx 1 - F_{\chi^2_1}\left(L^2 \frac{F^{-1}_{\chi^2_{m(n-1)}}(p)}{(m+1)(n-1)}\right) \tag{15}$$

The formula in (15) can be very useful in practice when one seeks a quick answer.

Table 1 shows the values of α_p and $1/\alpha_p$ for $p = 0.05$ (i.e., the 0.95 quantile of $CFAR$ and the 0.05 quantile of $CARL_0$) and $p = 0.1$ (the 0.9 quantile of $CFAR$ and the 0.1 quantile of $CARL_0$) for some values of m and n in Case KU and Case UU, respectively. For Case UU we show the exact values calculated using Equation (13) and a search method and the values obtained using the simple approximation given by Equation (15) in bold. In all cases, we consider $\alpha = 0.0027$ ($L = 3$). Table 1 shows that when m and/or n are small, the values of $CARL_0$ that are exceeded with a large probability of 95% or 90% are much smaller than the desired $1/\alpha$. For example, in Case UU, for $m = 25$ and $n = 5$ (values suggested in many textbooks, see Montgomery, 2009), $1/\alpha_{0.05} = 102.4$, which is more than 3 times smaller than the nominal ARL_0 of 370.4. This means that, for $m = 25$ and $n = 5$, if 3-Sigma limits are used, the variabilities of $CARL_0$ and $CFAR$ are quite large, so the realization of these random variables may be very different from the nominal in-control average run length and the nominal false alarm rate (370.4 and 0.0027, respectively). Also note that the proposed approximation works well for $m \geq 50$ in Case UU.

It is also interesting to note that the lower quantiles of $CARL_0$ (or higher quantiles of $CFAR$) in Case KU are larger (or smaller, for $CFAR$) than the quantiles in Case UU. This is due to the higher variability of the $CARL_0$ and $CFAR$ in Case UU since one additional parameter (the in-control process mean) is estimated in this case.

TABLE 1. 0.95 (0.90) quantiles of $CFAR$ (α_p) and 0.05 (0.1) quantiles of $CARL_0$ ($1/\alpha_p$).

Case	p	n	α_p		$1/\alpha_p$		α_p		$1/\alpha_p$	
Case KU		**$m \rightarrow$**	**25**				**50**			
	$p = 0.05$	**5**	0.0081		123.6		0.0059		168.7	
		10	0.0057		176.3		0.0046		218.1	
		20	0.0045		221.1		0.0039		256.7	
		25	0.0043		233.8		0.0037		267.1	
	$p = 0.10$	**5**	0.0065		154.4		0.0050		198.7	
		10	0.0049		205.8		0.0041		244.0	
		20	0.0041		246.7		0.0036		277.7	
		25	0.0039		258.0		0.0035		286.6	
		$m \rightarrow$	**100**				**300**			
	$p = 0.05$	**5**	0.0047		211.3		0.0037		267.1	
		10	0.0039		254.1		0.0034		297.6	
		20	0.0035		285.5		0.0031		318.5	
		25	0.0034		293.7		0.0031		323.8	
	$p = 0.10$	**5**	0.0042		238.0		0.0035		286.6	
		10	0.0036		275.5		0.0032		312.1	
		20	0.0033		302.0		0.0030		329.2	
		25	0.0032		308.9		0.0030		333.5	
Case UU		**$m \rightarrow$**	**25**				**50**			
	$p = 0.05$	**5**	0.0098	**0.0094**	102.4	**106.3**	0.0066	**0.0064**	152.5	**155.4**
		10	0.0071	**0.0067**	140.1	**149.7**	0.0052	**0.0050**	193.6	**199.9**
		20	0.0060	**0.0054**	167.8	**186.3**	0.0045	**0.0043**	223.1	**234.6**
		25	0.0057	**0.0051**	174.5	**196.6**	0.0043	**0.0041**	230.5	**244.0**
	$p = 0.10$	**5**	0.0078	**0.0076**	128.8	**131.7**	0.0055	**0.0055**	180.2	**182.4**
		10	0.0060	**0.0058**	167.2	**173.8**	0.0046	**0.0045**	218.8	**223.3**
		20	0.0051	**0.0048**	195.3	**207.1**	0.0041	**0.0039**	246.1	**253.5**
		25	0.0049	**0.0046**	202.4	**216.2**	0.0040	**0.0038**	252.9	**261.4**
		$m \rightarrow$	**100**				**300**			
	$p = 0.05$	**5**	0.0050	**0.0049**	200.7	**202.3**	0.0038	**0.0038**	262.5	**263.1**
		10	0.0042	**0.0041**	239.5	**242.9**	0.0034	**0.0034**	292.0	**293.0**
		20	0.0037	**0.0037**	266.7	**272.5**	0.0032	**0.0032**	311.9	**313.5**
		25	0.0037	**0.0036**	273.6	**280.3**	0.0032	**0.0031**	316.9	**318.7**
	$p = 0.10$	**5**	0.0044	**0.0044**	226.3	**227.6**	0.0035	**0.0035**	281.8	**282.2**
		10	0.0038	**0.0038**	260.7	**263.2**	0.0033	**0.0033**	306.5	**307.2**
		20	0.0035	**0.0035**	284.3	**288.2**	0.0031	**0.0031**	322.9	**324.0**
		25	0.0034	**0.0034**	290.3	**294.7**	0.0031	**0.0030**	327.0	**328.2**

Note: For Case UU, the values in bold were obtained using Eq. 15; the other values are exact, calculated numerically using Eq. 14).

6. Number of Phase I Samples Required for Guaranteed In-Control Performance

Having recognized that the Phase II false alarm rate (and the in-control average run length) are both random variables when parameters are estimated and therefore vary, sometimes substantially, from estimator to estimator (reference sample to reference sample, practitioner to practitioner), another important question is the amount of Phase I data that can ensure a “satisfactory” in-control performance of the $\bar{X}$ chart. Epprecht et al. (2015) formulated this problem in terms of the EPC for the one-sided S and S^2 charts, and gave formulas for the minimum number m of Phase I samples that guarantees, with a specified high probability $1-p$ (say, 0.9), that the $CFAR$ does not exceed the nominal α by more than a user specified (tolerated) percentage 100ε (e.g. 20%). Following this approach, in any case, KU or UU, this problem is formulated as:

Given the values of n, α, ε and p, find the minimum number of in-control Phase I samples, m, such that

$$P(CFAR \leq (1+\varepsilon)\alpha) = 1-p, \tag{16}$$

where 100ε is called the *tolerance factor*.

