3
Modelos de Gestão Estoque

O presente capítulo se inicia com uma reflexão sobre a importância do desenvolvimento de modelos na ciência moderna e para a Pesquisa Operacional. Apresenta-se em seguida uma revisão bibliográfica sobre os últimos desenvolvimentos relacionados à gestão de estoque de sobressalentes de manutenção. Por fim, expõem-se o modelo que será utilizado para análise de dados de estoque da RLAM.

3.1
Revisão bibliográfica

A filosofia da ciência tem discutido amplamente nos últimos anos a importância dos modelos em diversos contextos científicos. Essa discussão tem se dado principalmente em relação a aspectos ontológicos (o que são os modelos) e epistemológicos (como aprender com os modelos). Apesar das divergências nessas áreas, não há como negar que os modelos possuem duas funções fundamentais: são uma representação de um determinado objeto/fenômeno do mundo real, ou a representação de uma teoria. (Frigg e Hartman, 2006).

Partindo deste ponto, temos que os modelos são uma simplificação da realidade que tem como objetivo facilitar o entendimento da mesma, ou esclarecer como fatores que a influenciam se relacionam. Desta forma, os modelos podem assumir diversas naturezas: eles podem ser uma representação em tamanho diferente da realidade, uma idealização simplificada de uma realidade complexa, equações matemáticas, programas de computador, entre outras. De acordo com Frigg e Hartmann (2006), os modelos pertencem a uma classe especial de objetos formados por uma mistura de elementos pertencentes a diferentes categorias ontológicas.

O mais importante dos modelos, no entanto, são as suas funções cognitivas. O aprendizado a partir da utilização de modelos se dá em três estágios: a denotação, a demonstração e a interpretação. Na denotação, ao se estabelecer uma relação entre a realidade e o modelo, podem-se obter vários esclarecimentos sob a própria realidade. O mesmo ocorre na fase de demonstração, quando o modelo é manipulado. Aqui fica claro como os fatores que compõem o modelo se relacionam. Por fim, ao confrontar os resultados obtidos no modelo com a realidade, sua natureza torna-se mais clara ao pesquisador, que terá tido, ao final do percurso de modelagem, um entendimento maior sobre o fenômeno estudado (Frigg e Hartman, 2006).

Os modelos são amplamente utilizados não apenas para entender problemas científicos, mas para operar sobre a própria realidade. Esta sua utilização tem se tornado cada vez mais comum, desde o início do século XX, com o desenvolvimento da Pesquisa Operacional.

Fruto de um ponto de vista pragmático em relação à ciência (Ormerod, 2006), ou de uma realidade histórica específica (Rosenhead, 2006), a Pesquisa Operacional se tornou uma forma amplamente utilizada de se estudar e propor soluções para problemas complexos.

A utilização de modelos e técnicas desenvolvidas pela Pesquisa Operacional se estendeu ao estudo da gestão de estoques. Como visto anteriormente, o primeiro modelo sobre estoque foi publicado em 1913 por Harris, e os estudos sobre o assunto ganharam profundidade principalmente a partir da década de 1950. Uma série de modelos foi desenvolvida, cobrindo os diferentes tipos de estoque.

Atualmente, os estudos sobre a gestão de estoque de sobressalentes de manutenção desenvolve esforços em melhorar a compreensão e permitir tratamentos mais precisos em termos de controles gerenciais e otimização de custos, desenvolvendo tanto modelos para um item como para vários itens, assim como para problemas de um e vários níveis.

Aronis *et al.* (2004) apresentam uma abordagem para a otimização de estoques de sobressalentes de equipamentos eletrônicos utilizando uma abordagem Bayesiana para prever a taxa de demanda dos mesmos. O modelo baseia-se na política de estoque-base (s, s-1), onde os itens são encomendados em lotes unitários, e na qual a taxa de demanda segue uma distribuição conjunta das distribuições de Poisson e Gama. De acordo com os autores, a utilização da abordagem Bayesiana apresenta maior acurácia para previsão da demanda, resultando em níveis de estoques mais ajustados aos níveis de serviço requerido.

Chang *et al.* (2005) propõem um modelo baseado no sistema (r, q) no qual parte do estoque de sobressalentes deve ser reservada para unidades e equipamentos críticos. Os autores consideram dois tipos de demanda e apenas duas variáveis: a quantidade de ressuprimento q, e o ponto de ressuprimento r, que é definido em função do nível de criticidade dos equipamentos. No modelo apresentado, os sobressalentes são demandados ao estoque um por vez. Quando o estoque chega ao nível r, é emitida uma ordem de compra com q quantidade. A partir desse instante, todo o estoque em mãos é reservado para os equipamentos críticos e as outras demandas não críticas são atendidas posteriormente, quando do recebimento de um novo lote. Ainda são características do modelo o tempo de ressuprimento constante e a demanda normalmente distribuída.

Dadas essas características, Chang *et al*. mostram que a função de custo é convexa e que o resultado obtido pelo lote econômico não é o resultado ótimo.

