

## Propagação de Incertezas

Grande parte dos modelos hoje investigados estão sujeitos a incertezas, sejam estas nos parâmetros ou nas condições de contorno. Assim, para um modelo ser uma boa representação do fenômeno que ele deseja caracterizar, se torna essencial o tratamento destas incertezas que afetam a resposta final do modelo.

Quantificar incertezas é um processo que pode ser dividido em três etapas: representar as incertezas nos parâmetros de entrada de um sistema; propagação das incertezas através do modelo; e estimar o efeito estocástico na saída. A saída é usualmente um resultado determinístico com um erro médio para indicar os “ranges” de incerteza. Uma função de densidade é desejável para descrever as estatísticas da incerteza.

Em particular, modelos de reservatórios e simulação estão sujeitos a uma incerteza significativa, a qual surge da heterogeneidade da formação geológica e deficiência da informação medida (Li et al., 2011). Pelo que a quantificação de incertezas tem um papel de vital importância na simulação de reservatórios.

Para fazer uma análise de incertezas, precisamos um grande número de simulações. No entanto, isto tem um custo computacional, pois cada simulação do modelo consome muito tempo. Portanto, aproximações eficientes na quantificação de incertezas são necessárias.

Existem várias metodologias utilizadas para quantificar incertezas. Dentre estas técnicas cabe destacar o Método de Monte Carlo (MC). Esta metodologia é considerada a mais usual para quantificação de incertezas. Suas principais vantagens são: simplicidade; e fácil implementação (Fishman, 1996). O MC consiste em extrair um conjunto de amostras independentes e identicamente distribuídas de uma população e simulá-las. Em seguida, construímos um histograma a partir dos resultados, e, assim, calculamos as estatísticas desejadas. No entanto, o MC requer um número grande de simulações para obter resultados das estatísticas de uma forma precisa. O que se mostra como uma desvantagem na maioria de aplicações de simulação de reservatórios.

Por outro lado, para as técnicas que não dependem da amostragem, temos que as incertezas são representadas por aproximações. Uma vantagem destes

métodos em comparação ao MC é que em média eles exigem um número menor de simulações para obter os resultados desejados. Nestas técnicas, o simulador (i.e, um sistema de equações diferenciais parciais que descrevem o fenômeno) é tratado como uma “caixa preta”. Uma das principais ferramentas para o tratamento de incerteza para este caso é a Expansão por Caos Polinomial (ECP).

Dentre esse grupo encontramos o método não-intrusivo de Projeção Espectral. Esta metodologia, será o foco deste trabalho. Esta técnica nada mais é do que o cálculo independente dos coeficientes da ECP, usando a ortonormalidade da base desta expansão. Assim, a eficácia da técnica é diretamente proporcional à qualidade do método utilizado para estimar os coeficientes. Para resolver este problema, existem dois grupos de técnicas: técnicas de quadratura; e técnicas de simulação.

O primeiro grupo de técnicas, técnicas de quadratura, são extremamente eficientes para dimensões baixas. Porém, como o número de simulações necessárias cresce exponencialmente. No segundo grupo, técnicas de simulação, podemos destacar os métodos Quase-Monte Carlo (QMC). Os quais são métodos não aleatórios tipicamente mais eficientes que o Monte Carlo tradicional para estimar integrais. Segundo Blatman (1957) e Sarma & Xie (2011) os métodos de QMC são mais eficientes que Monte Carlo e que Hipercubo Latino na estimação numérica de integrais.

Propagar incertezas significa determinar estatísticas de uma determinada resposta referente a um modelo matemático ou computacional, a partir de distribuições de probabilidades nos parâmetros de entrada. Portanto, neste trabalho abordamos o problema de propagação de incertezas por meio do cálculo das estatísticas da resposta do modelo em questão, também conhecida como análise dos momentos. Para isto, vários métodos têm sido desenvolvidos. No presente capítulo, Capítulo 5, vamos apresentar as metodologias: método MC; método da quadratura; e um método QMC. Logo, no próximo capítulo, Capítulo 6, serão mostrados os resultados aplicados a simuladores de reservatórios de petróleo fictícios.

## 5.1

### Formulação Matemática

Considere que o modelo matemático de um sistema físico é descrito por uma função determinística desconhecida  $\mathcal{M}$ . A função  $\mathcal{M}$  geralmente não tem uma expressão explícita e só pode ser conhecida ponto a ponto, por exemplo, quando o código de um computador é executado. Neste sentido  $\mathcal{M}$  pode ser

entendido como uma função “caixa preta”.

