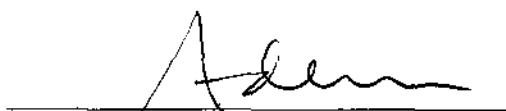


KRIGING DISJUNTIVO USANDO ANAMORFISMO GAUSSIANO - UM MÉTODO
ESTOCÁSTICO PARA PREDIÇÃO DE RESERVAS RECUPERÁVEIS

Este exemplar corresponde à redação final da tese devidamente corrigida e defendida por Paulo Sérgio Lucio e aprovada pela comissão julgadora.

Campinas, 15 de março de 1991



Ademir José Petenate
Orientador

Dissertação apresentada ao Instituto de Matemática, Estatística e Ciência da Computação, UNICAMP, como requisito parcial para obtenção do Título de Mestre em Estatística.

L963k

13651/BC

UNICAMP
BIBLIOTECA CENTRAL

UNIVERSIDADE ESTADUAL DE CAMPINAS
INSTITUTO DE MATEMÁTICA ESTATÍSTICA E CIÊNCIA DA COMPUTAÇÃO
DEPARTAMENTO DE ESTATÍSTICA

KRIGING DISJUNTIVO USANDO ANAMORFISMO GAUSSIANO - UM MÉTODO
ESTOCÁSTICO PARA PREDIÇÃO DE RESERVAS RECUPERÁVEIS

POR

PAULO SÉRGIO LUCIO

DISSERTAÇÃO DE MESTRADO, REQUISITO PARCIAL PARA A OBTENÇÃO DO GRAU DE
MESTRE EM ESTATÍSTICA

CAMPINAS, 15 DE MARÇO DE 1991

306103774

AGRADECIMENTOS

A todos os professores que participaram da minha formação, em especial ao Prof. Ademir José Petenate pela confiança, ao Prof. Armando Zaupa Remacre pelo apoio e ao Prof. Rubem Klein pelo tema.

À tia Maria pela dedicação, aos meus irmãos Carlos Frederico, que sempre chamarei de "Lico", pela paciência e Cinthia pelo carinho.

Aos colegas e amigos, pelo convívio.

Às pessoas que nunca aceitarão que o ser humano tenha limitações, pelo incentivo de terem a mim feito me superar do pouco que acreditava ser capaz.

*Aos meus pais
José Carlos
e Cely, dedico.*

A Rocio com muito carinho.

SUMÁRIO

	Página
INTRODUÇÃO	i-viii
SISTEMAS	SI
CAP.I - VARIÁVEIS REGIONALIZADAS	1-16
I.1 UM MODELO PARA VARIÁVEIS REGIONALIZADAS	1
I.2 HIPÓTESES DE ESTACIONARIDADE	2
I.3 MODELOS DE SEMI-VARIOGRAMAS	7
I.4 FUNÇÃO DE PREDIÇÃO	11
CAP.II - KRIGING LINEAR	17-45
II.1 O MÉTODO	18
II.2 KRIGING SIMPLES	19
II.3 KRIGING ORDINÁRIO	23
II.4 KRIGING UNIVERSAL	28
II.5 ALGUNS RESULTADOS COMPLEMENTARES	31
II.6 A TÉCNICA	34
II.7 CO-KRIGING	41
CAP.III - KRIGING NÃO-LINEAR	46-56
III.1 PROCESSO DE KRIGING DISJUNTIVO	47
III.2 REPRESENTAÇÃO FATORIAL	51
III.3 MODELO ISOFATORIAL	54

CAP.IV - ANAMORFISMO GAUSSIANO	57-67
IV.1 MODELO HERMITIANO	59
IV.2 SISTEMA DE KRIGING DISJUNTIVO HERMITIANO	62
IV.3 CÁLCULO APROXIMADO	65
CAP.V - PREDIÇÃO DA PROBABILIDADE CONDICIONAL	68-81
V.1 FUNÇÕES INDICATRIZES	68
V.2 KRIGING DISJUNTIVO BIGAUSSIANO	71
V.3 FUNÇÃO DE RECUPERAÇÃO	76
V.4 PREDIÇÃO DE RECUPERAÇÃO DE BLOCOS	77
V.5 PREDIÇÃO POR CONDICIONAMENTO UNIFORME	80
CAP.VI - PREDIÇÃO DE RESERVAS RECUPERÁVEIS	82-96
VI.1 RESERVAS RECUPERADAS	82
VI.2 RELAÇÃO DE CARTIER	86
VI.3 FUNÇÃO DE RECUPERAÇÃO PONTUAL	87
VI.4 FUNÇÃO DE RECUPERAÇÃO DE BLOCOS	88
VI.5 EFEITO DE SUPORTE	91
VI.6 MÉTODO PARA PREDIÇÃO DE RECUPERAÇÃO PONTUAL	92
VI.7 PREDIÇÃO POR DISTRIBUIÇÃO CONDICIONAL	94
CONSIDERAÇÕES FINAIS	97-101
APÊNDICE A	a1-a2
REFERÊNCIAS BIBLIOGRÁFICAS	rb1-rb3

LISTA DE SISTEMAS E FÓRMULAS

	pagina
(SK): Sistema de <i>Kriging Simples</i>	20
(OK): Sistema de <i>Kriging Ordinário</i>	25
(UK): Sistema de <i>Kriging Universal</i>	29
$\Gamma_o = \Delta_o \Lambda_o$: (OK)	40
$\Gamma_U = \Delta_U \Lambda_U$: (UK)	40
D=UA: Sistema de <i>Co-Kriging</i>	44
C=VA: Sistema de <i>Co-Kriging</i> (semi-variogrâmico)	44
(DK): Sistema de <i>Kriging Disjuntivo</i>	50
(DK_ϕ) : Sistema (DK) para ϕ Mensurável	51
(S): Representação Iso-Fatorial da Esperança Condicional	54
(DK_R) : Sistema (DK) Reescrito em Termos de (S)	55
(CK): Sistema para os Coeficientes de (DK_R)	55
(D): Mínima Variância de DK	63
(F): Condição de Ortogonalidade do Preditor de DK	63
(C-K): Sistema de <i>Co-Kriging Indicador</i>	70
(FF): Fórmula Fundamental	72
Relação de Cartier	86

INTRODUÇÃO

Sob sua conotação etimológica, *geoestatística* deveria denotar a aplicação da estatística em problemas de Ciências da Terra (Geociências). Mas em geologia, e áreas afins, esta é definida de maneira mais formal: "*Geoestatística* é a aplicação do formalismo de funções aleatórias ao reconhecimento e predição de fenômenos naturais.". A *geoestatística* está, desta forma, naturalmente envolvida em problemas de predição espacial.

Fenômenos naturais podem frequentemente ser caracterizados pela distribuição no espaço de um certo número de quantidades mensuráveis chamadas variáveis regionalizadas. Uma variável regionalizada é uma função $Z(x)$ de um ponto de coordenadas x em um espaço finito-dimensional. Uma das principais características dessas variáveis em geociências é que sua variação no espaço é localmente "caótica", o que impossibilita um estudo determinístico direto da função de interesse $z(x)$ e isso faz com que esta medida seja interpretada como uma realização de uma função aleatória $Z(x)$. O esforço estatístico se concentra em encontrar uma estrutura no "caos". Contudo, sob um ambiente aparentemente "caótico", uma certa estrutura na variabilidade espacial de $Z(x)$ é usualmente perceptível, que é determinada por um padrão de correlação entre as diversas variáveis aleatórias $Z(x)$.

Uma função aleatória é definida por suas distribuições finitamente-dimensional, isto é, pelo conjunto de todas as distribuições conjuntas acumuladas. É clara a impossibilidade de inferir sobre todas estas distribuições através de um número finito de dados. Mas esta informação completa, nunca é necessária; na prática o mínimo necessário para propiciar as informações desejáveis sobre o que

é de estudo, é o conhecimento da função semi-variograma, que é a variância da diferença de conteúdo em duas locações amostrais, ou seja um dos momentos de segunda ordem.

A análise estrutural de um fenômeno regionalizado é o processo de análise de dados que leva a um modelo de semi-variograma. Este age como uma quantidade sumária de toda informação avaliável na estrutura de variabilidade do fenômeno. Esta sumarização da informação é então um direcionamento de todos os passos subsequentes ao reconhecimento de reservas à predição. A modelagem e interpretação do semi-variograma requerem mais que programas computacionais. É essencial um bom conhecimento sobre a física do fenômeno e também processos criteriosos na coleta e análise dos dados .

Relativo a um modelo de covariância, um modelo de semi-variograma não pode ser uma função qualquer; deve ser assegurado a positividade no cálculo de todas as variâncias, isto é, este deve ser relacionado a um modelo de covariância, que por sua vez deve ser uma função "condicionalmente positiva-definida". Na prática investiga-se um modelo que seja uma combinação linear positiva de funções básicas que são positivas-definidas.