According to Equation (16), $(1+\varepsilon)\alpha$ is the $(1-p)$–quantile of $CFAR$, which plays the role of α_p in the previous section. The difference is that in the previous section, m was given and α_p was calculated, whereas now, α_p is specified (by the practitioner) to be $(1+\varepsilon)\alpha$ (as the upper bound to $CFAR$ that may be tolerable) and m is to be determined corresponding to that value. Note that, since m is an integer, a perfect match is generally not possible, so, we re-state the problem as finding the smallest integer m such that $P(CFAR \leq (1+\varepsilon)\alpha) \geq 1-p$, for given $\varepsilon,\ p,\ \alpha,\ m$ and n. Also, note that this problem is equivalent to finding the smallest integer m such that $P(CARL_0 \geq 1/[(1+\varepsilon)\alpha]) \geq 1-p$. A direct formula for m is not available because the c.d.f.’s involve a quantile of a chi-square variable whose number of degrees of freedom is itself a function of the unknown m. However, m can be found using a simple search method (as the Secant Method, for example) since, for large values of $(1-p)$, $F_{CFAR}\big((1+\varepsilon)\alpha\big)$ is a monotonic increasing function of m (see Figure 2). Basically, this means that for Case KU and UU, we need to solve, respectively,

$$F_{\chi^2_{m(n-1)}}\left(m(n-1)\left(-\frac{\Phi^{-1}\left(\frac{(1+\varepsilon)\alpha}{2}\right)}{L}\right)^2\right)=p \tag{17}$$

and

$$\int_{-\infty}^{\infty} F_{\chi^2_{m(n-1)}}\left(\frac{m(n-1)F^{-1}_{\chi^2_{1,\left[\frac{z^2}{m}\right]}}(1-(1+\varepsilon)\alpha)}{L^2}\right) f_Z(z)dz = p, \tag{18}$$

for the smallest m. Note that Equations (17) and (18) follow from Eq. (12) and Eq. (14), respectively, when we replace α_p by $(1+\varepsilon)\alpha$. Table 2 shows the minimum number of in-control Phase I samples, m, for $\varepsilon = 0.1, 0.2, 0.3, 0.4, 0.5$, $\alpha = 0.0027$, $p = 5\%, 10\%, 15\%$, and n = 5, 10, 20 and 25. As it can be seen, for small values of n, one needs a larger number of reference samples (m) to guarantee such conditional performance.

TABLE 2. Minimum number of Phase I reference samples, $\boldsymbol{m}$, required for $\boldsymbol{P(CFAR \leq (1+\varepsilon)\alpha) \geq 1-p}$ (or $\boldsymbol{P(CARL_0 \geq 1/[(1+\varepsilon)\alpha]) \geq 1-p}$) for $\boldsymbol{\alpha = 0.0027}$ ($\boldsymbol{L = 3}$) and various values of $\boldsymbol{\varepsilon}$, **n and** $\boldsymbol{p}$

		$\varepsilon = 10\%$			$\varepsilon = 20\%$			$\varepsilon = 30\%$			$\varepsilon = 40\%$			$\varepsilon = 50\%$		
Case	n	$\boldsymbol{p} \rightarrow$ 0.05	0.10	0.15	0.05	0.10	0.15	0.05	0.10	0.15	0.05	0.10	0.15	0.05	0.10	0.15
Case KU	5	3588	2185	1435	975	595	393	468	287	190	283	174	116	194	120	80
	10	1595	971	638	433	265	175	208	128	85	126	78	52	87	53	36
	20	756	460	303	206	126	83	99	61	40	60	37	25	41	26	17
	25	598	365	240	163	100	66	78	48	32	48	29	20	33	20	14
Case UU	5	3687	2285	1536	1029	649	446	507	324	226	314	203	144	219	144	103
	10	1701	1077	742	492	321	229	250	167	122	159	108	80	114	78	59
	20	871	571	409	270	185	138	145	102	77	97	68	52	72	51	39
	25	717	477	346	230	160	120	126	89	69	85	61	47	64	46	36

From Table 2 we see that in the majority of cases (for example, when $\varepsilon = 0.1$ and n = 5), the minimum numbers of reference samples required are larger than the 25 or 30 subgroups traditionally proposed in most manuals and textbooks (see Montgomery, 2009); they can also be larger than the 200 or 300 samples proposed by authors who focused on the unconditional ARL_0 (see Quesenberry, 1993 and others) and even larger than the recent numbers recommended by Saleh et al. (2015), who focused on the standard deviation of $CARL_0$ as an additional performance metric (they recommended using $m = 1200$ when $\alpha = 0.0027$ is used). One can see that, as might be expected, in Case UU more Phase I samples are needed than in Case KU. This happens because in Case UU the estimation of the in-control process mean adds more uncertainty (variation) in the performance of the $\bar{X}$ control chart.

In case the required m for the specified values of $(1+\varepsilon)\alpha$ and p is infeasible and yet relaxing the value of either ε or p is unacceptable on practical grounds, a possible solution is to change the value of the control limit factor L (instead of using $L = 3$, the common 3-sigma limits), given the values of m and n at hand, in order to satisfy the *EPC* in the in-control situation. This is discussed in the next section.

7. Adjustment of the Limits for a Guaranteed Conditional In-Control Performance

In the previous section, we saw that the minimum number of reference samples required to guarantee a desired in-control performance under the *EPC* can be very large and may be infeasible in many practical situations. Given this practical hurdle, in this section, we present an exact and an approximate adjusted control limit for the $\bar{X}$ chart (for any m and n) that guarantees a desired in-control performance in terms of *EPC*. The idea is to replace the control limit factor L in Equations (1) and (2) by $L(p,\varepsilon,\alpha)$, where $L(p,\varepsilon,\alpha)$ represents the value of the control limit factor which guarantees that $P(CFAR \leq (1+\varepsilon)\alpha) = 1-p$ (or $P(CARL_0 \geq 1/(1+\varepsilon)\alpha) = 1-p$) for given values of ε, α, m and n.

In Case KU, since the expression for the c.d.f of $CFAR$ is in a simple form given by Equation (8a), we can derive a closed-form expression for $L_{KU}(p,\varepsilon,\alpha)$. Using Equation (16), replacing L by $L(p,\varepsilon,\alpha)$ in Equation (8a) and rearranging the terms, one has:

$$L_{KU}(p,\varepsilon,\alpha) = \frac{\Phi^{-1}\left(\frac{(1+\varepsilon)\alpha}{2}\right)}{\sqrt{\frac{F^{-1}_{\chi^2_{m(n-1)}}(p)}{m(n-1)}}} \tag{19}$$

In Case UU, we can find $L_{UU}(p,\varepsilon,\alpha)$ by solving the following equation

$$\int_{-\infty}^{\infty} F_{\chi^2_{m(n-1)}}\left(\frac{m(n-1)F^{-1}_{\chi^2_{1,\left[\frac{z^2}{m}\right]}}(1-(1+\varepsilon)\alpha)}{L_{UU}(p,\varepsilon,\alpha)^2}\right) f_Z(z)dz = p \tag{20}$$

for $L_{UU}(p,\varepsilon,\alpha)$ using a search method. Equation (20) relates to the theory of Tolerance Intervals. Krishnamoorthy and Mathew (2009, p. 30) give an equation which is equivalent to (20) where S and $\bar{X}$ are used instead of S_p and $\bar{\bar{X}}$ respectively. But, they did not make the relationship with Statistical Process Control area. Alternatively, Jardim et al. (2017) derived the following approximate formula for $L_{UU}(p,\varepsilon,\alpha)$:

$$L_{UU}(p,\varepsilon,\alpha) \approx \sqrt{(n-1)(m+1)\frac{F^{-1}_{\chi^2_1}(1-(1+\varepsilon)\alpha)}{F^{-1}_{\chi^2_{m(n-1)}}(p)}} \tag{21}$$

Note that, as explained in Section 1, other authors have considered adjustments in terms of the *EPC* for the $\bar{X}$ chart constant in Case UU. For example, Saleh et al. (2015a) used bootstrapping and Goedhart et al. (2017a and 2018) provided approximate formulas for the *correction term* defined as $c = L_{UU}(p,\varepsilon,\alpha) - L$. Jardim et al. (2017) made a detailed comparison between all these methods [including the one we propose in Equation (20)] and conclude that all of them provide reasonably good and similar results (the one in Equation (20) being the most accurate one). The differences among these methods lie mainly in their complexity in terms of formulas and algorithms. So, here, we just show the approximation in (21) because, as shown by Jardim et al. (2017), this equation is simpler than the available approximations and provides good results — as it can be seen in Table 3. Nevertheless, since Equation (20) is based on the exact c.d.f., it provides the most accurate results, and since much of SPC is expected to be implemented with the help of software, we recommend it to be used and implemented in a SPC software.