O mesmo tipo de problema é tratado por Dekker *et al.* (1998). As demandas dos sobressalentes são divididas em críticas e não-críticas. Parte do inventário é reservado para atender as unidades e equipamentos críticos. A demanda tem uma distribuição de Poisson e o tempo de ressuprimento é constante. Contudo, no modelo apresentado os autores não consideraram os custos de fatores.

A influência de emissões de ordens de emergência foi tratada por Johansen e Thorstensen (1998). No modelo proposto pelos autores, ordens normais com tempo de ressuprimento constante são emitidas através da política de controle (r, q). Contudo, durante esse intervalo de tempo podem ser emitidas ordens de emergência que têm prazo de ressuprimento mais curto e custo de emissão diferente. Uma outra característica do modelo é que as demandas não atendidas imediatamente são satisfeitas quando há disponibilidade de sobressalentes em estoque.

Johansen e Thorstensen (1998) formulam os custos médios de longo-prazo que incluem os custos atrelados aos dois tipos de ordem, os custos de falta e custos de manutenção de estoque. Os autores apresentam ainda um algoritmo para achar o custo mínimo, e mostram que a utilização de ordens de emergência minimiza os custos totais principalmente quando os custos de falta são altos.

Dentre os modelos de gestão de estoque de sobressalentes em cadeias de suprimento de mais de um nível, podemos citar o trabalho de Al-Rifai e Rossetti

(2007). Nele, os autores apresentam um modelo com base no sistema de controle (*r, q*), que opera com dois níveis: o armazém central e seus revendedores. A função objetivo do modelo é minimizar o custo total de estoque, dadas restrições de freqüência de emissão das ordens de compra e o número de demandas não atendidas. O modelo é resolvido decompondo o problema por nível da cadeia através de um algoritmo desenvolvido pelos autores.

Kalschschmidt *et al.* (2003) também apresentam um modelo para gestão de inventários de sobressalentes em uma cadeia de suprimento de vários níveis, no qual os consumidores de diferentes tamanhos se situam no mesmo nível da cadeia. Neste modelo, as diferenças de porte entre os clientes geram variações da demanda ao longo do tempo. Os autores separam a demanda em duas categorias: estável e irregular. A partir da análise das referidas categorias, o modelo busca identificar o momento de emissão das ordens de compra e as quantidades ótimas a serem solicitadas que minimizem o custo de faltas e o custo da emissão das ordens, de acordo com a probabilidade de ocorrência de um pico de consumo.

Cohn e Barnhart (2006) apresentam uma abordagem alternativa para gestão de estoque de sobressalentes a partir do uso de variáveis compostas. Ou seja, trata-se de modelos de programação linear que contém variáveis que englobam múltiplas decisões. Para as autoras, uma modelagem que se utiliza das técnicas de programação linear facilita o cálculo de soluções ótimas para o seu modelo.

Uma outra abordagem que tem sido desenvolvida, é a otimização de estoque para vários itens. Wong *et al.* (2006), tratam o problema través de um modelo de estoque para vários itens, no qual há dois centros de armazenagem. Quando alguma demanda não pode ser atendida, o administrador pode solicitar que o outro centro envie o sobressalente, ou pode emitir ordens de compra de emergência. O objetivo do modelo é minimizar o custo total de manter o de estoque, os custo de transporte entre os centros de armazenagem, e os custos de emitir ordens emergenciais, dada uma restrição de tempo para o atendimento de todas as demandas. O resultado é obtido através de uma solução heurística.

Ainda em relação ao tratamento multi-item para controle de estoque, Panda *et al.* (2005), propõem a elaboração de modelos não-lineares de programação por metas (*goal programming*). O modelo busca identificar as quantidades ótimas que devem ser solicitadas para o estoque, de forma a minimizar os custos fixos e variáveis de emissão das ordens de compras. Para os autores, a utilização desta metodologia, torna os modelos mais flexíveis e realistas, apesar do modelo tornar-se maior, mais complexo e conter mais variáveis.

3.2
Sistema (*r, q*) com demanda estocástica

Os modelos citados anteriormente são exemplos do que tem sido desenvolvido nos últimos anos para a gestão de estoques de sobressalentes de manutenção. Em seguida, apresenta-se o modelo utilizado para tratar o problema de otimização de estoques de sobressalentes que se caracteriza por ter demandas estocásticas e tempo de ressuprimento constante, conforme estabelecido inicialmente nos objetivos do presente trabalho. Este modelo, baseado na obra de Zipkin (2000), apesar de ter aspectos restritivos na sua aplicação, pode ser aplicado em uma ampla variedade de sistemas, como os que possuem demanda correlacionadas ao longo do tempo, ou os sistemas no qual o intervalo de tempo entre as demandas seguem uma distribuição de Poisson, e a quantidade demandada é aleatória, por exemplo.