A função do modelo  $\mathcal{M}$  (i.e. o simulador) depende de colocar um único parâmetro de entrada  $x$ , ou um conjunto de parâmetros de entrada denotado por  $\mathbf{x} = \{x_1, x_2, \dots, x_n\}$ . No segundo caso, cada elemento  $x_i$  de  $\mathbf{x}$  representa uma propriedade do modelo, por exemplo, propriedades materiais, geológicas, etc. Neste trabalho, cada  $x_i$  representa uma variável aleatória que afeta o reservatório, por exemplo, a permeabilidade ou porosidade das rochas (isto será mais detalhado no próximo capítulo). Logo, cada modelo de avaliação do nosso simulador  $\mathcal{M}(x)$  retorna uma quantidade de saída  $y$ .

Como estamos assumindo que o parâmetro de entrada  $\mathbf{x}$  é incerto, necessitamos estabelecer um espaço probabilístico para este vetor. Assim, considere  $(\Theta, \mathcal{B}, P)$  um espaço de probabilidade, onde  $\Theta$  é o espaço de todos os possíveis resultados, equipado com uma  $\sigma$ -álgebra de Borel  $\mathcal{B}$  e uma medida de probabilidade  $P$ . Seja  $L^2 \equiv L^2(\Theta, \mathcal{B}, P)$  o espaço das variáveis aleatórias com segundo momento finito. Equivalentemente, se  $X \in L^2$ , onde  $X(\theta) : \Theta \rightarrow \mathcal{C} \subset \mathbb{R}$  é uma variável aleatória (com uma realização denotada com  $x$ ), então temos que

$$E[X^2] = \int_{\Theta} X^2(\theta) dP(\theta) = \int_{\mathcal{C}} x^2 f_X(x) dx < \infty, \quad (5-1)$$

onde  $E[\cdot]$  denota o operador esperança,  $f_X$  representa a função de densidade de probabilidade (pdf) da variável  $X$ , e  $\mathcal{C}$  o suporte de  $X$  (ou seja o domínio da função). Pelo visto na seção 2.4, sabemos que o espaço  $L^2$  é um espaço de Hilbert com respeito ao produto interno dado por

$$\langle X_1; X_2 \rangle = E[X_1 X_2] = \int_{\Theta} X_1(\theta) X_2(\theta) dP(\theta). \quad (5-2)$$

Logo, se nosso parâmetro de entrada do modelo físico  $\mathcal{M}$  é representado como um vetor aleatório  $\mathbf{X}(\theta)$  (onde cada uma de suas entradas são variáveis aleatórias contínuas independentes <sup>1</sup>) temos pela propagação de incerteza, desse vetor aleatório no modelo  $\mathcal{M}$ , que a saída do simulador é também uma variável aleatória  $Y(\theta) = \mathcal{M}(\mathbf{X}(\theta))$ , como pode ser visto na figura 5.1.

A partir disto, nosso objetivo é quantificar a incerteza propagada no modelo  $\mathcal{M}(\mathbf{X}(\theta))$  (ou seja, na resposta do modelo,  $Y$ ). Esta resposta pode ser completamente caracterizada por sua pdf conjunta denotada por  $f_Y(y)$ . No entanto,  $f_Y(y)$  geralmente não tem expressão analítica pois está em função de uma “caixa preta”  $\mathcal{M}$ . Consequentemente a análise que adotaremos nesta dissertação será o análise dos momentos da resposta. Especificamente,

<sup>1</sup>A suposição de independência é matematicamente expressada como:  $f_{\mathbf{X}} = f_{x_1} \cdots f_{x_n}$ , onde a pdf conjunta  $f_{\mathbf{X}}$  é o produto de as densidades marginais.



Figura 5.1: Propagação de Incertezas

estamos interessados no cálculo da média e variância os quais são definidos respectivamente por:

$$E[Y] = \int_{\mathcal{C}} z f_Y(z) dz, \quad (5-3)$$

$$Var[Y] = \int_{\mathcal{C}} (z - E[z])^2 f_Y(z) dz = E[Y^2] - (E[Y])^2. \quad (5-4)$$

Portanto, nas próximas seções serão apresentados os métodos que adotaremos para calcular estes dois momentos. Tais métodos têm uma abordagem simples para propagar incertezas em simulações numéricas. Suas metodologias envolvem simulações repetidas (também chamadas de realizações) com uma seleção própria de valores de entrada, onde todos os resultados são utilizados para gerar a caracterização do resultado.