Table 3 shows the exact values of $L_{KU}(p,\varepsilon,\alpha)$ and $L_{UU}(p,\varepsilon,\alpha)$ using Equations (18) and (19) and (in bold) the approximate values of $L_{UU}(p,\varepsilon,\alpha)$ given by Equation (20), for some values of m and n, $p = 10\%$, $\varepsilon = 0$ and 0.2, respectively, and $\alpha = 0.0027$ ($L = 3$). This means $P(CARL_0 \geq 370.4) = 90\%$ and $P(CARL_0 \geq (1/1.2)370.4 = 308.67) = 90\%$. So, for example, if the users have 25 reference samples each with 10 observations from a Phase I analysis, then to guarantee that $P(CARL_0 \geq 370.4) = 90\%$, they should replace L by $L_{KU}(10\%, 0\%, 0.0027) = 3.2$ in Case KU or by $L_{UU}(10\%, 0\%, 0.0027) = 3.27$ in Case UU, in Equations 1 and 2. Note that the approximation given by (21) is very accurate.

TABLE 3. Values of $\boldsymbol{L(10\%, 0\%, 0.0027)}$ and $\boldsymbol{L(10\%, 20\%, 0.0027)}$ for $\boldsymbol{P(CARL_0 > 370.4) = 90\%}$ for $\boldsymbol{n = 5, 10, 15, 20}$, $\boldsymbol{\varepsilon = 0, 0.20}$ and $\boldsymbol{m = 25, 50, 100, 300, 1000}$ in Case KU and UU

		m																			
		25				**50**				**100**				**300**				**1000**			
Case	***n*** $\varepsilon \rightarrow$	**0**		**0.2**		**0**		**0.2**		**0**		**0.2**		**0**		**0.2**		**0**		**0.2**	
Case KU	**5**	3.31		3.24		3.21		3.15		3.14		3.09		3.08		3.02		3.04		2.99	
	10	3.20		3.14		3.14		3.08		3.09		3.04		3.05		3.00		3.03		2.97	
	15	3.15		3.10		3.11		3.05		3.07		3.02		3.04		2.99		3.02		2.97	
	20	3.13		3.07		3.09		3.03		3.06		3.01		3.04		2.98		3.02		2.96	
Case UU	**5**	3.38	**3.37**	3.32	**3.31**	3.24	**3.24**	3.18	**3.18**	3.16	**3.16**	3.10	**3.10**	3.09	**3.09**	3.03	**3.03**	3.05	**3.05**	2.99	**2.99**
	10	3.27	**3.26**	3.21	**3.20**	3.17	**3.17**	3.11	**3.11**	3.11	**3.11**	3.05	**3.05**	3.06	**3.06**	3.00	**3.00**	3.03	**3.03**	2.97	**2.97**
	15	3.23	**3.22**	3.17	**3.16**	3.15	**3.14**	3.09	**3.08**	3.09	**3.09**	3.04	**3.03**	3.05	**3.05**	2.99	**2.99**	3.02	**3.02**	2.97	**2.97**
	20	3.21	**3.19**	3.15	**3.13**	3.13	**3.12**	3.07	**3.06**	3.08	**3.08**	3.03	**3.02**	3.04	**3.04**	2.99	**2.98**	3.02	**3.02**	2.97	**2.96**

Note: For Case UU, the values in bold were obtained using Equation (21); the other values are exact, calculated numerically using Equation (20).

From Table 3, it is seen that when $\varepsilon = 0$, the control limit factor is larger than 3 in all cases, making the control limits wider, and when $\varepsilon = 0.2$ this factor can be smaller than 3 (turning the limits narrower) only when m is quite large. It should be noted that these findings contrast with the results obtained by the authors who adjusted the $\bar{X}$ chart control limits focusing on a desired unconditional ARL_0, for example, Chakraborti (2006) and Goedhart et al (2016). With the unconditional in-control *ARL* as a performance criterion, they found, in most of the cases, adjusted limit factors smaller than 3 for $\alpha = 0.0027$.

Since the adjustment based on the conditional in-control performance of the $\bar{X}$ control chart in most cases results in a control limit factor greater than 3, which widens the interval between the control limits, the question arises about its impact on the out-of-control performance of the chart. In the next section, we examine this issue.

8. Out–of-control Analysis

In this section, we analyze the impact of the adjustments proposed in the previous section on the out-of-control performance of the $\bar{X}$ chart for Case UU. The results and conclusions for Case KU are similar and are omitted. As we noted earlier, in most cases the adjustment leads to widening the interval between the control limits (see Table 3 where $L_{UU}(p, \varepsilon, \alpha) > 3$ in most cases). In these cases, the out-of-control *conditional ARL* (i.e., the $CARL_{\delta,UU}$ with $\delta \neq 0$) will be larger with the adjusted limits than with the unadjusted limits. This is the price to pay for guaranteeing a desired in-control performance. So, it is important to assess the deterioration in the $CARL_{\delta,UU}$ due to the adjustment. This assessment will enable the user to choose an appropriate compromise, in terms of $m, n, \varepsilon,$ and $p,$ since the out-of-control performance deterioration is lesser with larger *m* and *n*, and also with larger values of ε and p. For example, for *m* = 25 and *n* = 5 (a typical amount of reference data in practice and according to traditional recommendations), the adjustments proposed in the last section enable achieving the desired conditional in-control performance (for example, $P(CARL_0 > 370.4) = 90\%$, as shown in Table 3) in terms of the *EPC*, however they may produce the undesirable effect of deteriorating the out-of-control performance. Still considering this typical amount of data (i.e., *m* = 25 and *n* = 5), to detect a shift in the process mean of the size of one process standard deviation (i.e., $|\delta| = 1$), with no adjustment, the chart will have $P\big(CARL_{1,UU} > 7.25\big) = 10\%$, which means that the average number of samples until a true alarm will be most likely below 10 samples. However, with the adjustment (in order to achieve $P(CARL_0 > 370.4) = 90\%$), the chart will have $P\big(CARL_{1,UU} > 15.98\big) = 10\%$: a difference of 8.23 (more than 100%) on the 0.9-quantile of the $CARL_{1,UU}$. Note that an out-of-control *ARL* of 15.98 may be unacceptable for the practitioner. However, with *m* = 50 and *n* = 5, with no adjustment $P\big(CARL_{1,UU} > 6.55\big) = 10\%$ and with adjustment, $P\big(CARL_{1,UU} > 9.99\big) = 10\%$. So, with *m* = 50 and *n* = 5, either with or without the adjustment, the $CARL_{1,UU}$ will most likely be below 10 samples, however, only with the adjustment one can guarantee that $P(CARL_0 > 370.4) = 90\%$. Also, the difference between the 0.9-quantiles of the $CARL_{1,UU}$ with and without the adjustment, is 3.44 (about 50%). So, a particular practitioner may consider adjusting the limits with *m* = 50 and *n* = 5 a good compromise solution between the number and size of subgroups to collect in Phase I, a desired nominal in-control performance and a reasonable out-of-control performance of the $\bar{X}$ chart.