O desenvolvimento e as demonstrações apresentados nesta seção podem ser encontradas em Zipkin (2000) e Hadley e Whitin (1963). O modelo apresentado considera uma sistema de controle (r,q) no qual a demanda é incerta. Assim como o tradicional modelo (r,q), este também permite que demandas de itens não atendidos imediatamente pelo estoque aguardem a chegada de um novo lote de compra para serem satisfeitas.

No presente modelo, o estoque disponível é consumido pelos usuários até que seja completamente exaurido. A partir de então, as demandas não atendidas são acumuladas. Podemos entender, destarte, que as faltas funcionam como um estoque negativo. Evidentemente, o não atendimento das solicitações dos clientes se converte em custos para a empresa, representados por perdas de produção ou não-atendimento a clientes. A depender da natureza dos itens e das necessidades dos clientes, o custo da falta pode ser mais ou menos punitivo.

Um outro pressuposto do modelo é que os itens possuem tempo de fornecimento constante. Ou seja, o tempo entre a emissão de uma ordem de compra até a entrada do material em estoque não varia. Além disto, os materiais são entregues com qualidade perfeita.

O último e mais importante pressuposto diz respeito à natureza da demanda, que segue um processo de Poisson. Um processo de Poisson é um conjunto $\{N(t)\colon t \geq 0\}$ de variáveis aleatórias, onde $N(t)$ é o número de eventos que ocorreram até o instante t. O número de eventos entre os instantes t_1 e t_2 é dado por $N(t_2)$ - $N(t_1)$ e segue a distribuição de Poisson.

O processo de Poisson possui duas propriedades fundamentais: a independência estatística e a regularidade dos eventos. A independência estatística assegura que o número de eventos que ocorrem após o instante t não depende dos eventos que ocorreram antes de t, assim como intervalo de tempo entre a ocorrência de dois eventos quaisquer não depende dos intervalos anteriores. Por sua vez, a propriedade da regularidade assevera que os eventos não ocorrem simultaneamente. Contudo, apesar da demanda em determinado momento ser estocástica, a sua média é estacionária ao longo do tempo.

A combinação de uma demanda estocástica com a existência do tempo de ressuprimento faz com que surjam algumas dificuldades para o planejamento de estoque. Neste sentido, há sempre um atraso entre as ações tomadas e os seus efeitos. Além disso, existência de incertezas quanto à demanda reduz a precisão do controle sobre o estoque. E quanto menos preciso for este controle, maior o inventário necessário para atender um determinado nível de serviço.

O objetivo central do modelo é, então, gerenciar os riscos de falta e as variações de demanda no longo prazo, buscando atingir um nível mínimo de custo de estoque. Isto pode ser feito de duas formas: reduzindo as oscilações das solicitações dos clientes e entendendo o processo de gestão de estoques através da utilização de modelos.

Define-se os seguintes parâmetros e variáveis do modelo:

λ = taxa de demanda

L= tempo de ressuprimento

q = quantidade solicitada em cada ordem de compra

t = tempo (variável contínua), $t \geq 0$

r = ponto de ressuprimento

Para todo $t \geq 0$, tem-se:

$D(t)$ = demanda acumulada durante o período t

$D(t,u]$ = demanda no intervalo $(t,u]$

$= D(u) - D(t),\ u \geq t$

O processo de demanda $D(t)$ segue uma distribuição de Poisson com média λt, ou seja:

$$\Pr\{D(t) = d\} = \frac{(\lambda t)^d \cdot e^{-\lambda t}}{d!}\ ,\quad d \geq 0 \tag{3.1}$$

Para uma dada política de estoque, as seguintes variáveis são usadas para descrever a evolução do modelo ao longo do tempo:

$I(t)$ = estoque disponível

$B(t)$ = demanda não atendida (faltas existentes)

$IN(t)$ = estoque líquido = $I(t)$-$B(t)$

$IO(t)$ = encomendas a serem entregues

$IP(t)$ = posição do estoque = $IN(t) + IO(t)$

$\overline{OF}$ = freqüência média de emissão das ordens de compra

3.2.1
Política de estoque base

Suponha agora o caso em que q = 1. Tal política é recomendada quando as economias de escala no sistema de suprimento são desprezíveis em relação a outros fatores. Por exemplo, quando o custo unitário do item e seu custo de falta dominam claramente o custo fixo de encomendar. Desde que q =1, a política passa a ser de apenas um parâmetro, r. É conveniente definir o nível de estoque base $s = r + 1$. As medidas de desempenho para o sistema são: (i) fração da demanda não atendida prontamente, $\overline{A}(s)$; (ii) o valor esperado de faltas em estoque em um instante qualquer, $\overline{B}(s)$; e (iii) o estoque médio, $\overline{I}(s)$. Definindo

$g(s) = \Pr(D = s)$ como sendo a função de probabilidade, Zipkin (2000) demonstra que:

$$\overline{A}(s) = \Pr\{IN \leq 0\} = \Pr\{D \geq s\} = \sum_{y=s}^{\infty} g(y) = G^0(s) \tag{3.2}$$

$$\overline{B}(s) = E\left[(IN)^-\right] = E\left[(D-s)^+\right] = \sum_{y=s}^{\infty}(y-s)g(y) = G^1(s) \tag{3.3}$$

$$\overline{I}(s) = E\left[(IN)^+\right] = E\left[(IN)+(IN)^-\right] = s - \lambda L - \overline{B}(s) \tag{3.4}$$