Como mencionado anteriormente, objetiva-se a predição local. Quando se prediz um valor desconhecido $Z(x)$ por alguma combinação $Z^*(x)$ dos dados disponíveis, um erro é cometido. Interpretado como uma variável aleatória, este erro é associado a uma função de distribuição cujas principais características de interesse prático são a média e a variância. O critério, em geral, utilizado para medir a qualidade de um preditor $Z^*(x)$ é o vício e a variância de predição. Sob estes aspectos, para verificar tal qualidade deve-se garantir a não-viciabilidade. Dentre os preditores não-viciados é desejável encontrar um cuja variância de predição seja mínima. Quando se lida com Ciências da Terra (não só), a prática tem mostrado que em casos "muito bem comportados" a distribuição dos erros experimentais, embora não sendo Normal, tende a ser simétrica em torno de uma média zero (não-viciabilidade), com erros que oscilam em torno de valores muito pequenos (o que, naturalmente,

poderia ser obtido de uma distribuição Normal com média zero e variância σ_E^2).

Uma técnica clássica de predição linear de processos estocásticos envolvendo variáveis regionalizadas é o *Kriging*, que proporciona o melhor preditor linear não-viciado (BLUP) de algum valor desconhecido $Z(x)$ ou valor médio $Z_V(x)$. As informações utilizadas são os n dados observados $z(x_\alpha)$ ou $z_V(x_\alpha)$ na vizinhança do valor a ser predito. *Kriging* é, em suma, um método de interpolação de processos espaciais aleatórios, chamado assim após seu precursor D. G. Krige. Apesar de se ter notícias de alguns trabalhos de Kolmogorov envolvendo tal técnica datados da primeira metade do século, a idéia fora desenvolvida na década de 50 como uma solução de importantes problemas práticos de como se fazer a predição do conteúdo médio de um painel, tais como, volumes de metal e/ou de rocha que deveriam ser extraídos de um painel se este fosse suficientemente rico. Subsequentemente (depois de ter tido acesso a alguns trabalhos de D. G. Krige) G. Matheron desenvolveu a Teoria de Variáveis Regionalizadas com o objetivo de proporcionar uma estrutura formal para a predição. A *Geoestatística* tornou-se assim a interpretação probabilística desta Teoria. Cabe, aqui um adendo, D. G. Krige era um Engenheiro de Minas das províncias auríferas do RAND (África do Sul), sendo quem primeiro introduziu o uso de "médias móveis" com o objetivo de evitar uma sobre predição de depósitos minerais.

O problema em si pode ser proposto da seguinte forma: "Deseja-se predizer o valor de Z_V , o conteúdo médio do atributo Z no interior de um painel V tal que $Z_V = (1/V) \int_V Z(x) dx$, onde $Z(x)$ é a função aleatória que descreve a densidade de minério na locação com coordenadas vetoriais x , e V é usado para denotar o painel e seu volume (em uma ou mais dimensões)". Os dados utilizados para predizer Z_V são n observações $Z(x_\alpha)$, em locações experimentais x_α onde $\alpha=1,2,\dots,n$. O painel V tem dimensões aproximadas de $100 \times 100 m^2$. Em termos práticos poderia ser um canteiro, uma frente de lavra. O bloco v tem dimensões aproximadas de $10 \times 10 m^2$. Em termos práticos seria a unidade de seleção

de lavra, unidade sobre a qual se toma a decisão do bloco ser ou não lavrável.

Se $Vx \in D$ (dimensão onde se está estudando Z) tem-se $E[Z(x)] = m$ ou seja a média é constante para todas as locações, e a covariância do processo é totalmente especificada, o preditor linear não-viciado de mínima variância é denominado preditor de *Kriging* à média conhecida. Se esta for constante então a técnica utilizada é a de *Kriging Simples*. Se a função de covariância não for especificada, podendo esta ser inferida, e supondo estacionaridade de segunda ordem, a técnica é a de *Kriging Ordinário*. A técnica de *Kriging Universal* permite a predição de um fenômeno não-estacionário. A função aleatória $Z(x)$ é, neste caso, a soma de componentes determinísticos, chamados tendência ou deriva, e uma flutuação intrinsecamente estacionária, com seu próprio padrão de variabilidade espacial caracterizado pelo semi-variograma.

Um preditor linear pode proporcionar, em geral, um preditor viciado em média, que em muitos casos pode ser corrigido. Uma visão que não requer o conhecimento prévio, ou hipóteses sobre a distribuição, usa aproximação por séries polinomiais ortogonais, como de Hermite ou de Laguerre e é conhecido como *Kriging Disjuntivo*, que permite a obtenção de preditores não-lineares e não-viciados, através de métodos numéricos e sequenciais.

Engenheiros de minas e geólogos estão, geralmente, interessados na predição do número de blocos em determinado painel ou zona com teor médio superior a um determinado "ponto de corte" e na quantidade de metal em tais blocos; que são questões de grande importância em tomadas de decisão em gerenciamento econômico. O *Kriging Disjuntivo* fornece uma grande contribuição ao estudo destes problemas. Quantos blocos v se recupera com teor acima de um ponto-de-corte? Qual a quantidade de metal contida nestes blocos assim recuperados?

O conjunto dos recursos de um depósito é em poucos casos explorável, em sua totalidade. Deve-se aplicar uma seleção sobre estes recursos para definir as reservas que serão efetivamente extraídas.

Fica claro que, o volume e as características destas reservas dependem:

- dos recursos "*in situ*" iniciais;
- dos parâmetros da seleção considerada, que são:
 - * critérios de corte;
 - * suporte da unidade de exploração (volume sobre o qual se toma a decisão);
 - * informações disponíveis no momento da seleção efetiva.

A seleção real dispõe, em geral, de uma informação muito mais refinada, que a que se dispõe no momento do estudo de reservas. Este estudo deve prever a efetividade da seleção real futura. Para evitar um vício grave na predição de reservas, é essencial considerar todos estes fatores de forma simultânea.

A decisão de se explorar uma mina é muito importante. Esta necessita não só da predição precisa dos recursos potenciais do depósito, mas também da fração destes recursos ou reservas, que poderão ser recuperados sob certo contexto técnico. Os recursos dependem somente do depósito, enquanto as reservas recuperadas dependem, além do depósito, de todos os parâmetros técnicos da exploração projetada.

Considere um depósito J de extensão geométrica perfeitamente definida. Chama-se recurso "*in situ*" ao conjunto de minério contido em J. Este minério no volume total J é caracterizado por certas propriedades. O minério "*in situ*" assim definido é, na maioria das vezes, inexplorável em sua totalidade. Do conjunto de reservas "*in situ*", somente será recuperado um subconjunto $j \subset J$; denomina-se este conjunto j, efetivamente recuperado pela exploração, por reservas recuperadas.

Sob o ponto de vista técnico, a seleção que permite a obtenção da fração de reservas que são recuperáveis, depende de três efeitos principais:

- i) efeito de suporte (função da seletividade do método de lavra a ser usado);
- ii) efeito de informação (função da malha de controle de lavra);

iii) efeito de restrição geométrica (função da distribuição espacial dos blocos ricos, de parâmetros geotectônicos, etc).

A predição de reservas recuperáveis envolve técnicas paramétricas. Esta portanto se faz a um nível global de um subconjunto do depósito mineral, baseado em um nível local que seja de interesse. A situação típica, neste caso, é a que se segue. O depósito é dividido, sistematicamente, em grandes painéis V , de dimensão correspondente à malha. Objetiva-se prever a reserva recuperável do total do painel, isto é, o minério e o metal que serão recuperados a partir da seleção de blocos (v) do painel. Na seleção livre (quando não se está sob o efeito de restrição geométrica), minério e metal recuperáveis, funções da fração, são, chamados "função de recuperação" (notada originalmente por G. Matheron como "função transfer"). Tais funções são ditas "diretas", se a seleção futura realiza-se sobre os teores reais dos blocos (supõe-se, então, ser perfeitamente contínua no momento da exploração) e aquelas que são ditas seleção "indireta", se são feitas sobre os preditores dos blocos (função do controle de lavra).

A proporção é considerada como uma variável aleatória. Surge, então, a necessidade de prever a distribuição dos teores reais dos blocos v sob um painel V . As informações quantitativas que se dispõe são geralmente constituídas por meio de uma malha de sondagem, de dimensões comparáveis ao painel. A partir dessa informação os métodos de *Kriging Não-Linear* permitem a predição do teor médio de reservas de um painel V . Pode-se preconizar a solução de prever por *Kriging* todos os blocos v de um painel. Porém, a predição de tais blocos é, em geral, muito imprecisa, ficando aquém para o uso em reservas recuperáveis, e ainda menos a prever quais blocos serão selecionados. Com o controle de lavra existente, sob o efeito de informação no momento da seleção se disporá de uma informação muito mais rica do que se dispõe no momento do atual conhecimento.