## 5.2

### Método Monte Carlo

Nesta seção, vamos definir o método de amostragem mais usado, o método de Monte Carlo. Para isso, estabelecemos antes algumas noções. Começamos com um dos principais resultados em teoria de probabilidade: Lei dos Grandes Números.

**Teorema 5.1 (Lei Forte dos Grandes Números)** *Se  $\{X_i\}_{i \geq 1}$  é uma sequência de variáveis aleatórias independentes e identicamente distribuídas (i.i.d.) entre si, e  $E[|X_1|] < \infty$ , então:*

$$\lim_{n \rightarrow \infty} \frac{1}{n} (X_1 + \cdots + X_n) = E[X_1] \quad a.s..$$

**Demonstração:** Ver, e.g., Durrett, 2010, p. 73.

□

O teorema anterior estabelece que conforme uma experiência é repetida várias vezes, a probabilidade observada converge para a probabilidade real. Uma consequência deste resultado é o Teorema do Limite Central.

**Teorema 5.2 (Teorema do Limite Central)** *Seja  $\{X_i\}_{i \geq 1}$  uma sequência de variáveis aleatórias i.i.d.. Se  $E[(X_1)^2] < \infty$ ,  $E[X_1] = \mu$ , então:*

$$\left(\frac{\sqrt{n}}{\sigma} \bar{X} - \mu\right) \text{ converge em distribuição para } \mathcal{N}(0, 1),$$

onde  $\sigma^2$  denota a variância de  $X_1$ .

**Demonstração:** Ver, e.g., Xiu, 2010, p. 23. □

O Teorema 5.2 implica que a média numérica de um conjunto de variáveis aleatórias i.i.d.  $\{X_i\}_{i=1}^n$  converge para uma variável aleatória com distribuição  $\mathcal{N}(\mu, \sigma^2/n)$ , isto quer dizer:

$$\bar{X} \sim \mathcal{N}(\mu, \sigma^2/n), \quad n \rightarrow \infty,$$

onde  $\mu$  e  $\sigma^2$  são respectivamente a média e a variância de variáveis aleatórias i.i.d..

O Método Monte Carlo (MC) é a aproximação amostral mais popular que se apresenta na literatura (ver, e.g., Fishman (1996) e as referências ali citadas). Este método gera um conjunto de realizações aleatórias de um conjunto de variáveis aleatórias  $X_i$  com certa distribuição de probabilidade. Logo, simuladores são utilizados para obter um conjunto de resultados associado com essas realizações. Estes resultados são processados para obter as estatísticas média e variância, obtidos a partir das equações: eq. (5-3) e eq. (5-4), respectivamente. O cálculo numérico das integrais dessas equações são aproximadas por somas sobre o número de amostras. Em outras palavras, como  $Y = \mathcal{M}(X)$  representa a saída do modelo e  $\{x_1, \dots, x_n\}$  representa uma amostra aleatória obtemos que:

$$\hat{\mu}_Y \approx \frac{1}{n} \sum_{i=1}^n \mathcal{M}(x_i) \tag{5-5}$$

$$\hat{\sigma}^2 \approx \frac{1}{n-1} \sum_{i=1}^n (\mathcal{M}(x_i) - \hat{\mu}_Y)^2 \tag{5-6}$$

Pela Lei dos Grandes Números (Teorema 5.1) temos que a média da eq. (5-5) com variáveis aleatórias independentes, com mesma média e variância finita, tende à média da eq. (5-3) isto é

$$\lim_{n \rightarrow \infty} \hat{\mu}_Y = E[Y], \text{ com probabilidade } 1.$$

Além disso, a taxa de convergência pode ser avaliada pelo Teorema do Limite Central:

$$\sqrt{n}(\hat{\mu}_Y - E[Y]) \rightarrow N(0, \sigma^2),$$

em distribuição, onde  $\sigma^2 = \text{Var}[Y]$ . Assim, o “erro” da aproximação por MC é  $\mathcal{O}(n^{-1/2})$  (Liu, 2001).

