

UNIVERSIDADE DE SÃO PAULO
ESCOLA DE ENGENHARIA DE SÃO CARLOS
DEPARTAMENTO DE ENGENHARIA DE ESTRUTURAS

META-MODELAGEM EM CONFIABILIDADE ESTRUTURAL

VERSÃO CORRIGIDA

A versão original encontra-se na Escola de Engenharia de São Carlos

HENRIQUE MACHADO KROETZ

Dissertação apresentada ao Departamento de Engenharia de Estruturas da EESC-USP como parte dos requisitos para obtenção do título de Mestre em Ciências, Programa de Engenharia Civil (Estruturas).

Orientador: Prof. Assoc. André Teófilo Beck

São Carlos

2015

AUTORIZO A REPRODUÇÃO TOTAL OU PARCIAL DESTE TRABALHO,
POR QUALQUER MEIO CONVENCIONAL OU ELETRÔNICO, PARA FINS
DE ESTUDO E PESQUISA, DESDE QUE CITADA A FONTE.

K93m Kroetz, Henrique Machado
Meta-Modelagem em Confiabilidade Estrutural /
Henrique Machado Kroetz; orientador André Teófilo Beck.
São Carlos, 2015.

Dissertação (Mestrado) - Programa de Pós-Graduação
em Engenharia de Estruturas -- Escola de Engenharia de
São Carlos da Universidade de São Paulo, 2015.

1. Confiabilidade Estrutural. 2. Meta-Modelos. 3.
Redes Neurais Artificiais. 4. Polinômios de Caos. 5.
Krigagem. I. Título.

FOLHA DE JULGAMENTO

Candidato: Engenheiro **HENRIQUE MACHADO KROETZ**.

Título da dissertação: "Meta-Modelagem em confiabilidade estrutural."

Data da defesa: 23/03/2015

Comissão Julgadora:

Resultado:

Prof. Associado **André Teófilo Beck (Orientador)**
(Escola de Engenharia de São Carlos/EESC)

APROVADO

Prof. Dr. **Edson Denner Leonel**
(Escola de Engenharia de São Carlos/EESC)

APROVADO

Prof. Dr. **Rafael Holdorf Lopes**
(Universidade Federal de Santa Catarina/UFSC)
PARTICIPAÇÃO POR VIDEO-CONFERÊNCIA

APROVADO

Coordenadora do Programa de Pós-Graduação em Engenharia Civil
(Engenharia de Estruturas):
Profa. Associada **Ana Lucia Homce de Cresce El Debs**

Presidente da Comissão de Pós-Graduação:
Prof. Associado **Paulo César Lima Segantine**

AGRADECIMENTOS

Agradeço a todos os seres humanos, que basicamente geraram todo o conhecimento e me deram todas as ferramentas para que eu pudesse fazer este trabalho.

Agradeço em especial aos meus pais pelo incondicional apoio, bem como meu irmão e meus avós. Ao professor e amigo André Beck pela atenciosa e competente orientação, ao Felipão, do qual não me livreii nem indo morar em outra cidade, ao Buzz, que me aguentou (e apoiou) por um ano e meio, e às pessoas que conheci ao longo da jornada, que se tornaram, à sua forma, uma família para mim em São Carlos: Ayrão, Barretos, Giovanni, Serjão, , Lara, Thaís, Karen, Jackson, Caballero, Sonia, todo o pessoal da sala do Rodolfo e tantos outros amigos do SET.

Agradeço àquelas pessoas que se fizeram especiais e me ajudaram de alguma forma: Minha querida Dayane por ser exatamente quem ela é, Ketson e Wellison pelas dicas valiosas no início do trabalho, Carlinhos, Arthur, Matheus e tantos outros veteranos pela reconfortante recepção que fizeram à minha turma, e certamente serei injusto ao parar a lista tão cedo, mas não tenho como mencionar todos mais do que já fiz no primeiro parágrafo.

Por fim, agradeço à CAPES pelo apoio financeiro e a todo contribuinte que ajudou a custear minha educação: farei o possível para retornar à sociedade muito mais do que me foi dado.

Muito obrigado!

*- Dedico este trabalho a todos aqueles que tentam fazer
do mundo um lugar um pouco melhor*

*"Eu preferiria ser um macaco sem cérebro a um monstro sem
coração."*

Son Goku

KROETZ, H. M. **Meta-modelagem em confiabilidade estrutural**. Dissertação (Mestrado – Engenharia de Estruturas),– Escola de Engenharia de São Carlos, Universidade de São Paulo, São Paulo, 2015.

RESUMO

A aplicação de simulações numéricas em problemas de confiabilidade estrutural costuma estar associada a grandes custos computacionais, dada a pequena probabilidade de falha inerente às estruturas. Ainda que diversos casos possam ser endereçados através de técnicas de redução da variância das amostras, a solução de problemas envolvendo grande número de graus de liberdade, respostas dinâmicas, não lineares e problemas de otimização na presença de incertezas são comumente ainda inviáveis de se resolver por esta abordagem. Tais problemas, porém, podem ser resolvidos através de representações analíticas que aproximam a resposta que seria obtida com a utilização de modelos computacionais mais complexos, chamadas chamados meta-modelos. O presente trabalho trata da compilação, assimilação, programação em computador e comparação de técnicas modernas de meta-modelagem no contexto da confiabilidade estrutural, utilizando representações construídas a partir de redes neurais artificiais, expansões em polinômios de caos e através de krigagem. Estas técnicas foram implementadas no programa computacional StRAnD – Structural Reliability Analysis and Design, desenvolvido junto ao Departamento de Engenharia de Estruturas, USP, resultando assim em um benefício permanente para a análise de confiabilidade estrutural junto à Universidade de São Paulo.

Palavras Chave: *Confiabilidade Estrutural. Meta-Modelos. Redes Neurais. Polinômios de Caos. Krigagem.*

KROETZ, H. M. **Meta-modeling Techniques in Structural Reliability**. Thesis (Masters – Structural Engineering), 2015 –São Carlos School of Engineering, São Paulo University, São Carlos, 2015.

ABSTRACT

The application of numerical simulations to structural reliability problems is often associated with high computational costs, given the small probability of failure inherent to the structures. Although many cases can be addressed using variance reduction techniques, solving problems involving large number of degrees of freedom, nonlinear and dynamic responses, and problems of optimization in the presence of uncertainties are sometimes still infeasible to solve by this approach. Such problems, however, can be solved by analytical representations that approximate the response that would be obtained with the use of more complex computational models, called meta-models. This work deals with the collection, assimilation, computer programming and comparison of modern meta-modeling techniques in the context of structural reliability, using representations constructed from artificial neural networks, polynomial chaos expansions and Kriging. These techniques are implemented in the computer program StRAnD - Structural Reliability Analysis and Design, developed at the Department of Structural Engineering, USP; thus resulting in a permanent benefit to structural reliability analysis at the University of São Paulo.

Palavras Chave: *Structural Reliability. Meta-Models. Artificial Neural Networks. Polynomial Chaos. Kriging.*

LISTA DE SIGLAS:

FAPESP	Fundação de Amparo à Pesquisa do Estado de São Paulo
F.D.A	Função de Distribuição Acumulada de Probabilidades
F.D.P	Função Densidade de Probabilidade
FORM	First Order Reliability Method
FOSM	First Order Second Moment
EPC	Expansão em Polinômios de Caos
HLRF	Algoritmo de Hassofer, Lind, Rackwits e Fiessler
MEFE	Método dos Elementos Finitos Estocásticos
MMQ	Método dos Mínimos Quadrados
MSR	Método de Superfície de Resposta
RNA	Rede Neural Artificial
SMC	Simulação de Monte Carlo
SORM	Second Order Reliability Method
StRAnD	Structural Reliability Analysis and Design
VA	Variável Aleatória

LISTA DE FIGURAS

Figura	Legenda	Página
3.1	Equação de estado limite de domínios de falha e segurança. (Beck, 2012)	29
3.2	Representação gráfica da probabilidade de falha. (Sagrilo e Lima, 1996)	30
3.3	Aproximação de primeira ordem, integração uni-dimensional. Adaptado de: (Beck, 2012).	32
3.4	Aproximações de primeira e segunda ordem para a equação de estado limite. (Beck, 2012).	36
3.5	Planos de Experiência. (Beck, 2012).	41
3.6	Amostragem por hipercubo latino. (Hurtado e Barbat, 1998)	43
4.1	Neurônio Biológico, <i>adaptado de</i> (Junior 2006)	48
4.2	Neurônio Artificial (Cardoso et al 2002)	49
4.3	de Ativação Sigmóide Logística	51
4.4	Função de Ativação Tangente Hiperbólica	52
4.5	Neural Multicamada (Kovács, 2006)	53
4.6	Geração Recursiva da Base da EPC. (Sudret et al 2006)	67
5.1	Treliça Estudada	82
5.2	Varição do Número de Épocas da RNA	83
5.3	Varição da Grau da EPC	84
5.4	Varição dos Hiperparâmetros na Krigagem	85
5.5	Comparação Entre as Diferentes Técnicas	86
5.6	Torre de Telecomunicação	87
5.7	Varição da quantidade de pontos de dados	88
5.8	Variância da previsão entre os dois primeiros pontos	89
5.9	Variância da previsão entre os próximos dois pontos	89
A.1	Número de peças remanescentes no estoque de uma loja	109
A.2	Temperatura em São Carlos no dia 16/02/2015	109
A.3	Problema de confiabilidade estrutural envolvendo variações no tempo (Beck, 2012)	110

LISTA DE TABELAS

Tabela	Legenda	Página
3.1	Quadrado Latino para $N = 5$	42
4.1	Polinômios de Legendre	62
4.2	Polinômios de Hermite	63
4.3	Polinômios e Graus Multi Índice	66
4.4	Funções de Auto-Correlação	73
5.1	Equações de estado limite estudadas	80
5.2	Variáveis aleatórias utilizadas	80
5.3	β para o número mínimo de pontos de suporte	81
5.4	β para o dobro do número mínimo de pontos de suporte	81
5.5	Propriedades das variáveis aleatórias envolvidas	83
5.6		86
5.7	Propriedades das variáveis aleatórias envolvidas	88
A.1	Sumário de distribuições contínuas de probabilidade. Fonte: (Beck 2012)	103

SUMÁRIO

1. INTRODUÇÃO.....	17
1.1. CONTEXTUALIZAÇÃO.....	17
1.2. OBJETIVOS.....	18
1.3. METODOLOGIA.....	19
1.4. JUSTIFICATIVA.....	20
1.5. ORGANIZAÇÃO DO CONTEÚDO.....	20
2. REVISÃO DA LITERATURA.....	22
3. CONFIABILIDADE ESTRUTURAL.....	27
3.1. INCERTEZAS NA ENGENHARIA DE ESTRUTURAS.....	27
3.2. ESTADOS LIMITES E PROBABILIDADE DE FALHA.....	27
3.3. MÉTODOS DE TRANSFORMAÇÃO E DE SIMULAÇÃO:.....	31
3.3.1. Métodos de transformação.....	31
3.3.2. Simulação de Monte Carlo.....	37
3.4. PLANEJAMENTO DE EXPERIMENTOS.....	39
4. META-MODELAGEM.....	44
4.1. REGRESSÃO POLINOMIAL.....	45
4.2. REDES NEURAIS ARTIFICIAIS.....	46
4.2.1. Neurônio biológico.....	47
4.2.2. Neurônio artificial.....	48
4.2.3. Funções de ativação.....	51
4.2.4. Arquiteturas.....	52
4.2.5. Processos de treinamento.....	54
4.2.6. Inicialização dos pesos.....	58
4.3. EXPANSÕES EM POLINÔMIOS DE CAOS (EPC).....	58
4.3.1. Ortonormalidade de polinômios.....	59
4.3.2. Famílias clássicas de polinômios.....	60
4.3.3. Transformação isoprobabilística.....	63
4.3.4. Polinômios multivariados.....	64

4.3.5.	Algoritmo para geração dos índices e construção da base	66
4.3.6.	Determinação dos coeficientes.....	68
4.3.7.	Pós-processamento dos coeficientes	70
4.4.	KRIGAGEM.....	70
4.4.1.	Regressão	71
4.4.2.	Funções de correlação	72
4.4.3.	Determinação dos hiperparâmetros.....	74
4.4.4.	Valor esperado e incerteza do modelo de Krigagem.....	75
5.	RESULTADOS	77
5.1.	META-MODELAGEM NO STRAND.....	77
5.2.	MAPEAMENTO DE DADOS	78
5.3.	EXEMPLOS ANALÍTICOS	79
5.4.	COMPARAÇÃO DAS TÉCNICAS PARA UMA TRELIÇA PLANA	82
5.4.1.	RNA	83
5.4.2.	EPC	84
5.4.3.	Krigagem	84
5.4.4.	Comparação	85
5.5.	AVALIAÇÃO DA EFICIÊNCIA DA KRIGAGEM SOBRE UMA TORRE DE TELECOMUNICAÇÃO.....	87
6.	CONCLUSÕES E SUGESTÕES	90
6.1.	CONCLUSÕES.....	90
6.2.	SUGESTÕES PARA TRABALHOS FUTUROS	92
	REFERÊNCIAS.....	94
	APÊNDICE – TÓPICOS DE PROBABILIDADE E ESTATÍSTICA.....	98
	CONCEITOS DE PROBABILIDADE	98
	VARIÁVEIS ALEATÓRIAS.....	101
	FUNÇÕES DE PROBABILIDADE	101
	VALOR ESPERADO DE UMA VARIÁVEL ALEATÓRIA	103
	DISTRIBUIÇÕES CONJUNTAS E MOMENTOS CONJUNTOS	106
	PROCESSOS ESTOCÁSTICOS	107

1. INTRODUÇÃO

1.1. CONTEXTUALIZAÇÃO

A segurança é um tema de central importância na engenharia. Sistemas projetados por engenheiros devem não apenas funcionar, mas também ser viáveis e seguros. Em especial na engenharia de estruturas, a segurança deve ser garantida não apenas por princípio, mas pela intensidade das consequências de uma possível falha. Uma vez que uma estrutura já tenha um nível de segurança suficiente, porém, pode ser inviável superdimensioná-la. Fazê-lo pode envolver grande consumo de recursos naturais, impactos ambientais e custos que aumentem rapidamente no sentido do aumento da segurança. A confiabilidade estrutural, então, quantifica a segurança de uma estrutura, e dá aos engenheiros as ferramentas para que elas sejam projetadas de maneira a cumprir todos os requisitos necessários.

A descrição do comportamento mecânico de uma estrutura é feita através da determinação de seus campos de tensões, deformações e deslocamentos. Muito se ganhou nos últimos anos com o advento dos computadores, que permitiram a aplicação em larga escala de técnicas numéricas. Tornou-se possível a utilização automática de métodos numéricos que trabalham com soluções aproximadas, como o *método dos elementos finitos*. A crescente capacidade de processamento dos computadores viabiliza a aplicação de soluções cada vez mais refinadas para problemas cada vez mais complexos, associados, porém, a um maior custo computacional. A estes, pode-se associar ferramentas de análise de confiabilidade, sendo possível determinar computacionalmente a segurança de uma estrutura.

Sólidos avanços têm sido alcançados nos últimos anos no estudo da confiabilidade estrutural. Os métodos de transformação FORM e SORM possibilitaram a análise de problemas antes impraticáveis, e os métodos de simulação (como as simulações de Monte Carlo) demonstraram ser uma

ferramenta muito poderosa. Ainda assim, estes podem ter custos de aplicação impraticáveis em problemas mais complexos, como problemas dinâmicos não lineares e problemas de otimização na presença de incertezas, enquanto aqueles podem não apresentar uma solução boa o suficiente para estes casos. Para contornar isso, diversas técnicas de redução de variância das amostras foram desenvolvidas, de modo a permitir que os métodos de simulação funcionem de maneira mais eficiente. Mesmo com a ajuda destas técnicas, porém, a solução de alguns problemas continua impraticável, levando dias ou até anos para que sejam realizadas com precisão aceitável. Nesse contexto, insere-se a ideia da utilização de técnicas de meta-modelagem, de modo que tais análises, em geral custosas para o computador, possam ser feitas sobre modelos mais simples e rápidos de se analisar, mas que representem de forma satisfatória o modelo original. Esta abordagem já foi bastante difundida e utilizada para o caso das superfícies de resposta polinomiais, que consistem em uma aproximação através de um ajuste pelo método dos mínimos quadrados. Nos últimos anos, porém, técnicas mais avançadas, precisas e eficientes vêm sendo propostas e aplicadas com sucesso, apesar de não serem tão amplamente utilizadas. O material disponível na literatura sobre estas técnicas e sua aplicação em confiabilidade estrutural não é abundante, e em português é quase inexistente. É neste contexto que se insere o presente trabalho, que busca produzir uma ferramenta para resolver problemas de confiabilidade estrutural usando algumas destas técnicas, e também descrevê-las de maneira prática, ajudando assim a difundi-las no Brasil.

1.2. OBJETIVOS

Este trabalho tem por objetivo compilar, assimilar, descrever, programar em computador e comparar meta-modelos construídos a partir de redes neurais artificiais, expansões em polinômios de caos e krigagem.

Objetivos específicos são:

1. Estudar e compreender as diferentes técnicas modernas de meta-modelagem.

2. Descrever as técnicas, as diferenças entre as mesmas e sua aplicação computacional.
3. Implementar as diferentes técnicas de meta-modelagem em um módulo do programa computacional StRAnD – *Structural Reliability Analysis and Design* – desenvolvido junto ao Departamento de Engenharia de Estruturas, USP.
4. Comparar a eficiência das diferentes técnicas de meta-modelagem na solução de problemas de confiabilidade estrutural.

1.3. METODOLOGIA

A metodologia consistiu em uma revisão da literatura com a finalidade de se estudar o que tem sido realizado nos últimos anos em relação à meta-modelagem. Estudos que deram origem às técnicas utilizadas neste trabalho, e nos quais elas foram aplicadas a problemas da engenharia de estruturas foram levantados, em especial para confiabilidade estrutural. Em seguida, programou-se um módulo de meta-modelagem no programa computacional StRAnD – *Structural Reliability Analysis and Design* contendo as técnicas prontas para sua utilização na análise de problemas de confiabilidade estrutural. Por fim, foi realizado um estudo comparativo do desempenho dos diferentes meta-modelos, de modo a identificar as vantagens e desvantagens de cada um deles.

1.4. JUSTIFICATIVA

O programa computacional **StRAnD** – *Structural Reliability Analysis and Design* – foi desenvolvido junto ao Departamento de Engenharia de Estruturas da Universidade de São Paulo, através de dois projetos de pesquisa individuais financiados pela FAPESP. Este programa tem sido fundamental para o desenvolvimento de projetos de pesquisa envolvendo análise de confiabilidade e otimização estrutural na presença de incertezas. O programa serviu como plataforma para obtenção dos resultados numéricos de diversas dissertações de mestrado e algumas teses de doutorado. Os recursos computacionais do programa StRAnD são comparáveis aos recursos dos melhores programas acadêmicos e comerciais da área desenvolvidos no mundo. Há módulos para análise de confiabilidade envolvendo variáveis aleatórias e para análise de confiabilidade dependente do tempo, envolvendo processos estocásticos. O usuário pode escolher entre 15 distribuições de probabilidade pré-programadas, e pode incluir correlação entre pares de variáveis aleatórias. Antes deste trabalho, porém, o programa estava bastante desatualizado no que diz respeito às técnicas modernas de meta-modelagem, contando apenas com superfícies de resposta polinomiais.

Através do presente projeto de mestrado, foram implementadas no StRAnD formas de aproximação das equações de estado limite através de técnicas modernas de meta-modelagem, que vem sendo aplicadas para estes fins apenas nos últimos dez anos, mantendo assim o programa atualizado com o estado da arte, tornando-o mais versátil e dando uma contribuição para que o programa possa continuar ajudando na produção acadêmica da Universidade de São Paulo.

1.5. ORGANIZAÇÃO DO CONTEÚDO

Após o panorama geral sobre o trabalho apresentado nesta introdução, o **capítulo 2** apresenta de forma resumida as técnicas abordadas neste trabalho, de sua origem até aplicações contemporâneas. Não é objetivo explicar as técnicas formalmente, nem fornecer equações, mas situar o leitor sobre o

estado da arte no campo de pesquisa, sem deixar de abordar também alguns aspectos históricos. O **capítulo 3** apresenta a teoria da confiabilidade estrutural, desde as ferramentas básicas da teoria da probabilidade necessárias à sua compreensão, até os métodos clássicos aplicados à resolução de problemas de confiabilidade. Este capítulo não tem por objetivo ser um tratado sobre os temas abordados, mas resgatar conceitos chave para a aplicação das técnicas de meta-modelagem. O **capítulo 4** apresenta o conceito de meta-modelos, e as técnicas abordadas neste trabalho são explicadas em sua teoria. Em oposição ao capítulo 2, a ideia é explanar, com didática e profundidade coerentes ao escopo deste trabalho, a definição e o funcionamento das diferentes técnicas. O **capítulo 5** apresenta alguns exemplos que envolvem a aplicação das técnicas estudadas. Tais resultados foram a partir do programa que foi desenvolvido pelo autor ao longo do presente trabalho. Por fim, no **capítulo 6** os resultados obtidos são discutidos e algumas sugestões para trabalhos futuros são apresentadas.

2. REVISÃO DA LITERATURA

Em 1989, a engenheira de estruturas italiana Lucia Faravelli propôs a utilização de superfícies de resposta polinomiais para a solução de problemas envolvendo confiabilidade estrutural, no trabalho “*A Response Surface Approach for Reliability Analysis*”. A publicação apresentava uma técnica de elementos finitos estocásticos aplicada a um vaso de pressão, cuja resposta estrutural era então representada pela superfície de resposta construída. Este artigo abriu o campo de pesquisa em meta-modelagem para aplicação em confiabilidade estrutural, e superfícies de resposta foram usadas desde então em diversas publicações. A aplicação desta técnica à confiabilidade estrutural é, então, relativamente recente, mas a técnica em si já vinha sendo desenvolvida há muito tempo. O método dos mínimos quadrados, ferramenta chave para a construção das superfícies de resposta, vem sendo usado pelo menos desde o século XVIII, quando astrônomos tentavam melhorar a precisão com que faziam suas previsões sobre os movimentos dos planetas, orientando as missões cumpridas pelas últimas grandes navegações da era das explorações (Stigler, 1986).

Da mesma forma, ainda que as aplicações das técnicas mais avançadas de meta-modelagem tenham surgido recentemente, é necessário voltar um pouco na história para identificar sua origem. Já no início dos anos 50, Danie Krige, o engenheiro de minas sul-africano que empresta seu nome à técnica da krigagem, criticava as técnicas estatísticas aplicadas à extração nos campos de mineração de ouro da África do Sul. Naquela época, estimava-se a quantidade de ouro que se deveria minerar em uma determinada área extensa considerando-se que a média da quantidade existente deveria corresponder à média encontrada em áreas menores, mais fáceis de estudar, que fizessem parte do terreno. Krige identificou erros nessa abordagem, e sugeriu que os pontos estudados estavam espacialmente correlacionados. Primeiramente, sugeriu que se levasse em consideração apenas as medições feitas nas adjacências da região sobre a qual a previsão deveria ser feita, e depois expandiu este conceito para as medições vizinhas a estas (Cressie, 1990). Foi apenas com o trabalho de Matheron (1963) que alguns erros da teoria de Krige

foram corrigidos, e a krigagem ganhou um aspecto matemático mais formal. Matheron é até hoje considerado o pai da geoestatística (Agterberg, 2005), mas o nome da técnica foi atribuída a Krige, devido à sua importante contribuição. A partir de então, a krigagem passou a ser utilizada nas mais diversas áreas: meteorologia, geodésica e física. Até uma versão primitiva da krigagem, baseada em séries temporais, já havia sido usada, sem muito sucesso, para prever o movimento de aviões inimigos na segunda guerra mundial, a partir de dados detectados por radar (Cressie, 1990).

Foi apenas em 1989 que Sacks et al. introduziram a ideia de usar a krigagem para ajustar um preditor global a dados gerados por algoritmos determinísticos em computador. As aplicações em simulações estocásticas começaram com o trabalho de Mitchell e Morris (1992). Desde então, a krigagem pode ser aplicada a problemas de confiabilidade estrutural. As aplicações da técnica como um meta-modelo, porém, datam do final da década de 90. Dubourg et. al. (2011) fazem um estudo sobre a utilização conjunta de krigagem com técnicas de amostragem por importância em simulação de Monte Carlo, e Soares et al. (2014) analisam a eficiência da Krigagem especificamente em problemas de confiabilidade estrutural.

Outra abordagem que se popularizou na década 90, mas vem sendo desenvolvida desde a metade do século XX, é a das chamadas redes neurais artificiais. O funcionamento do cérebro humano fascina a humanidade desde que se tem registro, e seu estudo tem sido objeto de atenção ao menos desde o antigo Egito. Ao final do século XIX o modernismo, que valorizava o racionalismo, moldava o pensamento do mundo ocidental e abria espaço para grandes avanços na área de ciência e de tecnologia. Na área da biologia, os estudos sobre a fisiologia do cérebro avançavam, e no início do século XX, o fisiologista russo Ivan Pavlov, ganhador do prêmio Nobel de medicina de 1904, descobria o reflexo condicionado, talvez a primeira abordagem científica sobre o comportamento. Na área da matemática, o matemático britânico Alan Turing lançava sua máquina universal – um modelo abstrato que daria origem ao computador moderno. Neste contexto, não demorou até que um neurofisiologista, Warren McCulloch, e um matemático, Walter Pitts, se unissem a tentar representar, de maneira artificial, o funcionamento do cérebro

humano. O resultado foi o trabalho considerado como o primeiro na área das redes neurais artificiais, McCulloch e Pitts (1943). Os pesquisadores usaram circuitos elétricos para criar uma unidade de processamento, e provaram que ela era capaz de realizar várias operações. A próxima contribuição importante veio em 1949, com o livro “A Organização do Comportamento”, de Donald Hebb, que sugeriu que as forças das sinapses seriam alteradas a cada vez que a informação passasse pelos neurônios. Com o surgimento dos computadores na década de 50, os modelos puderam avançar, e em 1957 o psicólogo americano Frank Rosenblatt propôs uma nova abordagem para o problema de reconhecimento de padrões, sugerindo o *perceptron*, que serve de base para a construção de muitas redes utilizadas até a atualidade. As pesquisas continuaram avançando até a publicação do livro “perceptron”, em 1969, pelo cientista cognitivo Marvin Minsky e o matemático Seymour Papert. Este livro demonstrava sérias limitações nos modelos de redes neurais propostos até então, e isso desanimou a comunidade científica, que praticamente abandonou este tipo de modelo. Os anos 80 trouxeram uma nova vida para a área, conforme o poder dos computadores aumentava de forma acelerada, e novos algoritmos adaptativos permitiam a representação de sistemas cada vez mais complexos. Em 1982 John Hopfield propõe um novo tipo de rede, e seu trabalho atraiu diversos matemáticos, físicos e engenheiros para a área, e, por fim, o desenvolvimento do treinamento por *backpropagation*, ao final dos anos 80, estabeleceu de vez as redes neurais artificiais como uma ferramenta sólida para desenvolvimento científico em inúmeras áreas do conhecimento. (Silva, 1998). Foi neste contexto que surgiu a ideia de aplicá-las a problemas de confiabilidade estrutural. Em 1993, Murotsu et al. estudaram a eficiência das redes neurais em problemas de confiabilidade e otimização estrutural, no trabalho “*Studies in assessment of structural reliability by response surface method and neural network. Reliability and Optimization of Structural System*”. Desde então, diversos trabalhos foram publicados, explorando o potencial das redes em aplicações relacionadas a engenharia de estruturas. Destacam-se Papadrakis et al. (1996), Gomes e Awruch (2004), Deng et al. (2005) e Elhewy et al. (2005).