As funções do minério e do metal que é recuperável acima de um corte z_c , sobre o teor $Z(v_i)$ de um bloco v_i do painel são, respectivamente:

$$T = \mathcal{I}(Z(v_1) \geq z_c)$$

e

$$Q = Z(v_1) \mathcal{I}(Z(v_1) \geq z_c),$$

onde $\mathcal{I}$ é a função indicadora.

A predição diretamente recuperável para os L blocos do painel, posto que a predição do bloco seja invariante, é dada por:

$$T' = \frac{1}{L} \sum_L \mathcal{I}(Z(v_1) \geq z_c)$$

e

$$Q' = \frac{1}{L} \sum_L Z(v_1) \mathcal{I}(Z(v_1) \geq z_c).$$

Para prever tais funções de recuperação (que são funções não-lineares dos teores dos blocos), desenvolveu-se métodos não-lineares adaptados à predição do recuperável. Naturalmente, o preditor do recuperável requer mais atenção que o da reserva "*in situ*". Por consequência métodos não-lineares são mais restritos teoricamente e portanto em compreensão que os lineares. E é uma das razões pela qual tais métodos são tão pouco utilizados.

Ao método não-linear aqui descrito, é atribuído uma hipótese muito estrita de estacionaridade, sob a qual surgiu a necessidade de se fazer uma adaptação.

A essência da *geoestatística* está, não em *Kriging*, mas em:

- i) explicitar a decisão de estacionaridade;
- ii) na escolha dos dados;
- iii) na transformada a ser usada e
- iv) na inferência do correspondente modelo de semi-variograma.

Aplicado à Geociências, muitas vezes o objetivo final é construir um mapa de iso-valores de $Z(x)$ para uma determinada malha x , bem como o de sua variabilidade. A qualidade deste mapa depende do método de predição e ainda da escolha das locações amostradas.

Outros métodos de interpolação espacial incluem o uso de regressão polinomial, e técnicas não estocásticas baseadas em celas de Dirichlet. Mas, a prática tem mostrado vantagens do preditor de *Kriging*, principalmente no que se refere aos métodos computacionais envolvidos.

O objetivo desta dissertação é apresentar um preditor não-linear que seja adequado à predição de certa propriedade espacial de interesse. Assim como fornecer um método simples para o cálculo de uma probabilidade condicional visando a aplicação de uma técnica *geoestatística* a uma decisão de gerenciamento.

A importância deste trabalho se concentra no esforço de formalizar estatística e matematicamente tal objetivo mencionado, bem como o de apresentar e sintetizar alguns métodos *geoestatísticos* numa abordagem inédita apresentada no país. O ponto que entenderia como crucial se concentra no estudo do anamorfismo gaussiano, bem como na sua existência e seu mérito na simplificação de problemas que envolvem a predição não-linear.

Cabe aqui uma nota final no que diz respeito a atentar ao cuidado e à capacidade de discernimento ao se trabalhar com predição não-linear de processos (que é o enfoque desta tese) e predição de processos não estacionários.

CAPÍTULO 1

VARIÁVEIS REGIONALIZADAS

As informações avaliadas num depósito são fragmentárias. Se é desejado obter conclusões sobre alguma quantidade que não tenha sido acessada, se faz necessário um modelo. Muitas vezes uma inspeção revela que várias variáveis não são "completamente aleatórias". Pontos próximos parecem ser relacionados por um complexo conjunto de correlações. Na totalidade existem zonas onde os valores destas correlações tendem a ser menores e outras onde estes são maiores. O termo "variável regionalizada", fora denotado por G. Matheron [1963] para enfatizar a particular natureza de tal variável em estudo, que concatena dois aspectos, um aleatório, contabilizado por irregularidades locais, outro estrutural, que reflete a característica geral do fenômeno.

Uma variável regionalizada é definida, de uma forma sintética, como uma função em um domínio do espaço.

1.1 Um modelo para variáveis regionalizadas

O modelo probabilístico predominante assume que toda a aleatoriedade é devida a "erros", e toda a estrutura é devido à relação funcional denominada "tendência". Para estudar esta alegada tendência, muitos ajustes de mínimos quadrados são usados. Contudo, há casos onde isto não é adequado com o objetivo de explicar esta estrutura.

Um melhor modelo para representar a realidade parece ser introduzir aleatoriedade em termos de flutuações em torno de uma localidade fixa. Flutuações não são erros, mas característica do fenômeno estudado, tendo uma estrutura própria. Deve-se primeiramente,

então, identificar estas propriedades estruturais. Tendo-se feito isto, estar-se-á apto a resolver vários problemas de predição de uma variável regionalizada baseado em uma amostragem fragmentada (conjunto de dados).

Em aplicações da Estatística em Geociências usualmente se modela a medida de um determinado experimento ou observação por meio de uma variável aleatória. No entanto, características peculiares de tais fenômenos levam a modelagens mais complexas tais como funções aleatórias.

1.2 Hipóteses de estacionaridade

Muitas ciências aplicadas estudam fenômenos com caráter aleatório no espaço. Estes fenômenos são geralmente representados por uma função aleatória, ou seja, um processo estocástico, cujo conjunto de índices tem uma ou mais dimensões.

Seja $(\Omega, \mathcal{F}, P)$, um espaço de probabilidades, $(S, \mathcal{S})$ um espaço mensurável e D um conjunto, arbitrários. Um processo estocástico em $(\Omega, \mathcal{F}, P)$ com espaço de estados $(S, \mathcal{S})$ e conjunto de índices D , é uma família de funções mensuráveis $Z(x): (\Omega, \mathcal{F}) \rightarrow (S, \mathcal{S})$; $x \in D$.

Se D for um conjunto enumerável, $S = \mathcal{R}$ e $\mathcal{S} = \mathcal{B}(\mathcal{R})$, o processo é uma sequência de variáveis aleatórias

Definição 1.2.1 Um processo estocástico real é um conjunto de variáveis aleatórias reais indexadas no espaço vetorial finito-dimensional, denotado por $\{Z(x): x \in D \subseteq \mathcal{R}^d\}$.

Uma realização de um processo estocástico espacial é, em verdade, uma trajetória que a um ponto do espaço faz corresponder uma variável aleatória.

Afim de dar consistência à teoria de variáveis regionalizadas faz-se necessário as seguintes notas:

Nota 1- Examinando alguns paralelos na geometria euclidiana finita-dimensional: um espaço linear é um conjunto de elementos, com operações associadas de adição de vetores e produto por escalar. Um produto interno num espaço vetorial é uma função real, denotada por $\langle X, Y \rangle$, de pares de vetores satisfazendo a $\langle X, Y \rangle = \langle Y, X \rangle$, $\langle aX_1 + bX_2, Y \rangle = a\langle X_1, Y \rangle + b\langle X_2, Y \rangle$ e $\langle X, X \rangle \geq 0$ se $X \neq 0$.

Nota 2- Em um espaço com produto interno, pode-se definir a norma de um vetor ou o tamanho deste por $\|X\| = \langle X, X \rangle^{1/2}$. Um espaço de Hilbert é um espaço vetorial com produto interno que é completo na métrica determinada pela norma. Isto significa que se $\{X_n\}$ for uma sequência de vetores, para a qual $\lim_{n,m \rightarrow \infty} \|X_n - X_m\| = 0$, então existe um vetor X para o qual $\lim_{n \rightarrow \infty} \|X_n - X\| = 0$.

$L^2(\Omega, \mathcal{F}, P)$ é um espaço de Hilbert, uma vez que este é composto por funções com quadrado integrável e posto que L^2 é provido de um produto interno igual à covariância não-centrada (que não é necessariamente estacionária),

$$\langle Z(x_\alpha), Z(x_\beta) \rangle = E\{Z(x_\alpha)Z(x_\beta)\},$$

onde $Z(x_\alpha), Z(x_\beta) \in L^2$. Esta covariância caracteriza a correlação espacial entre duas variáveis aleatórias, localizadas em dois diferentes pontos.

Se $E\{Z(x_\alpha)\} = m(x_\alpha)$, a covariância centrada é então dada por:

$$E\{[Z(x_\alpha) - m(x_\alpha)][Z(x_\beta) - m(x_\beta)]\} = E\{Z(x_\alpha)Z(x_\beta)\} - m(x_\alpha)m(x_\beta).$$

A distância entre dois elementos $Z(x_\alpha)$ e $Z(x_\beta)$ é definida como a norma do vetor $(Z(x_\alpha) - Z(x_\beta))$. A norma de um vetor $Z(x_\alpha)$, denotada por $\|Z(x_\alpha)\|$, é definida como $\langle Z(x_\alpha), Z(x_\alpha) \rangle^{1/2}$.

Um processo estocástico real em L^2 é uma família de variáveis aleatórias reais $Z(x)$, $x \in D$, tal que $Z(x)^2 = E|Z(x)|^2 < \infty$, $\forall x \in D$.