It becomes evident from the above example that knowing the prediction bound for the $CARL_{\delta,UU}$, with adjusted and with unadjusted limits, is useful for assessing the deterioration (increase) in the $CARL_{\delta,UU}$ due to the adjustments. The lower prediction bound for $CARL_{\delta,UU}$ can be calculated similarly as presented in Section 5 for bounds for $CARL_0$ in the in-control situation. That is, for a given δ, m and n, we can find the distribution of $CARL_{\delta,UU}$ following the same steps used for finding the c.d.f. of $CARL_{0,UU}$ presented in Section 4, and use that to find a lower bound (denoted $Q_{p_{OOC}}$) that has only a low (specified) probability p_{OOC} (e.g. 0.10) of being exceeded. Formally: for a given p_{OOC},

$$P\left(CARL_{\delta,UU} > Q_{p_{OOC}}\right) = p_{OOC} \tag{22}$$

Thus, $Q_{p_{OOC}}$ is the $(1 - p_{OOC})$-quantile of the $CARL_{\delta,UU}$ distribution. Since the $CARL_{\delta,UU}$ is the realized average number of samples until a true alarm, the smaller the $Q_{p_{OOC}}$, the better the chart's OOC performance. Table 4 presents the values of $Q_{p_{OOC}}$ in Case UU with the adjusted limits from Table 3, for $\varepsilon = 0$ and $p = 0.1$ (in grey) and with unadjusted limits, $L = 3$ (in white), for the same values of m and n presented in Table 3, for mean shifts $|\delta| = 0.5$, $|\delta| = 1$ and $|\delta| = 1.5$ and for $p_{OOC} = 0.05$ and $p_{OOC} = 0.1$. Finally, Table 4 also shows the differences (in bold) between the $Q_{p_{OOC}}$ values with the adjusted and the unadjusted limits, respectively, to enable a direct performance comparison.

An examination of Table 4 shows that for $|\delta| = 1$ (a shift in the mean of one standard deviation) and $p_{OOC} = 0.05$, $m = 25$ and $n = 5$, the difference between the $Q_{p_{OOC}}$ values, with and without the adjustment, is of 10.87 samples on average. This is a difference of about 100%, but note that we are considering a 0.95 quantile, a small sample size and a very small number of initial samples, in addition to the fact that a shift of 1 standard deviation in which case the efficacy of the $\bar{X}$ chart may be questionable. For a slightly larger shift, say $|\delta| = 1.5$, the difference between the values of $Q_{p_{OOC}}$ with and without adjustment is only 1.14 samples on average. So, for shifts of this magnitude or larger (i.e., $|\delta| \geq 1.5$), the impact of the adjustment on the out-of-control performance is small for any value of n and m. However, for a smaller shift, say $|\delta| = 0.5$, the $Q_{p_{OOC}}$ is large in most cases. For example, for $p_{OOC} = 0.05$, $m = 25$ and $n = 5$, $Q_{p_{OOC}}$ is 107.85 and 351.98 with unadjusted and adjusted limits, respectively, that is an increase of 224.13 samples on an average. This shows that there can be a negative impact of the adjustment on the OOC performance for smaller shifts. However, this is not a surprise since the $\bar{X}$ chart is usually not recommended for signaling mean shifts smaller than 1 standard deviation.

TABLE 4. The 0.95 and 0.9 quantiles of $CARL_{\delta,UU}$ with adjusted limits ($\alpha = 0.0027$, $p = 0.1$ and $\varepsilon = 0$) in grey and unadjusted limits ($L = 3$) in white for different values of m, n and δ

			m														
			25			**50**			**100**			**300**			**1000**		
δ	p_{OOC}	n	**unadj.**	**adj.** $p = 0.1$ $\varepsilon = 0$	**difference**	**unadj.**	**adj.** $p = 0.1$ $\varepsilon = 0.$	**difference**	**unadj.**	**adj.** $p = 0.1$ $\varepsilon = 0.$	**difference**	**unadj.**	**adj.** $p = 0.1$ $\varepsilon = 0$	**difference**	**unadj.**	**adj.** $p = 0.1$ $\varepsilon = 0$	**difference**
$\delta = \pm 1.5$	$p_{OOC} = 0.05$	**5**	2.21	3.36	**1.14**	1.97	2.47	**0.50**	1.83	2.09	**0.26**	1.71	1.82	**0.11**	1.64	1.69	**0.05**
		10	1.10	1.17	**0.07**	1.08	1.11	**0.03**	1.07	1.08	**0.02**	1.05	1.06	**0.01**	1.05	1.05	**0.00**
		15	1.01	1.01	**0.01**	1.01	1.01	**0.00**	1.00	1.01	**0.00**	1.00	1.00	**0.00**	1.00	1.00	**0.00**
		20	1.00	1.00	**0.00**	1.00	1.00	**0.00**	1.00	1.00	**0.00**	1.00	1.00	**0.00**	1.00	1.00	**0.00**
	$p_{OOC} = 0.1$	**5**	2.02	2.95	**0.93**	1.86	2.30	**0.44**	1.76	2.00	**0.23**	1.67	1.78	**0.10**	1.62	1.67	**0.05**
		10	1.08	1.14	**0.06**	1.07	1.10	**0.03**	1.06	1.08	**0.02**	1.05	1.06	**0.01**	1.05	1.05	**0.00**
		15	1.01	1.01	**0.01**	1.00	1.01	**0.00**	1.00	1.01	**0.00**	1.00	1.00	**0.00**	1.00	1.00	**0.00**
		20	1.00	1.00	**0.00**	1.00	1.00	**0.00**	1.00	1.00	**0.00**	1.00	1.00	**0.00**	1.00	1.00	**0.00**
$\delta = \pm 1$	$p_{OOC} = 0.05$	**5**	9.27	20.14	**10.87**	7.37	11.50	**4.12**	6.33	8.27	**1.95**	5.45	6.21	**0.77**	4.99	5.32	**0.34**
		10	2.46	3.32	**0.86**	2.21	2.62	**0.40**	2.06	2.28	**0.21**	1.93	2.02	**0.09**	1.85	1.90	**0.04**
		15	1.45	1.66	**0.22**	1.37	1.48	**0.11**	1.33	1.38	**0.06**	1.29	1.31	**0.03**	1.26	1.27	**0.01**
		20	1.15	1.23	**0.08**	1.13	1.16	**0.04**	1.11	1.13	**0.02**	1.09	1.10	**0.01**	1.09	1.09	**0.00**
	$p_{OOC} = 0.1$	**5**	7.75	15.98	**8.23**	6.55	9.99	**3.44**	5.84	7.56	**1.72**	5.21	5.93	**0.72**	4.87	5.19	**0.32**
		10	2.27	3.00	**0.73**	2.10	2.46	**0.36**	1.99	2.19	**0.20**	1.89	1.98	**0.09**	1.84	1.88	**0.04**
		15	1.39	1.58	**0.19**	1.34	1.43	**0.10**	1.31	1.36	**0.05**	1.27	1.30	**0.02**	1.26	1.27	**0.01**
		20	1.13	1.20	**0.07**	1.11	1.15	**0.03**	1.10	1.12	**0.02**	1.09	1.10	**0.01**	1.08	1.09	**0.00**
$\delta = \pm 0.5$	$p_{OOC} = 0.05$	**5**	107.85	351.98	**244.13**	75.24	151.07	**75.83**	58.80	90.45	**31.65**	46.05	57.15	**11.10**	39.75	44.30	**4.55**
		10	29.02	56.41	**27.39**	22.55	33.35	**10.80**	18.99	24.16	**5.17**	16.03	18.06	**2.04**	14.47	15.36	**0.89**
		15	13.34	21.55	**8.21**	10.90	14.40	**3.50**	9.50	11.25	**1.75**	8.30	9.01	**0.72**	7.65	7.97	**0.32**
		20	7.72	11.23	**3.51**	6.52	8.08	**1.56**	5.82	6.61	**0.80**	5.20	5.53	**0.33**	4.86	5.01	**0.15**
	$p_{OOC} = 0.1$	**5**	81.29	249.12	**167.82**	62.14	121.44	**59.30**	51.59	78.40	**26.82**	42.82	52.94	**10.13**	38.23	42.56	**4.33**
		10	23.89	45.11	**21.22**	19.77	28.88	**9.11**	17.35	21.95	**4.60**	15.24	17.15	**1.91**	14.08	14.94	**0.86**
		15	11.42	18.08	**6.65**	9.81	12.85	**3.04**	8.84	10.43	**1.59**	7.97	8.65	**0.68**	7.49	7.80	**0.31**
		20	6.78	9.71	**2.92**	5.98	7.35	**1.38**	5.48	6.21	**0.73**	5.03	5.35	**0.32**	4.78	4.92	**0.14**