Já as expansões em polinômios de caos provavelmente não teriam surgido, ou ao menos teriam demorado mais até serem pensadas, não fosse a revolução francesa. Pierre Simon de Laplace, matemático e astrônomo francês, além de seus dotes para a ciência, havia desenvolvido também uma grande habilidade política em meio ao contexto conturbado da época. Antes da revolução, ele foi recrutador para a artilharia real, e teve a promissora chance de avaliar um candidato de estatura menor do que a média dos demais candidatos, ninguém menos que Napoleão Bonaparte. Em 1789, quando foi instituída a assembleia nacional constituinte, Laplace foi considerado suspeito, mas declarou seu “ódio inextinguível à realeza”, e acabou por receber diversas honrarias na república. Quando Napoleão, seu conhecido, assumiu o império em 1808, Laplace mudou mais uma vez de lado, e ganhou o título de Conde (Mlodinov, 2008). Foi nessa época, usufruindo das regalias de sua posição, que escreveu seu trabalho, em francês, em que dava uma primeira definição para os polinômios de Hermite. O trabalho, em francês, foi estudado e aprofundado por Pafnuty Chebyshev (o idealizador da desigualdade de Chebyshev), e por fim, os polinômios foram nomeados homenageando o matemático francês Charles Hermite, que não os havia inventado, mas foi o primeiro a definir os polinômios multivariados. Este foi o desenvolvimento chave para que o filósofo e professor de matemática do MIT, Norbert Wiener, escrevesse o trabalho em que modelava processos estocásticos, um resultado que ficou conhecido posteriormente como “expansões em caos de Wiener”, ou ainda “expansões em polinômios de caos”.

Foi apenas na década de 90 que a primeira aplicação das expansões em polinômios de caos à mecânica foi publicada, com o pioneiro trabalho de Ghanem e Spanos (1991). Neste trabalho, as expansões são aplicadas a um problema de valor de contorno, resolvido pelo método dos elementos finitos estocásticos, no qual alguns parâmetros são descritos como um campo aleatório. Apenas os parâmetros de entrada eram assim modelados, e não a resposta estrutural do sistema. A ideia de utilizar a expansão como um meta-modelo surgiu apenas recentemente, no trabalho de Sudret e Der Kiureghian (2000), e se popularizou a partir da possibilidade de se calcular os coeficientes

da expansão de forma não intrusiva, com os trabalhos de Berveiller et al. (2004), Choi et al. (2004) e Berveiller et al. (2006).

Mesmo a abordagem não intrusiva torna-se inviável quando o problema abordado está definido em muitas dimensões estocásticas. Uma proposta eficaz de solução para este problema é fornecida nos trabalhos de Blatman e Sudret (2010, 2011), e Sudret (2014) que utilizam em problemas de confiabilidade estrutural uma técnica chamada expansão em polinômios de caos esparsa.

3. CONFIABILIDADE ESTRUTURAL

3.1. INCERTEZAS NA ENGENHARIA DE ESTRUTURAS

Diversos sistemas naturais podem ser descritos deterministicamente apenas de forma muito limitada. Em diversas situações cotidianas as descrições clássicas servem muito bem, mas quando os sistemas estudados se tornam mais complexos, ou quando a precisão das previsões precisa ser maior, é comum que elas se tornem inadequadas. Tanto a grande complexidade e caoticidade de inúmeros fenômenos, quanto o caráter naturalmente estocástico de outros, implicam em melhores descrições ao se utilizar ferramentas da probabilidade e da estatística.

Em engenharia de estruturas, é usual que os sistemas estudados apresentem diversas fontes de incertezas, das dimensões de uma viga concretada in loco, à resistência exata que seu concreto atingirá aos 28 dias. Das condições climáticas às quais a estrutura será submetida ao longo de sua vida útil, à falta de informação que se tem a partir do número reduzido de ensaios que se costuma fazer no solo antes da fundação. Desse modo, a resposta estrutural do sistema não pode ser prevista com certeza. Uma probabilidade de que o sistema não se comporte de forma satisfatória é um problema particularmente sério neste campo de atuação, pois falhas drásticas em estruturas civis podem estar associadas a grandes desastres. É importante então saber quantificar esta probabilidade e projetar os sistemas de forma consciente, de modo que ela seja tão pequena quanto seja viável. A determinação destas probabilidades é o problema central da confiabilidade estrutural.

3.2. ESTADOS LIMITES E PROBABILIDADE DE FALHA

A falha de uma estrutura pode ser uma situação difícil de caracterizar. Não é necessário que ela colapse por completo para que ela apresente um desempenho insatisfatório para determinada aplicação. Desse modo, é

preferível se trabalhar com o conceito de estado limite para a estrutura, que representa a passagem da estrutura de uma situação satisfatória para uma não satisfatória, dado um requisito que ela deva atender. Desse modo, para cada requisito, há um estado limite associado. Os estados limites costumam ser divididos em duas categorias (Beck, 2012):

- Estados limites últimos: São aqueles que dizem respeito à segurança e à estabilidade da estrutura. Suas violações incluem perda de equilíbrio global e a ação de solicitações que superam a resistência mecânica da estrutura.
- Estados limites de serviço: São aqueles que caracterizam situações indesejáveis de menor gravidade, que podem comprometer a utilização ou o desempenho da estrutura. Suas violações podem causar desconforto aos usuários da estrutura, prejuízos estéticos e a diminuição de sua vida útil.

Os estados limites podem ser representados de forma analítica através de funções chamadas *equações de estado limite* $g(\mathbf{X})$. A cada estado limite se associa uma equação de estado limite, que tem por domínio o espaço das variáveis aleatórias que são parâmetros de entrada do problema estudado, e por saída valores reais. Estes valores são positivos quando a estrutura apresenta desempenho satisfatório, e negativos quando o estado limite é violado. O conjunto de todos os pontos no espaço das variáveis de entrada que levam $g(\mathbf{X})$ a assumir valores negativos recebe o nome de *domínio de falha*, denotado por D_f , e dos pontos que levam a valores positivos, *domínio de segurança* ou *domínio de sobrevivência*, denotado por D_s . Desse modo, $g(\mathbf{X}) = 0$ define, no espaço das VAs, uma superfície limítrofe entre a falha e o sucesso da estrutura ao cumprir o requisito representado pelo respectivo estado limite. Este comportamento é ilustrado pela Figura (3.1) para o caso de duas VAs. Representada no eixo vertical, está a função conjunta de densidade de probabilidade dessas VAs.

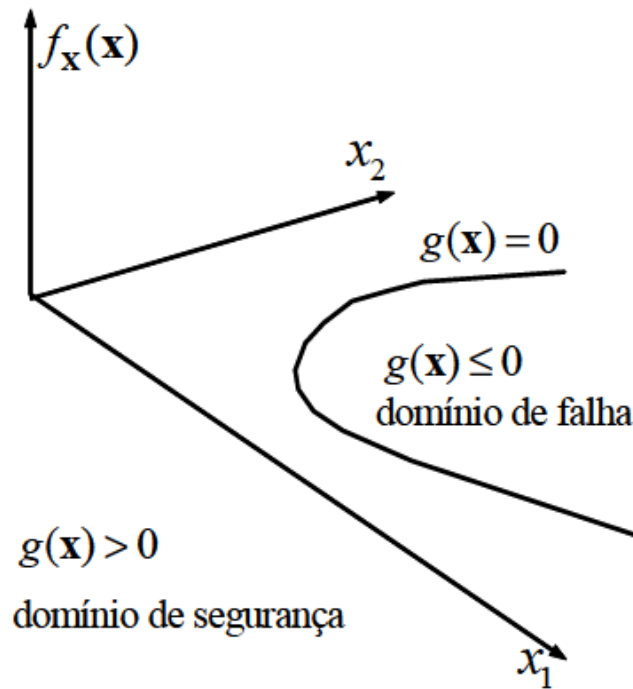


Figura 3.1 – Equação de estado limite de domínios de falha e segurança. Fonte: (Beck, 2012)

Conhecendo-se as características das variáveis aleatórias envolvidas em um dado problema, bem como a relação entre elas (basicamente, sua função conjunta de densidade de probabilidade), e considerando-se uma equação de estado limite, é possível calcular probabilidade de que tais variáveis adquiram conjuntamente valores que impliquem em $g(\mathbf{X})$ assumir valores negativos. Esta probabilidade é chamada *probabilidade de falha*, e denotada por P_f . Esta situação é retrada pela equação (3.1).

$$P_f = P[\{\mathbf{X} \in D_f\}] = P[\{g(\mathbf{X}) \leq 0\}] \quad (3.1)$$

A equação (A.21), do apêndice, pode ser aplicada à função conjunta de densidade de probabilidade das VAs de entrada para determinar o conteúdo de probabilidade no domínio de falha, e, assim, obtém-se uma expressão analítica para a determinação de P_f .

$$P_f = \int_{D_f} f(\mathbf{x}) d\mathbf{x} \quad (3.2)$$

No exemplo mais comum de equação de estado limite, que representa uma situação conhecida como *problema fundamental da confiabilidade*, duas variáveis aleatórias definem o estado do sistema, sendo elas R representando uma resistência ou capacidade, e S representando uma solicitação ou uma demanda. Desse modo, o estado limite é violado quando a demanda excede a capacidade, de modo que a equação de estado limite consiste na expressão (3.3).

$$g(R, S) = R - S \quad (3.3)$$

Esta situação é ilustrada pela Figura 3.2, que apresenta também hachurada a região onde está o conteúdo de probabilidades referente à probabilidade de falha para este estado limite, ilustrando também uma aplicação da equação (3.2).

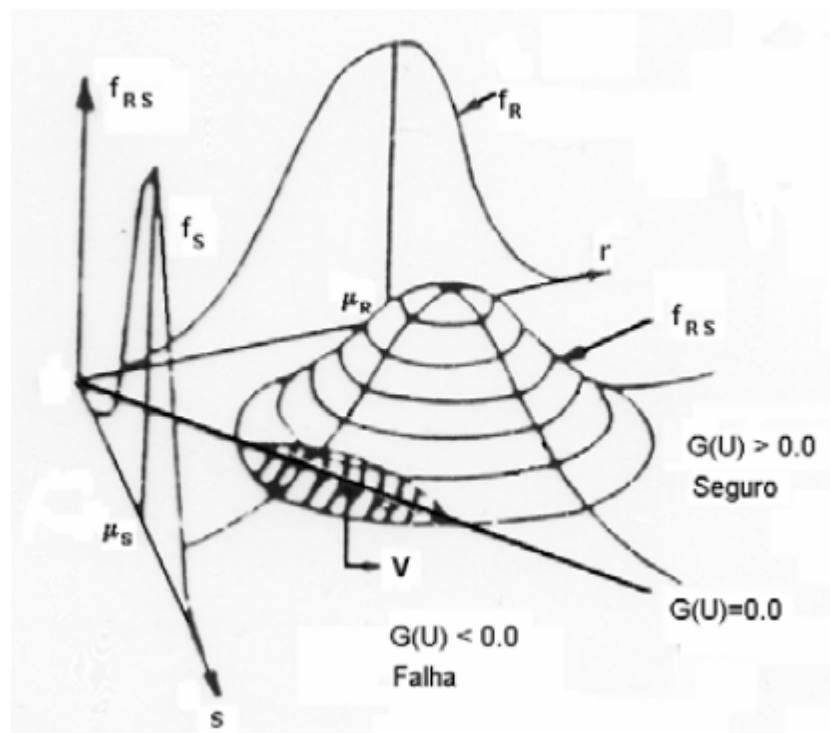


Figura 3.2 – Representação gráfica da probabilidade de falha. Fonte: (Sagrilo e Lima, 1996)

3.3. MÉTODOS DE TRANSFORMAÇÃO E DE SIMULAÇÃO:

A probabilidade de falha de uma estrutura pode ser determinada através da solução da integral expressa pela equação (3.2). Esta integral, porém, é em geral multidimensional e pode não ter solução analítica, ou ter uma solução analítica muito complexa. Além disso, o domínio de integração pode ser tal que a solução se torna ainda mais difícil. Buscando contornar este problema, diversos métodos para aproximar essa solução foram propostos e utilizados com sucesso ao longo do tempo. O desenvolvimento deste tópico é baseado na aclaração contida em Beck (2012).

3.3.1. Métodos de transformação

Um método bastante utilizado e relativamente simples para a determinação de P_f é o chamado *método de primeira ordem e segundo momento*, ou FOSM (*first order second moment*). Ele consiste na transformação do espaço das variáveis aleatórias do problema X para um espaço de variáveis independentes normal padrão Y , e na avaliação da probabilidade de falha a partir de uma linearização da equação de estado limite neste espaço, a partir do ponto de falha mais provável. Neste espaço, as distribuições de probabilidades das variáveis X_i são representadas pelas distribuições normais equivalentes a elas, atribuídas às variáveis Y_i . Tal representação é obtida através da transformação de Hassofer-Lind, conforme mostra a equação (3.4).

$$Y_i = \frac{X_i - \mu_{X_i}}{\sigma_{X_i}} \quad (3.4)$$

A função de distribuição de probabilidade conjunta de Y apresenta algumas propriedades importantes: Ela possui simetria radial e decai exponencialmente com relação à origem. Como as variáveis Y_i são normais e padronizadas, $f(\mathbf{y})$ fica centrada na origem do espaço Y . Por isso, a menor distância entre essa origem e a superfície que define o domínio de falha do problema será a norma de um vetor que vai até o ponto de falha mais provável da estrutura. Este ponto especial recebe o nome de *ponto de projeto*. Tal distância é por si só uma medida da probabilidade de falha da estrutura, e para uma equação de estado limite linear ou linearizada, pode ser interpretada como a quantidade de

desvios padrão que existem entre a média e a probabilidade de falha. A essa medida dá-se o nome de *índice de confiabilidade*, comumente representado por β . Esta aproximação é mostrada na Figura 3.3 para um caso unidimensional.

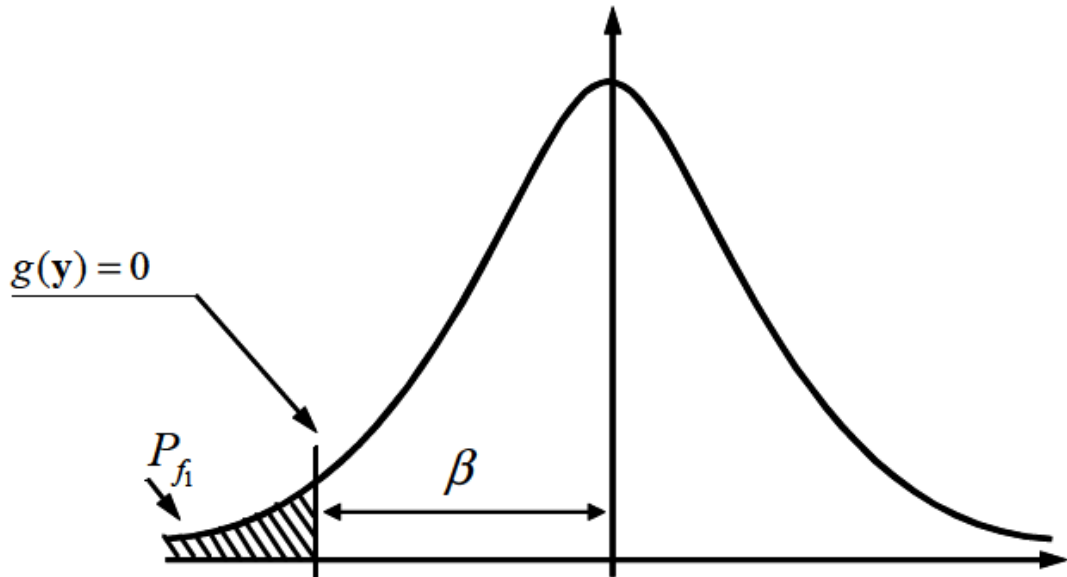


Figura 3.3 – Aproximação de primeira ordem, integração uni-dimensional. Adaptado de: (Beck, 2012).

Como o ponto de projeto reúne o maior conteúdo de probabilidade do domínio de falha, ele é o melhor ponto para se linearizar a equação de estado limite, pois assim, o erro cometido devido à aproximação é menor. Uma forma de enunciar o problema, preocupando-se em determinar o índice de confiabilidade, é tratá-lo como um problema de otimização. Deseja-se encontrar o ponto \mathbf{y}^* que minimiza a seguinte função da distância:

$$d = \sqrt{\mathbf{y}^T \mathbf{y}} \quad (3.5)$$

Sujeito à restrição de que.

$$g(\mathbf{y}) = 0 \quad (3.6)$$

Há um algoritmo que foi desenvolvido especificamente para endereçar este problema, que é conhecido como algoritmo de Hasofer, Lind, Rackwitz e Fiessler (HLRF). Trata-se de uma fórmula recursiva que está baseada na

aproximação de um ponto qualquer y à superfície $g(y) = 0$ e na perpendicularização entre o valor y e a tangente à superfície no ponto. Note que como a distribuição multi-normal apresenta simetria radial e está centrada na origem, um vetor perpendicular a uma linha de isoprobabilidade que esteja na direção da origem terá como norma a menor distância entre a origem e esta linha. Para um ponto inicial qualquer y_k , expande-se a equação de estado limite em série de Taylor até a primeira ordem, conforme equação (3.7).

$$g(y_{K+1}) = g(y_K) + \nabla g(y_K)^T (y_{K+1} - y_K) \quad (3.7)$$

na qual $\nabla g(y_k)$ é o gradiente da equação de estado limite no espaço normal padrão, avaliado em y_k . Para a equação linearizada deve-se encontrar um novo ponto y_{k+1} , de tal forma que $g(y_{k+1}) = 0$. O valor inicial do índice de confiabilidade é obtido como $\beta_k = \sqrt{\mathbf{y}^T \mathbf{y}}$ e o vetor de cossenos diretores da equação de estado limite α_k avaliado no ponto y_k é dado por:

$$\alpha_k = \frac{\nabla g(y_k)}{\|\nabla g(y_k)\|} \quad (3.8)$$

A ideia é caminhar sobre a equação linearizada buscando-se o próximo ponto, até que se obtenha uma variação pequena o suficiente entre iterações sucessivas, enquanto $\frac{g(y_k)}{\|\nabla g(y_k)\|}$ vai se aproximando de zero conforme o algoritmo vai convergindo (e a equação de estado limite se aproxima da superfície de falha). Portanto, o novo ponto pode ser encontrado através da expressão 3.9.

$$y_{K+1} = -\alpha_k \left[\beta_k + \frac{g(y_k)}{\|\nabla g(y_k)\|} \right] \quad (3.9)$$

Esta equação é então utilizada de forma iterativa até que o algoritmo convirja. Uma descrição detalhada do algoritmo HLRF melhorado com a inclusão de um ajuste para que o passo dado a cada iteração seja ótimo pode ser encontrada em Beck (2012).

Ainda que em métodos anteriores as transformações feitas até este ponto fossem suficientes, uma avaliação mais sofisticada de P_f é proposta pelo

método de confiabilidade de primeira ordem ou FORM (*first order reliability method*). Ela se baseia em uma função conjunta de probabilidade f_X construída não apenas com base nos momentos de cada variável aleatória independente, mas também nos coeficientes de correlação entre cada par de variáveis (o que costuma ser toda a informação possível de se obter na maioria dos casos práticos). Portanto, define-se uma matriz de correlação \mathbf{R}_X :

$$\mathbf{R}_X = \begin{bmatrix} 1 & \cdots & \rho_{1m} \\ \vdots & \ddots & \vdots \\ \rho_{m1} & \cdots & 1 \end{bmatrix} \quad (3.10)$$

sendo m o número de variáveis aleatórias envolvidas. A ideia é a determinação de uma função de distribuição de probabilidade multinormal $f(Y)$ que represente $f(X)$.

O primeiro passo consiste na transformação das distribuições marginais de probabilidade $f(X_i)$ em distribuições normais equivalentes. Essa transformação é pontual, e vale para o ponto analisado x^* . Essa transformação é isoprobabilística (conforme comentado em 4.3.3), e leva a:

$$z_i^* = \Phi^{-1}(F(x_i^*)) \quad (3.11)$$

Os parâmetros da distribuição normal equivalente são obtidos através das equações (3.12) e (3.13)

$$\sigma_{X_i}^{neq} = \frac{\phi(z_i^*)}{f(x_i^*)} \quad (3.12)$$

$$\mu_{X_i}^{neq} = x_i^* - z_i^* \sigma_{X_i}^{neq} \quad (3.13)$$

Nesta notação, Φ é a FDA da distribuição normal, e ϕ a FDP. Como esta transformação é feita para cada distribuição marginal, é conveniente escrevê-la de forma matricial, se utilizando de um vetor de médias \mathbf{M}^{neq} e uma matriz diagonal de desvios padrão \mathbf{D}^{neq} :

$$\mathbf{M}^{neq} = [\mu_{X_1}^{neq}, \mu_{X_2}^{neq}, \dots, \mu_m^{neq}]^T \quad (3.14)$$

$$\mathbf{D}^{neq} = \begin{bmatrix} \sigma_{X_1}^{neq} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma_{X_m}^{neq} \end{bmatrix} \quad (3.15)$$

As transformações então ficam:

$$\mathbf{z} = (\mathbf{D}^{neq})^{-1}(\mathbf{x} - \mathbf{M}^{neq}) \quad (3.16)$$

$$\mathbf{x} = \mathbf{D}^{neq} \mathbf{z} + \mathbf{M}^{neq} \quad (3.17)$$

Como as variáveis originais X_i podem apresentar correlação, uma correlação equivalente deve ser imposta também às variáveis Z_i . Nisso consiste o segundo passo, que se utiliza do *modelo de Nataf* para construir uma função conjunta de densidade de probabilidades aproximada, partindo-se da distribuição normal padrão multivariada e de uma matriz de correlação \mathbf{R}_z a ser determinada.

$$f_X(\mathbf{x}) = \phi_n(\mathbf{z}, \mathbf{R}_z) \frac{f_{X_1}(x_1)f_{X_2}(x_2) \dots f_{X_m}(x_m)}{\phi(z_1)\phi(z_2) \dots \phi(z_m)} \quad (3.18)$$

A determinação dos coeficientes de correlação equivalentes é feita de forma iterativa, utilizando a informação dos coeficientes de correlação das variáveis originais, através da expressão (3.19).

$$\rho_{X_{ij}} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} z_i z_j \phi_2(z_i, z_j, \rho_{Z_{ij}}) dz_i dz_j \quad (3.19)$$

sendo ϕ_2 a distribuição de probabilidade normal conjunta para duas variáveis. Esta equação é utilizada iterativamente até que sejam obtidos valores para $\rho_{Z_{ij}}$ que impliquem nos $\rho_{X_{ij}}$ conhecidos a priori. Esta etapa conclui a transformação do espaço das VAs de entrada para o espaço \mathbb{Z} .

A terceira e última etapa consiste na eliminação das correlações obtidas, obtendo-se assim um espaço \mathbb{Y} de variáveis normais padrão não correlacionadas, o que pode ser feito através da decomposição espectral da matriz de correlação ou da decomposição de Cholesky.

Isso conclui as transformações executadas no FORM. A partir daí, a probabilidade de falha pode ser determinada utilizando o algoritmo HLRF para

se encontrar o ponto de projeto e adotando-se $\beta = \|\mathbf{y}^*\|$, ou seja, a avaliação é feita no espaço \mathbb{Y} .

O FORM apresenta uma aproximação boa para muitos problemas, mas sua eficácia é comprometida quando o vetor \mathbf{X} possui dimensão muito alta, ou quando suas variáveis são muito correlacionadas ou fortemente não gaussianas. O método de confiabilidade de segunda ordem, ou SORM (*second order reliability method*) é uma tentativa de se obter uma aproximação mais precisa do que aquela obtida com o FORM, método com o qual se assemelha muito. O método segue exatamente o mesmo roteiro do FORM, porém a equação de estado limite não é aproximada por uma reta, mas por uma parábola, e o conteúdo de probabilidade dela é considerado como sendo a probabilidade de falha do sistema. Existem dois métodos para a determinação da aproximação parabólica, um baseado em curvas, e o outro baseado em pontos. A Figura 3.4 ilustra uma aproximação feita por FORM e por SORM para o mesmo problema, evidenciando o ganho de precisão que se obtém quando se adota o segundo método:

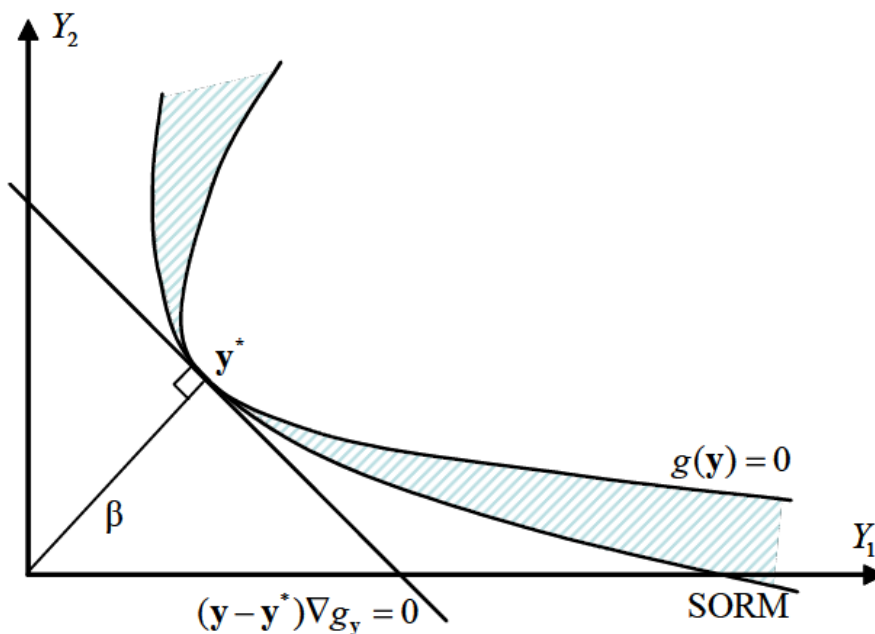


Figura 3.4 Aproximações de primeira e segunda ordem para a equação de estado limite. Fonte: (Beck, 2012).