Seja $\{Z(x): x \in D \subseteq \mathcal{R}^d\}$, um processo em L^2 . A função de covariância do processo é definida por $C(x,y) = \text{Cov}(Z(x), Z(y))$, $y \in D$. A $C(x,y)$ é o produto interno de $(Z(x) - E[Z(x)])$ e $(Z(y) - E[Z(y)])$ em $L^2(\Omega, \mathcal{F}, P)$, sendo que todo processo em L^2 tem uma função de covariância.

Uma condição necessária para a existência em L^2 de um processo com uma função de covariância específica é que $C(x,y)$ seja simetricamente Hermitiana, ou seja, $C(x,y) = C(y,x)$, $\forall x,y \in D$, e a matriz de covariâncias seja positiva semi-definida.

Seja x um vetor em $\mathcal{R}^d$ e $z(x)$ o valor da função de interesse em x ; esta função é dita uma variável regionalizada, que pode ser interpretada como uma realização de um processo estocástico $\{Z(x): x \in D \subseteq \mathcal{R}^d\}$.

Definição 1.2.2 Seja $\{Z(x): x \in D \subseteq \mathcal{R}^d\}$ um processo estocástico, suponha que este seja observado em n locações espaciais $\{x_i: i=1, \dots, n\}$, resultando uma variável aleatória multivariada $\underline{Z}(x) = (Z(x_1), \dots, Z(x_n))$. Desta forma assume-se que os dados, definidos como variáveis regionalizadas, e observados nas locações espaciais, sejam de fato uma realização de $\underline{Z}(x)$, cuja gênese está no processo estocástico.

Um processo estocástico é caracterizado por suas distribuições finitamente-dimensionais, isto é, a distribuição conjunta de qualquer conjunto de variáveis $Z(x_1), \dots, Z(x_k)$, $\forall k=1, \dots, n$. Certamente o modelo probabilístico permanecerá uma simples conjectura se não for possível constituir estas distribuições, ou pelo menos conhecer alguns de seus momentos. Em geral, quanto menos realizações disponíveis, a predição estatística requer um número maior de suposições. A suposição de variabilidade espacial pode ser feita diretamente em termos das hipóteses básicas sob o processo estocástico. Estas

hipóteses adicionais devem reduzir o número de determinações das quais o processo estocástico depende. A regra então é introduzir um número mínimo de hipóteses, tal que torne viável a predição em situações práticas.

O processo estocástico $\{Z(x): x \in D \subseteq \mathcal{R}^d\}$ é dito ser Gaussiano, se, e somente se, para cada $x_1, \dots, x_n \in D$, $(Z(x_1), \dots, Z(x_n))$ é conjuntamente gaussiano, ou seja, o vetor aleatório $(Z(x_1), \dots, Z(x_n))$ tem distribuição Normal multivariada.

Uma hipótese usual introduzida à teoria de processos estocásticos é a de estacionaridade. Atenta-se ao seguinte: a estacionaridade refere-se ao processo e não aos dados.

Estacionaridade estrita

$\{Z(x): x \in D \subseteq \mathcal{R}^d\}$ é dito estritamente estacionário se, para todo número finito de pontos $x_1, \dots, x_n$ e algum $h \in D$ tal que $h+x \in D$, a distribuição conjunta de $Z(x)$ é a mesma que a de $Z(x+h)$; ou seja o processo é invariante sob translação no conjunto de índices.

Esta não é a forma de estacionaridade que é usualmente referida em *geoestatística* (principalmente em técnicas de predição linear) pois esta é muito restrita, mas é essencialmente tal forma que é requerida para a aplicação de técnicas não-lineares tais como *Kriging Probabilístico*, *Kriging Indicador* e *Kriging Disjuntivo*. Note que esta forma de estacionaridade não implica na existência de médias, variâncias ou covariâncias.

Estacionaridade de segunda ordem

$\{Z(x): x \in D \subseteq \mathcal{R}^d\}$ é estacionário de segunda ordem ou de covariância estacionária se $\text{Cov}[Z(x+h), Z(x)]$ existe e é função apenas do vetor diferença de posição, ou seja, $\forall h \in D$, $\text{Cov}[Z(x+h), Z(x)] = C(h)$. Isto implica que a $\text{Var}[Z(x)]$ existe e não depende de x .

Notação: $\text{Var}[Z(x)] = C(0)$.

Para um processo Gaussiano a função de covariância determina completamente todas as distribuições finita-dimensionais. Assim, estacionaridade no sentido fraco é equivalente à estacionaridade estrita.

Como muitas vezes a variância de $Z(x)$ não é finita é conveniente em muitos casos, assumir que apenas os incrementos da função sejam estacionários de segunda ordem.

Estacionaridade intrínseca

A forma de estacionaridade implicada pela hipótese intrínseca é essencialmente estacionaridade de segunda ordem, não para o processo estocástico $\{Z(x): x \in D \subseteq \mathcal{R}^d\}$, mas para a diferença de primeira ordem $[Z(x+h)-Z(x)]$. Como definido por Matheron, $\{Z(x): x \in D \subseteq \mathcal{R}^d\}$ satisfaz a hipótese intrínseca se, $E[Z(x+h)-Z(x)] = 0$, $\forall h \in D$ e

$$\gamma(h) = \frac{1}{2} \text{Var}[Z(x+h)-Z(x)],$$

ou seja o segundo momento depende apenas da diferença da locação amostrada.

O $\gamma(h)$ define uma função denominada semi-variograma, que determina a estrutura espacial do processo estocástico, sendo assim de grande importância em *geoestatística*, no que diz respeito à variabilidade espacial de tal processo.

Estacionaridade de uma forma ou de outra é uma suposição característica dos processos estocásticos em *geoestatística*. Na prática frequentemente acontece que estas suposições não sejam satisfeitas. Quando o fenômeno mostra uma tendência sistemática, não pode ser assumido que a média seja uma constante. Veja maiores informações em G. Matheron [1983], onde se faz um estudo de função de covariância generalizada.

Teorema 1.2.1 Se $\{Z(x): x \in D \subseteq \mathbb{R}^d\}$ é um processo estocástico de covariância estacionária, então este é intrinsecamente estacionário.

Dem: $2\text{Cov}[Z(x), Z(x+h)] = \text{Var}[Z(x)] + \text{Var}[Z(x+h)] - \text{Var}[Z(x+h) - Z(x)]$

$$2C(h) = 2C(0) - \text{Var}[Z(x+h) - Z(x)]$$

$$2\gamma(h) = \text{Var}[Z(x+h) - Z(x)]. \blacksquare$$

Em geral, a recíproca deste teorema não é verdadeira. Seja $\{W(t): t \geq 0\}$ o processo de Wiener. A variável $[W(t+h) - W(t)]$ é normalmente distribuída com média zero e variância $\sigma^2 h$, onde $2\gamma(h) = |h|$. $\{W(t): t \geq 0\}$ é um processo intrinsecamente estacionário, mas, como $\text{Var}[W(x)] = \sigma^2 x$ e $\text{Cov}[W(t), W(t+h)] = \min(t, t+h)$ onde $h \geq 0$, esta não é uma função apenas de $|h|$, portanto este processo não é de covariância estacionária.

Teorema 1.2.2 Seja $\{Z(x): x \in D \subseteq \mathbb{R}^d\}$ um processo estocástico de covariância estacionária, então $\gamma(h) = C(0) - C(h)$.

Dem: $2\gamma(h) = E[Z(x+h) - Z(x)]^2 - \{E[Z(x+h) - Z(x)]\}^2 = E[Z(x+h) - Z(x)]^2$

$$2\gamma(h) = E[Z(x+h)]^2 - 2C(h) - 2\{E[Z(x)]\}^2 + C(0) + \{E[Z(x)]\}^2$$

$$2\gamma(h) = \text{Var}[Z(x+h)] + C(0) - 2C(h)$$

$$2\gamma(h) = 2C(0) - 2C(h). \blacksquare$$

1.3 Modelos de semi-variograma

A função semi-variograma $\gamma(h)$ é uma função do vetor h , módulo e direção. As propriedades do semi-variograma como uma estrutura são sumarizadas a seguir:

i) $\gamma(0) = 0$;

ii) $\gamma(h) = \gamma(-h)$, (para verificar basta substituir x por $x'-h$);

iii) $\gamma(\infty) = C(0)$, $\gamma(\infty)$ é dito *patamar*;

iv) $|\gamma(h)| \leq C(0)$, (para verificar usa-se a desigualdade de Schwarz).

Uma condição necessária e suficiente para C expressar a função covariância de $Z(\mathbf{x})$, é que $C(h)$ seja positiva semi-definida.