Apesar do MC ser simples de aplicar e não requer de nenhuma modificação nas ferramentas computacionais, ele tem limitações. Por exemplo, através dos parágrafos anteriores, podemos verificar que para conseguirmos uma boa aproximação, precisamos de um grande número de amostras. O que implica uma convergência lenta. Por outro lado, esta taxa de convergência não depende diretamente do número de variáveis do problema, i.e. não depende da dimensão do vetor no qual estejamos trabalhando. Pelo que o MC sempre nos dá a resposta correta, mas para isso precisamos de um grande número de realizações para estimar com exatidão essas respostas.

Devido às limitações desta técnica precisamos de outra abordagem que nos permita quantificar de uma forma mais eficiente os momentos da saída aleatória; o método da quadratura.

## 5.3

### Método da Quadratura

O segundo método proposto para o cálculo dos momentos de nossa saída aleatória é o método da quadratura. Esta é uma das duas categorias de técnicas usadas no método não intrusivo de Projeção Espectral. Em particular, o método em questão calcula os momentos descritos pelas equações eq. (5-3) e eq. (5-4) por meio de integração numérica dos coeficientes da ECP. Note que, a entrada de nosso modelo pode ser tanto uma única variável como um vetor aleatório. Portanto, esta seção, aborda ambos casos, e para isso a presente será dividida em duas sub-seções: o método da quadratura unidimensional; e o método da quadratura multidimensional. Em cada um destes métodos faremos ligações com a teoria desenvolvida no capítulo anterior.

#### 5.3.1

##### Método da Quadratura Unidimensional

Seja a saída de nosso modelo  $Y = \mathcal{M}(X) \in L^2$ , tal que  $X$  é uma variável aleatória. Pode-se verificar que existe uma base ortonormal formada

por polinômios unidimensionais que gera  $L^2$ , ou seja,  $\{\Phi_0, \Phi_1, \dots, \Phi_N\}$  forma uma base ortonormal para  $L^2$  tal que

$$\langle \Phi_i; \Phi_j \rangle = \int_{\mathcal{C}} \Phi_i(x) \Phi_j(x) f_X(x) dx = \delta_{ij}, \quad (5-7)$$

onde  $Y$  admite uma ECP finita do tipo

$$Y \approx \sum_{k=0}^N \hat{\alpha}_k \Phi_k(X),$$

onde os coeficientes  $\hat{\alpha}_k$  são determinados a partir da equação

$$\hat{\alpha}_k = \int_{\mathcal{C}} \mathcal{M}(x) \Phi_k(x) f_X(x) dx, \quad k \in 0, \dots, N. \quad (5-8)$$

Apesar da Expansão por Caos Polinomial ser uma ferramenta para representar a variável aleatória  $Y$ , nós estamos apenas interessados no cálculo da esperança e variância. Portanto, a representação da distribuição de  $Y$  foge ao escopo deste trabalho. No entanto, o Teorema 5.3 estabelece uma das principais vantagens da ECP mostrando como calcular os momentos estabelecidos nas equações eq. (5-3) e eq. (5-4).

**Teorema 5.3** *Se  $Y = \mathcal{M}(X) \in L^2$ , com  $X$  variável aleatória, então as afirmações abaixo são verdadeiras:*

- (1)  $E[Y] = \hat{\alpha}_0$ ,
- (2)  $Var[Y] = \sum_{k \geq 1} \hat{\alpha}_k^2$ .

onde  $\hat{\alpha}_k$  são dados pela eq. (5-8).

**Demonstração:**

Seja  $\{\Phi_0, \Phi_1, \dots, \Phi_N\}$  uma base polinomial ortonormal para  $L^2$ , com  $\Phi_0 = 1$ . Onde o valor esperado da saída do modelo é dado por

$$E[Y] = E[\mathcal{M}(X)] = \int_{\mathcal{C}} \mathcal{M}(x) f_X(x) dx = \int_{\mathcal{C}} \mathcal{M}(x) \Phi_0(x) f_X(x) dx. \quad (5-9)$$

Como  $Y \in L^2$ , segue que  $Y$  admite uma aproximação do tipo Caos Polinomial.