Outra forma de se obter uma aproximação parabólica para a equação de estado limite é a utilização de uma superfície de resposta quadrática. Este método será abordado no item 4.1.

3.3.2. Simulação de Monte Carlo

Nem sempre a adoção de abordagens analíticas é interessante para problemas de confiabilidade estrutural. Em alguns casos, os métodos de transformação apresentados no item anterior podem não convergir ou apresentar um resultado insatisfatório. Isso acontece sobretudo quando as equações de estado limite são fortemente não lineares, quando há mais de um ponto de projeto no problema e também em sistemas com muitas equações de estado limite. Outra abordagem possível é a utilização de simulações numéricas. Trata-se de técnicas que permitem a realização de experimentos em um computador, tomando por base um modelo que represente o sistema original a ser avaliado. A simulação de monte Carlo foi proposta por Metropolis e Ulam (1949), e aplicada nas mais diversas áreas desde então. Este método empresta seu nome das roletas de Monte Carlo, cidade situada no principado de Mônaco. O caráter aleatório do experimento de girar uma roleta não viciada faz um paralelo com o funcionamento deste método de simulação, que trabalha a partir da geração de números aleatórios. A simulação de Monte Carlo se baseia na repetição de um experimento para diversos valores aleatórios de seus parâmetros de entrada.

Primeiramente, gera-se uma amostra aleatória de tamanho n_{si} segundo a função conjunta de densidade de probabilidade das variáveis de entrada do problema, $f(\mathbf{X})$. Analisa-se então o modelo estudado (uma equação de estado limite) para cada amostra i dentre as n_{si} geradas usando como parâmetros de entrada os valores do vetor \mathbf{x}_i , $i = 1, 2, \dots, n_{si}$. Ou seja, uma realização de cada variável aleatória é realizada para cada i , e o modelo estudado é analisado n_{si} vezes. Em seguida, define-se uma função indicadora I , conforme representado nas expressões (3.20) e (3.21).

$$x \in D_f \Rightarrow I[x] = 1 \quad (3.20)$$

$$x \notin D_f \Rightarrow I[x] = 0 \quad (3.21)$$

Com o auxílio desta função, a probabilidade de falha pode ser obtida através de uma integração sobre todo o domínio. Trata-se de outra forma de escrever a expressão (3.2).

$$P_f = \int_{\Omega} I[x] f(x) dx \quad (3.22)$$

A expressão 3.22, porém, é por definição (equação A.12, do apêndice) o valor esperado da função I , $E[I]$. Sendo assim, é possível calcular uma probabilidade de falha aproximada \bar{P}_f para o problema considerado utilizando-se um estimador para o valor esperado da função indicadora. Este estimador é representado pela equação (3.23).

$$\bar{P}_f = \frac{1}{n_{si}} \sum_{i=1}^{n_{si}} I[x_i] = \frac{n_f}{n_{si}} \quad (3.23)$$

sendo n_f o número de vezes em que a amostra gerada correspondeu a um ponto sobre o domínio de falha. Este é um estimador não tendencioso para a P_f , portanto se aproxima de seu valor exato conforme o número de simulações aumenta.

Como na prática a amostra pode ter apenas um tamanho finito, esta aproximação estará sujeita a um erro estatístico, que é diluído conforme o tamanho da amostra cresce. Uma estimativa para este erro pode ser calculada através da expressão (3.24).

$$Var[\bar{P}_f] = \frac{1}{(n_{si} - 1)} \sum_{i=1}^{n_{si}} (I[x_i] - \bar{P}_f)^2 \quad (3.24)$$

A ordem de grandeza da probabilidade de falha também influencia na qualidade dos resultados. Quanto menor for a probabilidade de falha, maior o

número de simulações necessários para se obter um mesmo coeficiente de variação. Esse é um problema especialmente importante para a engenharia de estruturas, uma vez que a P_f dos sistemas analisados costuma ser pequena. Uma estimativa para o coeficiente de variação que evidencia este problema é fornecida pela equação (3.25).

$$CV_{P_f} = \frac{1}{\sqrt{n_{si}P_f}} \quad (3.25)$$

Este método é bastante conveniente para a solução de problemas de confiabilidade estrutural, pois a probabilidade de falha pode ser determinada a partir de um equacionamento bastante simples, uma vez que haja meios de se resolver o sistema estudado n_{si} vezes. Por outro lado, a complexidade dos modelos de algumas estruturas, como em problemas dinâmicos, física e geometricamente não lineares, de otimização na presença de incertezas, ou simplesmente que apresentam muitos graus de liberdade, pode tornar essa abordagem inviável, sobretudo levando em consideração a equação (3.25). Uma forma de se tentar superar este problema é a utilização de técnicas de amostragem inteligente. A partir de estratégias na seleção das amostras, é possível fazer com que as simulações de Monte Carlo convirjam para um resultado preciso com n_{si} menores. Este trabalho não irá abordar tais técnicas. Descrições detalhadas de várias delas podem ser encontradas no trabalho de Santos (2014).

3.4. PLANEJAMENTO DE EXPERIMENTOS

Quando se deseja estudar um sistema com base em observações de seu funcionamento, a escolha dos pontos observados pode influenciar em muito na qualidade das informações obtidas. A ideia principal do planejamento de experimentos é a obtenção do máximo de informação útil sobre o funcionamento de um sistema se realizando o mínimo de observações possível. É comum utilizar as informações obtidas desta forma para a construção de um modelo através de uma regressão ou outra técnica de aproximação. Também, a escolha inteligente de pontos de um sistema é a base para algumas técnicas de amostragem inteligente em simulações de

Monte Carlo. Quando se está interessado em resolver um modelo o menor número de vezes possíveis, como em situações que motivam a utilização de meta-modelos, este tópico ganha especial importância. Na estatística, mais importante do que a análise de dados propriamente dita, é o planejamento dos experimentos em que esses dados devem ser obtidos (Neto et al., 2012). Uma frase célebre atribuída ao estatístico e biólogo inglês R. A. Fisher ilustra bem este princípio: “Chamar o especialista em estatística depois que o experimento já foi feito pode ser o mesmo que pedir para ele fazer um exame post-mortem. Talvez ele consiga dizer de que foi que o experimento morreu.”.

Uma técnica bastante usada por engenheiros e cientistas é a chamada *abordagem do melhor palpite (best guess approach)*, que consiste na realização de um experimento após a alteração de um fator que o influencia, mantendo todos os outros constantes. Essa decisão costuma funcionar devido à experiência e ao grau de conhecimento sobre o fenômeno por parte dos especialistas envolvidos. Todavia, caso o experimento inicial não produza bons resultados, esta abordagem não dá nenhuma ferramenta para a decisão sobre os valores dos fatores para o próximo experimento, o que pode se estender indefinidamente. Ainda, caso o experimento inicial apresente um resultado bastante satisfatório, o pesquisador pode se convencer de que obteve sucesso e parar os experimentos, porém não há garantia alguma de que o melhor resultado possível já foi observado. Outra técnica bastante usada na prática é chamada *um fator a cada vez (one factor at a time)*, que consiste na variação sistemática de um fator para vários valores possíveis, fixando-se os valores de todos os outros. Em seguida, varia-se apenas um outro fator, e faz-se isso para diversas combinações, até se poder inferir sobre o comportamento do sistema. Esta abordagem, além de implicar na realização de um número muito grande de experimentos, conforme o número de fatores aumenta, não considera possíveis interações que existam entre tais variáveis de entrada (Montgomery, 2009). Faz-se necessária então a adoção técnicas mais criteriosas de planejamento.

Uma possibilidade de fácil aplicação é a utilização de planos de experiência. Trata-se de uma maneira pré-definida e sistemática de selecionar

os pontos a serem observados, tomando-se por base um ponto central. Este ponto pode ser a média das variáveis aleatórias envolvidas no problema, ou o ponto de projeto em problemas de confiabilidade estrutural. A Figura 3.5 ilustra alguns planos de experiência comuns. Em sua notação, V_1 e V_2 são os fatores.

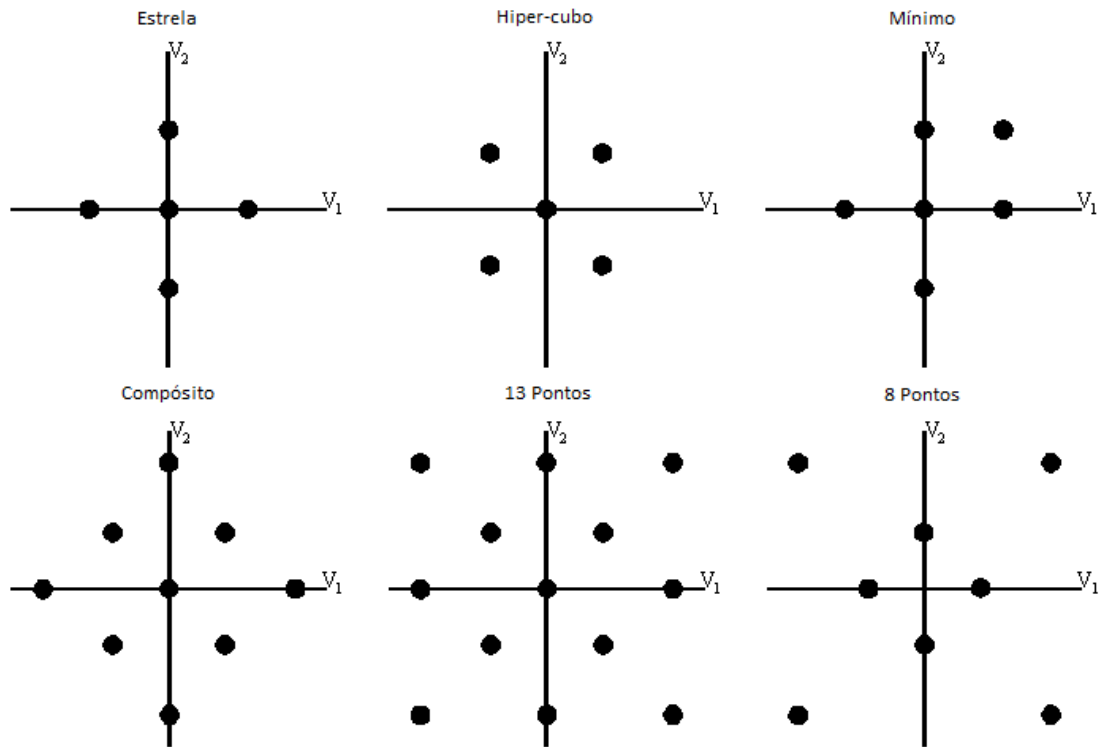


Figura 3.5 Planos de experiência. Fonte: (Leonel, 2009).

Algumas estratégias de planejamento de experimento se baseiam na ideia do preenchimento do espaço das variáveis de entrada de maneira esparsa, de modo a se obter informações sobre diferentes regiões do espaço sem redundância na seleção dos pontos. Neste contexto se insere a amostragem por *quadrados latinos*, matrizes de ordem N preenchidas por N tipos de símbolos, cada um podendo constar uma única vez em cada linha e em cada coluna. As linhas e as colunas dessa matriz representam fatores, e os símbolos, tipos de experimentos a serem realizados (por exemplo, diferentes formulações a serem testadas). Sendo assim, se apenas um tipo de experimento estiver sendo considerado, haverá apenas um símbolo, e ele aparecerá apenas N vezes, nunca repetindo uma linha ou coluna. O restante da matriz, então, ficará vazio. O preenchimento de um quadrado latino foi comparado por Montgomery (2009) a um jogo de *sudoku*, expressão que

significa “número único” em japonês. No jogo, deve-se preencher uma grade com números de 1 a 9, não sendo permitido repetir o mesmo número na mesma linha ou na mesma coluna. No caso dos quadrados latinos, porém, usualmente se utilizam letras para preencher a matriz. A tabela 3.1 mostra um quadrado latino com $N = 5$ completamente preenchido:

Tabela 3.1: Quadrado Latino para $N = 5$

A	D	B	E	C
D	A	C	B	E
C	B	E	D	A
B	E	A	C	D
E	C	D	A	B

A técnica dos quadrados latinos pode ser generalizada para muitas dimensões, dando origem ao chamado *hipercubo latino*. Selecionar a amostra de um problema de confiabilidade estrutural usando a lógica dos hipercubos latinos é bastante útil, consistindo em uma técnica de amostragem inteligente para a realização de simulações de Monte Carlo. A técnica de amostragem por hipercubo latino consiste na divisão do domínio de cada variável aleatória em um número finito de faixas, amostrando-se apenas uma vez um valor pertencente a tal faixa. Desse modo, o espaço das VAs é coberto de maneira esparsa e homogênea, conforme exemplificado pela Figura (3.6):

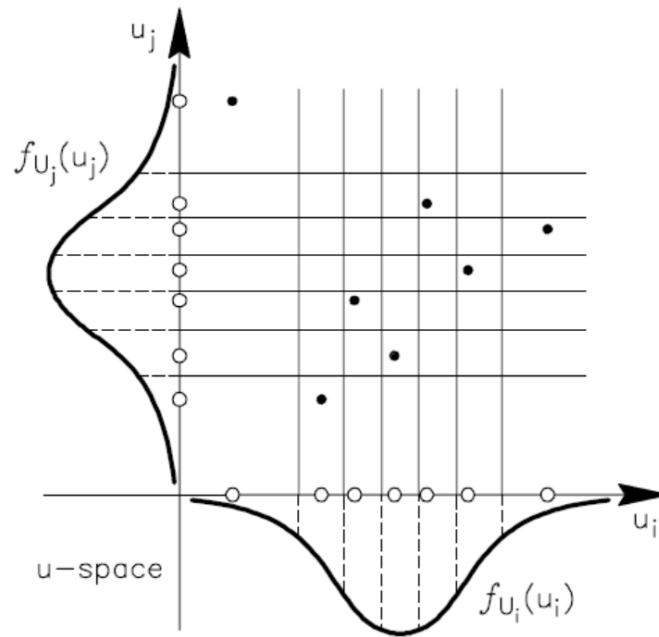


Figura 3.6 Amostragem por hipercubo latino. Fonte: (Hurtado e Barbat, 1998)

É possível montar uma amostra utilizando hipercubo latino a partir de alguns passos simples: para o caso de um problema com m variáveis aleatórias e n pontos amostrais, gera-se uma matriz \mathbf{P} de dimensões $(m \times n)$, na qual cada uma das colunas contém uma permutação aleatória de $1, \dots, n$. Define-se também outra matriz \mathbf{R} , com as mesmas dimensões de \mathbf{P} , preenchida com números aleatórios, gerados a partir de uma distribuição uniforme entre 0 e 1. A partir delas, define-se então uma matriz \mathbf{S} , conforme a expressão (3.26).

$$\mathbf{S} = \frac{1}{n}(\mathbf{P} - \mathbf{R}) \quad (3.26)$$

As amostras são então geradas a partir da matriz \mathbf{S} :

$$x_{ij} = F_{x_j}^{-1}(s_{ij}) \quad (3.27)$$

com $F_{x_j}^{-1}$ a função de probabilidade acumulada da variável X_j .

4. META-MODELAGEM

Meta-modelo (ou metamodelo) é o nome atual para o que classicamente se conhecia por superfície de resposta. Seu objetivo é fornecer, para conjuntos de dados de entrada específicos, uma resposta próxima o suficiente daquela que seria oferecida ao se avaliar o modelo que se está tentando representar. Basicamente, um meta-modelo é uma função analítica que aproxima a resposta de um modelo numérico complexo \mathcal{M} (como um modelo de elementos finitos ou uma equação de estado limite, que pode conter implicitamente modelos numéricos também), tal que:

$$\mathbf{y} = \mathcal{M}(\mathbf{x}) \quad (4.1)$$

onde \mathbf{y} é um vetor que reúne as respostas de interesse, e \mathbf{x} o vetor que reúne as variáveis aleatórias, parâmetros de entrada do problema considerado.

Um meta-modelo apresenta as seguintes características (Sudret, 2012):

- Pertence a uma determinada classe de funções, o que determina o seu tipo, e é completamente caracterizado por um conjunto de parâmetros, uma vez que a classe esteja estabelecida.
- O custo computacional de sua avaliação é consideravelmente menor do que o custo do modelo numérico original.
- O meta-modelo se ajusta ao modelo numérico a partir do conhecimento da resposta do modelo original em alguns pontos.

Desse modo, sua utilização se faz bastante útil em problemas nos quais um modelo precise ser resolvido muitas vezes, como no caso da determinação de uma probabilidade de falha através de simulações de Monte Carlo. Um modelo numérico complexo pode levar horas para ser avaliado por um computador, o que torna inviável resolvê-lo milhares de vezes. Este problema pode ser contornado ao se substituir tal modelo por um meta-modelo adequado nas simulações.

Em geral, os modelos são ajustados aos pontos conhecidos através de técnicas de minimização dos erros, sendo necessários m pontos conhecidos para um modelo com m parâmetros livres, ainda que costume-se avaliar mais pontos do que o estritamente necessário para se obter informações sobre os erros do modelo (Ditlevsen & Madsen, 1996, p. 137).

4.1. REGRESSÃO POLINOMIAL

Regressão polinomial, também denominada *Metodologia de Superfície de Resposta* (MSR), é uma forma de regressão linear que ajusta um modelo não linear a um conjunto de treinamento D , que pode ser definido da seguinte forma (Loshchilov, 2013)

$$D = \{(x_i, y_i), \quad i = 1, \dots, l\} \quad (4.2)$$

onde $x_i \in \mathbb{R}^n$ é um conjunto de parâmetros de entrada do sistema a ser aproximado, como aqueles reunidos pelo vetor de variáveis aleatórias X , e $y_i \in \mathbb{R}$ é um conjunto de respostas do sistema aos valores x_i . Um modelo de regressão polinomial pode ser expresso pela função aproximadora:

$$\hat{f} = \tilde{x}^T \boldsymbol{\beta} \quad (4.3)$$

sendo \tilde{x} um vetor base de funções linearmente independentes, e $\boldsymbol{\beta}$ reunindo parâmetros de ajuste da função aproximadora, que podem estar associados a pesos que ponderam a influência de cada variável utilizada no modelo. O tipo do modelo de regressão utilizada depende de \tilde{x} . Pode-se escolher, por exemplo, $\tilde{x} = (1, x_1, x_2, \dots, x_n)$ para um modelo de regressão linear, e $\tilde{x} = (1, x_1^2, x_2^2, x_1 x_2, \dots, x_{n-1} x_n, x_1, x_2, \dots, x_n)$ para um modelo quadrático com termos cruzados. Os coeficientes de $\boldsymbol{\beta}$ podem então ser determinados através de técnicas de minimização ou interpolação, como o método dos mínimos quadrados. A solução pelo método dos mínimos quadrados (MMQ) consiste em determinar os coeficientes que minimizam a diferença entre a resposta exata e a aproximação dada pela superfície de resposta:

$$\boldsymbol{\beta} = \min \left\{ \sum_{i=1}^l [(\tilde{\mathbf{x}}_i^T \boldsymbol{\beta}_i - y_i)^2] \right\} \quad (4.4)$$

cuja solução é dada pela expressão 4.5 (Faravelli, 1989).

$$\mathbf{A} = (\mathbf{Q}\mathbf{Q}^T)^{-1}\mathbf{Q}\mathbf{B} \quad (4.5)$$

sendo \mathbf{Q} uma matriz em cujas colunas estão dispostos os vetores $\tilde{\mathbf{x}}_i$, e \mathbf{B} um vetor que reúne as respostas $y_i, i = 1, 2, \dots, l$

O próximo passo consiste em validar a aproximação feita, testando-a em outros pontos de resposta conhecida do modelo original ou utilizando testes estatísticos. É possível ainda testar a significância da contribuição dos termos utilizados na aproximação. Por exemplo, em uma aproximação quadrática, pode-se testar a influência dos termos quadráticos nos resultados. Se ela for observada pequena, um modelo mais simples que não considera estes termos pode ser utilizado (Breitung & Faravelli, 1996).

4.2. REDES NEURAIAS ARTIFICIAIS

A ideia de redes neurais artificiais (RNA) foi desenvolvida a partir da década de 1940, refletindo os então recentes avanços científicos que estavam ocorrendo na época, como o estudo dos neurônios na área das ciências biológicas, e o da máquina de Turing na ciência da computação (Kovács, 2006). Redes Neurais Artificiais são dispositivos inspirados no funcionamento das redes neurais biológicas que, assim como faz o cérebro humano, possam reconhecer padrões e auxiliar na otimização e previsão da resposta de sistemas complexos. Estes modelos foram desenvolvidos tomando por base as seguintes hipóteses (Fausett, 1994):

- O processamento das informações não ocorre como na arquitetura tradicional da máquina de Von Neumann, com uma unidade central de processamento que executa todas as computações de forma sequencial, mas

em diversos elementos simples, denominados neurônios, unidades, células ou nós.

- Sinais são transmitidos entre os neurônios através de conexões.

- A cada uma destas conexões está associado um peso, que multiplica o valor do sinal transmitido através dela.

- O valor de saída de cada neurônio é o somatório dos seus valores de entrada, cada sinal enviado a ele ponderado pelo peso de sua respectiva conexão, submetido à função de ativação deste neurônio.

Uma rede neural artificial é caracterizada pela disposição e quantidade de neurônios e suas conexões ou sua arquitetura, pela forma como são definidos os pesos da conexão ou seu algoritmo de treinamento, e pela função de ativação utilizada por cada neurônio.

4.2.1. Neurônio biológico

Tanto os neurônios quanto a arquitetura das redes neurais artificiais são inspiradas nas redes neurais biológicas. Ainda que o funcionamento do cérebro humano não tenha sido até hoje totalmente compreendido, há muitos estudos sobre sua fisiologia e anatomia. Seu funcionamento de forma geral já foi identificado, bem como diversas partes do cérebro já foram catalogadas como responsáveis por funções executadas pelo corpo humano.

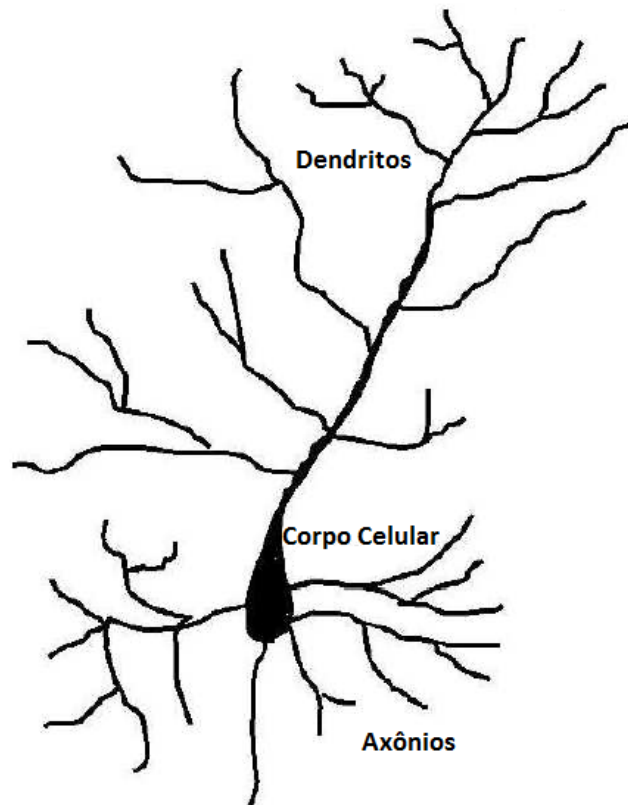


Figura 4.1 – Neurônio biológico, *adaptado de* (Junior 2006)

Em um neurônio biológico, representado de forma simplificada pela Figura 4.1, o corpo celular pode disparar um impulso elétrico que será transmitido através os axônios, os transmissores de informações de saída, até os dendritos das células com as quais tiverem contato. Os dendritos são as estruturas responsáveis pela transmissão da informação de entrada, e ao ponto em que um axônio se conecta a um dendrito dá-se o nome de sinapse. Nas sinapses, não há contato físico entre os dendritos e os axônios, elas ocorrem através da liberação de neurotransmissores responsáveis por ponderar a transmissão dos impulsos. Quando vários impulsos elétricos são recebidos pela região dendrítica de um neurônio, o resultado usual é um aumento na probabilidade do seu corpo celular disparar também um impulso (Dayhoff, 1990).

4.2.2. Neurônio artificial

Neurônios artificiais são modelos simplificados dos neurônios biológicos, inspirados em estudos como o aclamado trabalho de Hodgkin e Huxley (1952),

no qual a propagação e a transmissão de impulsos elétricos através da membrana celular de neurônios biológicos foram estudadas, e seus comportamentos foram descritos matematicamente.

O primeiro modelo proposto, ainda amplamente utilizado, foi o de McCulloth – Pitts (1943). Este modelo trabalhava apenas com inputs e outputs binários, e o processamento basicamente consistia em verificar se o somatório dos valores de entrada superavam ou não determinado limite. O output era então 1 caso superasse, e 0 caso não superasse. Ainda que muito simples, a maioria das aplicações de RNA hoje em dia se utilizam de um modelo muito parecido, um pouco mais desenvolvido, conforme representado na Figura 4.2. A Figura 4.2 representa um neurônio artificial m que recebe informações através de L canais de entrada.