Seja $\sum_1^n \lambda_\alpha Z(\mathbf{x}_\alpha)$, uma combinação linear de $Z(\mathbf{x}_\alpha)$'s. Sabe-se, que

$$\text{Var} \left[\sum_1^n \lambda_\alpha Z(\mathbf{x}_\alpha) \right] \geq 0.$$

$$\text{Var} \left[\sum_1^n \lambda_\alpha Z(\mathbf{x}_\alpha) \right] = \sum_1^n \text{Var} \left[\lambda_\alpha Z(\mathbf{x}_\alpha) \right] + \sum_{1 \leq \alpha \neq \beta \leq n} \sum_1^n \text{Cov} \left[\lambda_\alpha Z(\mathbf{x}_\alpha), \lambda_\beta Z(\mathbf{x}_\beta) \right]$$

$$\text{Var} \left[\sum_1^n \lambda_\alpha Z(\mathbf{x}_\alpha) \right] = \sum_1^n \lambda_\alpha^2 \text{Var} \left[Z(\mathbf{x}_\alpha) \right] + \sum_{1 \leq \alpha \neq \beta \leq n} \sum_1^n \lambda_\alpha \lambda_\beta \text{Cov} \left[Z(\mathbf{x}_\alpha), Z(\mathbf{x}_\beta) \right]$$

$$\text{Var} \left[\sum_1^n \lambda_\alpha Z(\mathbf{x}_\alpha) \right] = \sum_1^n \sum_1^n \lambda_\alpha \lambda_\beta \text{Cov} \left[Z(\mathbf{x}_\alpha), Z(\mathbf{x}_\beta) \right] = \sum_1^n \sum_1^n \lambda_\alpha \lambda_\beta C(\mathbf{x}_\alpha - \mathbf{x}_\beta) \geq 0 \quad \blacksquare$$

Então por definição $C(\cdot)$ deve ser do tipo positiva semi-definida.

Por outro lado, o Teorema de Bochner fornece a condição suficiente para uma função ser do tipo positiva: "Uma função é do tipo positiva se esta for a transformada de Fourier de uma medida positiva somável."

Se $C(0)$ não for finita, assume-se a hipótese que o processo estocástico seja intrinsecamente estacionário e no lugar da variância de $\sum_1^n \lambda_\alpha Z(\mathbf{x}_\alpha)$, seja definida uma função sob a condição de que $\sum_1^n \lambda_\alpha = 0$ tal que

$$-\sum_1^n \sum_1^n \lambda_\alpha \lambda_\beta \gamma(\mathbf{x}_\alpha - \mathbf{x}_\beta) \geq 0.$$

Como o argumento h do semi-variograma é um vetor, o semi-variograma deve ser calculado ao longo de várias direções no espaço. É suficiente avaliar apenas os pares de pontos que têm uma orientação definida.

A variabilidade observada pode ser descrita por mais de um tipo de semi-variograma, ou seja uma composição destes.

Modelo de Anisotropia

Diz-se que um semi-variograma é anisotrópico, se este não exibe o mesmo comportamento em todas as direções. Um modelo é isotrópico se o semi-variograma associado a este não depende da direção do vetor $h \in D \subseteq \mathcal{R}^d$, mas apenas de seu comprimento, isto é $\gamma(h) = \gamma(|h|)$.

Os modelos anisotrópicos são divididos em dois grupos:

- i) modelo geométrico: pode ser reduzido a um modelo isotrópico por meio de uma transformação linear; $\gamma(h) = \gamma(h')$, onde $h' = Ah$ e A é a matriz de transformação;
- ii) modelo zonal ou estratificado: é o tipo mais complexo de anisotropia a ser tratada, pois na impossibilidade de se encontrar uma transformação linear, tenta-se a sua correção através de uma composição de modelos geométricos, tratados separadamente.

Efeito pepita

É um modelo teórico que procura descrever fenômenos tais que em torno de distâncias muito pequenas haja uma brusca queda da covariância entre os conteúdos estocásticos (as variáveis apresentam variabilidade não percebida pelo suporte amostral), ou seja, no semi-variograma é detectado uma descontinuidade na origem, que é causada apenas por um parâmetro de escala; este é denotado por C_0 .

O domínio onde a correlação ainda está em mutação é chamado zona de influência. Os semi-variogramas com esta propriedade são modelados como de transição. A diferença entre os modelos de transição são expressas por seu comportamento na origem, posto que a modelagem é feita tendo em vista pequenas distâncias.

Modelando a variabilidade espacial do processo

$$\text{Seja a definição: } \mathcal{F}_{(a,b)}(h) = \begin{cases} 1 & \text{se } h \in (a,b) \\ 0 & \text{c.c.} \end{cases},$$

onde (a,b) é um intervalo real.

Os modelos de transição mais comuns são: ("a" é o *alcance* ou seja o menor valor de h ao qual o semi-variograma atinge o *patamar*)

i) comportamento linear na origem (variáveis com maior variabilidade espacial)

i1) modelo esférico

$$\gamma(h) = \{C(0)[1.5h/a + 0.5h^3/a^3] + C_0\} \mathcal{I}_{[0,a]}(h) + \{C(0) + C_0\} \mathcal{I}_{(a,\infty)}(h);$$

i2) modelo exponencial

$$\gamma(h) = \{C(0)[1 - e^{-h/a}] + C_0\} \mathcal{I}_{(0,\infty)}(h);$$

i3) modelo potência

$$\gamma(h) = \{C(0)|h|^w + C_0\} \mathcal{I}_{(0,\infty)}(|h|) \quad e \quad w \in (0,2);$$

i4) modelo linear

$$\gamma(h) = \{C(0)w(h) + C_0\} \mathcal{I}_{(0,\infty)}(|h|).$$

ii) comportamento parabólico na origem (variáveis com alta continuidade)

ii1) modelo gaussiano

$$\gamma(h) = \{C(0)[1 - e^{-h^2/a^2}] + C_0\} \mathcal{I}_{(0,\infty)}(h);$$

ii2) modelo logarítmico

$$\gamma(h) = \{C(0)[\log|h|] + C_0\} \mathcal{I}_{(0,\infty)}(|h|);$$

Como formulados anteriormente, os modelos exponencial e

gaussiano não atingem o *patamar*. Entretanto numericamente no primeiro caso $\gamma(3a) \cong C(0)+C_0$ e no segundo $\gamma(3a^{1/2}) \cong C(0)+C_0$, o que implica em se tomar *alcances* aproximados igual a $3a$ e $3\sqrt{a}$ respectivamente.

Raramente modela-se estruturas com "efeito de buracos", que é caracterizado pela evidência de um aumento não monótono de $\gamma(h)$, que em muitos casos são confundidos com a presença de uma deriva.

A covariância $C(h)$ pode ser padronizada ao seu valor na origem $C(0)$, definindo o correlograma ou função de auto-correlação, $\rho(h) = [C(h)/C(0)]$; $\rho(h) \in [-1,1]$. O $\rho(h)$ é simplesmente o coeficiente de correlação entre duas variáveis aleatórias $Z(x)$ e $Z(x+h)$, $\rho(h) = 1-[\gamma(h)/C(0)]$, com $\rho(0) = 1$.

Como mencionado na parte introdutória, um dos principais eixos de pesquisas em *geoestatística*, é o desenvolvimento de métodos para a predição de reservas.

Nota: Seja $Z(x_0)$ um valor incógnito, e que se deseja predizer. Seja $\hat{Z}$ o preditor de $Z(x_0)$. O erro de predição será $\hat{Z}-Z(x_0)$ e é desejável fazer com que este seja tão pequeno quanto possível, em algum sentido. Em geral, $\hat{Z}-Z(x_0)$ é uma quantidade aleatória, tal que se deva entender "pequeno" no sentido de que o valor esperado de alguma função do erro seja menor. Medir-se-á a qualidade de um preditor por seu erro quadrático médio.

1.4 Função de predição

Problema: Quando se tem a realização de um processo estocástico $\{Z(x): x \in D \subseteq \mathbb{R}^d\}$, em n locações espaciais, um problema que surge naturalmente é a predição de $Z_V(x)$ em blocos onde não se tenha resultados de observações.

O preditor de uma variável aleatória Z é caracterizado como uma função de alguma propriedade física de tal variável de interesse.

Sabe-se que o melhor preditor é a esperança condicional $E[Z_v(x)|Z(x_1), \dots, Z(x_n)]$. Seu cálculo requer o conhecimento da distribuição $(n+1)$ -variada de $Z(x)$. Entretanto, é razoável especificar parcialmente a distribuição do processo estocástico, ou conhecer apenas seus dois primeiros momentos.

Serão utilizados para prever $Z(x_0)$, os pontos $x_1, \dots, x_n$ e as observações associadas a estes, e, naturalmente, uma função de predição $\pi[Z(x_1), \dots, Z(x_n)]$. O interesse, aqui, é determinar esta função de predição.

Sejam $Z(x)$ realizações de $\{Z(x): x \in D \subseteq \mathcal{R}^d\}$. Suponha que se observa $Z(x_1), \dots, Z(x_n)$, $x_\alpha \neq x_\beta$ para $\alpha \neq \beta = 1, \dots, n$. Há interesse em prever o valor $Z(x_0)$, $x_0 \in D$, baseado nas informações $Z(x_1), \dots, Z(x_n)$. Desta forma constrói-se uma função real $\pi[Z(x_1), \dots, Z(x_n)]$ que é denominada função de predição, denotando-se por $Z^*(x_0) = \pi[Z(x_1), \dots, Z(x_n)]$ o preditor de $Z(x_0)$.