Dáí, a eq. (5-9) pode ser reescrita como

$$\begin{aligned}\int_{\mathcal{C}} \mathcal{M}(x) \Phi_0(x) f_X(x) dx &= \int_{\mathcal{C}} \sum_{k=0}^N \hat{\alpha}_k \Phi_k(x) \Phi_0(x) f_X(x) dx \\ &= \sum_{k=0}^N \hat{\alpha}_k \int_{\mathcal{C}} \Phi_k(x) \Phi_0(x) f_X(x) dx,\end{aligned}$$

isto pela linearidade do operador esperança. Por outro lado, pela ortonormalidade de  $\Phi_0$  com os outros polinômios da base, obtemos que:

$$E[Y] = \sum_{k=0}^N \hat{\alpha}_k \int_{\mathcal{C}} \Phi_k(x) \Phi_0(x) f_X(x) dx = 1 \cdot \hat{\alpha}_0.$$

Repetindo os mesmos procedimentos anteriores só que para a definição de variância eq. (5-4), temos que

$$\begin{aligned}Var[Y] &= E[(Y - E[Y])^2] \\ &= \int_{\mathcal{C}} \left( \sum_{k=0}^N \hat{\alpha}_k \Phi_k(x) - \hat{\alpha}_0 \right)^2 f_X(x) dx \\ &= \int_{\mathcal{C}} \left( \sum_{k=1}^N \hat{\alpha}_k \Phi_k(x) \right)^2 f_X(x) dx \\ &= \sum_{k \geq 1} \hat{\alpha}_k^2 \int_{\mathcal{C}} \Phi_k(x) f_X(x) dx \\ &= \sum_{k \geq 1} \hat{\alpha}_k^2 \langle \Phi_k; \Phi_k \rangle \\ &= \sum_{k \geq 1} \hat{\alpha}_k^2.\end{aligned}\tag{5-10}$$

□

Em outras palavras, o teorema acima nos diz que a média de  $Y$  é simplesmente o primeiro coeficiente determinístico do ECP (ou seja  $\hat{\alpha}_0$ ), e a variância de  $Y$  é dada por a soma ponderada dos coeficientes do ECP ao quadrado. Por outro lado, mesmo que nosso interesse seja só o valor esperado e a variância, expressões similares para os momentos de ordem superior para  $Y$  podem ser encontradas em termos dos coeficientes do ECP; mas estas não são tão simples (Hagan, 2013).

A partir do Teorema 5.3, temos que para calcular a esperança e variância da saída do modelo basta obtermos os coeficientes do ECP. Assim, veremos que a técnica numérica da quadratura para o caso unidimensional nos ajudará



na aproximação destes coeficientes, que se reduz, a avaliar as integrais (sobre o espaço do parâmetro) realizadas através de métodos de integração numérica.<sup>2</sup>

Considere uma variável aleatória  $X$  com função de densidade de probabilidade (pdf)  $f_X(x)$  e suporte  $\mathcal{C}$ . Tome  $\mathcal{M}$  uma função de uma variável aleatória  $X$  quadraticamente integrável com respeito à PDF  $f_X(x)$ . Logo, a quadratura unidimensional nos permite aproximar os coeficientes que aparecem no Teorema 5.3 por meio da seguinte soma ponderada:

$$\begin{aligned}\hat{\alpha}_k &= \int_{\mathcal{C}} \mathcal{M}(x) \Phi_k(x) f_X(x) dx \\ &\approx \sum_{j=1}^N \omega_j \mathcal{M}(x_j) \Phi_k(x_j), \quad k \in \{0, \dots, N\}.\end{aligned}\quad (5-11)$$

Na expressão anterior, eq. (5-11),  $N$  é chamado o nível do esquema da quadratura, enquanto, o conjunto de pares ordenados  $(x_j, \omega_j)$  são respectivamente os zeros e pesos da quadratura, onde  $j = 1, \dots, N$ . Daí, por exemplo, a esperança de uma variável aleatória  $\mathcal{M}(X)$  seria

$$E[\mathcal{M}(X)] \approx \sum_{j=1}^N \omega_j \mathcal{M}(x_j). \quad (5-12)$$

Em termos práticos, os métodos de quadratura para propagação de incertezas requerem a avaliação dos zeros e dos pesos para uma base para uma família de funções ortogonais. Uma vantagem deste método é que estes pares ordenados podem ser calculados e salvados previamente, que favorece ao tempo computacional. Por outro lado, um conjunto de  $N$  cálculos independentes são realizados para avaliar os zeros em nossa função  $\mathcal{M}$ , e os resultados são combinados para obter as estatísticas de interesse. Portanto, o método da quadratura é uma técnica que mantém a mesma vantagem do método MC sobre a não modificação da ferramenta computacional existente.