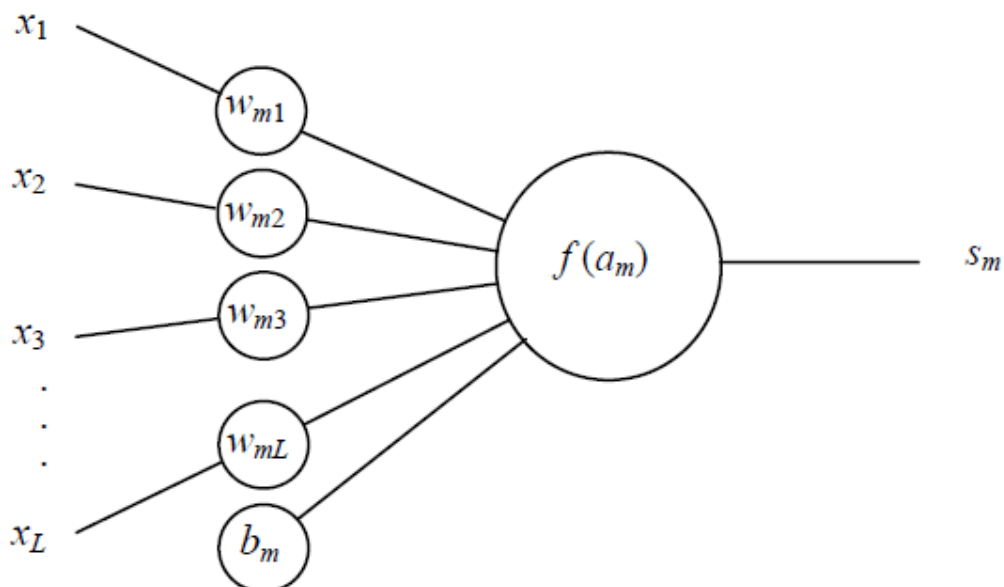


Figura 4.2 – Neurônio artificial (Cardoso et al., 2002).

Dados de entrada podem ser aqueles vindos de outros neurônios ou aqueles vindos do próprio vetor de variáveis aleatórias do problema que se está tratando. São análogos aos impulsos elétricos externos que são recebidos pelos dendritos do neurônio biológico, e costumam ser representados pelo conjunto $\{x_1, x_2, \dots, x_L\}$. Pesos que irão ponderar a conexão são representados pelo conjunto $\{w_{m1}, w_{m2}, \dots, w_{mL}\}$, e se relacionam às ponderações exercidas

pelos neurotransmissores das sinapses biológicas. A informação recebida é então processada, processo análogo ao ocorrido no corpo celular biológico, e é então transmitida a informação de saída.

Silva et al. (2010) mostram de forma didática que um neurônio artificial pode ser dividido em sete elementos básicos, a saber:

- Sinais de entrada: dados de entrada do problema, informação de saída de outros neurônios, valores assumidos pelas variáveis de alguma aplicação, entre outras possibilidades.
- Pesos sinápticos: valores que ponderam cada sinal de entrada, representando a relevância de cada um deles para cada situação específica.
- Combinador linear: a primeira etapa do processamento, os sinais de entrada ponderados pelos pesos sinápticos são submetidos a um somatório.
- Limiar de ativação (bias): representado na Figura 4.2 por b_m , trata-se de um termo do qual se subtrai o resultado do combinador linear, fornecendo um limiar para a ativação ou não daquele neurônio.
- Potencial de ativação: é o valor obtido através da subtração entre o limiar de ativação e o resultado do combinador linear.
- Função de ativação: é uma função característica do neurônio, que limita os valores de saída que o neurônio pode produzir a um intervalo razoável.
- Sinal de saída: Definido pelo potencial de ativação submetido à função de ativação, é o valor que o neurônio irá transmitir.

Sendo assim, o valor de saída de um neurônio m , a partir dos seus valores de entrada, pode ser expresso através da expressão (4.6).

$$y_m = f \left(b_m - \sum_{i=1}^L w_i x_i \right) \quad (4.6)$$

4.2.3. Funções de ativação

Existem diversos tipos de funções de ativação que podem ser utilizados nas redes neurais artificiais; sua escolha influenciando de forma importante na eficiência da rede. Elas podem ser divididas em funções parcialmente diferenciáveis e funções totalmente diferenciáveis. As duas funções de ativação mais utilizadas na prática (Gomes e Ludermir, 2008), a sigmoide logística e a tangente hiperbólica, são totalmente diferenciáveis.

Funções de ativação parcialmente diferenciáveis são, conforme o nome sugere, funções que não apresentam derivada de primeira ordem em ao menos um ponto de seu domínio. Estas funções são, em geral, bastante simples, e costumam ser utilizadas para problemas de classificação de padrões. Entre as mais utilizadas, estão a função degrau (*heavyside*), a função sinal e a função rampa simétrica (Silva et al., 2010).

As funções de ativação totalmente diferenciáveis possuem derivadas de primeira ordem em todos os pontos do seu domínio. Funções de ativação lineares e gaussianas se enquadram nesta categoria, e podem ser utilizadas em redes neurais. A função sigmoide logística produz sempre valores entre 0 e 1, e sua expressão matemática é dada pela expressão 4.7.

$$f(x) = \frac{1}{1 + e^{-\lambda x}} \quad (4.7)$$

na qual λ é um termo que influencia no nível de inclinação da curva logística. A Figura 4.3 mostra um gráfico desta função.

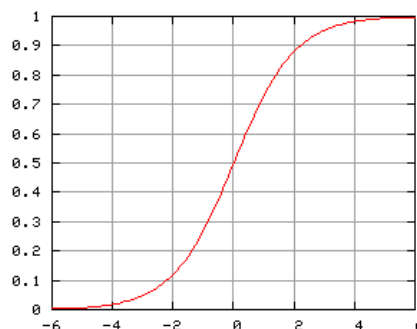


Figura 4.3 – Função de ativação Sigmóide Logística

A função tangente hiperbólica produz, para qualquer entrada, uma resposta que se situa entre os valores -1 e 1. Sua expressão matemática é expressa pela equação 4.8, e sua representação gráfica é apresentada na Figura 4.4. Novamente, o parâmetro λ define o grau de inclinação da curva.

$$f(x) = \frac{1 - e^{-\lambda x}}{1 + e^{-\lambda x}} \quad (4.8)$$

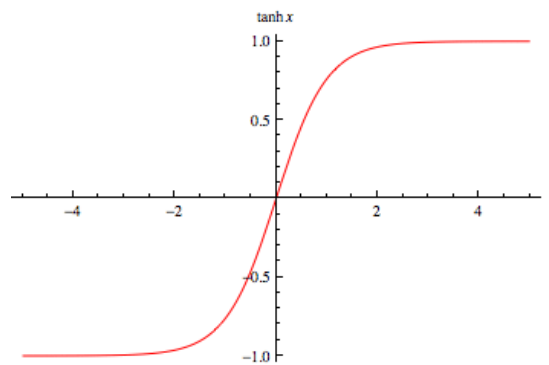


Figura 4.4 – Função de Ativação Tangente Hiperbólica

4.2.4. Arquiteturas

Para formar uma rede neural, um determinado número de neurônios (que pode inclusive ser 1) é disposto em um arranjo, determinando a partir de onde cada neurônio obterá seus dados de entrada, e para onde enviará sua informação de saída. Este arranjo define a arquitetura da rede neural. Já a topologia da rede é a composição que a rede assumirá dentro de uma arquitetura já definida (Silva et al., 2010). Sendo assim, questões como haver ou não retroalimentação dos dados da rede dizem respeito à sua arquitetura, enquanto a quantidade de neurônios em uma camada diz respeito à sua topologia.

Nas redes *feedforward*, a informação segue sempre o mesmo sentido, caminhando a partir da entrada da rede, passando de forma progressiva pelas camadas de neurônios intermediárias até chegar à camada de saída, a partir da qual a informação não volta mais para a rede. O exemplo mais simples é a rede *perceptron*. Idealizada por Frank Rosenblatt em 1957, esta rede é

composta de apenas uma camada neural, que contém apenas um neurônio. Sendo assim, a Figura 4.2 representa também uma rede perceptron, e a equação 4.9 representa seu funcionamento, caso a função de ativação representada seja uma função Degrau.

Redes feedforward de várias camadas podem representar problemas complexos. Foi provado por Hornik et al. (1990) que uma rede composta por apenas uma camada de entrada, uma camada escondida (*hidden-layer*) e uma camada de saída podem aproximar qualquer função derivável, desde que a curva gerada pela função de ativação dos neurônios seja suave. Este tipo de arquitetura tem sido usada para resolver diversos tipos de problemas de classificações de padrões, otimização de sistemas, aproximação de funções e confiabilidade estrutural. Uma representação gráfica de uma rede feedforward de várias camadas é mostrada na Figura 4.5.

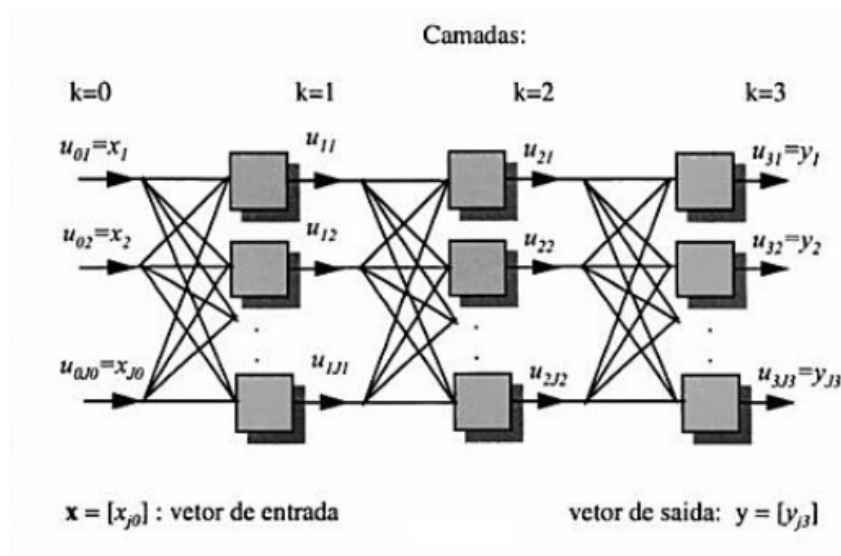


Figura 4.5 – Rede neural multicamada (Kovács, 2006)

É possível realimentar a rede neural com sua própria informação de saída, ou realimentar neurônios com a informação de saída de outros neurônios (ou deles mesmos). Este tipo de prática origina a chamada arquitetura recorrente. Este tipo de arquitetura é utilizada quando se tem sistemas dinâmicos, e vem sendo aplicada em problemas de previsão de séries temporais, controle de processos, entre outros (Silva et al., 2010).

4.2.5. Processos de treinamento

Redes neurais artificiais tem a capacidade de reconhecer padrões. A partir de alguns exemplos, a rede é capaz de representar o funcionamento de um sistema, e daí sua capacidade de ser utilizada como um meta-modelo. Como visto nos itens anteriores, as operações executadas no interior de uma rede são simples, portanto é mais rápido para um computador trabalhar com a RNA em comparação a trabalhar com um modelo numérico complexo.

Uma vez definidas a arquitetura e a topologia da RNA, é necessário ajustar seus pesos e limiares de ativação ao sistema que se está tentando representar. A este processo de ajuste dá-se o nome de Treinamento ou Processo de Aprendizado. Em geral, a informação disponível para treinar a rede é limitada e precisa ser dividida em dois conjuntos. O primeiro é o Conjunto de Treinamento, utilizado pelos algoritmos de aprendizagem para ajustar os parâmetros da rede. O segundo é denominado Conjunto de Teste, e serve para validar se os pesos e limiares de ativação obtidos através do treinamento estão oferecendo respostas suficientemente boas.

4.2.5.1. Paradigmas de Treinamento

Existem diversas estratégias ou paradigmas de treinamento nas quais pode-se agrupar diferentes algoritmos de aprendizagem. Três abordagens amplamente utilizadas são o treinamento supervisionado, não supervisionado e por reforço.

No treinamento supervisionado, a informação disponível é na forma de pares de valores de entrada e saídas correspondentes. A partir destes dados a rede poderá observar qual é o comportamento que se pretende representar. A cada teste, o valor obtido através do processamento da rede será comparado com o valor exato que ela deveria ter produzido caso representasse perfeitamente o modelo original. Desse modo, os pesos sinápticos e os limiares de ativação vão sendo corrigidos pelo algoritmo de aprendizagem, considerando a diferença entre a resposta desejada e a resposta obtida pela rede a cada iteração do algoritmo. Algum critério de parada precisa ser definido para o algoritmo, em geral, uma diferença pequena o suficiente entre a

resposta desejada e a obtida. Um número de iterações máximo limita o algoritmo, caso a precisão desejada não se verifique.

No paradigma de treinamento não supervisionado, os dados fornecidos para o treinamento da rede não incluem informações de saída correspondentes. Sendo assim, a própria rede precisa se organizar levando em consideração as particularidades do conjunto de amostras fornecido. Uma abordagem comum é a definição de uma função peso, levando em consideração tanto os dados de entrada quanto a resposta produzida a partir deles pela própria rede. O algoritmo de treinamento tenta então minimizar o valor da função peso, corrigindo os parâmetros da rede neste sentido a cada iteração.

Por fim, o paradigma de treinamento com reforço pode ser entendido como uma variação do treinamento supervisionado. Sutton e Barto (1998) abordam em profundidade este paradigma, que, em linhas gerais, consiste em ajustar os parâmetros da rede com base em respostas obtidas do sistema mapeado a partir de saídas geradas pela rede ou pelo algoritmo de treinamento. As respostas obtidas podem ser consideradas satisfatórias ou não, e assim, a rede é recompensada quando ajusta seus parâmetros da maneira desejada, o que a incentiva a continuar ajustando seus parâmetros de determinada maneira, ou mudar a direção dos ajustes.

4.2.5.2. Algoritmos de Treinamento

Nesta seção, descreve-se o funcionamento de alguns algoritmos de treinamento básicos. Estes algoritmos serviram de base para o desenvolvimento de outros, e são ainda utilizados atualmente em diversas aplicações. Outros métodos são ainda comentados no final da seção.

Regra de Hebb:

Proposta por Donald Hebb em 1949, esta regra é utilizada para treinar redes do tipo perceptron em problemas de classificação de padrão. Se a saída produzida pela rede coincidir com a saída correta (aquela correspondente à

entrada utilizada nos pares de dados fornecidos), os pesos e limiares da rede são incrementados – e são decrementados caso contrário. Este processo se repete até que todas as amostras tenham sido utilizadas, por um número de vezes definido por uma variável chamada *época*. Para programar o algoritmo, é conveniente reunir os pesos sinápticos do perceptron em um vetor. O limiar de ativação pode também ocupar uma posição deste vetor, pois o tratamento que ele recebe durante o treinamento é semelhante ao recebido pelos pesos. Desse modo, o vetor de pesos \mathbf{w} pode ser iniciado aleatoriamente. Define-se também o conjunto de treinamento, com amostras de entrada \mathbf{x} e saída \mathbf{d} . Define-se ainda o número de épocas e uma taxa de aprendizagem η , em geral entre 0 e 1. Repete-se então para cada época:

$$\begin{aligned} \mathbf{y} &= \text{sign}(\mathbf{w}^T \cdot \mathbf{x}) \\ \mathbf{w} &= \mathbf{w} + \eta(\mathbf{d} - \mathbf{y})\mathbf{x} \end{aligned} \quad (4.9)$$

Regra Delta:

A regra delta é utilizada desde a década de 1960 para o treinamento de redes do tipo Adaline, também constituídas de apenas um neurônio artificial, que eram aplicadas a problemas de chaveamento de circuitos telefônicos (Silva et al., 2010). Sendo também um treinamento assistido, esta regra baseia-se na minimização do erro quadrático entre o valor de saída da rede em treinamento e o valor exato que ela produziria se estivesse funcionando perfeitamente. Havendo um número p de amostras disponíveis para o treinamento, define-se um erro quadrático E :

$$E(\mathbf{w}) = \frac{1}{2} \sum_{i=1}^p [d^{(i)} - (\mathbf{w}^T \cdot \mathbf{x}^{(i)} - b)]^2 \quad (4.10)$$

sendo b o limiar de ativação, e $\mathbf{w}^T = [w_1, w_2, \dots, w_L]$. Aplica-se então o operador gradiente sobre o erro quadrático com relação aos pesos sinápticos:

$$\nabla E(\mathbf{w}) = \frac{\partial E(\mathbf{w})}{\partial \mathbf{w}} = - \sum_{i=1}^p (d^{(i)} - u) \cdot \mathbf{x}^{(i)} \quad (4.11)$$

sendo u o resultado do combinador linear do neurônio. Assim, a correção dos pesos é feita no sentido oposto ao do gradiente obtido, a fim de se minimizar o erro quadrático. Por esse motivo, este método também é conhecido como Método do Gradiente Descendente. Realiza-se então uma correção do valor dos pesos, novamente ponderada por uma taxa de aprendizagem η , e itera-se novamente até que se realizem todas as épocas ou algum critério de parada seja disparado.

$$\Delta \mathbf{w} = -\eta \cdot \nabla E(\mathbf{w}) \quad (4.12)$$

Backpropagation ou Regra Delta Generalizada

O método de Backpropagation é utilizado para treinar redes perceptron de muitas camadas. Este método consiste em duas fases distintas, chamadas de fase *forward* e fase *backward*. Na primeira, uma amostra do conjunto de treinamento é fornecida para a rede, e são computados os pesos e bias de cada neurônio. As respostas emitidas por cada neurônio da camada de saída são então comparadas às respostas desejadas para aquele conjunto de treinamento, e determina-se qual foi o erro obtido. Inicia-se então a segunda etapa, na qual as informações dos erros é *retropropagada*, ou seja, é utilizada para ir corrigindo os parâmetros dos neurônios desde a camada de saída, progressivamente, até a primeira camada. Este procedimento é análogo ao realizado na regra Delta, porém considera-se agora não apenas o erro apresentado na saída da rede, mas também as informações da camada adjacentes ao neurônio.

Destacam-se ainda o método de Levenberg–Marquardt, que faz também uma retropropagação dos desvios através da rede, porém considerando expressões de segunda ordem para aproximar o erro, e os algoritmos genéticos, que iteram através da criação de populações de dados a partir dos dados já utilizados (Ghaffari et al., 2006).

4.2.6. Inicialização dos pesos

O valor inicial que os pesos e bias da rede assumem tem um papel importante na taxa de convergência dos algoritmos de treinamento, pois representam o passo inicial das iterações. Uma abordagem comum e bastante simples é a definição aleatória dos pesos, desde que dentro de certos limites de razoabilidade. Sendo assim, os valores de todos os pesos são sorteados.

Diversas abordagens para uma melhor escolha dos pesos iniciais da rede já foram propostas, inclusive considerando aspectos do modelo a ser emulado a priori. Nguyen e Widrow (1990) propuseram uma distribuição sem conhecimento a priori, ainda aleatória, mas que permitiria que os algoritmos de treinamento operassem com bom desempenho. Basicamente, os pesos são inicializados com valores entre -0,5 e 0,5. Define-se então um parâmetro β da seguinte forma:

$$\beta = 0,7P\frac{1}{n} \quad (4.13)$$

sendo P o número de neurônios na camada intermediária, e n o número de neurônios na camada de entrada. O novo valor do peso w_1 é atualizado a partir de seu valor anterior w_0 , da seguinte forma:

$$v_1 = \beta \frac{w_0}{|w_0|} \quad (4.14)$$

sendo w_0 o vetor de pesos.

4.3. EXPANSÕES EM POLINÔMIOS DE CAOS (EPC)

Sistemas físicos em geral podem ser expressos através de uma relação entre um conjunto de parâmetros de entrada e uma resposta correspondente. Esta relação nem sempre é explícita, podendo não haver para ela forma analítica. Nestes casos uma representação numérica costuma se fazer necessária, como a utilização do método dos elementos finitos para representar o comportamento do sistema. Quando se leva em consideração as incertezas envolvidas nos fenômenos estudados, os parâmetros de entrada

costumam ser representados por variáveis aleatórias ou processos estocásticos. Tais incertezas se propagam através do sistema, de modo que sua resposta será também estocástica. É este tipo de resposta que se busca expressar através de uma expansão em polinômios de caos.

A ideia básica da EPC é considerar que a resposta do sistema possa ser mapeada através de suas coordenadas em um espaço gerado pela base de polinômios de caos. Para formarem tal base, os polinômios precisam respeitar algumas propriedades, conforme será visto a seguir. A resposta Y do sistema pode ser então expressa através da equação 4.15, na qual os Ψ são os polinômios multivariados que compõem a base, e os a são os coeficientes que mapeiam a resposta.

$$Y \approx \sum_{i=1}^{\infty} a_i \Psi_i \quad (4.15)$$

Ainda que esta expansão convirja para um número infinito de termos, por razões práticas, costuma-se truncar a série, após alguns termos. No desenvolvimento que se fará a seguir, as variáveis aleatórias apresentadas serão consideradas independentes, por uma questão de simplicidade, mas esta técnica pode ser também utilizada para casos em que as variáveis sejam dependentes.

4.3.1. Ortonormalidade de polinômios

Uma condição para que os polinômios façam parte da base procurada é que eles sejam ortonormais. Há ortogonalidade entre dois polinômios P_n e P_m com relação a alguma medida real positiva ζ se a igualdade expressa pela equação 4.16 for válida (Xiu, 2010).

$$\int_S P_n(x) P_m(x) d\zeta(x) = \gamma \delta_{nm} \quad (4.16)$$

onde S é o suporte da medida ζ , γ é um inteiro positivo denominado constante de normalização, e δ_{nm} é o delta de kroencker. Os polinômios são considerados também ortonormais se $\gamma = 1$. Se ζ possuir uma densidade, a expressão pode ser escrita como:

$$\int_S P_n(x)P_m(x)f(x)dx = \gamma\delta_{nm} \quad (4.17)$$

Para uma variável aleatória, esta densidade é a função densidade de probabilidade, e os polinômios pertencentes a uma base que possa representá-la devem respeitar a ortonormalidade com relação à sua FDP. Em geral, define-se um produto interno da forma:

$$\langle P_n, P_m \rangle_f = \int_S P_n(x)P_m(x)f(x)dx \quad (4.18)$$

e se expressa a relação de ortonormalidade por:

$$\langle P_n, P_m \rangle_f = \delta_{nm} \quad (4.19)$$

Uma facilidade que existe para se trabalhar com polinômios ortogonais de uma variável real é que todos necessariamente respeitarão alguma regra de recorrência. Desse modo, um polinômio de um grau posterior pode ser gerado a partir dos polinômios de graus anteriores. Sendo assim, pode-se gerar computacionalmente um polinômio de qualquer grau finito a partir de poucas linhas de código. A regra de recorrência geral pode ser expressa pela equação 4.20.

$$-xP_n(x) = b_nP_{n+1}(x) + a_nP_n(x) + c_nP_{n-1}(x) \quad n \geq 1 \quad (4.20)$$

sendo $b_n, c_n \neq 0$, $\frac{c_n}{b_{n-1}} > 0$, $P_{-1}(x) = 0$ e $P_0(x) = 1$. (Xiu, 2010).

4.3.2. Famílias clássicas de polinômios

Conforme visto nos itens anteriores, uma variável aleatória pode ser representada através de uma expansão em série, cujos termos são polinômios ortogonais entre si com relação à FDP desta variável. Sendo assim, faz-se necessário saber que polinômios podem ser utilizados para diferentes distribuições de probabilidade.

Diversas distribuições de probabilidade já foram associadas a famílias de polinômios clássicos que respeitam a condição de ortogonalidade. É o caso

dos polinômios de Laguerre com relação à distribuição exponencial, e os polinômios de Jacobi com relação à distribuição Beta. Ainda assim, os polinômios mais utilizados para a expansão de sistemas em problemas de confiabilidade estrutural são aqueles que dizem respeito à distribuição uniforme e à distribuição normal (Sudret 2014).

4.3.2.1. Polinômios de Legendre

Os polinômios de Legendre estão definidos no domínio $[-1,1]$, e se associam em ortogonalidade a variáveis com distribuição uniforme sobre o mesmo domínio. Sendo assim, são ortogonais à FDP:

$$f(x) = \frac{1}{2} \quad (4.21)$$

A relação de recorrência particular que permite gerar polinômios de Legendre de um grau qualquer, pode ser expressa por:

$$(n + 1)P_{n+1}(x) = (2n + 1)xP_n(x) - nP_{n-1}(x) \quad n \geq 1 \quad (4.22)$$

A aplicação do produto interno definido em (4.18) de um polinômio de Legendre sobre ele mesmo não resulta na unidade, mas na seguinte norma:

$$\langle P_n, P_n \rangle = \frac{1}{2n + 1} \quad (4.23)$$

Desta forma, a ortonormalidade é obtida fazendo-se a seguinte correção:

$$P'_n(x) = \sqrt{(2n + 1)}P_n(x) \quad (4.24)$$

Os seis primeiros polinômios de Legendre estão reunidos na tabela 4.1, bem como sua correspondente forma corrigida:

Tabela 4.1: Polinômios de Legendre

n	$P_n(x)$	$P'_n(x)$
0	1	1
1	x	$\sqrt{3}P_1$
2	$(3x^2 - 1)/2$	$\sqrt{5}P_2$
3	$(5x^3 - 3x)/2$	$\sqrt{7}P_3$
4	$(5x^4 - 30x^2 + 3)/8$	$\sqrt{9}P_4$
5	$(63x^5 - 70x^3 + 15x)/8$	$\sqrt{11}P_5$

4.3.2.2. Polinômios de Hermite

A família de polinômios de Hermite probabilísticos está relacionada à variáveis com distribuição normal padrão, isto é, $\mathcal{N} \sim (0,1)$, sendo ortogonal à sua FDP. É também usual na literatura se referir aos polinômios de Hermite físicos, que são ortogonais não a $\frac{e^{-x^2}}{\sqrt{2\pi}}$, mas apenas a e^{-x^2} . Apesar das semelhanças, serão abordados aqui os primeiros, mais comuns em problemas de quantificação de incertezas e confiabilidade estrutural. A obtenção destes polinômios através da formulação recorrente pode ser obtida através da expressão 4.25.

$$He_{n+1}(x) = xHe_n(x) - nHe_{n-1}(x) \quad (4.25)$$

Também neste caso, a aplicação do produto interno (4.18) demonstra que há ortogonalidade, porém não ortonormalidade:

$$\langle He_n, He_n \rangle = n! \quad (4.26)$$

Uma correção simples é requerida para obtenção da ortonormalidade:

$$He'_n = \frac{He_n}{\sqrt{n!}} \quad (4.27)$$

A Tabela 4.2 reúne os polinômios de Hermite de grau $n = 0$ até 5:

Tabela 4.2: Polinômios de Hermite

n	$He_n(x)$	$He'_n(x)$
0	1	1
1	x	He_1
2	$x^2 - 1$	$He_2/\sqrt{2}$
3	$x^3 - 3x$	$He_3/\sqrt{6}$
4	$x^4 - 6x^2 + 3$	$He_4/\sqrt{24}$
5	$x^5 - 10x^3 + 15x$	$He_5/\sqrt{120}$

4.3.3. Transformação isoprobabilística

É comum em problemas de confiabilidade estrutural que os parâmetros de entrada do sistema sejam representados por variáveis aleatórias ou processos estocásticos de diferentes distribuições de probabilidade. A área de uma seção transversal, por exemplo, não pode assumir valores negativos, enquanto cargas ambientais costumam estar associadas a distribuições de extremos. Esta característica dificulta a atribuição de polinômios ortogonais à FDP de cada variável, e o problema se agrava quando se quer não apenas resolver um problema específico, mas programar um software que resolva diferentes problemas através de EPC.