Define-se o erro de predição da forma usual: $[Z(x_0) - Z^*(x_0)]$. Uma medida de aderência, ou de qualidade de $Z^*(x_0)$ é dada pelo erro quadrático médio $MSE[Z^*(x_0)] = E[Z(x_0) - Z^*(x_0)]^2$.

Definição 1.4.1 Seja Π o conjunto de todas as funções de $Z(x_1), \dots, Z(x_n)$, que são quadrados integráveis. A função $\pi_0 \in \Pi$ é definida como a melhor função de predição (BPF) de $Z(x_0)$ baseado em $Z(x_1), \dots, Z(x_n)$, se

$$E\{\pi_0[Z(x_1), \dots, Z(x_n)] - Z(x_0)\}^2 \leq E\{\pi[Z(x_1), \dots, Z(x_n)] - Z(x_0)\}^2; \forall \pi \in \Pi.$$

Deve-se encontrar uma função $\pi[Z(x_1), \dots, Z(x_n)]$ que minimize o $MSE[Z^*(x_0)]$.

Teorema 1.4.1 A melhor função de predição de $Z(x_0)$ baseada em

$Z(x_1), \dots, Z(x_n)$, é dada pela esperança condicional

$$E[Z(x_0) | Z(x_1), \dots, Z(x_n)].$$

Dem.: Graybill [1976].

A BPF $\pi_0[Z(x_1), \dots, Z(x_n)]$ é a projeção ortogonal de $Z(x_0)$ no subespaço vetorial Π .

Função de Predição Linear

Com o objetivo de determinar a BPF é necessário conhecer a distribuição conjunta de $Z(x_0), Z(x_1), \dots, Z(x_n)$, o que em geral não é acessível, e não pode ser inferida a partir dos dados. Isto dificulta a obtenção de informações sobre a BPF. É usual, então, em casos como o de *Kriging Clássico*, restringir o conjunto de funções de predição Π a uma classe ξ_{n+1} , das combinações lineares de $Z(x_1), \dots, Z(x_n)$; $\xi_{n+1} \subseteq \Pi$.

Definição 1.4.2 A função $\pi_0[Z(x_1), \dots, Z(x_n)] \in \xi_{n+1}$ é definida como a melhor função de predição linear (BLPF) de $Z(x_0)$, onde

$$\pi_0[Z(x)] = \lambda_0 + \sum_{\alpha=1}^n \lambda_{\alpha} Z(x_{\alpha}), \lambda_{\alpha} \in \mathcal{R}, \text{ se}$$

$$E\{\pi_0[Z(x_1), \dots, Z(x_n)] - Z(x_0)\}^2 \leq E\{\pi^*[Z(x_1), \dots, Z(x_n)] - Z(x_0)\}^2; \forall \pi^* \in \xi_{n+1}.$$

Sejam $E[Z(x_{\alpha})] = m(x_{\alpha}) \forall \alpha=1, \dots, n$; $\mathcal{K}$, k_0 e μ como definidos a seguir:

$$\mathcal{K} = [k_{\alpha\beta}]_{\alpha, \beta=1, \dots, n}; \quad k_{\alpha\beta} = \text{Cov}[Z(x_{\alpha}), Z(x_{\beta})]$$

$$k_0 = [k_{\alpha 0}]_{n \times 1}; \quad k_{\alpha 0} = \text{Cov}[Z(x_{\alpha}), Z(x_0)]$$

$$\mu = [m(x_{\alpha})]_{n \times 1};$$

$$\forall \alpha, \beta=1, \dots, n, \text{ e ainda, } \mu_0 = m(x_0).$$

Teorema 1.4.2 Se K , k_0 e $m(x_\alpha)$, $\alpha=0,1,\dots,n$, são conhecidos para um conjunto de pontos $x_0, x_1, \dots, x_n$, então a função BLP de $Z(x_0)$ baseada em $Z(x)$ é dada por:

$$\pi_0^*[Z(x)] = \mu_0 + k_0' K^{-1} [Z(x) - \mu].$$

Dem.: Graybill [1976].

Afim de se obter um preditor ótimo deve-se verificar a não-viciabilidade.

Definição 1.4.3 Seja $\pi \in \Pi$, diz-se que $\pi[Z(x)]$ é um preditor não-viciado de $Z(x_0)$ se $E\{\pi[Z(x)] - Z(x_0)\} = 0$.

Corolário (teo.1.4.2) O BLP de $Z(x_0)$ é não-viciado.

$$\begin{aligned} \text{Dem.:} \quad E\{\pi_0^*[Z(x)] - Z(x_0)\} &= E\{k_0' K^{-1} [Z(x) - \mu] + \mu_0 - Z(x_0)\} \\ E\{\pi_0^*[Z(x)] - Z(x_0)\} &= k_0' K^{-1} E[Z(x)] - k_0' K^{-1} \mu + \mu_0 - \mu \\ E\{\pi_0^*[Z(x)] - Z(x_0)\} &= k_0' K^{-1} \mu - k_0' K^{-1} \mu = 0. \quad \blacksquare \end{aligned}$$

A função $\pi_0^*[Z(x)]$ é a melhor função de predição linear não-viciada de $Z(x_0)$ baseada em $Z(x)$.

Pela condição de não-viciabilidade tem-se:

$$E\{\pi_0^*[Z(x)]\} = E\{Z(x_0)\} = m(x_0)$$

$$\lambda_0 + \sum_{\alpha=1}^n \lambda_\alpha m(x_\alpha) = m(x_0) \Rightarrow \lambda_0 = \sum_{\alpha=1}^n \lambda_\alpha m(x_\alpha) + m(x_0).$$

Se $E[Z(x)] = m \forall x$, então $\lambda_0 + \sum_{\alpha=1}^n \lambda_\alpha = m$, isto é satisfeito se, e somente se, $\lambda_0 = 0$ e $\sum_{\alpha=1}^n \lambda_\alpha = 1$, onde, em geral, m é desconhecido.

Se $E[Z(\mathbf{x})] = m(\mathbf{x}) = \sum_{k=1}^p \alpha_k f_k(\mathbf{x})$, onde os α_k 's são constantes desconhecidas e os f_k 's são funções conhecidas de $\mathbf{x}$, então $\lambda_0 = \sum_{k=1}^p \alpha_k f_k(\mathbf{x}_0) - \sum_{\alpha=1}^n \lambda_{\alpha} \sum_{k=1}^p \alpha_k f_k(\mathbf{x}_{\alpha})$, ou seja $\lambda_0 + \sum_{\alpha=1}^n \lambda_{\alpha} \sum_{k=1}^p \alpha_k f_k(\mathbf{x}_{\alpha}) = \sum_{k=1}^p \alpha_k f_k(\mathbf{x}_0)$, e isto é satisfeito se, e somente se $\lambda_0 = 0$ e $f_k(\mathbf{x}_0) = \sum_{\alpha=1}^n \lambda_{\alpha} f_k(\mathbf{x}_{\alpha})$; $\forall k=1, \dots, p$.

Função de Predição Não-Linear

Considerou-se, anteriormente, apenas preditores lineares, isto é, a procura de um preditor tinha sido limitada ao espaço vetorial gerado pelas combinações lineares das n variáveis $Z(\mathbf{x})$. Mas esta limitação pode ser removida e o problema se concentrará então em encontrar um preditor não-linear como uma função dos n dados observados. Isto é, permite-se que se considere um espaço de preditores que seja maior que o das combinações lineares. As condições de mínima variância e de não-viciabilidade devem ainda ser garantidas.

Sabe-se que o melhor preditor de $Z(\mathbf{x}_0)$ não-viciado, que minimiza o erro quadrático médio é $E[Z(\mathbf{x}_0) | Z(\mathbf{x}_1), \dots, Z(\mathbf{x}_n)]$. Em geral para o cálculo desta esperança condicional torna-se necessário o conhecimento da distribuição finitamente-dimensional do processo $\{Z(\mathbf{x}) : \mathbf{x} \in D \subseteq \mathcal{R}^d\}$. Contudo, para processos Gaussianos as esperanças condicionais são funções lineares dos dados. Então os preditores de menor erro quadrático médio são lineares, e portanto facilmente determinados através de técnicas clássicas de predição. Se for possível encontrar uma transformação ϕ , tal que $T(\mathbf{x}) = \phi^{-1}(Z(\mathbf{x}))$ seja um processo Gaussiano, então torna-se possível prever $T(\mathbf{x})$ e usar $\phi^*[T(\mathbf{x}_0)]$ como um preditor de $Z(\mathbf{x}_0)$.