$G^0(d)$ é denominada função de probabilidade complementar cumulativa e $G^1(r)$ é denominada função de perda (*loss function*) de primeira ordem. Incidentalmente, para a distribuição de Poisson, tem-se que:

$$G^1(d) = -(d - \lambda L)G^0(d) + \lambda L g(d) \tag{3.5}$$

3.2.2
Política para qualquer tamanho de lote

Nesse caso as medidas de desempenho para o sistema são: (i) fração da demanda não atendida prontamente, $\overline{A}(r,q)$; (ii) o valor esperado de faltas em estoque em um instante qualquer, $\overline{B}(r,q)$; e (iii) o estoque médio, $\overline{I}(r,q)$. Definindo $g(r) = \Pr(D = r)$ como sendo a função de probabilidade, Zipkin (2000) demonstra que:

$$\overline{A}(r,q) = \Pr\{IN \leq 0\} = \sum_{s=r+1}^{r+q}\left(\tfrac{1}{q}\right)\Pr\{IN \leq 0 \mid IP = s\}$$

$$= \left(\tfrac{1}{q}\right)\sum_{s=r+1}^{r+q}\overline{A}(s) \tag{3.6}$$

$$\overline{B}(r,q) = (1/q)\sum_{s=r+1}^{r+q}\overline{B}(s) \tag{3.7}$$

$$\overline{I}(r,q) = (1/q)\sum_{s=r+1}^{r+q}\overline{I}(s) \tag{3.8}$$

Onde $\overline{A}\,(s)$, $\overline{B}\,(s)$ e $\overline{I}\,(s)$ representam as medidas de desempenho para o modelo de estoque base. Desenvolvendo as equações 3.6 a 3.8, para o modelo geral, temos:

$$\overline{A}(r,q)=\left(\frac{1}{q}\right)\left[G^{1}(r)-G^{1}(r+q)\right] \tag{3.9}$$

$$\overline{B}(r,q)=\left(\frac{1}{q}\right)\left[G^{2}(r)-G^{2}(r+q)\right] \tag{3.10}$$

$$\overline{I}(r,q)=\frac{1}{2}(q+1)+r-\lambda L+\overline{B} \tag{3.11}$$

Onde $G^2(r)$ é denominada de função de perda de segunda ordem

$$G^{2}(r)=\frac{1}{2}E\left[(X-r)^{+}(X-r-1)^{+}\right]$$

$$G^{2}(r)=\frac{1}{2}\sum_{y=r}^{\infty}(y-r)(y-r-1)g(y)$$

$$G^{2}(r)=\sum_{y=r}^{\infty}(y-r)G^{0}(y)$$

$$G^{2}(r)=\sum_{y=r+1}^{\infty}G^{1}(y) \tag{3.12}$$

Incidentalmente, para a distribuição de Poisson, tem-se que:

$$G^{2}(d)=\frac{1}{2}\left\{\left[(d-\lambda L)^{2}+d\right]G^{0}(d)-\lambda L(d-\lambda L)g(d)\right\}$$

Finalmente, a freqüência de emissão das ordens é dada pela relação entre a taxa de demanda λ e a quantidade demandada q, ou seja:

$$\overline{OF}(r,q)=\frac{\lambda}{q} \tag{3.13}$$

Estes resultados mostram como se comporta o modelo em relação aos parâmetros λ e L e às variáveis r e q. Primeiramente, fica claro que o desempenho do modelo depende de λ e L apenas através da demanda durante o tempo de ressuprimento, dada pelo produto λL. Ou seja, sistemas com grande demanda e baixo tempo de ressuprimento se comportam de forma semelhante a sistemas com baixa demanda e alto tempo de ressuprimento.

Os efeitos das variações de r sobre o modelo, para um dado valor fixo de q, também podem ser deduzidos pelos resultados acima. Primeiramente, temos que $\overline{B}$ é decrescente e convexo em relação a r. Isso significa que à medida que aumenta-se o ponto de ressuprimento, reduz-se o valor esperado de faltas. Contudo, isto ocorre a taxas decrescentes. Analogamente, $\overline{I}$ é crescente e convexo em relação a r. À medida que aquele cresce, este aumenta à taxas cada vez maiores. Quanto ao comportamento de $\overline{A}$, temos que o mesmo é decrescente em relação a r: aumentar o ponto de ressuprimento reduz a probabilidade de ruptura do estoque. Contudo, só podemos assegurar sua convexidade quando $\lambda L \leq 1$ (Zipkin, 2000).

Na figura 3.1 é apresentado um gráfico que mostra o comportamento das variáveis de desempenho para um sistema hipotético, operando com os parâmetros λ=12, L=1 e q=10.