Uma transformação isoprobabilística pode ser feita de modo a contornar esta dificuldade, sem que a precisão dos resultados seja prejudicada. Trata-se de um mapeamento simples que leva do espaço das variáveis originais do problema a um espaço mais conveniente para se trabalhar as expansões. Pode-se, por exemplo, associar todos os parâmetros de entrada a variáveis normais padrão, que assumem seus respectivos valores para o mesmo conteúdo de probabilidade assumido pela variável original. Deste modo, pode-se alimentar a EPC com as variáveis normalizadas, e se utilizar polinômios de Hermite para mapear a resposta desejada. Uma forma simples de fazer isso é submeter o valor das variáveis a serem transformadas à inversa de sua FCP, e este resultado então ser utilizado para alimentar a FCP normal padrão.

Esta abordagem permite ainda trabalhar com parâmetros de entrada correlacionados, bastando-se utilizar o modelo de Nataf para fazer a transformação, e utilizar as variáveis não correlacionadas resultantes no modelo de EPC (Blatman e Sudret, 2010).

4.3.4. Polinômios multivariados

Os polinômios abordados até agora formam bases para a representação de variáveis aleatórias simples, associadas a funções de probabilidade unidimensionais. A resposta Y do sistema que se quer simular, porém, em geral depende de diversas variáveis, e por isso, os polinômios estudados até agora não são capazes de representá-la sozinhos.

Polinômios de caos multidimensionais podem ser obtidos através da tensorização dos polinômios unidimensionais que expandem as variáveis de entrada do problema, da seguinte forma:

$$\Psi_{\alpha}(x) = \Psi_{\alpha_1, \alpha_2, \dots, \alpha_M} = P_{\alpha_1}^1(X_1) \dots P_{\alpha_M}^M(X_M) \quad (4.28)$$

Nesta notação, o vetor α é um multi-índice que reúne os graus de todos os polinômios unidimensionais que foram utilizados na multiplicação que compõem o polinômio multivariado $\Psi_{\alpha}(x)$. Assim, para um problema de três variáveis aleatórias X_1, X_2, X_3 , o polinômio multidimensional Ψ_{201} seria composto pela multiplicação do polinômio unidimensional de primeiro grau associado à variável X_1 pelo polinômio de grau zero associado à variável X_2 e pelo polinômio de segundo grau associado à variável X_3 .

Foi demonstrado por Soize e Ghanem (2004) que o conjunto dos polinômios multivariados com índices formados por todas as combinações possíveis de α é uma base adequada para representar a resposta Y . Ressalta-se que α tem dimensão igual ao número de variáveis do problema, e que o índice em cada uma de suas posições varia entre 0 e o grau máximo de cada polinômio individual, definido pelo número de termos em que se decidiu truncar a expansão para cada variável.

A série que expande a resposta do sistema também precisa ser truncada. Em geral, define-se um grau máximo para a expansão, e são utilizados para ela apenas os polinômios multidimensionais com *grau multi-índice* menor ou igual a ele. O grau multi-índice é obtido somando-se os graus de todos os polinômios que contribuíram na tensorização para formar o respectivo polinômio multidimensional, ou seja:

$$g^o = \sum_{i=1}^M \alpha_i \quad (4.29)$$

Desse modo, uma base para um problema com M variáveis e grau multi-índice máximo p é formada por P termos, sendo:

$$P = \frac{(M + p)!}{M! p!} \quad (4.30)$$

Para deixar mais claro, adapta-se aqui um exemplo a partir de Sudret (2011):

Seja um problema cuja resposta Y é função de duas variáveis de entrada: $X_1 \sim \mathcal{N}(0,1)$, $X_2 \sim \mathcal{N}(0,1)$. Como as variáveis são normais, os polinômios mais adequados para realizar a expansão são os de Hermite. Suponha-se ainda que o grau máximo definido para a expansão seja 3. Sendo assim, a expansão ficaria:

$$\begin{aligned} Y = & a_0 + a_1 X_1 + a_2 X_2 + a_3 \frac{X_1^2 - 1}{\sqrt{2}} + a_4 X_1 X_2 + a_5 \frac{X_2^2 - 1}{\sqrt{2}} + \\ & + a_6 \frac{X_1^3 - 3X_1}{\sqrt{6}} + a_7 \frac{(X_1^2 - 1)X_2}{\sqrt{2}} + a_8 \frac{(X_2^2 - 1)X_1}{\sqrt{2}} + a_9 \frac{X_2^3 - 3X_2}{\sqrt{6}} \end{aligned} \quad (4.31)$$

Na expressão, foram utilizados todos os termos de grau multi-índice menores ou iguais a 3. A Tabela 4.3 reúne os índices e os respectivos polinômios a eles associados no exemplo:

Tabela 4.3 – Polinômios e graus multi-índice

α	Ψ_α	grau multi-índice
[0,0]	1	0
[1,0]	X_1	1
[0,1]	X_2	1
[2,0]	$\frac{X_1^2 - 1}{\sqrt{2}}$	2
[1,1]	X_1X_2	2
[0,2]	$\frac{X_2^2 - 1}{\sqrt{2}}$	2
[3,0]	$\frac{X_1^3 - 3X_1}{\sqrt{6}}$	3
[2,1]	$\frac{(X_1^2 - 1)X_2}{\sqrt{2}}$	3
[1,2]	$\frac{(X_2^2 - 1)X_1}{\sqrt{2}}$	3
[0,3]	$\frac{X_2^3 - 3X_2}{\sqrt{6}}$	3

Ainda que no exemplo as duas variáveis do problema sigam a mesma distribuição de probabilidade, distribuições diferentes poderiam ser consideradas. Polinômios unidimensionais de famílias diferentes seriam então utilizados e multiplicados para compor a base de polinômios multidimensionais, desde que cada um fosse coerente à FDP da variável unidimensional que está representando.

4.3.5. Algoritmo para geração dos índices e construção da base

Sudret et al. (2006) propõe um algoritmo para gerar todas as combinações de índices que α deve assumir para a construção da base. Consiste em uma analogia com o problema de distribuir $(M - 1)$ bolas em $(M + q - 1)$ caixas, sendo q o grau múlti índice máximo para a base e M a

dimensão do problema (leia-se, o número de variáveis aleatórias que são parâmetros de entrada). O algoritmo descreve como as bolas devem ser trocadas a partir de uma posição inicial, de modo que todas as combinações possíveis de posições sejam contempladas. Faz-se então, uma correspondência entre as posições das bolas nas caixas e os índices para cada posição de α , da seguinte forma: da esquerda para a direita, para cada inteiro α_i , pula-se α_i caixas vazias e coloca-se uma bola na próxima caixa depois delas. Ou seja, para cada bola encontrada, o número α_i a ser considerado é o número de espaços entre duas bolas consecutivas. Este algoritmo consiste nas seguintes etapas:

- 1- Coloca-se todas as bolas nas $(M - 1)$ primeiras caixas.
- 2- Caso seja possível, move-se a bola que esteja mais à direita uma posição à direita. Caso não seja, identifica-se qual é a bola mais à direita que pode ser movida uma posição para a direita, e move-se ela uma posição para a direita. Então, todas as bolas à direita dela são movidas uma posição para a esquerda.
- 3- Repete-se a etapa 2 até que todas as bolas estejam ocupando as $(M - 1)$ últimas caixas.

A Figura 4.6 ilustra o funcionamento do algoritmo, etapa por etapa, para o caso de $M = 4$ e $p = 2$.

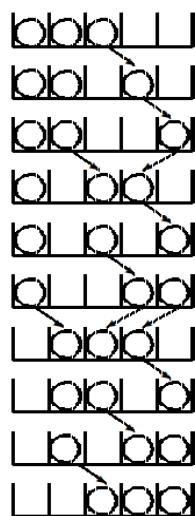


Figura 4.6 – Geração recursiva da base da EPC. Fonte: (Sudret et al 2006)

4.3.6. Determinação dos coeficientes

Uma vez definidos os termos que compõem a base, a próxima etapa consiste na determinação dos coeficientes, reunidos pelo vetor α na equação (4.15). Classicamente, esta determinação era feita de forma *intrusiva*, em problemas utilizando o método dos elementos finitos estocásticos. Uma formulação baseada na minimização das equações de equilíbrio, a partir da abordagem de Galerkin, levava à solução de um sistema de equações lineares acopladas à formulação MEFE, cujas dimensões dependiam do número de graus de liberdade do modelo, muitas vezes bastante alto (Sudret, 2008). Esta abordagem vem perdendo espaço nos últimos anos para a abordagem *não intrusiva*, na qual os coeficientes são determinados de forma determinística, utilizando informações obtidas a partir de um número finito de soluções do modelo original a ser aproximado. Uma vez que estes dados sejam obtidos, os coeficientes são calculados unicamente a partir deles, sem utilização subsequente do modelo original. Esta segunda abordagem é mais interessante para aplicações de meta-modelagem, pois permite que o sistema original seja utilizado na forma de uma “caixa preta”, ao longo de toda a resolução do problema considerado.

Entre os métodos não intrusivos, duas classes principais podem ser identificadas, a saber, métodos de regressão e métodos de projeção.

4.3.6.1. Método da Projeção

Os métodos de projeção se utilizam da propriedade de ortogonalidade dos polinômios envolvidos na base da EPC para determinar os coeficientes de forma direta. Basicamente, multiplica-se os dois lados da expressão (4.15) por Ψ_α e toma-se seu valor esperado. O resultado é o vetor de coeficientes α .

$$\alpha = E[Y(\mathbf{X})\Psi_\alpha(\mathbf{X})] = \int_{\mathcal{D}_x} Y(\mathbf{X})\Psi_\alpha(\mathbf{X}) f_x(\mathbf{X}) d\mathbf{X} \quad (4.32)$$

Em geral esta integral não é resolvida analiticamente, e o nome do método de projeção que se está considerando varia em função do modo como ela é resolvida. Em geral, a integral é substituída por um somatório finito de N pontos ponderados por uma função peso, na forma:

$$\mathbf{a} \approx \sum_{i=1}^N w_i Y(\mathbf{X}^{(i)}) \Psi_{\mathbf{a}}(\mathbf{X}^{(i)}) \quad (4.33)$$

SMC e integração por quadratura de Gauss são técnicas comumente utilizadas para a obtenção de \mathbf{a} .

4.3.6.2. Método da Regressão

É possível tratar o problema da determinação dos coeficientes da EPC como um problema de minimização do erro entre a aproximação e as respostas exatas obtidas em alguns pontos. Este processo é realizado analogamente ao método dos mínimos quadrados utilizado nas superfícies de resposta polinomiais. Sejam $\mathbf{a} = [a_0, a_1, \dots, a_{p-1}]^T$ e $\Psi(\mathbf{X}) = [\Psi_0(\mathbf{X}), \Psi_1(\mathbf{X}), \dots, \Psi_2(\mathbf{X})]^T$, os índices dos Ψ identificando os membros da base que são acompanhados pelo coeficiente de mesmo índice na expansão. Pode-se escrever:

$$Y^{PC}(\mathbf{X}) = \mathbf{a}^T \Psi(\mathbf{X}) \quad (4.34)$$

Resolve-se então o problema um número finito de vezes, a partir de um conjunto de q valores arbitrários para as variáveis de entrada $\mathbf{X} = [\mathbf{X}^{(1)}, \mathbf{X}^{(2)}, \dots, \mathbf{X}^{(q)}]$, associados às respostas $\mathbf{y} = [Y^{(1)}, Y^{(2)}, \dots, Y^{(q)}]$. Procura-se então o vetor de coeficientes que minimize o erro:

$$\hat{\mathbf{a}} = \arg \min \sum_{i=1}^k [\mathbf{a}^T \Psi(\mathbf{X}^{(i)}) - Y^i]^2 \quad (4.35)$$

Um problema cuja solução é bem conhecida, e análoga à expressão 4.5. A solução é mostrada pela equação 4.36.

$$\hat{\mathbf{a}} = (\Psi^T \Psi)^{-1} \Psi^T \mathbf{y} \quad (4.36)$$

sendo Ψ uma matriz cujos termos são expressos por: $\Psi_{ij} = \Psi_j(\mathbf{X}^{(i)})$. Para esta solução, o número mínimo para q , ou seja, o número mínimo de vezes que o modelo original precisa ser resolvido é P , expresso pela equação (4.30). Em geral, a adoção de uma amostragem por Hiper Cubo Latino e $q = 2P$ leva a bons resultados (Sudret, 2011).

4.3.7. Pós-processamento dos coeficientes

Estando definida a base e determinados os coeficientes, a aproximação pode ser substituída diretamente nas equações de estado limite dos problemas de confiabilidade estrutural estudados. Isso significa que a equação de estado limite original está sendo substituída por uma superfície de resposta estocástica. Os momentos de primeira e segunda ordem da variável aleatória Y podem ser facilmente determinados a partir dos coeficientes α , devido à sua propriedade de ortonormalidade, conforme mostram as expressões 4.37 e 4.38.

$$E[Y] = a_0 \quad (4.37)$$

$$Var[Y] = \sum_{i=1}^{P-1} a_i^2 \quad (4.38)$$

4.4. KRIGAGEM

A Modelagem por Processos Gaussianos, ou Krigagem, é uma poderosa técnica de meta-modelagem, à qual tem-se dado crescente atenção nas últimas duas décadas (Echard et. al. 2013). Classicamente utilizada na geoestatística, sua ideia básica consiste em representar uma função a partir do conhecimento do seu valor em alguns poucos pontos, interpretados como realizações de um processo estocástico gaussiano. Os espaços que serviam como domínio para as funções modeladas eram, em geral, sistemas físicos de duas ou três dimensões espaciais. As aplicações da krigagem tornaram-se mais abrangentes a partir do trabalho de Sacks et al. (1989), que expandiu a ideia para o uso da técnica em dados obtidos a partir de um modelo computacional, ao invés de medidas realizadas em campo. Ainda, os pontos da função nos quais as informações são conhecidas são não mais coordenadas físicas espaciais, mas um espaço abstrato de dimensão M (Sudret, 2011).

4.4.1. Regressão

Os meta-modelos obtidos com Krigagem se inserem no contexto das variáveis regionalizadas, portanto sua variação é expressa por uma soma de três termos, representando uma parcela determinística, uma estocástica e um erro. (Burrough, 1987). Utilizando-se a notação de processos estocásticos, mais comum na literatura sobre confiabilidade estrutural, é possível escrever uma equação de estado limite $g(\mathbf{X})$ que se queira aproximar conforme mostrado na equação 4.39.

$$g(\mathbf{X}) = \mathfrak{F}(\boldsymbol{\beta}, \mathbf{X}) + z(\mathbf{X}) \quad (4.39)$$

O primeiro termo do lado direito desta equação é uma parcela determinística obtida a partir de um modelo de regressão, muito similar às superfícies de resposta polinomiais estudadas no item 4.1. Este termo fornece uma média para a aproximação de $g(\mathbf{X})$. Basicamente, uma vez os coeficientes reunidos pelo vetor $\boldsymbol{\beta}$ definidos, obtém-se este termo através da equação 4.40.

$$\mathfrak{F}(\boldsymbol{\beta}, \mathbf{X}) = f_1(\mathbf{X})\beta_1 + \dots + f_{ne}(\mathbf{X})\beta_{ne} = \mathbf{f}(\mathbf{x})^T \boldsymbol{\beta} \quad (4.40)$$

Um vetor de funções linearmente independentes $\mathbf{f}(\mathbf{x})$ é aqui utilizado, análogo ao vetor $\tilde{\mathbf{x}}$ da equação 4.3. O segundo termo do lado direito da equação 4.39 representa um processo estocástico gaussiano estacionário (ou um campo aleatório homogêneo, para \mathbf{X} multidimensional). Sua média é 0, e sua covariância é fornecida pela equação 4.41. Este termo corrige a aproximação $\mathfrak{F}(\boldsymbol{\beta}, \mathbf{X})$, interpolando os erros nos pontos amostrais. Sendo assim, um meta-modelo de Krigagem apresenta, quando avaliado sobre um ponto amostral utilizado para a sua construção, o valor exato que seria fornecido pelo modelo original (Gaspar et al., 2014).

$$Cov[z(\mathbf{X}), z(\mathbf{X}')] = \sigma_z^2 \mathfrak{R}(\boldsymbol{\theta}; \mathbf{X}, \mathbf{X}') \quad (4.41)$$

Aqui, \mathfrak{R} é uma função de correlação arbitrária, conforme discutido no item 4.4.2, e $\boldsymbol{\theta}$ é um hiperparâmetro relacionado à correlação espacial entre dois pontos, cuja determinação é discutida no item 4.4.3. Sendo assim, um modelo de Krigagem é completamente definido através da variância do

processo estocástico σ_z^2 , do vetor $\boldsymbol{\beta}$ e de um vetor de parâmetros de correlação $\boldsymbol{\theta}$. Para os dois primeiros parâmetros, seleciona-se um conjunto de pontos de suporte (por exemplo, através de planos de experiência ou utilizando um hipercubo latino), e ajusta-se os parâmetros através do método dos mínimos quadrados. A determinação dos coeficientes $\boldsymbol{\beta}$ se assemelha bastante àquela utilizada para determinar os coeficientes das superfícies de resposta polinomiais, representada pela equação 4.5. A diferença principal é que, neste caso, considera-se na regressão a correlação espacial entre os pontos do modelo, conforme mostrado pela equação 4.42.

$$\boldsymbol{\beta} = (\mathbf{F}^T \mathbf{R}^{-1} \mathbf{F})^{-1} \mathbf{F}^T \mathbf{R}^{-1} \mathbf{y} \quad (4.42)$$

A variância do processo é fornecida pela equação 4.43.

$$\sigma_z^2 = \frac{1}{np} (\mathbf{y} - \mathbf{F}\boldsymbol{\beta})^T \mathbf{R}^{-1} (\mathbf{y} - \mathbf{F}\boldsymbol{\beta}) \quad (4.43)$$

\mathbf{F} é a matriz de regressão, montada da mesma forma que a matriz \mathbf{Q} da equação 4.5. np é o número de pontos de suporte utilizados para calibrar o modelo e \mathbf{R} é uma matriz que define a correlação espacial entre cada par, dentre os pontos de suporte, segundo uma função de auto-correlação. Diferentes funções de auto-correlação são apresentadas na Seção 4.4.2. Isso considerado, a matriz \mathbf{R} fica:

$$\mathbf{R} = \begin{bmatrix} R(\mathbf{x}_1, \mathbf{x}_1) & \cdots & R(\mathbf{x}_1, \mathbf{x}_m) \\ \vdots & \ddots & \vdots \\ R(\mathbf{x}_m, \mathbf{x}_1) & \cdots & R(\mathbf{x}_m, \mathbf{x}_m) \end{bmatrix} \quad (4.44)$$

Como ambos $\boldsymbol{\beta}$ e σ_z^2 dependem ainda de $\boldsymbol{\theta}$, este parâmetro precisa ser conhecido.

4.4.2. Funções de correlação

O campo aleatório descrito na equação 4.39 tem como característica uma função de correlação (ou função de auto-correlação) $R(\mathbf{x}_i, \mathbf{x}_j)$. Como para a Krigagem é requisito que tal campo aleatório seja homogêneo, a correlação entre os pontos depende apenas da distância entre eles, ou seja:

$$R(\mathbf{x}_i, \mathbf{x}_j) = R(|\mathbf{x}_i - \mathbf{x}_j|) = R(\tau) \quad (4.45)$$

$R(\mathbf{x}_i, \mathbf{x}_j)$ tem um papel importante na qualidade da aproximação obtida, e influencia consideravelmente na suavidade do meta-modelo resultante (Kaymaz, 2005). Dentre todos os parâmetros que compõem um meta-modelo de krigagem, nota-se que apenas a função de auto-correlação depende de \mathbf{x} . Sendo assim, a escolha da de $R(\mathbf{x}_i, \mathbf{x}_j)$ irá definir também a diferenciabilidade do meta-modelo (Xie et al., 2010).

De forma geral, $R(\mathbf{x}_i, \mathbf{x}_j)$ terá um valor maior para pontos mais próximos entre si no espaço das variáveis aleatórias, e irá decaindo conforme \mathbf{x}_i e \mathbf{x}_j se distanciam. Este comportamento representa um importante princípio para a krigagem: o de que pontos mais próximos estão mais fortemente correlacionados entre si. O enunciado deste princípio costuma ser atribuído ao cartógrafo americano Waldo Tobler, que enunciou em seu trabalho (Tobler, 1970): “Todas as coisas estão relacionadas entre si, no entanto, as coisas mais próximas estão mais relacionadas que as distantes”. Este enunciado ficou posteriormente conhecido como a “primeira lei da geografia de Tobler”.

Existem muitas opções que podem ser utilizadas como funções de auto-correlação, sua forma exata podendo ser escolhida pelo usuário do modelo. A opção mais comumente utilizada é a função gaussiana (4.48). A tabela 4.4 mostra algumas opções de funções de autocorrelação:

Tabela 4.4 – Funções de auto-correlação

Nome	Função	#
Linear	$\max\{0, 1 - \theta_j d_j \}$	4.46
Exponencial	$e^{-\theta_j d_j }$	4.47
Gaussiana	$e^{-\theta_j d_j^2 }$	4.48
Esférica	$1 - 1,5\xi_{ij}^k + 0,5\xi_{ij}^{k^3} / \xi_{ij}^k = \min\{1, \theta_k d_{ij}^k \}$	4.49

Em um problema multidimensional, a abordagem mais comum - e a adotada neste trabalho - é fazer o somatório dos expoentes de cada dimensão

da função gaussiana, de modo que, em geral, a função que se utiliza em problemas resolvidos por krigagem é da forma:

$$R(\mathbf{x}_i, \mathbf{x}_j) = \exp \left[- \sum_{k=1}^M \theta_k |\mathbf{x}_i^k - \mathbf{x}_j^k|^2 \right] \quad (4.50)$$

As funções de autocorrelação são também chamadas de função núcleo (*kernel functions*), e podem ser compostas para formar novos núcleos. Tanto a soma quanto o produto de funções de autocorrelação resultam em funções que também são funções de autocorrelação.

4.4.3. Determinação dos hiperparâmetros

Se uma variável funcionar como parâmetro para um modelo, porém precisar ser definida a priori, ela é um hiperparâmetro. No caso da krigagem, por exemplo, $\boldsymbol{\beta}$ e σ_Z^2 são parâmetros que influenciam o comportamento do meta-modelo, e precisam ser definidos. Ambos dependem, porém, do vetor de parâmetros de distância da correlação $\boldsymbol{\theta}$, de modo que este precisa ser definido primeiro, para que então os demais parâmetros sejam escolhidos. Sendo assim, as componentes de $\boldsymbol{\theta}$ são hiperparâmetros.

Desse modo, $\boldsymbol{\theta}$ deve ser estimado primeiro, para então as equações 4.42 e 4.43 se tornarem aplicáveis. Uma possível abordagem para esta definição é através de tentativas criteriosas, baseadas em valores adotados em problemas parecidos, estudando-se os resultados obtidos com cada escolha. Cada iteração é então escolhida na direção de minimizar o erro. Este método implica em já haver uma referência para o resultado do problema estudado, e necessita de várias tentativas, o que na prática reduz sua aplicabilidade. Uma abordagem mais eficiente é a utilização de um *método da máxima possibilidade*. Este método consiste em avaliar os parâmetros de um sistema a partir da observação de uma amostra apenas. A ideia é procurar valores para tais parâmetros que produziram a amostra que foi observada. Define-se então uma função de probabilidades envolvendo os parâmetros e as amostras, e busca-se seus valores extremos, o que produz um sistema cuja solução leva a valores para tais parâmetros. Para o problema de krigagem abordado, esta técnica se trata da minimização de $\mathcal{L}(\boldsymbol{\theta})$ (Kaymaz, 2005):

$$\hat{\boldsymbol{\theta}} = \operatorname{argmin}\{\mathcal{L}(\boldsymbol{\theta})\} \quad (4.51)$$

sendo:

$$\mathcal{L}(\boldsymbol{\theta}) = \frac{np \ln(\sigma_z^2) + \ln(\det \mathbf{R})}{2} \quad (4.52)$$

Note que a minimização depende apenas dos parâmetros do modelo e dos pontos de suporte selecionados.

4.4.4. Valor esperado e incerteza do modelo de Krigagem

Uma vez os hiperparâmetros $\boldsymbol{\theta}$ obtidos, são definidos os parâmetros $\boldsymbol{\beta}$ e σ_z^2 através das equações 4.42 e 4.43. É possível então prever um valor futuro do processo, ou se estimar o valor da resposta que seria dada pelo modelo original em qualquer ponto do espaço das variáveis aleatórias. Assim, pode-se aproximar uma equação de estado limite utilizando a expressão 4.53.

$$g(\mathbf{x}) = \mathbf{f}(\mathbf{x})^T \boldsymbol{\beta} + \mathbf{r}(\mathbf{x})^T \boldsymbol{\omega} \quad (4.53)$$

Nesta expressão, tem-se que $\mathbf{r}(\mathbf{x})$ é um vetor que reúne as correlações entre o ponto no qual se está fazendo a previsão e cada um dos pontos que foram utilizados como suporte na construção do modelo de krigagem. Ou seja:

$$\mathbf{r}(\mathbf{x}) = [R(\mathbf{x}, \mathbf{x}_1), R(\mathbf{x}, \mathbf{x}_2), \dots, R(\mathbf{x}, \mathbf{x}_m)] \quad (4.54)$$

e

$$\boldsymbol{\omega} = \mathbf{R}^{-1}(\mathbf{Y} - \mathbf{F}\boldsymbol{\beta}) \quad (4.55)$$

Uma vantagem da Krigagem é a possibilidade de se determinar facilmente a variância que existe na aproximação, para um dado ponto estimado, ou seja, uma medida da incerteza da previsão. Esta característica possibilita que o modelo seja refinado, caso a incerteza esteja maior do que a desejada, e modelos adaptativos podem ser desenvolvidos. O trabalho de (Bect, 2011) explora diversos destes modelos. A incerteza sempre será zero quando a previsão for feita sobre um ponto de suporte. Esta medida é expressa pela equação 4.56.

$$\sigma_g^2(\mathbf{X}) = \sigma_Z^2 \left[1 + \frac{(\mathbf{F}^T \mathbf{R}^{-1} \mathbf{r}(\mathbf{X}) - \mathbf{f}(\mathbf{X}))^T (\mathbf{F}^T \mathbf{R}^{-1} \mathbf{F})^{-1} (\mathbf{F}^T \mathbf{R}^{-1} \mathbf{r}(\mathbf{X}) - \mathbf{f}(\mathbf{X}))}{\mathbf{r}(\mathbf{X})^T \mathbf{R}^{-1} \mathbf{r}(\mathbf{X})} \right] \quad (4.56)$$

Desse modo, os valores de $\sigma_g^2(\mathbf{X})$ podem ser plotados em um gráfico, mostrando os diversos pontos sobre os quais se fez a aproximação, e regiões de maior e menor incerteza podem ser identificadas.