### 5.3.2

#### Método da Quadratura Multidimensional

Seja a saída de nosso modelo  $Y = \mathcal{M}(\mathbf{X}) \in L^2$  tal que  $\mathbf{X} = (X_1, \dots, X_n)$  é um vetor aleatório com componentes independentes. A função de densidade de probabilidade conjunta  $f_{\mathbf{X}}(\mathbf{x})$  de  $\mathbf{X}$  é:

$$f_{\mathbf{X}}(\mathbf{x}) = f_{X_1}(x_1) \times f_{X_2}(x_2) \times \dots \times f_{X_n}(x_n) \quad (5-13)$$

<sup>2</sup>Integração numérica e quadratura são usualmente sinônimos para funções de uma dimensão. Em dimensões maiores o termo de cubatura é normalmente usado em lugar de quadratura.

Além disso, seja  $\{\Phi_{\mathbf{i}}(\mathbf{X})\}_{|\mathbf{i}| \leq p}$  a base dos polinômios multidimensionais de dimensão  $n$  de ordem  $p$  para o ECP. Isto é:

$$E[\Phi_{\mathbf{i}}\Phi_{\mathbf{j}}] = \int_{\mathcal{C}} \Phi_{\mathbf{i}}(x)\Phi_{\mathbf{j}}(x)f_{\mathbf{X}}(x)dx = \delta_{\mathbf{ij}}, \quad (5-14)$$

onde  $\delta_{\mathbf{ij}} = \delta_{i_1j_1} \cdots \delta_{i_nj_n}$ . Logo,  $Y$  admite uma ECP finita do tipo

$$Y = \sum_{|\mathbf{i}| \leq p} \hat{\alpha}_{\mathbf{i}} \Phi_{\mathbf{i}}(\mathbf{X}),$$

onde os coeficientes  $\hat{\alpha}_{\mathbf{i}}$  são determinados a partir da seguinte equação:

$$\hat{\alpha}_{\mathbf{i}} = E[\mathcal{M}(\mathbf{X})\Phi_{\mathbf{i}}(\mathbf{X})] = \int_{\mathcal{C}} \mathcal{M}(\mathbf{x})\Phi_{\mathbf{i}}(\mathbf{x})f_{\mathbf{X}}(x)d\mathbf{x}. \quad (5-15)$$

O Teorema 5.3 (que representa o caso de uma parâmetro de entrada como variável aleatória) pode ser generalizado para um vetor aleatório como parâmetro de entrada  $\mathbf{X} = (X_1, \dots, X_n)$ .

**Teorema 5.4** *Se  $Y = \mathcal{M}(\mathbf{X}) \in L^2$  e  $\mathbf{X} = (X_1, \dots, X_n)$  um vetor aleatório, então as afirmações abaixo são verdadeiras:*

- (1)  $E[Y] = \hat{\alpha}_{\mathbf{0}}$ .
- (2)  $Var[Y] = \sum_{0 \leq |\mathbf{i}| \leq p} \hat{\alpha}_{\mathbf{k}}^2$ ,

onde os  $\hat{\alpha}_{\mathbf{i}}$  são dados pela eq. (5-15).

**Demonstração:** Ver, e.g., Xiu, 2010, prova similar a do Teorema 5.3.  $\square$

A fórmula da quadratura multivariada pode ser obtida de cada pdf marginal  $f_{X_i}$ . Considere uma função quadraticamente integrável  $\mathcal{M}(\mathbf{X})$  em termos do vetor aleatório  $\mathbf{X}$ . Assim, os coeficientes para as expressões do Teorema 5.4 podem ser calculados como:

$$\begin{aligned} \hat{\alpha}_{\mathbf{i}} &= \int_{\mathcal{C}} \mathcal{M}(\mathbf{x})f_{\mathbf{X}}(\mathbf{x})\Phi_{\mathbf{i}}(\mathbf{x})d\mathbf{x} \\ &\approx \sum_{j_1=1}^{N_1} \cdots \sum_{j_n=1}^{N_n} \prod_{m=1}^n \omega_{j_m} \mathcal{M}(x_{j_1}, \dots, x_{j_n}) \Phi_{\mathbf{i}}(x_{j_1}, \dots, x_{j_n}), \end{aligned} \quad (5-16)$$

onde  $|\mathbf{i}| \leq p$ . Em particular, a esperança pode ser estimada como:

$$E[\mathcal{M}(\mathbf{X})] \approx \sum_{j_1=1}^{N_1} \cdots \sum_{j_n=1}^{N_n} \omega_{j_1} \cdots \omega_{j_n} \mathcal{M}(x_{j_1}, \dots, x_{j_n}). \quad (5-17)$$