Finally, note that a small difference in $Q_{p_{OOC}}$ values means that the adjustment guarantees the in-control performance as specified and does not significantly deteriorate the OOC performance of the chart. If we consider $Q_{p_{OOC}} \leq 10$ to be an acceptable OOC performance, Table 4 shows that both the unadjusted and the adjusted limits do not work well for $|\delta| = 1$ when $n = 5$ and $m = 25$. But in all other cases, for example, when $|\delta| \geq 1$ and $n \geq 10$ or when $|\delta| \geq 1$ and $m \geq 10$, the $Q_{p_{OOC}}$ values are either less than or close to 10 with the adjusted limits, which means the adjustment works well. The analysis can be easily replicated for other values of α, p, p_{OOC} and ε. It was done for $\varepsilon = 0.20$ and the conclusions were similar, so these results are omitted for space considerations.

Hence, we recommend the adjusted limits be used for $n \geq 10$ and $m \geq 25$, or for $n \geq 5$ and $m \geq 50$, in order to guarantee a high probability (such as 0.9) that the conditional in-control average run length is greater than a nominal in-control average run length value (370.4) and that a $Q_{p_{OOC}} \leq 10$.

9. Summary and Recommendations

For the two-sided Shewhart $\bar{X}$ control chart with estimated parameters, we derive the exact c.d.f. of the in-control conditional average run length ($CFAR$) and of its reciprocal, the conditional false alarm rate ($CARL_0$) assuming normality. These expressions, unavailable in the literature until now, enable us to examine the performance of the control chart more closely, in terms of guaranteeing a high probability of the $CARL_0$ (or the $CFAR$) being at least (or at most) equal to some specified nominal value [this is known as the Exceedance Probability Criterion (*EPC*)]. This helps the user better understand the impact of parameter estimation and the amount of Phase I data that should be used for establishing an $\bar{X}$ control chart when parameters are estimated. In order to avoid unacceptably low (high) $CARL_0$ ($CFAR$) values, exact expressions are proposed, based on the c.d.f. derived here, in order to calculate the value of m (the number of Phase I samples) required in order to guarantee a desired in-control performance in terms of the *EPC* for some Phase I sample sizes (n) values. We

show that depending on the practitioner's tolerances and on the subgroups size, m can be very large, such as 2000 samples of size 5 (i.e., a total of 10,000 Phase I data points). This number is even larger than the numbers recommended by some recent authors.

Given the unpractically large numbers (m) of Phase I samples required, we propose (adjusted) control limit factors and tabulate them so that some desired in-control nominal performance in terms of the *EPC* is achieved. Unlike other authors, who use approximations or bootstrapping to propose adjustment factors, our results are based on the exact c.d.f. of the $CARL_0$. Moreover, according to our detailed analysis of the impact of these adjustments on the out-of-control performance of the $\bar{X}$ chart, we recommend using the adjusted limits for at least $n = 10$ and $m = 25$ or $n = 5$ and $m = 50$, that is, for at least 250 reference data points (note that this required minimum total number of data points is much smaller than in the case of unadjusted limits). With these recommended amounts of data and the adjusted limits, the user can strike a balance between a desired nominal in-control conditional performance and a reasonable out-of-control shift detection capability.

Acknowledgements

This research was supported by the CNPq (Brazilian Council for Scientific and Technological Development) through projects numbers 201172/2016-0 (international split PhD scholarship for the first author), 401523/2014-4 (travel grant for the second author), 308677/2015-3 (research grant for the 3rd author) and CAPES (Brazilian Coordination for the Improvement of Higher Education Personnel) through a national PhD scholarship for the first author.

REFERENCES

Albers, W. and Kallenberg, W. C. M. (2004a). “Are Estimated Control Charts in Control?” *Statistics* 38(1), pp. 67-79.

Albers, W. and Kallenberg, W. C. M. (2004b). “Estimation in Shewhart control charts: effects and corrections”. *Metrika* 59(3), pp. 207–234.

Albers, W. and Kallenberg, W. C. M. (2005). “New Corrections for Old Control Charts”. *Quality Engineering* 17, pp. 457–473.

Albers, W., Kallenberg, W. C. M. and Nurdiati S. (2005) “Exceedance probabilities for parametric control charts”. *Statistics* Vol. 39 Issue 5, pp. 429-443.

Aly, A. A., Saleh, N. A., Mahmoud, M. A. and Woodall, W. H. (2015) “A Reevaluation of the Adaptive Exponentially Weighted Moving Average Control Chart When Parameters are Estimated.” *Quality and Reliability Engineering International* 31, pp. 1611–1622

Bischak, D. P. and Trietsch, D. (2007). “The Rate of False Signals in X Charts with Estimated Limits”. *Journal of Quality Technology* 39(1), pp. 54 - 65.

Capizzi, G. and Masarotto, G. (2003). “An adaptive exponentially weighted moving average control chart.” *Technometrics*, 45(3), pp.199-207.

Chakraborti, S. (2000). “Run Length, Average Run Length and False Alarm Rate of Shewhart X-bar Chart: Exact Derivations by Conditioning”. *Journal of Applied Statistics*, 33(4), pp. 439-459.

Chakraborti, S. (2006). “Parameter Estimation and Design Considerations in Prospective Applications of the $\bar{X}$ Chart”. *Communication in Statistics - Simulation and Computation* 29(1), pp. 61-81.

Chakraborti, S.; Graham, M. A.; and Human, S. W. (2009): “Phase I Statistical Process Control Charts: An Overview and Some Results”. *Quality Engineering*, 21, pp. 52-62.

Chen, G. (1997). “The mean and Standard Deviation of the Run Length Distributions of $\bar{X}$ Charts When Control Limits are Estimated”. *Statistica Sinica* 7, pp. 789-798.

Epprecht, E. K.; Loureiro, L. D.; Chakraborti, S. (2015). “Effect of the Amount of Phase I Data on the Phase II Performance of S^2 and S Control Charts”. *Journal of Quality Technology* 47(2). pp. 139-155.

Faraz A.; William H. Woodall W. H.; and Heuchenne C. (2015). “Guaranteed conditional performance of the S2 control chart with estimated parameters”, *International Journal of Production Research*, 53:14, pp. 4405-4413.

Faraz A.; Heuchenne C. H.; and Saniga E. (2015). “An exact method for designing Shewhart $\bar{X}$ and S^2 control charts to guarantee in-control performance”, *International Journal of Production Research*, doi: 10.1080/00207543.2017.1384580

Feller, W. (1966). “An Introduction to Probability Theory and Its Applications.”, Vol. 2. 2nd ed. New York: Wiley.

Gandy, A. and Kvaløy, J. T. (2013). “Guaranteed Conditional Performance of Control Charts via Bootstrap Methods”. *Scandinavian Journal of Statistics* 40, pp. 647-668.

Ghosh, B. K.; Reynolds, M. R., Jr.; and Hui, Y. V. (1981). “Shewhart $\bar{X}$-Chart with Estimated Process Variance”. *Communications in Statistics – Theory and Methods* 10(18), pp. 1797-1822.