5. RESULTADOS

5.1. META-MODELAGEM NO STRAND

As três técnicas de meta-modelagem abordadas neste trabalho foram implementadas no programa StRAnD, segundo o padrão de desenvolvimento da orientação a objeto. Desse modo, módulos correspondentes às técnicas foram adicionados e compatibilizados com as funcionalidades já presentes no software. Os meta-modelos podem ser utilizados em combinação com diferentes técnicas de amostragem inteligente em SMC, estratégia que permite uma redução ainda maior dos custos computacionais da análise de confiabilidade.

O programa conta com redes neurais do tipo multi-layer perceptron, sendo possível o treinamento através do algoritmo de backpropagation e de Levenberg–Marquardt. Em ambos os casos, o número de épocas de treinamento pode ser definido pelo usuário, assim como o número de pontos de treinamento e de verificação, e a taxa de aprendizado da rede. Todas as redes geradas no StRAnD contam com três camadas. O número de neurônios da primeira corresponde automaticamente ao número de parâmetros de entrada da rede, e da última, ao de parâmetros de saída. O número de neurônios na camada escondida pode ser escolhido pelo usuário, ou é definida como padrão pelo teorema Kolmogorov–Nielsen. É possível ainda escolher entre dois algoritmos de inicialização de pesos, aleatório ou Nguyen e Widrow. As funções de ativação das camadas de interface são lineares, e funções tangentes sigmóides estão programadas para a camada intermediária.

Para os meta-modelos com EPC basta o usuário definir o grau da expansão desejada e o número de pontos de treinamento. O StRAnD faz a transformação isoprobabilística de todas as variáveis para normais padrão e gera os índices que compõem a base da expansão conforme detalhado na seção 4.3.5. Em seguida, a base da expansão é montada através de um algoritmo que gera os polinômios de Hermite. Para cada posição do multi-índice α , é feita a geração recursiva do polinômio unidimensional correspondente, e o resultado de cada posição é armazenado e multiplicado pelo resultado da próxima posição,

gerando-se assim o polinômio multivariado. A determinação dos coeficientes da expansão é feita de forma não intrusiva através do método da regressão, e seus valores são armazenados, podendo ser utilizados no pós-processamento.

Na Krigagem, o grau da superfície de resposta que corresponde à média do processo estocástico é definido pelo usuário, bem como o tamanho do conjunto de treinamento do modelo. A matriz de correlação é então gerada automaticamente a partir da função gaussiana, com a estratégia descrita pela equação 4.50. Na versão atual, os valores dos hiperparâmetros ainda precisam ser fornecidos pelo usuário. A interpolação é realizada então para cada simulação, sendo definido a cada vez um novo vetor de correlação. A variância da aproximação pode ser obtida em qualquer ponto do domínio do meta-modelo.

Em todos os casos, simulações de Monte Carlo são realizadas sobre o meta-modelo obtido para a determinação da probabilidade de falha.

5.2. MAPEAMENTO DE DADOS

O desempenho de muitas técnicas de meta-modelagem é severamente afetado pelo formato que os parâmetros de entrada do problema analisado assumem. Em confiabilidade estrutural, é comum que as variáveis aleatórias de entrada admitam valores de ordens de grandeza muito diferentes entre si. Por exemplo, ao se trabalhar no sistema internacional de unidades, o módulo de elasticidade de um aço pode assumir valores da ordem de 10^{11} , enquanto a área da seção transversal de uma viga do mesmo material, facilmente assume valores da ordem de 10^{-3} . Esta disparidade mostrou-se relevante, e uma técnica de mapeamento precisou ser adotada nos exemplos estudados. No caso das EPC, o tratamento recebido pelos parâmetros de entrada quando da transformação isoprobabilística abordada em 4.3.3 se mostrou suficiente, não sendo necessária a adoção de nenhum tratamento adicional. A utilização dos dados em sua forma bruta, porém, causaram problemas numéricos nas aplicações de superfícies de resposta polinomiais e krigagem, pois geraram matrizes mal condicionadas. No caso das redes neurais artificiais, os valores de entrada estão sujeitos à função de ativação f . Esta função tem sua natureza

em formato de S, o que implica em que suas derivadas sejam praticamente constantes quando a amplitude dos dados é muito grande. Este problema atrapalha a convergência dos algoritmos de treinamento.

Para superar este problema, um algoritmo bastante simples foi desenvolvido, visando não apenas a solução dos exemplos testados, mas visando dar generalidade ao software programado. Assim, ele deve ser capaz de resolver diferentes problemas que futuros usuários apresentem, ainda que estes não tenham a experiência necessária para utilizar unidades compatíveis para os parâmetros de entrada. A seguinte rotina é aplicada ao vetor de variáveis aleatórias do problema. Isso é feito logo antes dele ser alimentado para o meta-modelo considerado, tanto durante o treinamento dos meta-modelos, quanto durante cada simulação de Monte Carlo feita sobre o meta-modelo já calibrado

```

repita para cada componente de X
  se a variável admitir valores fora do intervalo [-1,1]
    se a média da variável for diferente de 0
      divida o valor da variável pela sua média
    se não
      divida o valor da variável pelo seu desvio padrão
    fim se
  fim se
fim repita

```

Este procedimento se mostrou suficiente para a obtenção de um treinamento satisfatório e para a convergência dos resultados de todos os exemplos estudados.

5.3. EXEMPLOS ANALÍTICOS

Algumas equações analíticas de ordem de complexidade crescente foram selecionadas para testar a eficácia das técnicas programadas. As três primeiras equações são de conhecimento comum. A equação (5.4) apresenta a função de Ishigami, proposta por Ishigami (1990) para ser usada em problemas

de quantificação de incertezas e análise de sensibilidade. Ela se faz interessante para estas aplicações, devido a seu caráter fortemente não linear e não monotônico. Os valores adotados para seus coeficientes foram os mesmos usados por Marrel et al. (2009), e foi somada à equação uma constante para deixá-la mais semelhante a uma equação de estado limite usual (transladar seu valor esperado para a direita). Para exemplos analíticos, uma comparação entre o tempo necessário para a realização das simulações entre o modelo original e os metamodelos se faz secundária. A ideia central deste exemplo é, basicamente, verificar a capacidade dos meta-modelos de representar os diferentes modelos. As equações estudadas estão reunidas na tabela 5.1, e as características de suas variáveis aleatórias, na tabela 5.2.

Tabela 5.1 – Equações de estado limite estudadas.

<i>Tipo de Equação</i>	$g(x)$	<i>Eq. #</i>
<i>Linear</i>	$2X_1 - X_2 + 5$	(6.1)
<i>Polinomial</i>	$4X_1 + 2X_2^2 - X_1X_3^3 + 10$	(6.2)
<i>Exponencial</i>	$e^{(0.2X_1+1.4)} - X_2$	(6.3)
<i>Ishigami</i>	$\sin X_4 + 7 \sin^2 X_5 + 0.1X_6^4 \sin X_4 + 8$	(6.4)

Tabela 5.2 – Variáveis aleatórias utilizadas.

<i>Variável</i>	<i>Distribuição</i>
X_1, X_2, X_3	<i>Normal Padrão</i>
X_4, X_5, X_6	<i>Uniforme</i> $\sim [-\pi, \pi]$

As características dos meta-modelos foram programadas conforme discutido no tópico 5.1. Foram adotados parâmetros de modo a deixá-los equiparáveis ao encontrado na literatura (Elhewy e Pu 2006; Kaymaz 2005; Sudret 2008; Blatman e Sudret 2010; Gaspar et. al., 2014). Os números de pontos de treinamento das redes neurais referidos nos exemplos já leva em consideração o fato de que 20% deles foram utilizados como informação de

validação. O método de Levenberg–Marquardt foi usado para treinar as redes, com 1000 épocas.

Duas expansões em polinômios de caos foram utilizadas, com graus múlti-índice máximos 4 e 7, para comparação, com seus coeficientes obtidos através do método não intrusivo da regressão, discutido em 4.3.5.2.

Nos modelos de Krigagem, polinômios de grau até 2, sem a utilização de termos cruzados, foram adotados como funções de base para a construção da superfície de resposta. Valores ótimos foram adotados para os hiperparâmetros.

Em todos os casos, os pontos de suporte/treinamento foram definidos a partir de uma amostragem por Hipercubo Latino. Uma vez os meta-modelos calibrados, simulações de Monte Carlo com 10^6 amostras foram realizadas sobre cada meta-modelo estudado. Isso foi feito para duas situações: utilizando o número mínimo de pontos de suporte necessários para a calibração do modelo que necessita de mais pontos, no caso, dado pela equação (4.30), e utilizando-se o dobro deste número. Simulações de Monte Carlo com 10^6 amostras também foram realizadas sobre os modelos originais, para servir como resultado de referência. Os resultados de probabilidade de falha para cada caso foram computados e os índices de confiabilidade β correspondentes foram calculados. Estes resultados são exibidos nas tabelas 5.3 e 5.4.

Tabela 5.3 – β para o número mínimo de pontos de suporte

<i>Eq.</i>	<i>Ref.</i>	<i>RNA</i>	<i>EPC₄</i>	<i>EPC₇</i>	Krigagem
5.1	2,23	2,23	2,23	2,23	2,23
5.2	2,11	2,07	2,11	2,11	2,11
5.3	3,36	3,38	3,36	3,36	3,36
5.4	1,00	1,00	0,88	0,91	1,00

Tabela 5.4 – β para o dobro do número mínimo de pontos de suporte

<i>Eq.</i>	<i>Ref.</i>	<i>RNA</i>	<i>EPC₄</i>	<i>EPC₇</i>	Krigagem
5.1	2,23	2,23	2,23	2,23	2,23
5.2	2,11	2,11	2,11	2,11	2,11
5.3	3,36	3,36	3,36	3,36	3,36
5.4	1,00	1,00	0,99	0,99	1,00

5.4. COMPARAÇÃO DAS TÉCNICAS PARA UMA TRELIÇA PLANA

Uma treliça composta por 23 barras e 13 nós servirá de base para a avaliação das técnicas de meta-modelagem consideradas. Ela foi estudada por Blatman et al. (2008) na análise de meta-modelagem por expansões em polinômios de caos, e vem sendo utilizada em diversos trabalhos desde então. Esta treliça está submetida a 6 cargas verticais estocásticas em seus nós superiores, descritas por variáveis aleatórias que seguem uma distribuição de Gumbell, e suas barras apresentam propriedades geométricas e mecânicas descritas por VAs de distribuições log-normais, coerentemente evitando-se valores negativos. O estado limite de interesse diz respeito ao deslocamento vertical V_1 sofrido pelo ponto inferior que está no meio do vão, representando uma falha da estrutura quando este valor ultrapassar 10cm. A equação de estado limite examinada fica então definida pela expressão 5.5.

$$g(X) = 10 - |V_1(X)| \quad (5.5)$$

Uma representação da treliça estudada está esquematizada na Figura 5.1:

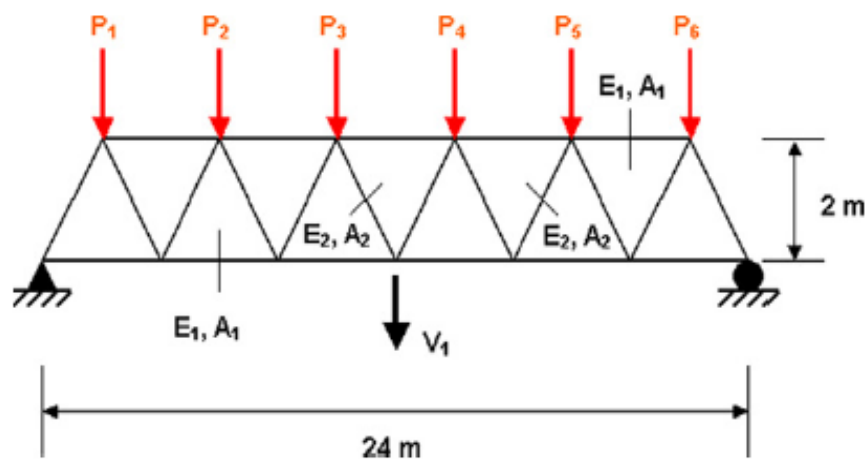


Figura 5.1 – Treliça estudada

As propriedades das variáveis aleatórias estão reunidas na tabela 5.5. Cada técnica de meta-modelagem é estudada separadamente, a partir da variação de um parâmetro relevante para seu desempenho, e, por fim, compara-se o

desempenho de todas elas simultaneamente. A referência foi adotada a partir de um modelo de elementos finitos, e o índice de confiabilidade obtido a partir de 10^4 simulações de Monte Carlo para este exemplo tem valor $\beta = 1,7082$. O mesmo número de simulações é feita sobre cada meta-modelo para a obtenção de cada resultado. Nas Figuras 5.2 a 5.5, a linha vermelha representa o valor de referência.

Tabela 5.5 – Propriedades das variáveis aleatórias envolvidas

VA	Distribuição	Média	Desvio Padrão
E_1, E_2 (Pa)	Lognormal	2.1×10^{11}	2.1×10^{10}
A_1 (m ²)	Lognormal	2×10^{-3}	2×10^{-4}
A_2 (m ²)	Lognormal	1×10^{-3}	1×10^{-4}
$P_1 - P_6$ (N)	Gumbel	5×10^4	7.5×10^3

5.4.1. RNA

Uma RNA semelhante á descrita no tópico 5.2 é considerada neste exemplo. A rede é treinada a partir do algoritmo de Levenberg–Marquardt, com um número de iterações que varia entre 10 e 100 épocas. Os resultados obtidos estão representados pela Figura 5.2. Para fins de comparação, será mantido o número de pontos de suporte mínimo para que as EPC funcionem. Este valor é fornecido pela equação 4.30, e vale 286 para este exemplo.

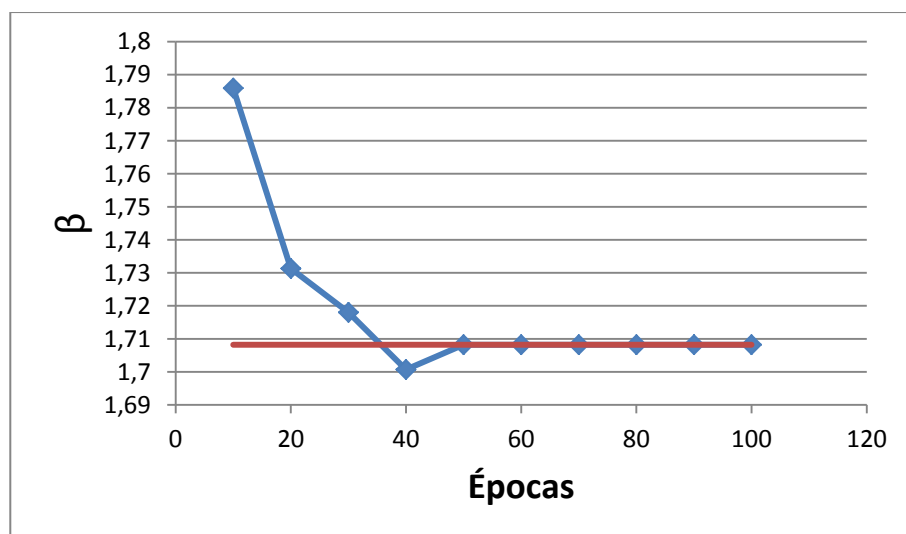


Figura 5.2 – Variação do número de épocas da RNA.

5.4.2. EPC

Uma expansão em polinômios de caos de grau multi-índice variando entre 0 e 4 é considerada neste exemplo, tendo sido a expansão de terceiro grau a avaliada por Blatman et al. (2008). Os resultados estão resumidos na Figura 5.3.

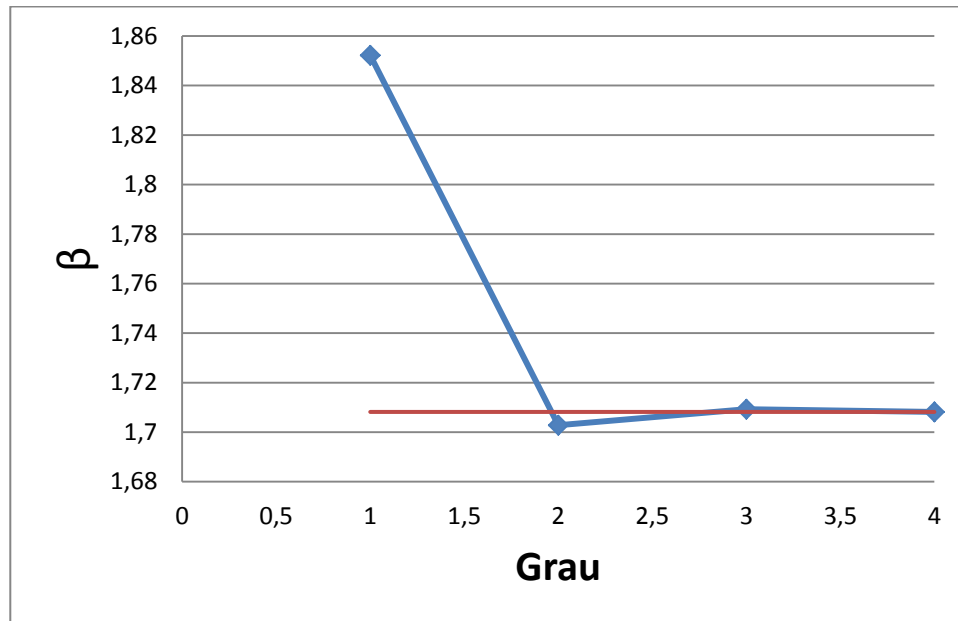


Figura 5.3 – Variação do grau da EPC.

Para o polinômio de grau 0 (uma constante), o resultado obtido não pode ser representado neste gráfico por refletir um erro muito grande.

5.4.3. Krigagem

Para a Krigagem, foram testadas superfícies de resposta de graus 0, 1 e 2, porém em todos os casos a resposta obtida foi a mesma. Considerou-se então um modelo construído a partir de uma superfície quadrática sem termos cruzados, e a influência do hiperparâmetro θ foi avaliada. Para tal, o modelo de correlação adotado foi o gaussiano isotrópico (Gaspar et al. 2014), que conta com função de auto-correlação definida pela equação (4.48), porém com $\theta_i = \theta_0 (i = 1, 2, \dots, m)$. Os resultados obtidos estão dispostos na Figura 6.4.

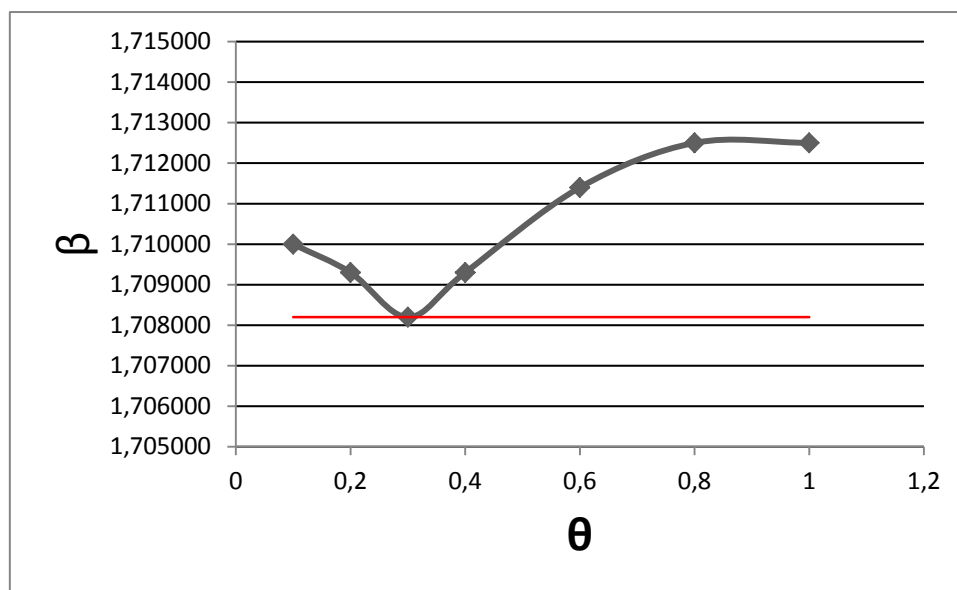


Figura 5.4 – Variação dos hiperparâmetros na krigagem.

5.4.4. Comparação

Para comparar o desempenho das diferentes técnicas de meta-modelagem no mesmo exemplo, buscou-se adotar em sua construção as já citadas práticas comuns na literatura, mantendo cada técnica tão simples quanto possível. A RNA utilizada foi construída conforme descrito, porém com um valor fixo de 100 épocas, conforme a Figura 6.2 mostra ser suficiente. Os polinômios de caos foram mantidos com grau multi-índice máximo de 3, o mesmo utilizado por de Blatman et al. (2008). O modelo baseado em krigagem também foi utilizado conforme descrito, porém considerando-se para o parâmetro θ_0 seu valor ótimo, $\theta_0 = 0,3$. Nesta comparação, variou-se o número de pontos de suporte utilizados na construção de cada modelo. Os resultados são apresentados na Figura 6.5, na qual pode-se observar que as três técnicas apresentam convergência conforme o número de pontos de treinamento aumenta. Erros da ordem de 2% são observados a partir dos 50 pontos de suporte, e os resultados se estabilizam a partir dos 700 pontos. O meta-modelo de krigagem apresenta os resultados que oscilam de forma mais brusca, porém são obtidos mais rapidamente do que aqueles das RNA, que apresentam um comportamento mais suave. Os resultados alcançados através das EPC equiparam-se em qualidade aos obtidos com as outras técnicas, porém só podem ser computados a partir dos 286 pontos de suporte.

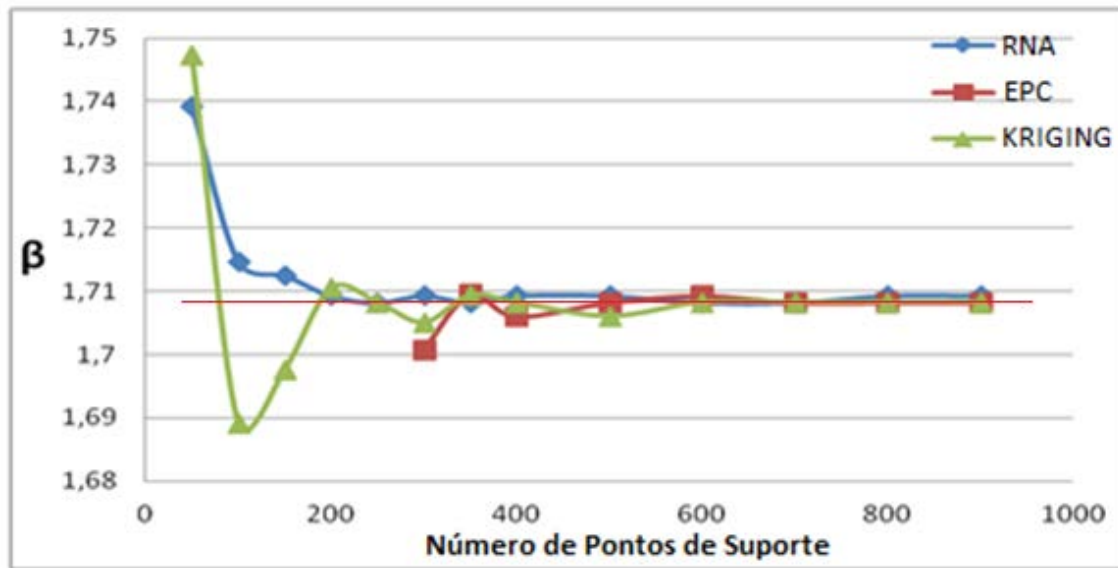


Figura 5.5 – Comparação entre as diferentes técnicas

As técnicas foram aplicadas também para a solução do mesmo problema com uma alteração na equação de estado limite, conforme representado pela equação 5.6:

$$g(X) = 14 - |V_1(X)| \quad (5.6)$$

Para este caso, o índice de confiabilidade, obtido da mesma forma, tem valor $\beta = 4,13$. Foram utilizados 600 pontos de suporte para a construção de todos os modelos, valor suficiente para a recuperação do resultado exato, no que diz respeito à análise da equação 5.5. Os resultados obtidos estão resumidos na tabela 5.6.

Tabela 5.6 – Propriedades das variáveis aleatórias envolvidas

	RNA	EPC	Krigagem
β	4,06	4,13	4,13

Nota-se que apenas para as RNA houve perda de precisão com a diminuição da probabilidade de falha, com um erro obtido da ordem de 2%.

5.5. AVALIAÇÃO DA EFICIÊNCIA DA KRIGAGEM SOBRE UMA TORRE DE TELECOMUNICAÇÃO

Uma torre de telecomunicações com 68 barras, conforme representado pela Figura 5.6, é avaliada com o auxílio de um meta-modelo construído com krigagem. As características do modelo são semelhantes às aquelas descritas para o exemplo 5.4. A falha da estrutura é considerada quando o nó central do trecho mais alto da torre apresenta um deslocamento horizontal superior a 60 centímetros, conforme representado pela equação 5.7.

$$g(\mathbf{X}) = 0,6 - |d_x| \quad (5.7)$$

As seis variáveis aleatórias que descrevem o problema estão reunidas na tabela 5.7. Todas as barras da torre são consideradas com a mesma área e módulo de elasticidade. As quatro cargas aplicadas no topo da torre são horizontais.