Uma idéia geral seria determinar transformadas de distribuições bidimensionais que possam ser expressas através de fatores ortogonais (que não são necessariamente polinômios). Fatores ortogonais, sendo não correlacionados, podem ser submetidos a *Kriging* independentemente para dado suporte a ser predito.

O problema crucial é encontrar uma distribuição bivariada que possa ser expressa em termos de tais fatores ortogonais. Com este objetivo faz-se a transformação do processo, obtendo-se um que seja Gaussiano, que se sabe pode ser expandido em séries polinomiais de Hermite .

Atenta-se ao fato que, quando aplicado um anamorfismo gaussiano (transformada gaussiana), o *Kriging Disjuntivo* (método de *Kriging Não-Linear*) pode apresentar problemas sérios de predição quando se dispõe de dados agrupados, ou se trata de variáveis discretas, ou mesmo quando se amostra um grande número de conteúdos nulos. Estes são casos especiais, e raros de aplicabilidade no que se refere ao objetivo maior desta dissertação, porém estes são estudados por G. Matheron e M. Armstrong [1986].

CAPÍTULO 2

KRIGING LINEAR

Em mineração um grande problema é determinar o melhor preditor possível do grau médio de um bloco, levando-se em consideração os valores ensaiados de diferentes amostras retiradas de dentro ou fora do bloco a ser predito. Em 1951 D. G. Krige propôs um procedimento de regressão para determinar uma ponderação a cada amostra analisada. O preditor do teor do bloco seria então uma combinação linear das observações. Este procedimento de regressão foi originalmente chamado por G. Matheron [1963] de *Kriging*, que o formalizou e generalizou.

Kriging é um método de predição que leva em consideração os seguintes fatores: (i) o número de dados e a qualidade destes em cada bloco; (ii) a posição dos dados com relação ao campo; (iii) a distância entre os blocos e a área de interesse e (iv) a continuidade espacial das variáveis interpoladas.

De acordo com a definição dada por G. Matheron, *Kriging* é um processo probabilístico de obtenção do melhor preditor linear não-viciado de uma variável desconhecida. "Melhor", aqui, é no sentido de minimização da variância do erro de predição.

Kriging, é ultimamente, o método de predição local mais desenvolvido em áreas que envolvem Geociências, pois com o objetivo de estudo e avaliação de depósitos minerais, implementou-se grandes estruturas no que se refere às condições de aplicação desta técnica.

O problema de predição local é envolvido por determinar o melhor preditor do valor médio de um processo estocástico sobre um domínio limitado, também chamado suporte. A técnica utilizada é aquela que fornece o BLUP.

Kriging Linear é uma técnica de predição que proporciona o melhor preditor linear não-viciado (BLUP). Dado um conjunto de realizações, $Z(x_1), \dots, Z(x_n)$, de um processo estocástico $\{Z(x): x \in D \subseteq \mathbb{R}^d\}$, deseja-se prever $Z(x_0)$ em $x_0 \in D$, ponto não amostrado.

A limitação à classe dos preditores lineares é natural, posto que isto significa que apenas os momentos de segunda ordem do processo estocástico será necessário.

Atenta-se ao risco de não adequabilidade quando se usa *Kriging Linear* para resolver problemas de predição funcionais não-lineares dos dados, tais como a proporção recuperada e o metal contido. Nestes casos utiliza-se uma técnica não-linear de *Kriging*, que será assunto nesta dissertação em capítulos subsequentes, o *Kriging Disjuntivo*.

2.1 O Método

Seja ξ um espaço vetorial provido de produto interno. Sejam, ainda, $Z(x_0) \in \xi$ e ξ' um subespaço vetorial de ξ . Pelo teorema da projeção, existe um e somente um elemento Z^* de ξ' que satisfaz à condição que a distância $Z^* - Z$ seja mínima com $Z \in \xi$, onde este elemento Z^* , é a projeção de $Z(x_0)$ no subespaço ξ' .

O método de *Kriging Linear* não é nada além que a projeção de $Z(x_0)$ no subespaço gerado por combinações lineares. Considere a classe dos preditores lineares, isto é, o subespaço vetorial $\xi_{n+1} \subset \xi$, onde ξ_{n+1} tem dimensão $(n+1)$ e é gerado pelas combinações lineares $\lambda_0 + \sum_{\alpha=1}^n \lambda_{\alpha} Z_{\alpha}$, das n variáveis $\{Z_{\alpha} = Z(x_{\alpha}): \alpha=1, \dots, n\}$ realizadas, mais a constante 1; desta forma o processo de *Kriging Linear* é definido como o processo de projeção da variável desconhecida $Z(x_0)$ em ξ_{n+1} ou em alguma variedade linear $\zeta \subset \xi_{n+1}$.

Um subespaço ou variedade linear de um espaço linear, é um subconjunto deste que é também um espaço vetorial; no caso de variedade linear aquele é provido de translação.

Objetivo: Prever o valor médio de $Z(x)$ sobre $V(x_0)$, um domínio

centrado em $x_0 \in D$, $Z^*(x_0) = Z_V^*(x_0) = (1/V) \int_{V(x_0)} Z(x) dx$. Os dados sobre os quais esta predição é baseada, são definidos no suporte v , ou são teores médios $Z_V(x_\alpha)$ definidos no suporte V centrado em $x_\alpha \in D$.

Formar um preditor global de um depósito D combinando preditores de *Kriging* locais, é feito ponderando os preditores pelos seus respectivos suportes.

Se as unidades são pontos de uma malha regular, ou seja, se for possível determinar uma regularidade estrutural na disposição dos dados amostrados, o preditor de um depósito D é obtido tomando a média dos pontos preditos por *Kriging* dentro do contorno poligonal que melhor aproxima a forma do depósito a ser predito.

O preditor de *Kriging* de $Z_V(x_\alpha)$, denotado por Z_K^* , é considerado não-viciado para $Z_V(x_0)$ se, e somente se, $E[Z_K^* - Z_V(x_0)] = 0$, isto é, $\langle Z_K^* - Z_V(x_0), 1 \rangle = 0$.

2.2 Kriging Simples

Seja $\{Z(x)\}$, o processo sob estudo, com x definido em um suporte D . No caso em que todas as esperanças são conhecidas, bem como a função de covariância centrada ou não (não há relevância se estas são ou não estacionárias), onde $Z_K^* = \lambda_0 + \sum_{\alpha=1}^n \lambda_\alpha Z_\alpha \in \xi_{n+1}$, então $E[Z_K^*] = \lambda_0 + \sum_{\alpha=1}^n \lambda_\alpha m_\alpha$, com $m_\alpha = E[Z(x_\alpha)]$ conhecida, a não-viciabilidade é caracterizada por $\lambda_0 = m(x_0) - \sum_{\alpha=1}^n \lambda_\alpha m_\alpha$, onde $m(x_0) = E[Z(x_0)]$.

O preditor de SK é da forma $Z_{SK}^* = \lambda_0 + \sum_{\alpha=1}^n \lambda_\alpha Z_\alpha$ que é a projeção de $Z(x_0)$ no espaço vetorial ξ_{n+1} ; esta projeção é única e caracterizada pela ortogonalidade do vetor $[Z(x_0) - Z_{SK}^*]$ a cada dos $(n+1)$ vetores que geram ξ_{n+1} , ou seja $\langle Z_{SK}^* - Z(x_0), Z_\alpha \rangle = 0$, $\forall \alpha = 1, \dots, n$.

Pela relação de ortogonalidade:

$$\langle Z_{SK}^* - Z(x_0), Z_\beta \rangle = 0 \Rightarrow \langle Z_{SK}^*, Z_\beta \rangle - \langle Z(x_0), Z_\beta \rangle = 0 \Rightarrow \langle Z_{SK}^*, Z_\beta \rangle = \langle Z(x_0), Z_\beta \rangle ;$$

$$\langle Z_{SK}^*, Z_\beta \rangle = \langle \lambda_0 + \sum_{\alpha=1}^n \lambda_\alpha Z_\alpha, Z_\beta \rangle = \langle \lambda_0, Z_\beta \rangle + \sum_{\alpha=1}^n \lambda_\alpha \langle Z_\alpha, Z_\beta \rangle = \lambda_0 \langle Z_\beta, 1 \rangle + \sum_{\alpha=1}^n \lambda_\alpha \sigma_{\alpha\beta}$$

$$\langle Z_{SK}^*, Z_\beta \rangle = \lambda_0 m_\beta + \sum_{\alpha=1}^n \lambda_\alpha \sigma_{\alpha\beta} ;$$

$$\langle Z(x_0), Z_\beta \rangle = \sigma_{x_0\beta} ; \quad \forall \beta=1, \dots, n .$$

Pela não-viciabilidade:

$$\langle Z_{SK}^* - Z(x_0), 1 \rangle = 0 \Rightarrow \langle Z_{SK}^*, 1 \rangle - \langle Z(x_0), 1 \rangle = 0$$

$$m(x_0) - \langle \lambda_0, 1 \rangle - \sum_{\alpha=1}^n \lambda_\alpha \langle Z_\alpha, 1 \rangle = m(x_0) - \lambda_0 - \sum_{\alpha=1}^n \lambda_\alpha m_\alpha = 0 .$$