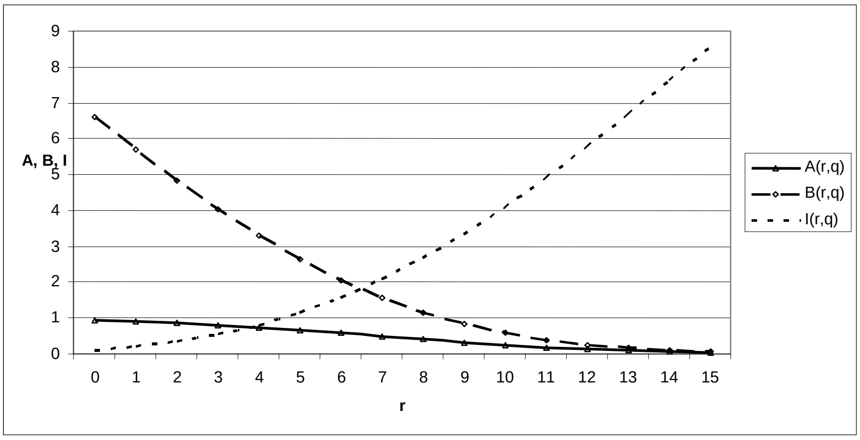

Figura 3.1 – Gráfico de medidas de desempenho em função de *r*

Pelos resultados acima, pode-se inferir o comportamento das variáveis de desempenho $\overline{A}$, $\overline{B}$ e $\overline{I}$ em função de variações de q. A equação (3.13) deixa claro a relação direta entre o aumento de q e a redução de $\overline{OF}$. Da mesma forma,

$\overline{A}$ e $\overline{B}$ são decrescentes em relação à q, visto que quanto maior forem as quantidades solicitadas para compor o estoque, haverá menores rupturas e menos demandas não atendidas. Por sua vez, $\overline{I}$ é crescente em q, pois quanto mais unidades chegarem ao estoque maior será o inventário carregado.

Todavia, as inferências feitas anteriormente não retratam o que é mais importante na relação entre as variáveis de desempenho e q. É preciso levar em consideração o efeito que variações em q causam sobre estas medidas quando o próprio ponto de ressuprimento r varia. A incerteza quanto ao comportamento das variáveis de desempenho quando r e q variam simultaneamente cria dificuldades para o controle do estoque e, consequentemente, uma tendência de elevação do inventário. Contudo, isto implica em custos maiores de estoque.

De posse dessas relações demonstradas acima, pode-se escolher uma política de estoque que atenda às necessidades da empresa, sempre levando em consideração que as variáveis de desempenho são conflitantes entre si. Há a possibilidade de se estabelecer um limite para cada uma das variáveis de desempenho através do controle de r e q, ou mesmo estabelecendo os níveis de serviços.

Contudo, há sempre um nível ótimo para operar o sistema, que é aquele que minimiza os custos totais de estoques. Neste ponto, haverá o equilíbrio entre os custos de encomendar, de manter o estoque e os custos de falta.

3.3
Otimização

A questão que se coloca, então, é como identificar os valores de r e q que minimizem o custo total de estoque.

Nesse sentido, os seguintes parâmetros representam os custos de estoque:

k = custo fixo para emitir uma ordem de compra

c = custo unitário do item

h = custo de manter uma unidade em estoque por uma unidade de tempo

b = custo de falta por unidade de tempo

O custo fixo *k* representa os custos que independem da quantidade solicitada na ordem de compra. Ele inclui os custos administrativos, os custos de transporte e de recebimento das mercadorias. O parâmetro *k* reflete, portanto, as economias de escalas obtidas no processo. O custo variável *c* engloba os custos unitários de compra, assim como todos os outros custos que variam com a quantidade solicitada.

Por sua vez, o parâmetro *h* é formado por dois tipos de custos. O primeiro inclui todos os custos ligados diretamente ao estoque como custo de manuseio, seguros, aluguel e depreciação de armazéns, entre outros. Esses custos podem ser denotados como $\underline{h}$. O segundo representa o custo financeiro de se manter o estoque, ou seja, seu custo de oportunidade. O custo financeiro pode ser representado como αc.

O custo *b* representa os custos da ruptura de estoque. Ou seja, quanto custa não atendimento de uma demanda imediatamente.

O custo total é então dado por:

$$C(r,q) = k\overline{OF}(r,q) + h\bar{I}(r,q) + b\bar{B}(r,q) \tag{3.14}$$

Não existe, contudo, um método simples de otimização para calcular os valores de *r* e *q* que minimizam o custo total de estoque, uma vez que se tratam várias variáveis discretas. Como aproximação, pode-se fazer a otimização considerando *r* e *q* como variáveis contínuas, sendo necessário, portanto, realizar aproximações quando os valores obtidos não fizerem parte do conjunto dos números naturais. Cabe lembrar que o resultado obtido muitas das vezes que for utilizado este recurso não será o resultado ótimo para as variáveis, mas um resultado que se espera estar próximo do ótimo. Entretanto, Federgruen e Zheng (1992) apresentam algoritmo que permite a busca por valores ótimos de *r* e *q* e que é utilizado em um estudo de caso apresentado no Capítulo 4.