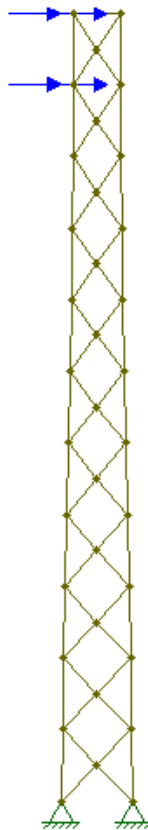


Figura 5.6 – Torre de Telecomunicação

Tabela 5.7 – Propriedades das variáveis aleatórias envolvidas

VA	Distribuição	Média	Desvio Padrão
E (Pa)	Lognormal	2.1×10^{11}	2.1×10^{10}
A (m ²)	Lognormal	3×10^{-3}	3×10^{-4}
$P_1 - P_4$ (N)	Gumbel	15×10^3	15×10^2

Variou-se o número de pontos utilizados para a construção do modelo para se verificar a melhora na precisão dos resultados conforme mais informação é utilizada, de modo semelhante ao representado na figura 5.5. Estes resultados estão sintetizados na Figura 5.7. A referência para o índice de confiabilidade, $\beta = 1,334$ foi obtida através da realização de 10^6 SMC, com amostragem por hipercubo latino.

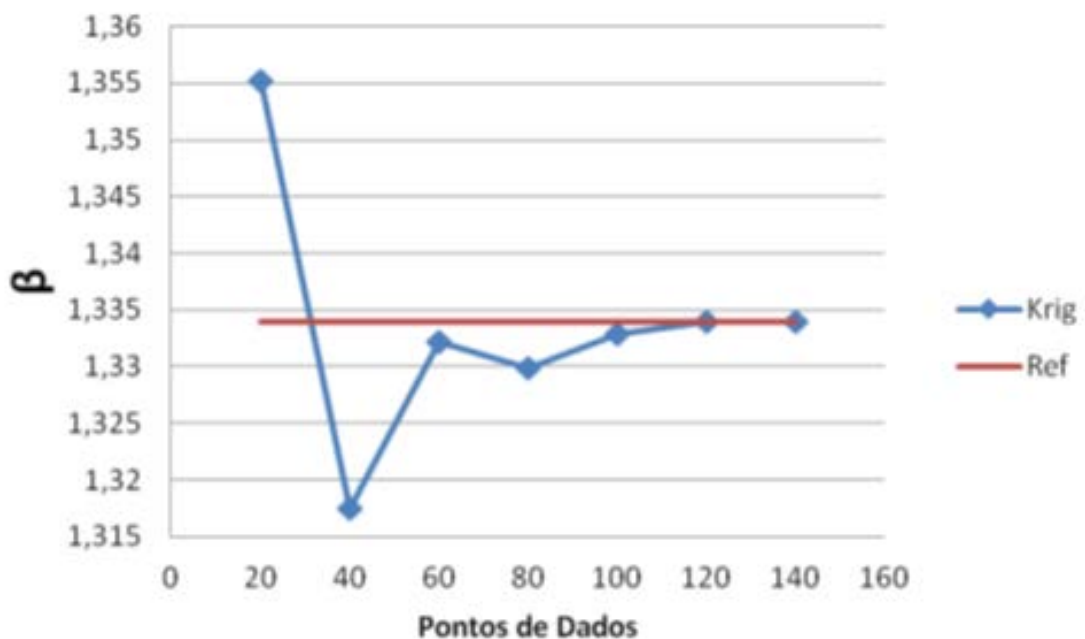


Figura 5.7 – Variação da quantidade de pontos de dados.

A variância das previsões foram também computadas para alguns pontos do espaço das variáveis aleatórias, conforme a equação 4.56. A figura 5.8 mostra a evolução desta grandeza conforme caminha-se neste espaço,

partindo-se do primeiro ponto utilizado para a construção do modelo de krigagem, até se chegar ao segundo. Esta progressão se dá em linha reta no espaço hexa-dimensional. O mesmo foi feito entre o segundo e o terceiro pontos, e os resultados são mostrados na figura 5.9.

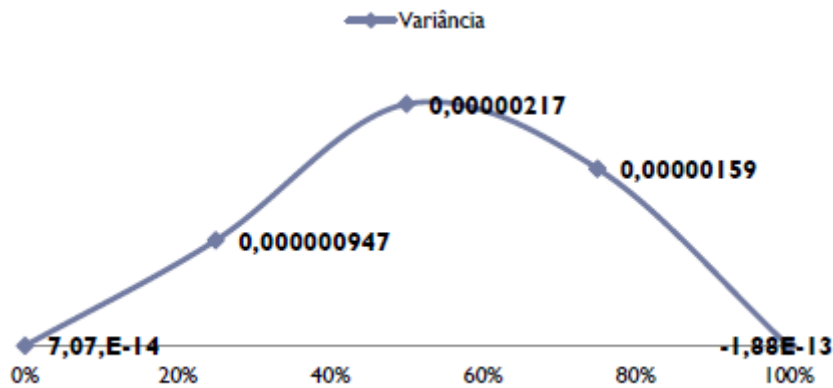


Figura 5.8 – Variância da previsão entre os dois primeiros pontos

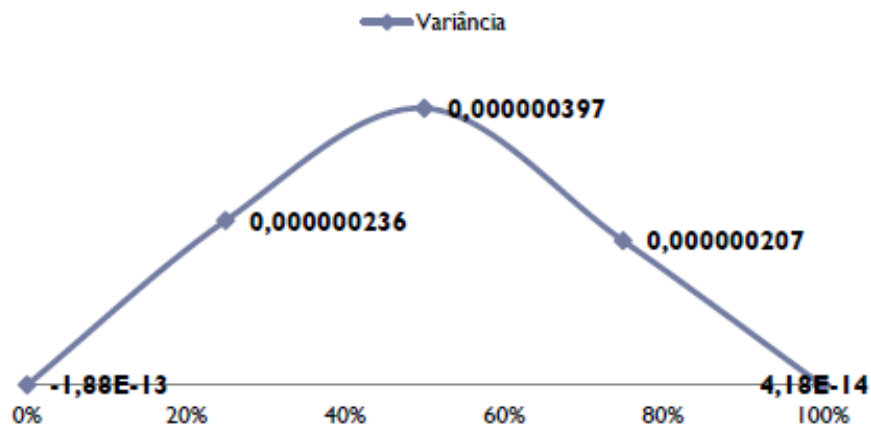


Figura 5.9 – Variância da previsão entre os próximos dois pontos

O tempo gasto para a construção e avaliação, com 10^6 SMC, do modelo de krigagem foi de 7,7 segundos. Isso equivale a aproximadamente 0,2% do tempo gasto para as 10^6 SMC sobre o modelo de elemento finitos original, de 1 hora e 10 minutos.

6. CONCLUSÕES E SUGESTÕES

6.1. CONCLUSÕES

Três técnicas modernas de meta-modelagem foram programadas e testadas para alguns exemplos, com diferentes valores de parâmetros. Em todos os casos, elas foram capazes de representar o modelo original proposto, com crescente qualidade conforme se ajustavam os parâmetros.

Algumas dificuldades foram encontradas para trabalhar com as Redes Neurais Artificiais, uma vez que uma pequena variação no número de pontos de suporte, ou na proporção entre pontos de treinamento e pontos de validação, implicaram em um considerável aumento ou diminuição da probabilidade de falha encontrada. Para as expansões em polinômios de caos, a matriz de dados obtida estava mal condicionada em alguns casos, sendo necessário selecionar um número diferente de pontos de suporte para que o resultado obtido fosse coerente. Ainda assim, em ambos os casos os erros parecem ser efeitos locais, e a precisão das técnicas aumenta com o número de pontos de suporte considerados.

O número de épocas adotado no processo de treinamento da rede neural adotada no exemplo numérico mostrou-se um parâmetro importante para a precisão do modelo, que melhora conforme mais épocas são utilizadas. Apenas 50 épocas de treinamento foram suficientes para fazer a rede apresentar o resultado exato. O mesmo pode ser observado para o número de elementos e a complexidade da expansão em polinômios de caos realizada para o mesmo exemplo. Verificou-se ainda a importância do valor do hiperparâmetro θ no modelo de krigagem, que leva a resultados piores conforme vai se afastando de seu valor ótimo.

Ao se variar, entre 0 e 2, o grau da superfície de resposta utilizada no modelo de Krigagem, não houve alteração na resposta apresentada. Isso mostra que a interpolação realizada pela parcela do processo estocástico estacionário influencia de maneira significativa a qualidade dos resultados obtidos.

Ao serem comparadas para resolver o mesmo problema, todas as técnicas convergiram para o valor de referência $\beta = 1,7082$ conforme o número de pontos de suporte utilizados para a construção dos modelos aumentava. As expansões em polinômios de caos apresentam bons resultados e parecem convergir rapidamente, porém necessitam de, no mínimo, 286 pontos de suporte, enquanto os modelos de RNA e krigagem apresentam resultados razoavelmente precisos a partir de apenas 50 pontos de suporte. A convergência mais rápida foi a apresentada pelas RNA, ainda que seu processo de treinamento consuma um tempo consideravelmente maior do que aquele necessário para a definição do meta-modelo por Krigagem. A variância calculada para o modelo de Krigagem apresentou resultados compatíveis com o esperado.

Uma sugestão para um mapeamento conveniente das variáveis aleatórias, de modo a não incidir em instabilidades numéricas ou comprometer a eficiência dos meta-modelos foi proposta, e aplicada a todos os exemplos analisados com bons resultados. Sintetizando-se algumas dificuldades encontradas ao longo do trabalho, acompanhadas de sugestões para superá-las, destacam-se:

- Nas regressões realizadas para os meta-modelos de EPC e Krigagem, a matriz de dados pode ser não inversível em alguns casos, devido à natureza aleatória da obtenção dos dados. Regularizá-la adicionando quantidades pequenas à sua diagonal principal, a utilização de um número diferente de pontos de suporte, ou a substituição da inversa da matriz por sua pseudo-inversa são maneiras de se superar este problema.
- Para modelos mecânicos com um número grande de parâmetros de entrada, o uso das EPC torna-se inviável, uma vez que o modelo original precisa ser resolvido um número grande de vezes para a obtenção dos coeficientes da base. A utilização de uma expansão esparsa em polinômios de caos (Sudret e Blatman, 2010) pode mitigar este problema.

- As Redes Neurais Artificiais apresentaram uma grande precisão na resposta para todos os exemplos estudados, mas foram também a técnica que demandou mais tempo de processamento no treinamento. A simples divisão dos pontos de suporte entre pontos de treinamento e de validação, para um mesmo número total de pontos, pode influenciar de forma significativa na qualidade da resposta. A utilização de outro algoritmo de treinamento pode superar estes problemas. Ainda que o modelo precise ser treinado novamente, para uma nova proporção entre pontos de treinamento e verificação, repetir o treinamento algumas vezes ainda é mais barato que a solução direta por simulação de Monte Carlo em problemas complexos com probabilidade de falha pequena. Ainda que o número de simulações utilizado no exemplo numérico não tenha sido tão alto, apenas 0,5% deles foram necessários para treinar a rede com precisão.

6.2. SUGESTÕES PARA TRABALHOS FUTUROS

Neste trabalho foi demonstrado que as técnicas de meta-modelagem representam satisfatoriamente alguns exemplos simples. Sugere-se testar as mesmas técnicas em problemas mais complexos, de forma a verificar sua versatilidade e eficiência. Outras técnicas de meta-modelagem também podem ser programadas e incluídas na comparação. Sugere-se a aproximação da equação de estado limite utilizando o método dos mínimos quadrados móveis e utilizando *support vector machines*.

Problemas de otimização estrutural, sobretudo na presença de incertezas e de otimização de risco, são também bastante custosos de se analisar. Certamente estes podem se beneficiar muito da aproximação, por meta-modelos, de sua equação de estado limite ou de sua função objetivo. As técnicas programadas podem ser utilizadas e testadas também na solução deste tipo de problema.

REFERÊNCIAS

- Agterberg, F., P. 2005: *Georges Matheron, founder of spatial statistics*. Geological Survey of Canada.
- Bailey, D., E. 1971: *Probability and Statistics – models for research*, Editora John Wiley and sons – ISBN 0-471-04160-2
- Beck, A. T., 2012: *Curso de confiabilidade estrutural: notas de aula*. São Carlos: EESC-USP
- Bect, J., Ginsbourger, D., Li, L., Picheny, V., Vazquez, E., 2011: *Sequential design of computer experiments for the estimation of a probability of failure*. Submitted to Journal of Statistics and Computing (arXiv:1009.5177).
- Berveiller, M., B. Sudret, e M. Lemaire 2004: *Presentation of two methods for computing the response coefficients in stochastic finite element analysis*. Proc. 9th ASCE Specialty Conference on Probabilistic Mechanics and Structural Reliability, Albuquerque, USA.
- Berveiller, M., B. Sudret, and M. Lemaire 2006: *Stochastic finite elements: a non intrusive approach by regression*. European Journal of Computational Mechanics, v15, p. 81–92.
- Blatman G, Sudret B, Berveiller M. 2007. *Quasi-random numbers in stochastic finite element analysis*. Mécanique & Industries 8:289_97.
- Blatman, G., Sudret, B. 2010: *An adaptive algorithm do build up sparse polynomial chaos expansions for stochastic finite element analisys*. Probabilistic Engineering Mechanics, v.25, p. 187-197
- Blatman, G., Sudret, B. 2011: *adaptive sparse polynomial chaos based on least angle regression*. Journal of Computational Physics, v.230, p. 2345-2367
- Breitung, K., Faravelli, L., 1996: “*Response Surface Methods and Asymptotic Approximations*”, in: *Mathematical Models for Structural Reliability Analisis*, Fabio Casciati and Brian Roberts, NW.
- Burrough, P. A. 1987: *Principles of geographical information systems for land resources assessment*. Oxford, Clarendon Press 193p.
- Cardoso, J. B., Almeida, J., R., Dias, J., M., Coelho, P., G. 2008: *Coelho: Structural reliability analysis using Monte Carlo simulation and neural networks*, Advances in Engineering Software, Volume 39, Issue 6, Pg 505-513
- Choi, S., R. Grandhi, and R. Canfield 2004: *Structural reliability under non-Gaussian stochastic behavior*. Computers & Structures v.82, p.1113–1121.
- Clarke, A., B., Disney, R., L. 1979: *Probabilidade e processos estocásticos*, editora S. A., ISBN: 85-216-0026-7
- Cressie, N. 1990: *The Origins of Kriging - Mathematical Geology*, Vol. 22, No. 3
- Dayhoff, J. 1990: *Neural Network Architectures*, Van Nostrand Reinhold
- Deng J, Gu D, Li X, Yue Z. 2005: *Structural reliability analysis for implicit performance functions using artificial neural network*. Structural Safety; 27:25–48.
- Ditlevsen, O., Madsen, H. O., 1996: *Structural Reliability Methods*, John Wiley and Sons, Chichester.

- Dubourg, V., Deheeger, w., Sudret, B, 2011: *Metamodel-based importance sampling for structural reliability analysis*. Probabilistic Engineering Mechanics. ,33, pp. 47–57.
- Echard, B., Gayton, N., Lemaire, M., Relun, R. 2013: *A combined Importance Sampling and Kriging reliability method for small failure probabilities with time-demanding numerical models* – Reliability Engineering and System Safety v 111, p. 232 -240
- Elhewy, A., H., Mesbahi, E., Pu, Y. 2005: *Reliability analysis of structures using neural network method*, Probabilistic Engineering Mechanics v.21 p. 44–53
- Faravelli, L. 1989: *A Response Surface Approach for Reliability Analisis*, J. Eng. Mechanics Division ASCE 115, 2763-2781
- Fausett, L., V.,1994: *Fundamentals of Neural Networks*. Prentice Hall, 3 Ed.
- Gaspar, B., Teixeira, A. P., Guedes Soares, C. 2014: *Assessment of the efficiency of Kriging surrogate models for structural reliability analysis*. Probabilistic Engineering Mechanics, 37, 24-34
- Ghaffari, A. 2006: *Performance comparison of neural network training algorithms in modeling of bimodal drug delivery*, International Journal of Pharmaceutics, v. 327, p. 126-138.
- Ghanem, R. e P. Spanos 1991: *Stochastic finite elements – A spectral approach*. Springer Verlag. (Reeditado por Dover Publications, 2003).
- Gomes HM, Awruch AM. 2004: *Comparison of response surface and neural network with other methods for structural reliability analysis*. Structural Safety 26:49–67.
- Gomes, S. S. G., Ludermir T. B., 2008: *Redes neurais artificiais com funções de ativação complemento log-log e probit para aproximar funções na presença de observações extremas*. Revista da Sociedade Brasileira de Redes Neurais (SBRN), Vol. 6, N.2, p.142-153, 2008
- Hodhkin, A. L., Huxley, A. F. 1952: *A quantitative description of membrane current and its application to conduction and excitation in nerve*, Journal of Psychology, vol 117, p. 500-544
- Hornik K, Stinchcombe M, White H. (1990) *Universal approximation of an unknown mapping and its derivatives using multilayer feedforward networks.* Neural Networks 3,551–560
- Hurtado, J.E., Barbar, A.H. 1998: *Monte Carlo Techniques in Computational Stochastic Mechanics*. Archives of Computational Methods in Engineering State of the art reviews, v.5, n.1, p.3-30.
- Ishigami, T., Homma, T. 1990: *An importance quantification technique in uncertainty analysis for computer models*. Uncertainty Modeling and Analysis, Proceedings., First International Symposium p. 398-403
- Junior, R., A., C. 2006: *Simulações Computacionais Biologicamente Plausíveis de Neurônios do Córtex Somestésico Primário*, Dissertação – Departamento de Física e Matemática da Universidade de São Paulo.
- Kaymaz, I. 2005: *Application of kriging method to structural reliability problems*, Structural Safety, v. 27, p. 133-151
- Kovács, Z. L., 2006: *Redes Neurais Artificiais – Fundamentos e Aplicações*, Livraria da Física, Quarta Edição.

- Leonel, E. D. 2009: *Modelos não lineares do método dos elementos de contorno para análise de problemas de fratura e aplicação de modelos de confiabilidade e otimização em estruturas submetidas à fadiga* – Tese de Doutorado, EESC/USP.
- Loshchilov, I. G, 2013: *Surrogate-Assisted Evolutionary Algorithms*, Tese de PhD, Ecole Doctorale d'Informatique, ED 427 Université Paris-Sud 11
- Marrel, A., Iooss, B., Laurent, B., e Roustant, O. 2009: *Calculations of sobol indices for the gaussian process metamodel*. Reliability Engineering & System Safety, v.94(3), p.742-751.
- Matheron, G, 1963: *Principles of Geostatistics, Economic Geology Vol. 58, p. 1246-1266*
- McCulloch, W. S., Pitts, W. 1943: *A logical calculus of the ideas immanent in nervous activity*. Bulletin of Mathematical Biophysics, vol 5, p. 115-133.
- Metropolis, S, N.; Ulam, S. 1949: *The Monte Carlo method*. Journal of the American statistical association, v.44, n.247, p. 335-341
- Mitchell, T. J., and M. D. Morris. 1992. *The spatial correlation function approach to response surface estimation*. - Proceedings of the Winter Simulation Conference, ed. R. C. Crain, J. R. Wilson, J. J. Swain, and D. Goldsman, 565–571: Piscataway, N.J.: Institute of Electrical and Electronics Engineers.
- Mlodinov, L. 2009: *O Andar do Bêbado*, Editora Zahar.
- Montgomery, D., C. 2013: *Design and Analysis of Experiments*. Wiley & Sons, 8ª edição.
- Neto, B., B., Scarminio, I., S., Bruns, R., E 2001: *Como fazer experimentos: pesquisa e desenvolvimento na ciência e na indústria*. Editora Unicamp, 2ª Edição
- Neto, P., Cymbalista, M. 1974: *Probabilidades*, Editora Edgard Blucher LTDA
- Papadrakis M, Papadopoulos V, Lagaros N. 1996: *Structural reliability analysis of elastic-plastic structures using neural networks and Monte Carlo simulation*. Computer methods in applied mechanics and engineering 136:145–63.
- Sacks, J., S. B. Schiller, e Welch W. J. 1989: *Designs for computer experiments*. Technometrics, 31
- Sagrilo, L., V., S., Lima, E., C., P. 1996: *Confiabilidade Estrutural: Métodos Analíticos, FORM e SORM*, notas de aula, COPPE/UFRJ
- Santos, K., R., M, 2014: *Técnicas de amostragem inteligente em simulação de Monte Carlo*. Dissertação de mestrado, EESC/USP
- Silva, I. N., Spatti, D. H., Flauzino, R. A. 2010: *Redes Neurais Artificiais para engenharia e ciências aplicadas*, Altiber Editora Ltda.
- Silva, L., N., C. 1998: *Análise e síntese de estratégias redes neurais artificiais*, dissertação de mestrado, UNICAMP.
- Soize, C., Ghanem, R., 2004: *Physical systems with random uncertainties: Chaos representations with arbitrary probability measure*. SIAM Journal on Scientific Computing, Society for Industrial and Applied Mathematics, 26 (2), pp.395-410
- Stark, H., Woods, J., W., 1994: *Probability, Random Processes, and estimation theory for engineers*. Segunda Edição, Prentice Hall, ISBN: 0-13-728791-7

- Stigler, S., M. 1986. *The History of Statistics: The Measurement of Uncertainty Before 1900*. Belknap Press of Harvard University Press. ISBN 0-674-40340-1.
- Sudret B., Berveiller M., Lemaire M., 2006: *A stochastic finite element procedure for moment and reliability analysis*, *Eur. J. Comp. Mech.*, vol. 15(7-8), p. 825-866
- Sudret, B., Blatman, G., Berveiller, M. 2011: *Construction Reliability: Safety, Variability and Sustainability – Capítulo 8: Response surfaces based on polynomial chaos expansions*. Wiley, editado por: Julien Baroth, Franck Schoefs, Denys Breysse
- Sudret, B. e Der Kiureghian, A., 2000: *Stochastic finite elements and reliability: a state-of-the-art report*. *Technical Report UCB/SEMM-2000/08*, University of California, Berkeley.
- Sudret, B. 2008: *Global Sensivity Analysis using Polynomial Chaos Expansions*. *Reliability Engineering and System Safety*, v. 93, p; 964–979
- Sudret, B, 2011: *Meta-models for structural reliability and uncertainty quantifications*, 5th Asian-Pacific Symposium on Structural Reliability and its Applications, 53-76.
- Sudret, B., 2012: *Meta-models for structural reliability and uncertainty quantifications*, Phoon KK, Beer M, Quek ST, Pang SD, Proc. Of the 5th Asian-Pacific Symposium on Structural Reliability and its Applications, 53-76.
- Sudret, B. 2014: *Polynomial Chaos Expansions, Theory, Numerical Methods and Applications*, MNMUQ Summer School Presentation.
- Sutton, R. S., Barto, A. G. 1998: *Reinforcement learning: an introduction*. MIT Press, Cambridge, Massachusetts, EUA.
- Tobler, W. 1970: *A computer movie simulation urban growth in the Detroit region*, *Economic Geography*, International Geographical Union. Commission on Quantitative Methods Vol. 46 p. 234-240.
- Xie, W., Nelson, B., Staum, J. 2010: *The Influence of Correlation Functions on Stochastic Kriging Metamodels*. Winter Simulation Conference.
- Xiu, D. 2010: *Numerical Methods for Stochastic Computations*, Princeton University Press.

APÊNDICE – TÓPICOS DE PROBABILIDADE E ESTATÍSTICA

CONCEITOS DE PROBABILIDADE

A teoria da probabilidade serve para descrever fenômenos aleatórios, aqueles cujos resultados não podem ser previstos com certeza. Ao longo da maior parte da história da humanidade se tentou descrever o mundo de maneira determinística, pois não era concebível a ideia de que o Criador teria feito o universo de maneira incerta, especula Mlodinov (2009). Foi apenas no Século XVI que os primeiros trabalhos importantes descrevendo probabilidades surgiram, como o “Jogos de Azar” do algebrista Jeronimo Cardano. A teoria foi aprofundada nos anos posteriores por grandes nomes como Blaise Pascal, Pierre de Fermat e Laplace, mas apenas depois dos trabalhos de Carl Friedrich Gauss, já no século XIX, a teoria começou a ser sistematicamente aplicada às ciências.

Diversos exemplos cotidianos que envolvem experiências aleatórias podem ser citados, do lançar de um dado à previsão do tempo, dos números da loteria à intensidade do vento. Para cada caso, pode-se definir um conjunto cujos elementos são cada um dentre todos os resultados possíveis do experimento que se está estudando. A este conjunto, usualmente denotado por S , dá-se o nome de *Espaço Amostral*. Por exemplo, para o experimento “lançamento de um dado de 6 faces”, o espaço amostral é:

$$S = \{1,2,3,4,5,6\} \tag{A.1}$$

Pois os números naturais entre 1 e 6 são todos os resultados possíveis após o lançamento. Denomina-se *Evento* um subconjunto do espaço amostral, que se refere a um resultado cuja veracidade só pode ser determinada após a realização da experiência (Clarke e Disney, 1979). Então, para o lançamento de um dado, pode-se definir um evento A como “o resultado é um número par”:

$$A = \{2,4,6\} \tag{A.2}$$

À ocorrência de cada possível evento pode-se associar uma medida, denominada probabilidade do evento, que se costuma denotar por P . A probabilidade de ocorrência do evento A é representada então por $P[A]$. Esta medida está sujeita aos seguintes axiomas:

a) $P[A] \geq 0$

A probabilidade de qualquer evento é um número maior ou igual a zero, não existe probabilidade negativa.

b) $P[S] = 1$

O espaço amostral tem probabilidade unitária, ou seja, se um evento ocorrerá com certeza, ele tem probabilidade 1.

c) $P[A + B] = P[A] + P[B]$

Esta propriedade é válida desde que A e B sejam eventos mutuamente exclusivos, ou seja, a ocorrência de um deles implica na não ocorrência do outro, ou $P[A \cap B] = 0$.

Interpreta-se ainda que a probabilidade de um evento ocorrer é a quantidade de eventos favoráveis a ele (N_A) com relação à quantidade total de eventos possíveis (N) em dado espaço amostral. Esta interpretação é conhecida como *definição clássica de probabilidade* (Beck, 2012), e é enunciada pela equação A.3:

$$P[A] = \frac{N_A}{N} \tag{A.3}$$

Nem sempre é possível determinar a probabilidade utilizando esta abordagem. Utiliza-se ainda a chamada *definição frequentista*, na qual se calcula a probabilidade de um evento a posteriori, e espera-se que ela convirja para a probabilidade exata conforme se aumenta o número de realizações do experimento:

$$P[A] = \lim_{n \rightarrow \infty} \frac{n_A}{n} \quad (\text{A.4})$$

A ocorrência de um evento pode influenciar na probabilidade de ocorrência de outro. Por exemplo, pode-se estudar a probabilidade de uma criança escolhida ao acaso crescer até os 1,80m de altura. Esta probabilidade aumenta, porém, quando considerada a informação que ambos os pais dela são muito altos. Costuma-se denotar a probabilidade de A ocorrer dado que B ocorreu por $P[A|B]$. Isto é análogo a calcular a probabilidade de A não mais sobre o espaço amostral inteiro, mas sobre a parcela de S que diz respeito ao evento B . A definição de probabilidade condicional é mostrada na expressão (A.5):

$$P[A|B] = \frac{P[A \cap B]}{P[B]} \quad (\text{A.5})$$

desde que $P[B] > 0$.