Na eq. (5-16) o número total de pontos da quadratura é  $d = \prod_{m=1}^n N_m$ . Usando uma fórmula de quadratura isotrópica (isto é, uma fórmula na qual  $N_1 = \dots = N_n = N$ ), o número total de pontos é  $d = N^n$ . Este crescimento exponencial com respeito ao número de variáveis aleatórias poderia levar a cálculos intratáveis no caso da demanda computacional da função do modelo. Este problema no crescimento exponencial é chamado de “Maldição da dimensionalidade”.

## 5.4

### Maldição da dimensionalidade

O termo “Maldição da dimensionalidade” foi introduzido por primeira vez por Bellman (1957), quando estava considerando problemas em otimização dinâmica. Se analisarmos individualmente cada um dos componentes, o termo *Dimensionalidade* refere-se ao número de dimensões (ou parâmetros de entrada) em nosso conjunto de informação. Por outro lado, o termo *Maldição* refere-se as dificuldades que surgem na análise quando o número de parâmetros de entrada aumenta. Portanto, o termo completo denota os problemas que surgem na investigação de fenômenos físicos em espaços com dimensão alta.

Segundo Novak & Ritter (1997) um aumento do número de dimensões implica no crescimento do custo computacional no tratamento do problema. Particularmente, como foi exposto na seção anterior, nosso problema surge quando aproximamos as integrais de um alto número de dimensões; onde o crescimento exponencial com respeito ao número de variáveis aleatórias nos leva a cálculos intratáveis no caso da demanda computacional da função do modelo.

Segundo Niederreiter (1978) atualmente o problema da *Maldição da dimensionalidade* é bastante investigado, por exemplo, em campos como: análise numérica; mecânica estatística; machine learning; dentre outros. No entanto, ainda não se tem um método definitivo para a solução do problema em questão. Apesar disso, várias técnicas têm obtidos resultados promissores. Dentre as quais podemos destacar: métodos Quase-Monte Carlo; e as quadraturas esparsas.

### 5.4.1 Método Quase-Monte Carlo

Como foi visto na Seção 5.3, uma das duas categorias de técnicas usadas no método não-intrusivo de Projeção Espectral é o método da quadratura. Além da categoria de técnica de simulação, onde destaca o método Quase-Monte Carlo (QMC).

O método QMC é uma aplicação que usa a ECP e que tem como fim (similar ao método da quadratura gaussiana) calcular os coeficientes de um modelo físico aleatório  $Y = \mathcal{M}$  em um subespaço de dimensão finita de  $L^2$ . Relembrando, estes coeficientes são calculados por meio da seguinte equação:

$$\hat{\alpha}_i \approx \sum_{j=1}^N \omega_j \mathcal{M}(x_j) \Phi_i(x_j). \quad (5-18)$$

O que diferencia o método QMC do método da quadratura é que o primeiro método depende da escolha de  $n$  nodos ( $x_j$ ) aleatórios e dos pesos definidos por  $\omega_1 = \omega_2 = \dots = \omega_n = \frac{1}{N}$ , onde os coeficientes são calculados por meio da seguinte equação

$$\hat{\alpha}_i \approx \frac{1}{N} \sum_{j=1}^N \mathcal{M}(x_j) \Phi_i(x_j). \quad (5-19)$$

A escolha dos nodos,  $x_j$ , é feita através de sequências determinísticas conhecidas como quase-aleatórias ou sequências de baixa discrepância (ver, e.g., Niederreiter, 1992). A aproximação amostral faz com que o número de simulações seja quase independente do número de variáveis aleatórias, e, assim, o resultado é mais eficiente para problemas de grandes escalas.

Alguns métodos QMC populares são as sequências de Halton, Faure, Sobol e Niederreiter-Xing. Uma metodologia que tem obtido bons resultados são as sequências de Sobol. Segundo Morokoff & Caflisch (1995) tal método se mostra eficiente para estimar integrais de dimensões  $n \geq 10$ . Particularmente, esta sequência foi usada no campo de reservatórios de petróleo por Sarma & Xie (2011) onde foram obtidos resultados satisfatórios.