Goedhart, R.; Schoonhoven, M. and Does, R.J.M.M. (2016). Correction factors for Shewhart X and $\bar{X}$ control charts to achieve desired unconditional ARL. *International Journal of Production Research* 54:24, pp: 7464-7479.

Goedhart, R.; Schoonhoven, M. and Does, R.J.M.M. (2017a). “Guaranteed In-Control Performance for the X and $\bar{X}$ Control Charts*”. Journal of Quality Technology* 49(2). pp. 155-171.

Goedhart, R.; da Silva, M. M.; Schoonhoven, M.; Epprecht, E. K.; Chakraborti, S.; Does, R. J.; & Veiga, Á. (2017b). “Shewhart control charts for dispersion adjusted for parameter estimation”. IISE Transactions, 1-11. (http://dx.doi.org/10.1080/24725854.2017.1299956)

Goedhart, R.; Schoonhoven, M. and Does, R.J.M.M. (2018). “Recent Developments and Future Directions of Estimating Control Chart Limits*”. Journal of Quality Technology* 50(1). *Paper Accepted.*

Guo, B.; and Wang, B. X. (2017). “The Design of the S^2 Control Charts Based on Conditional Performance via Exact Methods”. *Quality and Reliability Engineering International.*

Jardim, F. S.; Chakraborti, S; and Epprecht, E. K. (2017) “Two perspectives for designing a Phase II control chart with estimated parameters: The case of the Shewhart $\bar{X}$ Chart”. *Technical Memorandum 06/2017, Department of Industrial Engineering, PUC-Rio.*

Jensen, W. A.; Jones-Farmer, L. A.; Champ, C. W.; and Woodall, W. H. (2006) “Effects of Parameter Estimation on Control Chart Properties: A Literature Review”. *Journal of Quality Technology*, 38(4), pp. 349-364.

Jones-Farmer, L. A.; Woodall, W. H.; Steiner, S. H.; and Champ, C. W. (2014). “An Overview of Phase I Analysis for Process Improvement and Monitoring”. *Journal of Quality Technology* 46(3), pp. 265 280.

Krishnamoorthy, K. and Mathew, T. (2009). *Statistical Tolerance Regions: Theory, Applications, and Computation*. Wiley, New York.

Kumar, N. and Chakraborti, S. (2014) “Phase-II Shewhart-type control charts for monitoring time between events and effects of parameter estimation”. *Quality and Reliability Engineering International*. DOI: 10.1002/qre.1752

Mahmoud, M. A.; Henderson, G. R.; Epprecht, E. K.; and Woodall, W. H. (2010) “Estimating the Standard Deviation in Quality Control Applications”. *Journal of Quality Technology*, 42(4), pp. 348-357.

Montgomery, D.C. (2009). “*Introduction to Statistical Quality Control*”, 7th ed. John Wiley & Sons, Hoboken, NJ.

Moskowitz, H.; Plante, R. D. and Wardell, D. G. (1994). “Using run-length distributions of control charts to detect false alarms”. *Production and Operations Management*, 3(3), pp. 217–239.

Psarakis, S.; Vyniou, A. K.; and Castagliola, P. (2014). “Some Recent Developments on the Effects of Parameter Estimation on Control Charts”. *Quality and Reliability Engineering International* 30. pp. 641-650.

Quesenberry, C. P. (1993). “The Effect of Sample Size on Estimated Limits for $\bar{X}$ and X Control Charts”. *Journal of Quality Technology* 25(4), pp. 237-247.

Saleh, N. A.; Mahmoud, M. A.; Keefe, M. J.; and Woodall, W. H. (2015a) “The Difficulty in Designing Shewhart $\bar{X}$ and X Control Charts with Estimated Parameters”. *Journal of Quality Technology* 47(2). pp. 127-138.

Saleh, N. A.; Mahmoud, M. A.; Jones-Farmer L. A.; Zwetsloot I. and Woodall W.H.; (2015b) “Another look at the EWMA control chart with estimated parameters”. *Journal of Quality Technology* 47(4) pp. 363–382.

Schoonhoven, M. (2016). “On the Use of the Expected ARL Metric for Control Charts Based on Estimated Parameters”. *Quality Reliability Engineering International* 33 pp. 745–749. doi: 10.1002/qre.2052.

Schroeder, R.G.; Linderman, K. and Zhang, D. (2005). Evolution of quality: first fifty issues of production and operations management. Production and Operations Management 14 (4), pp. 468–481.

Trietsch, D. and Bischak, D. (1998). “The Rate of False Signals for Control Charts with Limits Estimated from Small Samples”. In: *Proceedings of the 33rd Annual Conference, The Operational Research Society of New Zealand*, pp. 232-241

Annex C - Paper Accepted in The American Statistician

Higher Order Moments Using the Survival Function: The Alternative Expectation Formula

S. Chakraborti
Department of Information Systems, Statistics and Management Science,
University of Alabama,
Tuscaloosa, AL, USA

F. Jardim
Department of Industrial Engineering
Pontifical Catholic University of Rio de Janeiro,
Rio de Janeiro, Brazil

E. Epprecht
Department of Industrial Engineering
Pontifical Catholic University of Rio de Janeiro
Rio de Janeiro, Brazil

Abstract

Undergraduate and graduate students in a first-year probability (or a mathematical statistics) course learn the important concept of the moment of a random variable. The moments are related to various aspects of a probability distribution. In this context, the formula for the mean or the first moment of a non-negative continuous random variable is often shown in terms of its c.d.f. (or the survival function). This has been called the alternative expectation formula. However, higher order moments are also important, for example, to study the variance or the skewness of a distribution. In this note, we consider the *r*th moment of a non-negative random variable and derive formulas in terms of the c.d.f. (or the survival function) paralleling the existing results for the first moment (the mean) using Fubini's theorem. Both continuous and discrete non-negative integer-valued random variables are considered. These formulas may be advantageous, for example,

when dealing with the moments of a transformed random variable, where it may be easier to derive its c.d.f. using the so-called c.d.f. method.

Keywords: Non-Negative Random Variables, Cumulative Distribution Function, Fubini's Theorem, Discrete and Continuous Random Variables

1. INTRODUCTION

Moments of random variables play a key role in describing and understanding of probability distributions. For a continuous non-negative random variable X, it is well known that the mean or the first moment, when it exists, can be expressed as

$$E(X) = \int_0^\infty \left(1 - F_X(x)\right) dx, \tag{1}$$

where $F_X(x)$ is the c.d.f. of X. The function $1 - F_X(x)$, which is the probability that X exceeds x, is commonly known as the survival function of X, which has a long history in the analysis of life tables, and has long been used in the actuarial, bio-statistical, demographic, and engineering applications [see, for example, Keyfitz (1968, page 6)]. Recently, there has been a good bit of interest in Formula (1), see for example, Hong (2012, 2015), who calls it the alternative expectation formula. It turns out that Formula (1) is a special case of a well-known property of the distribution function in harmonic analysis [see, for example, Stein (1970), as quoted in Hong (2012)]. Because the mean is one of the most important characteristics of a distribution, Formula (1) has received a lot of attention in the literature. It can be proved using Fubini's theorem [i.e., interchanging the order of integration in the basic definition of the expected value in terms of the p.d.f. [see for example, Ross (2010)], or using integration by parts, again using the basic definition of the expected value in terms of the p.d.f. [see for example, Hong (2012)].