Outro conceito de bastante relevância é a independência entre eventos, ou seja, quando a ocorrência de um evento não é de forma alguma influenciada pela ocorrência ou não de outro. Por exemplo, ao se jogar uma moeda e outra logo em seguida, o resultado do segundo lançamento, em condições normais, não dependerá do que ocorreu com o primeiro. Considera-se que dois eventos são independentes quando:

$$P[A \cap B] = P[A]P[B] \quad (\text{A.6})$$

Note que essa expressão decorre da equação (A.5), pois a probabilidade do evento A deve se manter a mesma quando a avaliarmos sujeita à ocorrência do evento B . Substituindo-se (A.6) em (A.5), obtém-se:

$$P[A|B] = P[A] \quad (\text{A.7})$$

VARIÁVEIS ALEATÓRIAS

Uma variável aleatória X é uma grandeza cujo valor é determinado através de um experimento aleatório. Para cada evento no espaço amostral S , associa-se um valor específico para a quantidade X (Bailey, 1971). Pode-se definir uma variável aleatória se utilizando do conceito de função – uma regra que relaciona cada evento do espaço amostral a um valor de X . Desse modo, S é o domínio de X . Denota-se usualmente com uma letra maiúscula uma variável aleatória X , e com a letra minúscula correspondente, x , uma realização desta variável aleatória. A variável é dita discreta quando seu domínio for composto por um número contável de pontos e contínua quando for composto por um número infinito e incontável. Na linguagem habitual da estatística, uma variável aleatória que serve como parâmetro de entrada para um experimento pode ser também denominada um fator.

Pode-se fixar um número real x e, a partir daí, definir eventos envolvendo a variável e o valor fixado. Por exemplo: $A = \{X = x\}$ ou $B = \{X < x\}$. Estes eventos podem ser associados à probabilidades, e denota-se por $P[A] = P[\{X < x\}]$ a probabilidade da variável aleatória X assumir um valor menor que o número x .

FUNÇÕES DE PROBABILIDADE

Esta última probabilidade recebe uma atenção especial, pois define a chamada *função de distribuição acumulada de probabilidades* (FDA):

$$F(x) = P[\{X < x\}] \quad (\text{A.8})$$

Esta função não é a variável aleatória X , mas uma característica dela. Seu domínio é o conjunto dos números reais, e seu contra-domínio é um valor entre 0 e 1, pois é uma probabilidade. Note que esta definição se aplica tanto a variáveis discretas quanto a variáveis contínuas. As FDA exibem ainda as seguintes propriedades (Beck, 2012):

- a) $F(-\infty) = 0, F(\infty) = 1$; Como seu domínio são os números reais, quanto mais à esquerda na reta um número, menor a probabilidade de que X seja menor que ele, e é um evento certo que X estará à esquerda de algum número.
- b) F é monotonicamente crescente - $F(x_1) \leq F(x_2)$ para qualquer $x_1 \leq x_2$
- c) F é contínua pela direita.
- d) Se $x_1 \leq x_2$, então: $P[\{x_1 < X \leq x_2\}] = F(x_2) - F(x_1)$
- e) Se F for descontínua em x_1 , a probabilidade de X assumir o valor x_1 é numericamente igual ao salto que ocorre neste ponto.
- f) Se F é contínua: $P[\{X = x\}] = 0$. Nestes casos contínuos, costuma-se trabalhar com a probabilidade de X pertencer a um determinado intervalo (uma vez que não há saltos na função)

Outra função de suma importância para o estudo das variáveis aleatórias é aquela que representa o conteúdo de probabilidade que há em cada ponto x . A chamada função densidade de probabilidade (FDP) se relaciona à FDA através da equação (A.9), para variáveis aleatórias discretas, em que a FDP é denotada por $P()$:

$$F(x) = \sum_{x_i < x} P(x_i) \quad (\text{A.9})$$

Para VAs contínuas, em que se costuma denotar a FDP por $f()$, tem-se a equação (A.10)

$$F(x) = \int_{-\infty}^x f(x) dx \quad (\text{A.10})$$

Existem diversas distribuições de probabilidade clássicas que descrevem variáveis aleatórias. Algumas delas, bem como os parâmetros que as definem, estão reunidas na Tabela A.1:

Tabela A.1: Sumário de distribuições contínuas de probabilidade. Fonte: (Beck, 2012)

#	Distribuição	$f_X(x)$	p_1	p_2	p_3	p_4
0	Determinística	$\delta(x_0)$	x_0	-	-	-
1	Uniforme	$\frac{1}{b-a}$	a	b	-	-
2	Normal	$\frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right]$	μ	σ	-	-
3	Log-Normal	$\frac{1}{\xi x\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{\ln(x)-\lambda}{\xi}\right)^2\right]$	λ	ξ	-	-
4	Exponencial	$v \exp[-v(x-\varepsilon)]$	v	-	ε	-
5	Rayleigh	$\frac{(x-\varepsilon)}{\eta^2} \exp\left[-\frac{1}{2}\left(\frac{x-\varepsilon}{\eta}\right)^2\right]$	η	-	ε	-
6	Logística	$\frac{e^{-\frac{\pi}{\sqrt{3}}\frac{(x-\mu)}{\sigma}}}{\left(1+e^{-\frac{\pi}{\sqrt{3}}\frac{(x-\mu)}{\sigma}}\right)^2}$	μ	σ	-	-
7	Gumbel mínimos	$\beta \exp[\beta(x-u_1) - e^{\beta(x-u_1)}]$	u_1	β	-	-
8	Gumbel máximos	$\beta \exp[-\beta(x-u_n) - e^{-\beta(x-u_n)}]$	u_n	β	-	-
9	Frechet mínimos	$\frac{\beta}{u_1} \left(\frac{x}{u_1}\right)^{\beta+1} \exp\left[-\left(\frac{x}{u_1}\right)^\beta\right]$	u_1	β	-	-
10	Frechet máximos	$\frac{\beta}{u_n} \left(\frac{u_n}{x}\right)^{\beta+1} \exp\left[-\left(\frac{u_n}{x}\right)^\beta\right]$	u_n	β	-	-
11	Weibull mínimos	$\frac{\beta}{u_1-\varepsilon} \left(\frac{x-\varepsilon}{u_1-\varepsilon}\right)^{\beta-1} \exp\left[-\left(\frac{x-\varepsilon}{u_1-\varepsilon}\right)^\beta\right]$	u_1	β	ε	-
12	Weibull máximos	$\frac{\beta}{\varepsilon-u_n} \left(\frac{\varepsilon-x}{\varepsilon-u_n}\right)^{\beta-1} \exp\left[-\left(\frac{\varepsilon-x}{\varepsilon-u_n}\right)^\beta\right]$	u_n	β	-	ε

VALOR ESPERADO DE UMA VARIÁVEL ALEATÓRIA

A média de uma variável aleatória, também denominada expectância, esperança matemática ou simplesmente valor esperado, é definido pela equação A.11 para o caso de VAs discretas, e pela A.12 para VAs contínuas.

$$\mu = E(X) = \sum_i x_i P(x_i) \tag{A.11}$$

$$\mu = E(X) = \int_{-\infty}^{+\infty} xf(x)dx \quad (\text{A.12})$$

Este valor é usado para caracterizar o centro da distribuição de probabilidades, e representa qual seria a média dos valores assumidos pela variável aleatória após a realização de um grande número de experimentos. O operador valor esperado exhibe ainda as seguintes propriedades (Neto e Cymbalista, 1978):

a) $E(k) = k$

A média de uma constante é a própria constante.

b) $E(kX) = kE(X)$

Se os valores de uma VA forem multiplicados por uma constante, sua média também será.

c) $E(X + Y) = E(X) + E(Y)$

A média de uma soma de VAs é igual à soma de duas médias.

d) $E(X + k) = E(X) + k$

Se uma constante for somada aos valores de uma VA, sua média será acrescida dessa constante.

A mediana, que é o ponto que divide a distribuição de probabilidade em duas partes equiprováveis, e a moda, que é o ponto de maior probabilidade da distribuição, assim como a média, são medidas de tendência central. A variância e o desvio padrão de uma variável aleatória são medidas de dispersão, ou seja, representam o quão espalhados em torno do valor esperado estão os valores da VA. Define-se a variância σ^2 como:

$$\sigma^2 = E[(X - \mu)^2] \quad (\text{A.13})$$

Uma forma conveniente de escrever a mesma variância é dada pela equação (A.14)

$$\sigma^2 = E(X^2) - [E(X)]^2 \quad (\text{A.14})$$

onde $E(X^2)$ é dado, para VAs discretas e contínuas, como:

$$E(X^2) = \sum_i x_i^2 P(x_i) \quad (\text{A.15})$$

$$E(X^2) = \int_{-\infty}^{+\infty} x^2 f(x) dx \quad (\text{A.16})$$

Fica claro aqui que o valor esperado $E()$ se comporta como um operador. A variância também exhibe algumas propriedades de interesse, conforme listadas abaixo (Neto e Cymbalista, 1978):

a) $\sigma^2(k) = 0$

A variância de uma constante é nula.

b) $\sigma^2(kX) = k^2 \sigma^2(X)$

Ao se multiplicar os valores de uma VA por uma constante, sua variância é multiplicada pelo quadrado dessa constante.

c) $\sigma^2(X + Y) = \sigma^2(X) + \sigma^2(Y)$

A variância da soma de duas VAs independentes é igual à soma de suas variâncias.

d) $\sigma^2(X + k) = \sigma^2(X)$

Se uma constante for somada aos valores de uma VA, sua variância não é afetada.

O desvio padrão de uma VA, σ , é a raiz quadrada de sua variância. Ainda, seu coeficiente de variação é definido como a razão entre seu desvio padrão e sua média:

$$cv = \frac{\sigma}{\mu} \quad (\text{A.17})$$

DISTRIBUIÇÕES CONJUNTAS E MOMENTOS CONJUNTOS

Para cada evento no espaço amostral de um experimento se associa um número que será assumido pela VA X . Da mesma forma, uma VA Y pode ser definida sobre o mesmo espaço. A probabilidade de que o valor assumido por cada uma das variáveis seja menor do que valores específicos ao mesmo tempo é descrita por sua *função conjunta de distribuição cumulativa probabilidades*, análoga à FDA, porém para mais de uma variável concomitantemente:

$$F(x, y) = P[\{X \leq x, Y \leq y\}] \quad (\text{A.18})$$

Este evento pode ser obtido através do produto dos dois eventos marginais (Beck 2012):

$$\{X \leq x\} \cdot \{Y \leq y\} = \{X \leq x, Y \leq y\} \quad (\text{A.19})$$

Assim como ocorre com as funções que envolvem apenas uma VA, a *função conjunta de densidade de probabilidade*, análoga à FDP, pode ser obtida derivando-se a função conjunta de distribuição cumulativa probabilidades, porém uma vez em cada variável envolvida:

$$f(x, y) = \frac{\partial^2 F(x, y)}{\partial x \partial y} \quad (\text{A.20})$$

E a partir dela, pode-se determinar o conteúdo de probabilidade em qualquer domínio D :

$$P[\{(x, y) \in D\}] = \iint_D f(x, y) dx dy \quad (\text{A.21})$$

Assim como acontece com dois eventos, duas variáveis aleatórias podem apresentar ou não dependência entre si. Elas serão independentes quando a probabilidade de uma delas assumir qualquer valor não seja influenciada, em nada, pelo valor que a outra assume, qualquer que ele seja (Baley, 1971). Ou seja, para quaisquer (x, y) :

$$P[\{X = x|Y = y\}] = P[\{X = x\}] \quad (\text{A.22})$$

Quando as variáveis são independentes, as funções de distribuição conjuntas podem ser obtidas multiplicando-se as funções de distribuição marginais:

$$\begin{aligned} F(x, y) &= F(x)F(y) \\ f(x, y) &= f(x)f(y) \end{aligned} \quad (\text{A.23})$$

No caso de variáveis aleatórias dependentes, pode ser importante saber qual é a relação entre elas. Em casos práticos, é difícil ter essa informação detalhada. Uma medida de correlação mais comum de se ter disponível é a covariância. Trata-se de uma medida linear de correlação entre duas VAs. A covariância é definida como:

$$\text{Cov}[X, Y] = E[(X - \mu_x)(Y - \mu_y)] \quad (\text{A.24})$$

É importante ressaltar que, por se tratar de uma medida linear, é possível que duas variáveis aleatórias sejam dependentes de forma não linear, porém a covariância falhe em indicar isso. A correlação de uma variável aleatória com ela mesma resulta em sua própria variância. É possível ainda definir momentos conjuntos para pares de variáveis aleatórias. Os momentos conjunto de ordem $(k + l)$ de duas VAs X e Y são definidos como:

$$\mu^{kl} = E[X^k Y^l] = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} x^k y^l f(x, y) dx dy \quad (\text{A.25})$$

Os momentos centrais de ordem $(k + l)$ são:

$$m^{kl} = E[(X - \mu_x)^k (Y - \mu_y)^l] = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} (x - \mu_x)^k (y - \mu_y)^l f(x, y) dx dy \quad (\text{A.26})$$

PROCESSOS ESTOCÁSTICOS

Diversas situações de interesse envolvem observações de fenômenos aleatórios feitas não apenas pontualmente, mas ao longo de um contínuo. Quando um fenômeno varia de forma incerta ao longo do tempo, ele pode ser considerado um processo estocástico (Clarke e Disney, 1979). Ainda que seja mais comum pensar em processos estocásticos em função do tempo, a teoria se aplica a qualquer parâmetro indexador, de modo que um processo

estocástico pode ser definido sobre uma distância, um ângulo, ou um parâmetro abstrato contínuo ou discreto (Stark e Woods, 1994). Aplicações comuns que podem ser descritas por processos estocásticos incluem sinais de áudio e de vídeo transmitidos a um ponto distante, a quantidade remanescente de um determinado produto no estoque de uma loja ao fim de cada dia e a intensidade de cargas ambientais que atuam sobre estruturas offshore, entre outros exemplos.

Dado um espaço amostral S , sobre o qual estão definidas probabilidades, define-se um mapeamento que leva de S a um espaço de funções contínuas de um parâmetro (como o tempo). Estas funções são chamadas de *funções amostrais*, e este mapeamento é um processo estocástico. Para um valor fixo do parâmetro, o processo se reduz a uma variável aleatória, e para um ponto fixo do espaço de probabilidades, a uma função do parâmetro. Ainda, se o processo estiver definido sobre não apenas um parâmetro unidimensional, mas um espaço paramétrico multidimensional, ele pode ser também denominado campo aleatório. As Figuras (A.1) a (A.3) ilustram exemplos de realizações de processos estocásticos. Em (A.1) está representado um processo indexado sobre um parâmetro discreto, e, portanto, um *processo estocástico de estado discreto*. Processos deste tipo são também chamados de *cadeias aleatórias*. Trata-se de um exemplo hipotético no qual se observa o número de peças que estavam no depósito de uma loja ao fim de cada dia de um mês. Em (A.2), o parâmetro tempo é tomado como contínuo, sendo assim um *processo estocástico de estado contínuo*. A Figura (A.3) mostra um exemplo de processos estocásticos típicos no contexto da engenharia de estruturas.

Dada uma coleção finita de n valores do parâmetro indexador (a partir daqui, o parâmetro será considerado como sendo o tempo), observa-se um conjunto de n variáveis aleatórias X_1, X_2, \dots, X_n que seguem uma distribuição conjunta de probabilidades. A ideia básica de se analisar um processo estocástico é a determinação destas distribuições conjuntas, e a utilização dessas informações para prever estados futuros do sistema ($t > n$) com base nas observações já realizadas (Clarke e Disney, 1979).

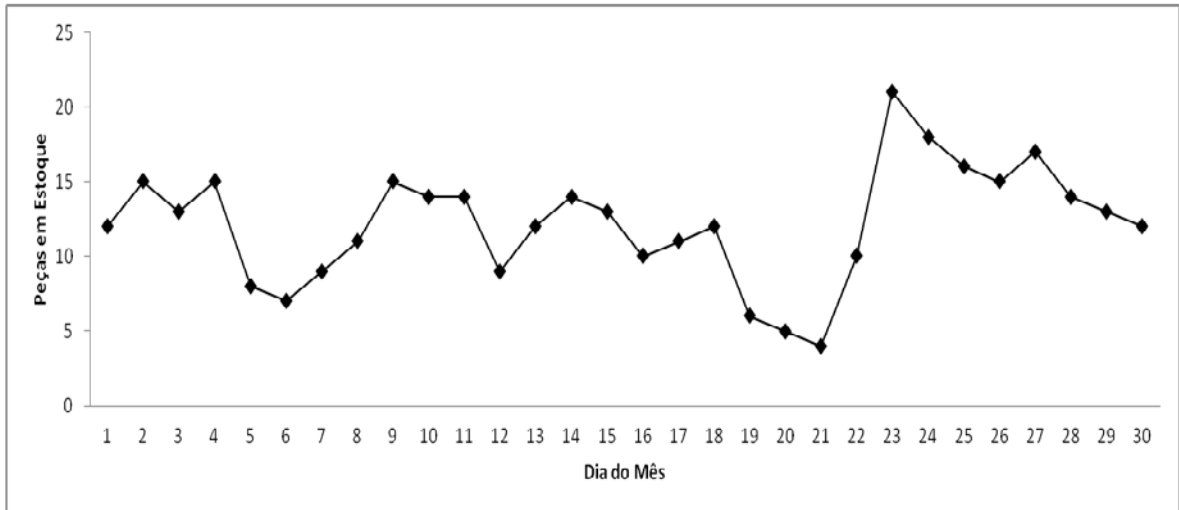


Figura A.1 – Número de peças remanescentes no estoque de uma loja

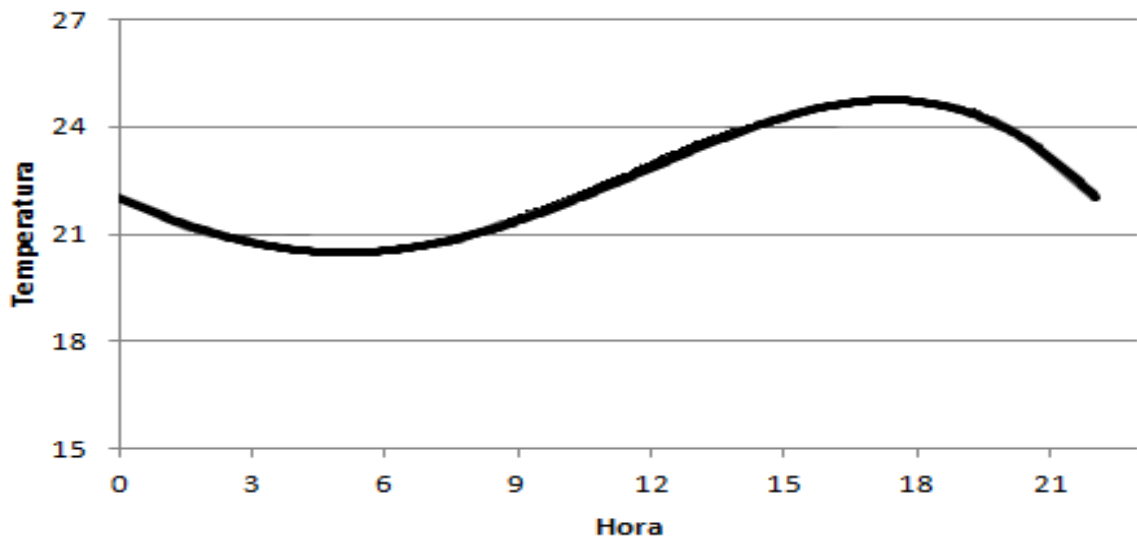


Figura A.2 – Temperatura em São Carlos no dia 16/02/2015

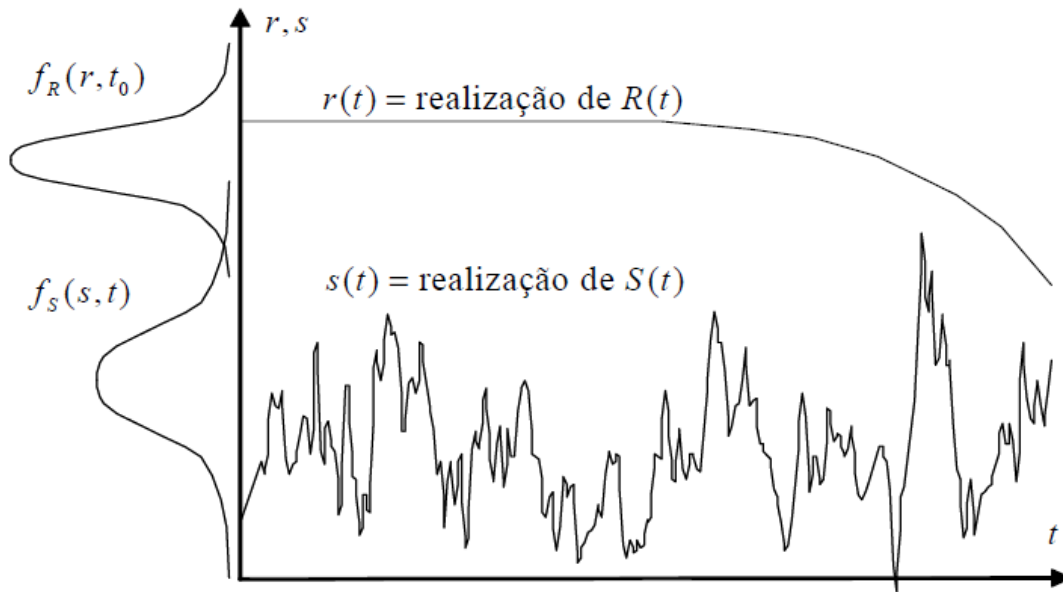


Figura A.3 – Problema de confiabilidade estrutural envolvendo carregamento estocástico e variação da resistência no tempo – Fonte: (Beck, 2012)

Assim como as variáveis aleatórias, os processos estocásticos têm momentos definidos, de forma bastante semelhante. A diferença básica é o fato de que as funções de probabilidade das VAs analisadas ao se fixar um tempo dependem dele. Ou seja:

$$F(x, t) = P\{X(t) \leq x\} \quad (\text{A.27})$$

Da mesma forma, a distribuição conjunta das variáveis que são definidas ao se fixar dois instantes, irá depender de cada um deles:

$$F(x_1, x_2; t_1, t_2) = P\{X(t_1) \leq x_1, X(t_2) \leq x_2\} \quad (\text{A.28})$$

Desse modo, o momento de ordem k de um processo estocástico é escrito da forma:

$$\mu^k(t) = \int_{-\infty}^{+\infty} x^k f(x, t) dx \quad (\text{A.29})$$

Note que para $k = 1$, tem-se o valor esperado.

Existem ainda outras duas funções importantes no que diz respeito ao estudo dos processos estocásticos, que tratam da relação que existe entre pares de variáveis aleatórias, a *função de auto-correlação* e a *função de auto-covariância*. A primeira se refere ao grau de correlação de uma variável

aleatória em um dado instante com outra produzida pelo mesmo processo estocástico ao ser observado instantes depois. Sua determinação consiste no cálculo do momento conjunto das duas variáveis aleatórias, obtidas a partir da fixação de dois pontos no tempo:

$$R_{XX}(t_1, t_2) = E[X(t_1), X(t_2)] \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} x_1 x_2 f(x_1, x_2; t_1, t_2) dx \quad (\text{A.30})$$

Já a função de auto-covariância costuma ser denotada por C_{xx} , e representa a covariância entre $X(t_1)$ e $X(t_2)$. Esta função pode ser obtida a partir da função de auto-correlação:

$$\begin{aligned} C_{XX}(t_1, t_2) &= E[(X(t_1) - \mu(t_1))(X(t_2) - \mu(t_2))] = \\ &= R_{XX}(t_1, t_2) - \mu(t_1)\mu(t_2) \end{aligned} \quad (\text{A.31})$$

Alguns processos estocásticos apresentarão o mesmo valor de auto-correlação para quaisquer dois instantes que sejam fixados, desde que a distância entre eles seja fixa. Estes constituem um caso particular de especial importância, e recebem o nome de processos estacionários. Para o caso multidimensional, é também usual denominá-los campos homogêneos. A expressão (A.32) mostra esta relação:

$$|t_1 - t_2| = |t_3 - t_4| = \tau \Rightarrow R_{XX}(t_1, t_2) = R_{XX}(t_3, t_4) \quad (\text{A.32})$$

Outro caso particular de especial importância são os processos nos quais todas as suas variáveis aleatórias, para qualquer tempo fixado, são conjuntamente normais. Isso significa que a FDP de uma variável qualquer deve ser da forma:

$$f(x, t) = \frac{1}{\sqrt{2\pi\sigma(t)}} \exp\left[-\frac{1}{2}\left(\frac{x - \mu(t)}{\sigma(t)}\right)^2\right] \quad (\text{A.33})$$

É importante ressaltar que todas as $X(t)$ respeitarem a expressão (A.33) não é suficiente para que o processo estocástico seja considerado gaussiano. Para isso, todas as VAs precisam apresentar função conjunta de densidade de probabilidade gaussiana. Duas variáveis $X(t_1), X(t_2)$, por exemplo, devem necessariamente obedecer:

$$f(x_1, x_2; t_1, t_2) = \frac{1}{2\pi\sigma(t_1)\sigma(t_2)\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2(1-\rho^2)}\lambda\right) \quad (\text{A.34})$$

Com:

$$\lambda = \left[\frac{(x_1 - \mu(t_1))^2}{\sigma(t_1)^2} + \frac{(x_2 - \mu(t_2))^2}{\sigma(t_2)^2} - \frac{2\rho(x_1 - \mu(t_1))(x_2 - \mu(t_2))}{\sigma(t_1)\sigma(t_2)} \right] \quad (\text{A.35})$$

Este tipo de processo é de especial importância para o desenvolvimento de modelos de Krigagem, conforme será abordado no item 4.4.