As sequências de Sobol, que chamaremos simplesmente de Sobol, são construídas para se obter uma uniformidade melhor do que as amostras aleatórias. Por exemplo, em um espaço bi-dimensional de  $[0, 1]$ , esta propriedade é ilustrada pelas grades na figura 5.2, onde 256 pontos e 512 pontos de QMC são comparados com 256 pontos e 512 pontos de MC. Onde pode-se ver uma melhor uniformidade da forma na sequência de Sobol.

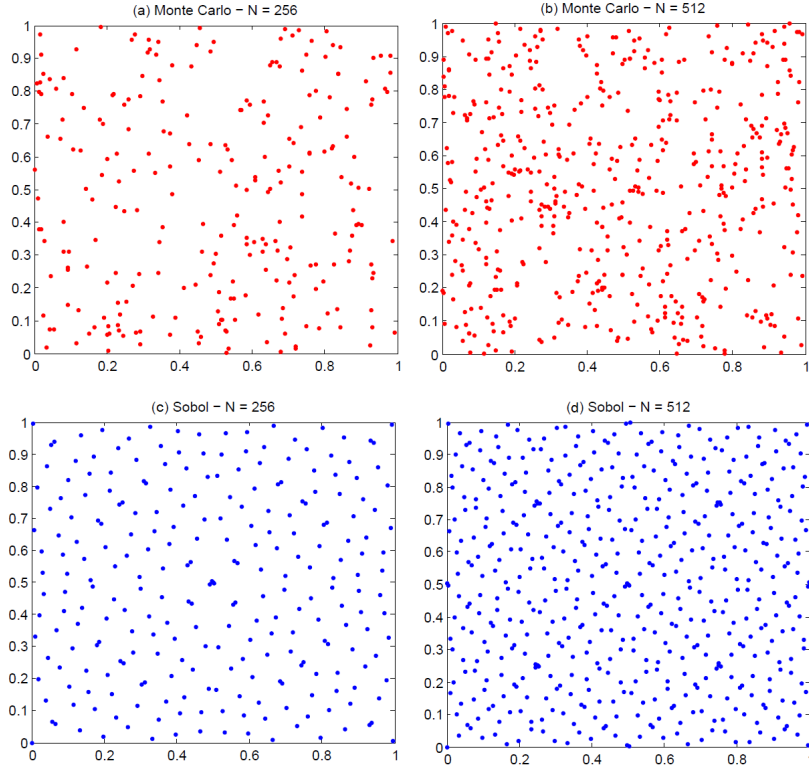


Figura 5.2: Conjunto de amostras aleatórias no quadrado unitário

A sequência de Sobol em uma dimensão é gerada expandindo o conjunto de inteiros  $\{1, 2, \dots, N\}$  em notação base 2. O  $i$ -ésimo termo da sequência é definida por:

$$u_j = \frac{b_0}{2} + \frac{b_1}{2^2} + \dots + \frac{b_m}{2^{m+1}}, \quad (5-20)$$

onde os  $b_k$ 's são inteiros tomados da expansão base 2 de o número  $j - 1$ , isto é:

$$[j - 1]_2 = b_m b_{m-1} \dots b_1 b_0, \quad b_k \in \{0, 1\}. \quad (5-21)$$

Vamos a denotar por  $(U_1, \dots, U_N)$  o conjunto de  $N$  números quase-aleatórios. É necessário transformar estes números em realizações de um vetor aleatório  $X$  de entrada. Para isto, vamos assumir que as variáveis aleatórias  $\{X_1, \dots, X_n\}$  são independentes. Assim, poderíamos obter os valores de  $x_j$  tais que

$$x_j^i = F_X[F_U(u_j^i)], \quad i = 1, \dots, N, \quad j = 1, \dots, n, \quad (5-22)$$

onde  $F_U$  é a função de acumulada de uma distribuição uniforme em  $[0, 1]$ . Logo, os coeficientes estimados do ECP que são obtidos colocando os nodos

de integração quase-aleatórios  $x_j$  na eq. (5-19) são chamados de quase-Monte Carlo. Onde a taxa de convergência é de  $\mathcal{O}(\frac{(\log N)^n}{N})$ .

Portanto, no próximo capítulo usaremos esta técnica para comparar os resultados com o método da quadratura e Monte Carlo, e além disso, faremos alguns testes com funções explícitas que dependem de um número considerável de variáveis.