Higher order (greater than the first) moments are also of great interest in the same context. These are needed to describe aspects of the distribution other than the location, for example, the variance, the skewness and the kurtosis of a distribution. To this end, it is known that when it exists, the rth moment about the origin, $E(X^r)$, of a *continuous non-negative* random variable X, is given by

$$E(X^r) = \int_0^\infty r x^{r-1}\big(1 - F_X(x)\big)\,dx, \quad r \geq 1. \tag{2}$$

This formula is given in Feller (1966), Hong (2012) and Nadarajah and Mitov (2003), the latter also derived the multivariate analogue. Formula (2) is as important as Formula (1), particularly for r up to 4, which yields the first four moments needed to define the variance, the skewness and the kurtosis. Hong (2012) sketched a proof of Formula (2), extending an argument in harmonic analysis that he “translated” in order to prove Formula (1). We feel this is beyond the background expected in a first course, both for undergraduate and graduate students on probability, mathematical statistics or statistical theory. The first and the second authors have taught these types of courses for many years for students in business and engineering and experience suggests that these students do not have the background or the training in advanced topics such harmonic analysis to fully grasp such proofs. Motivated by all this, we provide a simple derivation for this result using Fubini’s theorem, which has been used to prove Formula (1). Thus, our derivation can be covered in the classroom, in the same context, in a first-year course as indicated earlier, while discussing a derivation of Formula (1) for the first moment.

Moreover, much to our surprise, we did not find an analogue of Formula (2) for a *discrete non-negative integer-valued* random variable. This should be of considerable importance in practice as many applications involve such variables, such as the binomial and the Poisson. To this end, suppose that Y is a *discrete non-negative integer-valued* random variable with c.d.f. F_Y. We show that the rth moment of Y is given by

$$E(Y^r) = \sum_{i=0}^{\infty}((i+1)^r - i^r)\big(1 - F_Y(i)\big), \qquad r \geq 1. \tag{3}$$

For $r = 1$, from (3),

$$E(Y) = \sum_{i=0}^{\infty}\big(1 - F_Y(i)\big) = \sum_{i=0}^{\infty} P(Y > i). \tag{4}$$

This last Formula (4), for the first moment, is available, for example, in Karlin and Taylor (1975; page 33) however, the general Formula (3) is not.

It is interesting to note the similarities between the Formulas (2) and (3), for the continuous and the discrete random variables. Both involve the function

$(1 - F)$ which is the survival function of the underlying random variable. In the continuous case the term rx^{r-1}, is the first derivative of x^r, which can be written as $\lim_{h\to 0}\left(\frac{(x+h)^r - x^r}{h}\right)$. This looks very similar to the term $((i+1)^r - i^r)$ in the discrete case, in Formula (3), which may be viewed as $\lim_{h\to 1}\left(\frac{(i+h)^r - i^r}{h}\right)$. This makes sense, since h represents an increment in the value of the random variable, which in the case of a discrete integer-valued variable is equal to 1.

The rest of the note is structured as follows. In Section 2, we prove Formula (2) and in Section 3, we prove Formula (3). Finally, in Section 4, we illustrate the use of these formulas with one example for each of them. We hope instructors and students will benefit from these general formulas and their simple derivations.

2. PROOF OF FORMULA (2)

Assuming that it exists, the rth moment of a *continuous* non-negative random variable X, with p.d.f. $f_{X,}$, is given by:

$$E(X^r) = \int_0^\infty t^r f_X(t)dt, \quad r \geq 1. \tag{5}$$

Since $t^r = r\int_0^t x^{r-1}dx$, the integral in (5) can be written as

$$E(X^r) = r\int_0^\infty \left(\int_0^t x^{r-1}dx\right) f_X(t)dt. \tag{6}$$

Again, assuming that $E(X^r)$ exists (i.e., $E(X^r) < \infty$), one can interchange the order of integrations in (6) using Fubini's theorem. This yields

$$E(X^r) = r\int_0^\infty \int_x^\infty x^{r-1} f_X(t)\, dt\, dx \ = r\int_0^\infty x^{r-1}\left(\int_x^\infty f_X(t)dx\right)dt$$

$$= r\int_0^\infty x^{r-1}(1 - F_X(t))\, dx. \tag{7}$$

The interchange of the order of integration in the second equality in (7) can be explained as follows. The domain of the double integral in Equation (6) in the x-t plane is the region A, highlighted in grey, as shown in Figure 1, where t varies from 0 to ∞ and x varies from 0 to t. It is easy to see that this is the same region where t varies from x to ∞ and x varies from 0 to ∞. Thus, we obtain the right side of the second equality in Equation (7). The third equality in (7) is merely a reorganization of the terms. The proof is now completed.

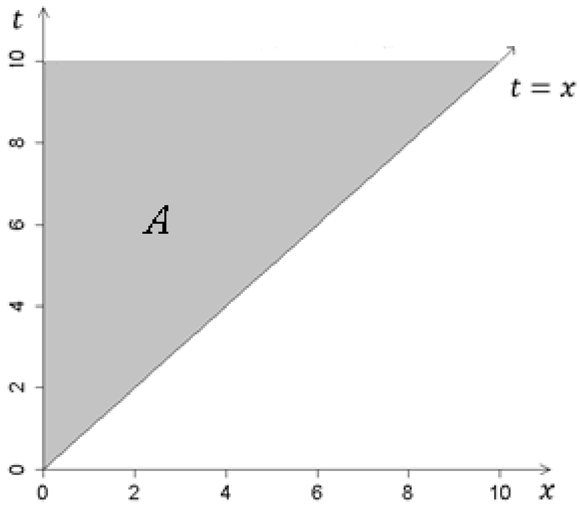


Figure 18. The domain of the double integrals showed in Equation (7).

For $r = 1$, from (7) we get the familiar Formula (1), which is available in the literature, including in Hong (2012) and Karlin and Taylor (1975; page 38).

3. PROOF OF FORMULA (3)

Assuming that it exists, the r-th moment of a discrete non-negative integer-valued random variable Y, with p.m.f. p_Y, is given by

$$E(Y^r) = \sum_{i=1}^{\infty} i^r p_Y(i), \qquad r \geq 1. \tag{8}$$

Using the identity $i^r = \sum_{j=1}^{i}(j^r - (j-1)^r)$, (8) reduces to

$$E(Y^r) = \sum_{i=1}^{\infty} i^r p_Y(i) = \sum_{i=1}^{\infty}\sum_{j=1}^{i}(j^r - (j-1)^r)\, p_Y(i). \tag{9}$$

Similarly for the integrals, Fubini's theorem gives sufficient conditions for the interchange of the summation in Equation (9) [see Hunter (1983, page 27) and references therein], which yields

$$E(Y^r) = \sum_{i=1}^{\infty}\sum_{j=1}^{i}(j^r - (j-1)^r)\, p_Y(i)$$

$$= \sum_{j=1}^{\infty}\sum_{i=j}^{\infty}(j^r - (j-1)^r) p_Y(i)$$

$$= \sum_{j=1}^{\infty}(j^r - (j-1)^r)\sum_{i=j}^{\infty} p_Y(i) = \sum_{j=1}^{\infty}(j^r - (j-1)^r) P(Y \geq j)$$

$$= \sum_{j=1}^{\infty} (j^r - (j-1)^r)\big(1 - F_Y(j-1)\big) = \sum_{i=0}^{\infty} (i+1)^r - \mathrm{i}^r)\big(1 - F_Y(i)\big). \quad (10)$$

The change in the order of the summations in the second equality of Equation (10) can be explained as follows. Write all the terms (elements) in the double summation in (9) or in the first equality in Equation (10) in a matrix form as shown in Figure 2. If we sum down the columns of this matrix, we get (as shown) each of the terms in the sum in (8) or the first sum in (9), which defines the required expected value. Hence the sum of these sums is the answer. On the other hand, if we sum across the rows of this matrix, we get (as shown) each of the terms in the last summation of (10) and the sum of these sums gives the final answer on the right side of the last equality in (10). The point is that we are summing the same elements, whether across the rows or down the columns, and hence their sum must be the same. The proof is now completed.