3.3.1 Otimização com o modelo de estoque base

Para a otimização com o modelo de estoque base tem-se que função de custo médio anual é dada por:

$$C(s) = h\bar{I}(s) + b\bar{B}(s) \qquad (3.15)$$

Onde:

$s = r + 1$ e as medidas de desempenho $\bar{B}(s)$e $\bar{I}(s)$ são dadas pelas Equações (3.3) e (3.4):

$$\bar{B}(s) = \sum_{y=s}^{\infty}(y-s)g(y) = G^1(s)$$

$$\bar{I}(s) = s - \lambda L - \bar{B}(s)$$

3.3.2 Otimização com variáveis contínuas

Para os casos de variáveis contínuas as equações das medidas de desempenho são as seguintes (Zipkin, 2000, pg 211):

$$\bar{A}(r,q) = \left(\frac{1}{q}\right)\left[F^1(r) - F^1(r+q)\right] \qquad (3.16)$$

$$\bar{B}(r,q) = \left(\frac{1}{q}\right)\left[F^2(r) - F^2(r+q)\right] \qquad (3.17)$$

$$\bar{I}(r,q) = \tfrac{1}{2}(q+1) + r - \lambda L + \bar{B} \qquad (3.18)$$

Onde, $F^1(.)$ e $F^2(.)$ são as funções de perda de primeira e segunda ordem, respectivamente. Tem-se que:

$$F^1(r) = E\left[(X-r)^+\right] = \int_r^\infty (x-r) f_X(x)dx = \int_r^\infty F^0(x)dx \qquad (3.19)$$

$$F^2(r) = 1/2\,E\left[(X-r)^+\right] = 1/2\int_r^\infty (x-r)^2 f_X(x)dx = \int_r^\infty F^1(x)dx \qquad (3.20)$$

$$F^0(r) = \int_r^\infty f_X(x)dx \qquad (3.21)$$

Onde, $F^0(.)$ é a função complementar cumulativa e $f_X(.)$ é a função densidade de probabilidade.

Zipkin (2000) afirma que $\overline{B}(r,q)$, $\overline{I}(r,q)$ e $\overline{OF}(r,q)$ são funções convexas e, conseqüentemente, $C(r,q)$ também é convexa. Adicionalmente $C(r,q)$ é continuamente diferenciável em todo seu domínio. Portanto, qualquer método de otimização não linear, encontrado nos softwares comerciais, pode ser usado.

Uma distribuição bastante utilizada para aproximações contínuas do modelo é a distribuição gama. Quando se considera esta distribuição, temos que:

$$f(x) = \frac{(x)^{\alpha-1}e^{-x/\beta}}{\beta^\alpha\Gamma(\alpha)} \qquad (3.22)$$

$$F^1(x) = (\alpha\beta - x)F^0(x) + \beta x f(x) \qquad (3.23)$$

$$F^2(x) = \frac{1}{2}\left\{\left[(\alpha\beta - x)^2 + \alpha\beta^2\right]F^0(x) + \left(\alpha\beta^2 - \beta x + \beta^2\right)x f(x)\right\} \qquad (3.24)$$

Cabe lembrar que as funções $1 - F^0(.)$ e $f_X(.)$ podem ser obtidas na planilha Excel com a função DISTGAMA(.) e , conseqüentemente, $F^1(.)$ e $F^2(.)$.

Para a distribuição normal com média μ e desvio padrão σ, as variáveis de desempenho são dadas por:

$$\overline{A}(r,q) = \left(\frac{\sigma}{q}\right)\left[\Phi^1(z_r) - \Phi^1(z_{r+q})\right] \qquad (3.25)$$

$$\overline{B}(r,q) = \left(\frac{\sigma^2}{2q}\right)\left[\Phi^2(z_r) - \Phi^2(z_{r+q})\right] \quad (3.26)$$

$$\overline{I}(r,q) = \frac{q}{2} + \sigma z_r + \overline{B} \quad (3.27)$$

Onde,

$$z_r = \frac{r-\mu}{\sigma} \quad \text{e} \quad z_{r+q} = \frac{r+q-\mu}{\sigma}$$

e $\Phi^1(.)$ e $\Phi^2(.)$são as funções de perda de primeira e segunda ordem da distribuição normal padronizada, respectivamente.

Incidentalmente, para a distribuição normal padronizada tem-se que:

$$\Phi^1(z) = -z\Phi^0(z) + \phi(z) \quad (3.28)$$

$$\Phi^2(z) = \frac{1}{2}\left[\left(z^2+1\right)\Phi^0(z) - z\phi(z)\right] \quad (3.29)$$

Onde $\Phi^0(z)$ e $\phi(z)$ são, respectivamente, a função complementar cumulativa e a função densidade de probabilidade da distribuição normal padronizada.