$$\begin{array}{c}
\left[\begin{array}{cccc}
(1^r - 0^r)p_Y(1) & (1^r - 0^r)p_Y(2) & (1^r - 0^r)p_Y(3) & \cdots \\
0 & (2^r - 1^r)p_Y(2) & (2^r - 1^r)p_Y(3) & \cdots \\
0 & 0 & (3^r - 2^r)p_Y(3) & \cdots \\
\vdots & \vdots & \vdots & \ddots
\end{array}\right]
\begin{array}{l}
\rightarrow (1^r - 0^r)\big(1 - F_Y(0)\big) \\
\rightarrow (2^r - 1^r)\big(1 - F_Y(1)\big) \\
\rightarrow (3^r - 2^r)\big(1 - F_Y(2)\big) \\

\end{array} \\
\begin{array}{cccc}
\downarrow & \downarrow & \downarrow & \\
1^r p_Y(1) & 2^r p_Y(2) & 3^r p_Y(3) &
\end{array}
\end{array}$$

Figure 19. Matrix with elements being summed in Equations (9) and (10).

4. EXAMPLES

In this section, we illustrate the use of Formulas (2) and (3) with two simple examples.

4.1. Example 1

For the first example, suppose that the c.d.f. of X is given by

$$F_X(x) = \begin{cases} 0, & \text{for } x < 1; \\ ln\, x, & \text{for } 1 \le x < e; \\ 1, & \text{for } x \ge e. \end{cases}$$

In order to find the variance of X, $V(X)$, we need to find the first and second moments of X, i.e., $E(X)$ and $E(X^2)$. Note that since X is non-negative, according to Equation (1), we have

$$E(X) = \int_0^\infty \left(1 - F_X(x)\right) dx = \int_0^\infty \left(1 - F_X(x)\right) dx$$

$$= \int_0^1 dx + \int_1^e (1 - ln\, x) dx + 0$$

$$= [x]_0^1 + \ [x]_1^e - [xln\, x - x]_1^e = 1 + \ e - 1 - 1 = e - 1.$$

According to Formula (2), the second moment of X is

$$E(X^2) = 2\int_0^\infty x\left(1 - F_X(x)\right)dx = 2\int_0^1 xdx + 2\int_1^e x(1 - ln\, x)dx + 0$$

$$= 2\int_0^1 xdx + 2\int_1^e xdx - 2\int_1^e xln\, xdx$$

$$= 2\left[\frac{x^2}{2}\right]_0^1 + 2\left[\frac{x^2}{2}\right]_1^e - 2\left[\frac{x^2 ln\, x}{2} - \frac{x^2}{4}\right]_1^e = \frac{1}{2}(e^2 - 1).$$

So that finally, $V(X)$ is equal to

$$V(X) = E(X^2) - \left(E(X)\right)^2 = \frac{1}{2}(e^2 - 1) - (e - 1)^2 = -\frac{1}{2}e^2 + 2e - \frac{3}{2}$$

$$= 0.2420.$$

Note that in order to find the moments using our formulas, it was not necessary to find the p.d.f. of X.

For the students in the first course and for the readers in general, we would like to draw attention to the fact that even though the domain of the distribution of X in this example is between 1 and e, in order to correctly apply Formula (2) (or Formula (1)), the lower limit of the integral must be 0 and not 1. This is because the survival function is not equal to 0 between 0 and 1 (in fact it is equal to 1 and thus contributes to the result). The same comment applies for discrete non-negative integer values random variables, in which case the sum in Formula (3) must start at 0 as shown. Students should take note of this and thus avoid making a mistake in applying these useful formulas.

4.2. Example 2

As our second example, consider the following c.d.f. of a discrete random variable Y:

$$F_Y(i) = \begin{cases} 0, & \text{for } i < 0; \\ 1/2\,, & \text{for } 0 \le i < 1; \\ 3/4\,, & \text{for } 1 \le i < 2; \\ 1, & \text{for } i \ge 2. \end{cases}$$

In order to find $V(Y)$, one just needs to use Formula (3), since Y is non-negative. So, the first moment of Y, $E(Y)$, is

$$E(Y) = \sum_{i=0}^{\infty} \big(1 - F_Y(i)\big) = \big(1 - F_Y(0)\big) + \big(1 - F_Y(1)\big) + \big(1 - F_Y(2)\big) + \cdots.$$

Since $1 - F_Y(0) = 1 - 1/2 = 1/2$, $1 - F_Y(1) = 1 - 3/4 = 1/4$, $1 - F_Y(2) = 1 - 1 = 0$ and $F_Y(i) = 1$ for $i \geq 2$, the rest of the terms in the summation are equal to 0. Thus

$$E(Y) = \sum_{i=0}^{1} \big(1 - F_Y(i)\big) = 1/2 + 1/4 = 3/4.$$

According to Formula (3), the second moment of Y is

$$E(Y^2) = \sum_{i=0}^{\infty} ((i+1)^2 - i^2)\big(1 - F_Y(i)\big) = \sum_{i=0}^{1} (2i+1)\big(1 - F_Y(i)\big)$$
$$= \big(1 - F_Y(0)\big) + 3\big(1 - F_Y(1)\big) = 1/2 + 3/4 = 5/4.$$

Finally, $V(Y)$ is

$$V(Y) = E(Y^2) - \big(E(Y)\big)^2 = 5/4 + (3/4)^2 = 29/16.$$

Again, for these calculations, finding the p.m.f. of Y was not necessary, thanks to Formula (3).

5. SUMMARY AND CONCLUSIONS

In this note we provide proofs for some alternative formulas for the *r*th moment of a non-negative continuous and a discrete random variable, respectively, in terms of its c.d.f. The proofs use Fubini's theorem and should be useful in the classroom while discussing moments of random variables.

Acknowledgements

We are grateful to two anonymous reviewers, an associate editor and the editor in chief for their insightful comments which improved the presentation. This work was supported in part by the CNPq (Brazilian Council for Scientific and Technological Development) through projects 401523/2014-4 (for the first author), 201172/2016-0 (for the second author), 308677/2015-3 (for the third author) and also by CAPES (Brazilian Coordination for the Improvement of Higher Education Personnel) with a scholarship for the second author.

6. REFERENCES

Feller, W. (1966). “An Introduction to Probability Theory and Its Applications.”, Vol. 2. 2nd ed. New York: Wiley.

Hong, L. (2012), “A Remark on the Alternative Expectation Formula”, *The American Statistician*, 66, 232–233.

Hong, L (2015), “Another Remark on the Alternative Expectation Formula”, *The American Statistician*, 69:3, 157-159.

Hunter, J. (1983), “Mathematical Techniques of Applied Probability - Discrete Time Models: Basic Theory”, Vol. 1, Academic Press: New York.

Karlin, S. and Taylor, H. M. (1975). “A First Course in Stochastic Processes”, Academic Press: New York.

Keyfitz, N. (1968). “Introduction to the Mathematics of Population”. Reading, MA: Addison-Wesley.

Nadarajah, S. and Mitov, K.V. (2003). “Product Moments of Multivariate Random Vectors”. Communications in Statistics: Theory and Methods, 32(1), 47-60

Ross, S. (2010), “A First Course in Probability”, 8th Edition, Upper Saddle River, NJ: Prentice Hall.

Stein, E. M. (1970). “Singular Integrals and Differentiability Properties of Functions”, Princeton, NJ: Princeton University Press.