Cabe lembrar que as funções $1-\Phi^0(z)$ e $\phi(z)$ podem ser obtidas na planilha Excel através da função DISTNORMAL(.) e, conseqüentemente, obter $\Phi^1(.)$ e $\Phi^2(.)$

Estas aproximações da normal são extremamente acuradas, exceto para aqueles casos em que λL é muito baixo ou q possui um valor pequeno. Muitos autores não se sentem confortáveis com esta representação da normal por causa de possíveis valores negativos para a demanda. Devido a esse fato, a formulação para o caso da normal é considerada boa quando o coeficiente de variação $cv = \sigma/\mu$ é menor do que 0,5, o que torna altamente improvável uma demanda negativa.

Vale ressaltar também, que a aproximação para a distribuição normal perde um pouco de precisão quando são avaliados valores extremos. Isto ocorre porque a distribuição normal não permite estimativas assintóticas precisas das probabilidades dos seus valores extremos.

Na verdade, deve-se avaliar a propriedade circunstancial da utilização das aproximações. Se houver razões para acreditar que λL não é suficientemente grande, ou que um item possuiu necessidades de nível de serviços tão altas que forcem a análise nos valores extremos da distribuição, deve-se, então, utilizar-se as fórmulas originais.

Contudo, vale a pena levar em consideração os ganhos que se pode obter através da utilização de aproximações com variáveis contínuas, as quais permitem facilidade neste cálculo.

A fim de reduzir a busca por valores ótimos quando se trata o modelo a partir de variáveis contínuas, Zheng (1992) apresenta o teorema abaixo que determina limitantes para as variáveis, reduzindo significativamente os valores procurados, permitindo que com pouco recurso computacional, se obtenha facilmente r e q ótimos.

Teorema:

$$r^* \le s^* \le r^* + q^* \tag{3.30}$$

$$q_{\text{mod}} \le q^* \le \left(q_{\text{mod}}^2 + q_{\max.aprox.}^2\right)^{1/2} \tag{3.31}$$

$$C_{\text{mod}} \le C^* \le \left(C_{\text{mod}}^2 + C_{\max.aprox.}^2\right)^{1/2} \tag{3.32}$$

$$q_{\text{mod}} \le q^* \le q_{est.base} + q_{\text{mod}} \tag{3.33}$$

$$\left(C_{\text{mod}}^2 + C_{est.base}^2\right)^{1/2} \le C^* \le C_{est.base} + C_{\text{mod}} \tag{3.34}$$

Onde,

s^* = estoque base ótimo para o modelo com q = 1

e s^* é tal que $F^0(s^*) = \dfrac{h}{b+h}$

C_{mod}= custo ótimo no modelo clássico do lote econômico (EOQ) com faltas

$C_{mod} = \sqrt{\dfrac{2k\lambda b}{b+h}}$

q_{mod}= quantidade ótima no modelo EOQ com faltas

$$q_{mod} = \sqrt{\frac{2k\lambda(b+h)}{bh}}$$

$C_{est.base}$= custo ótimo para modelo com estoque base

$$C_{est.base} = h\bar{I}(s^*) + b\bar{B}(s^*)$$

$q_{est.\ base}$ = quantidade ótima no modelo de estoque base

$$q_{est.base} = C_{est.base} \cdot \frac{b+h}{bh}$$

$$C_{\max.aprox} = \sqrt{bh}\sigma$$

$$q_{\max.aprox.} = C_{\max.aprox} \cdot \frac{b+h}{bh}$$

Exemplo

Λ	h	k	μ	σ	p
2500	4,00	5,00	500	100	20,00

Resolvendo com auxilio do Solver da planilha Excel obtém-se:

r = 555,44, q = 158,10, com um custo médio total de $C(r,q)$ = 547,05.

Como dito anteriormente, as aproximações permitem fazer inferências importantes sobre o modelo. A primeira inferência que se pode fazer é que as variáveis de desempenho dependem diretamente do desvio-padrão da demanda. Isso significa que reduções do tempo de ressuprimento podem melhorar significativamente os resultados.

Porém, todos os esforços que forem realizados para reduzir o desvio-padrão da demanda terão impacto atenuado, já que os mesmos incidem sobre os resultados através da sua raiz quadrada $\sqrt{\lambda L}$. O mesmo vale para impactos que ocorrem na taxa de demanda. Se, por exemplo, a demanda por unidade de tempo dobrar, e for feito um ajuste em r para manter o mesmo z_r, o valor de estoque irá aumentar apenas $\sqrt{2}$ vezes.

Os resultados mostram com base nos limites estimados dos parâmetros no teorema 1, que o modelo herda as principais características do EOQ e do modelo de estoque base. Isso significa que o modelo (r, q) com demanda estocástica também é robusto em relação a mudanças nos parâmetros.

A quantidade ótima q^* varia no mesmo sentido da mudança da taxa média de demanda. Contudo, a mudança relativa no lote ótimo de compra é menor do que a mudança das taxas de demanda. Na verdade, a quantidade ótima tende a variar na proporção da raiz quadrada da taxa de demanda ($\sqrt{\lambda}$). Esta característica, especialmente, tem uma grande importância prática, uma vez que a taxa de demanda é apenas uma estimação da verdadeira taxa, sujeita, portanto, aos erros de inferência estatística. Todavia, uma vez que o erro não seja grande, o valor estimado de q^* é próximo do verdadeiro valor.

O modelo depende da mesma maneira de k e h. Ou seja, uma vez que seus valores estimados estejam relativamente próximos dos seus valores reais, os valores estimados das variáveis r e q serão muito próximos dos valores ótimos.

Deve-se atentar também para o fato de que o modelo depende diretamente da relação entre os custos de manutenção do estoque e do custo de faltas. Esses termos, juntamente com o custo médio de encomendar, c, representam os fatores econômicos por trás do problema.

Pode-se afirmar também que a dinâmica física do processo de demanda e suprimento afeta o desempenho do sistema indiretamente através do desvio-padrão da demanda durante o tempo de ressuprimento, σ. O aumento de σ leva a uma piora da solução do modelo. Qualquer melhora obtida em σ melhora os resultados e as variáveis de desempenho.

3.4
Modelo especificando o nível de serviço

Muitas vezes, pode ser muito difícil atribuir valores numéricos ao custo de falta de estoque (b). O procedimento descrito nesta seção consiste em minimizar o valor esperado do custo anual de encomendar mais o de manter, sujeito à restrição de que a fração da demanda não atendida prontamente (f) não seja maior do que um valor fixo atribuído. No caso de demanda contínua, foi visto que as medidas de desempenho são:

$$\overline{A}(r,q) = \frac{1}{q}\left[F^1(r) - F^1(r+q)\right]$$

$$\overline{B}(r,q) = \frac{1}{q}\left[F^2(r) - F^2(r+q)\right]$$

$$\overline{I}(r,q) = \tfrac{1}{2}(q+1) + r - \lambda L + \overline{B}$$

$$\overline{OF}(r,q) = \frac{\lambda}{q}$$

Então, o problema consiste em minimizar o valor esperado custo anual de encomendar mais o valor esperado custo anual manter o estoque, sujeito à restrição de que a fração da demanda não atendida prontamente (*f*) não seja maior do que um valor fixo atribuído. Ou seja:

Minimizar

$$k\overline{OF}(r,q) + h\overline{I}(r,q) \tag{3.35}$$

Sujeito a

$$\overline{A}(r,q) \le f \tag{3.36}$$

Observe-se que, para minimizar o custo (3.35), a fração da demanda não atendida prontamente deve ser a maior possível. Portanto, a restrição (3.36) será ativa, isto é, $\overline{A}(r,q) = f$. Considerando este fato e substituindo os valores de $\overline{OF}(r,q)$, $\overline{I}(r,q)$ e $\overline{A}(r,q)$ a formulação do problema fica:

Minimizar

$$\frac{k\lambda}{q} + h\left(\frac{q+1}{2} + r - \lambda L + \frac{1}{q}\left[F^2(r) - F^2(r+q)\right]\right) \tag{3.37}$$

Sujeito a

$$\frac{1}{q}\left[F^1(r) - F^1(r+q)\right] = f \tag{3.38}$$

Em situações reais, usualmente é verdade que os termos $F^1(r+q)$ e $F^2(r+q)$ podem ser considerados desprezíveis. Isto significa dizer que a probabilidade da demanda durante o tempo de reposição ser maior que $r + q$ é muito pequena e $F^1(r+q)$ e $F^2(r+q)$ podem ser desconsiderados. Caso a demanda durante o tempo de reposição seja maior do que $r + q$, pode-se dizer que haverá falta em estoque mesmo após a encomenda chegar. Na prática, raramente deverá ser ótimo um caso em que haja uma probabilidade significativa de que a encomenda não seja suficiente para suprir as faltas, a não ser que o custo de falta seja muito pequeno em relação ao custo de manter estoque, o que geralmente não ocorre, como será o estudo de caso feito no Capítulo 4.

Portanto, desconsiderando os termos $F^1(r+q)$ e $F^2(r+q)$, a formulação do problema fica:

Minimizar

$$\frac{k\lambda}{q} + h\left(\frac{q+1}{2} + r - \lambda L + \frac{1}{q}\left[F^2(r)\right]\right) \tag{3.39}$$

Sujeito a

$$\frac{1}{q}\left[F^1(r)\right] = f \tag{3.40}$$

Da equação (3.40) tem-se que:

$$q = \frac{F^1(r)}{f}$$

Substituindo q em (3.39) a formulação do problema fica:

Minimizar:

$$H(r) = \frac{k\lambda f}{F^1(r)} + h\left(\frac{\left[F^1(r)/f\right]+1}{2} + r - \lambda L + \frac{f}{F^1(r)}\left[F^2(r)\right]\right) \tag{3.41}$$

Agora, tem-se que minimizar a função $H(r)$, que depende apenas de r. No Capítulo 4 será mostrado como resolver este problema com uma aplicação real.