Obtém-se, assim, as (n+1) equações lineares, com as (n+1) incógnitas $\lambda_0, \lambda_1, \dots, \lambda_n$; dando origem ao sistema de *Kriging Simples*:

$$(SK): \begin{cases} m(x_0) = \lambda_0 + \sum_{\alpha=1}^n \lambda_\alpha m_\alpha \\ \sigma_{x_0\beta} = \lambda_0 m_\beta + \sum_{\alpha=1}^n \lambda_\alpha \sigma_{\alpha\beta} ; \quad \forall \beta=1, \dots, n; \text{ onde } \sigma_{\alpha\beta} \text{ é a covariância não centrada.} \end{cases}$$

Multiplicando os elementos da primeira equação do sistema (SK) por m_β e subtraindo esta da ($\beta+1$)-ésima equação, tem-se:

$$\sum_{\alpha=1}^n \lambda_\alpha (m_\beta m_\alpha - \sigma_{\alpha\beta}) = m_\beta m(x_0) - \sigma_{x_0\beta}, \quad \forall \beta=1, \dots, n .$$

Agora, escrevendo-se o sistema em termos da covariância centrada, $\sigma_{\alpha\beta}^c = \sigma_{\alpha\beta} - m_\alpha m_\beta$, obtém-se o sistema de *Kriging Simples* reduzido, dado por:

$$\sum_{\alpha=1}^n \lambda_\alpha \sigma_{\alpha\beta}^c = \sigma_{x_0\beta}^c ; \quad \forall \beta=1, \dots, n ;$$

ou seja $\rho' = \Lambda' K$,

onde $\rho: [\sigma_{x_0\beta}^c]_{n \times 1}$; $\Lambda: [\lambda_\alpha]_{n \times 1}$ e $K: [\sigma_{\alpha\beta}^c]_{n \times n}$.

Uma vez determinadas as n ponderações Λ , pelo sistema de SK reduzido, λ_0 é obtido através da primeira equação de (SK), assim o preditor de SK pode ser escrito:

$$Z_{SK}^* = \lambda_0 + \sum_{\alpha=1}^n \lambda_{\alpha} Z_{\alpha} - m(x_0) - \sum_{\alpha=1}^n \lambda_{\alpha} m_{\alpha} + \sum_{\alpha=1}^n \lambda_{\alpha} Z_{\alpha}$$

$$Z_{SK}^* = \sum_{\alpha=1}^n \lambda_{\alpha} [Z_{\alpha} - m_{\alpha}] + m(x_0),$$

$$\text{ou } Z_{SK}^* = \Lambda' [Z - \mu] + \mu_0,$$

$$\text{onde } Z: [Z_{\alpha}]_{n \times 1}; \mu: [m_{\alpha}]_{n \times 1}; \mu_0 = m(x_0); \text{ e } \Lambda' = \rho' K^{-1}.$$

Precisamente, $[Z_{SK}^* - m(x_0)]$ é a projeção de $Z(x_0) - m(x_0)$ no subespaço vetorial gerado pelas combinações lineares dos n dados centrados.

Variância de predição

A variância de predição ou variância de extensão, $\sigma_E^2 = \text{Var}[Z_{SK}^* - Z(x_0)]^2$, não é nada além que a distância quadrática $Z_{SK}^* - Z(x_0)$ ².

$$\sigma_E^2 = Z_{SK}^* - Z(x_0) \text{ }^2 = \sum_{\alpha=1}^n \lambda_{\alpha} [Z_{\alpha} - m_{\alpha}] + m(x_0) - Z(x_0) \text{ }^2$$

$$\sigma_E^2 = \sum_{\alpha=1}^n \lambda_{\alpha} [Z_{\alpha} - m_{\alpha}] \text{ }^2 - 2 \langle \sum_{\alpha=1}^n \lambda_{\alpha} [Z_{\alpha} - m_{\alpha}], Z(x_0) - m(x_0) \rangle + Z(x_0) - m(x_0) \text{ }^2.$$

$$\sum_{\alpha=1}^n \lambda_{\alpha} [Z_{\alpha} - m_{\alpha}] \text{ }^2 = \langle \sum_{\alpha=1}^n \lambda_{\alpha} Z_{\alpha} - m_{\alpha}, \sum_{\alpha=1}^n \lambda_{\alpha} Z_{\alpha} - m_{\alpha} \rangle = \sum_{\alpha=1}^n \lambda_{\alpha} \sum_{\beta=1}^n \lambda_{\beta} \langle Z_{\alpha} - m_{\alpha}, Z_{\beta} - m_{\beta} \rangle = \sum_{\alpha=1}^n \lambda_{\alpha} \sum_{\beta=1}^n \lambda_{\beta} \sigma_{\alpha\beta}^c;$$

$$2 \langle \sum_{\alpha=1}^n \lambda_{\alpha} [Z_{\alpha} - m_{\alpha}], Z(x_0) - m(x_0) \rangle = 2 \sum_{\alpha=1}^n \lambda_{\alpha} \sigma_{\alpha x_0}^c;$$

$$Z(x_0) - m(x_0) \text{ }^2 = \langle Z(x_0) - m(x_0), Z(x_0) - m(x_0) \rangle = \sigma_{x_0 x_0}^c.$$

$$\sigma_E^2 = \sum_{\alpha=1}^n \lambda_{\alpha} \sum_{\beta=1}^n \lambda_{\beta} \sigma_{\alpha\beta}^c - 2 \sum_{\alpha=1}^n \lambda_{\alpha} \sigma_{\alpha x_0}^c + \sigma_{x_0 x_0}^c,$$

ou em termos matriciais:

$$\sigma_E^2 = \Lambda' K \Lambda - 2\Lambda' \rho + C(0).$$

Problema: Minimizar σ_E^2 sujeito a $m(x_0) = \lambda_0 + \sum_{\alpha=1}^n \lambda_{\alpha} m_{\alpha}$, ou seja minimizar a distância quadrática sujeito a $\mu_0 = \lambda_0 + \Lambda' \mu$.

O método clássico de resolução deste sistema é usar o formalismo de Lagrange, reescreve-se a variância de extensão como $\sigma_E^2 = \Lambda' K \Lambda - 2\Lambda' \rho + C(0) - 2\delta(\lambda_0 + \Lambda' \mu - \mu_0)$, onde δ é dito multiplicador de Lagrange. Deve-se, então encontrar (λ_0, Λ') que minimize σ_E^2 . Denota-se por σ_K^2 o $\min(\sigma_E^2)$.

$$\frac{\partial \sigma_E^2}{\partial \Lambda'} = 2K\Lambda - 2\rho - 2\delta\mu; \quad \frac{\partial^2 \sigma_E^2}{\partial \Lambda'^2} = 2K > 0$$

($K=0$ obtém-se uma otimalidade estatística irreal, isto é, um absurdo!)

$$2K\hat{\Lambda} - 2\rho - 2\delta\mu = 0$$

$$K\hat{\Lambda} = \rho + \delta\mu \Rightarrow \hat{\Lambda} = K^{-1}\rho + \delta K^{-1}\mu.$$

$$\text{Então, } \lambda_0 + \hat{\Lambda}'\mu = \lambda_0 + \rho'K^{-1}\mu + \delta\mu'K^{-1}\mu \Rightarrow \delta\mu'K^{-1}\mu = \mu_0 - \lambda_0 + \rho'K^{-1}\mu$$

$$\Rightarrow \delta = -\lambda_0(\mu'K^{-1}\mu)^{-1} + \mu_0(\mu'K^{-1}\mu)^{-1} - \rho'K^{-1}\mu(\mu'K^{-1}\mu)^{-1}, \text{ isto é,}$$

$$\delta = [(\mu_0 - \lambda_0) - \rho'K^{-1}\mu](\mu'K^{-1}\mu)^{-1},$$

$$\text{logo, } \hat{\Lambda} = K^{-1}\rho - K^{-1}\mu[(\mu_0 - \lambda_0) - \rho'K^{-1}\mu](\mu'K^{-1}\mu)^{-1},$$

$$\text{fazendo-se } \eta = [(\mu_0 - \lambda_0) - \rho'K^{-1}\mu](\mu'K^{-1}\mu)^{-1},$$

$$\text{tem-se } \sigma_K^2 = C(0) + [K^{-1}\rho - K^{-1}\mu\eta]'K[K^{-1}\rho - K^{-1}\mu\eta] - 2[K^{-1}\rho - K^{-1}\mu\eta]'\rho$$

$$\sigma_K^2 = C(0) - \rho'K^{-1}\rho - \rho'K^{-1}\mu\eta - \eta\mu'K^{-1}\rho - \eta\mu'K^{-1}\mu\eta + 2\eta\mu'K^{-1}\rho$$

$$\sigma_K^2 = C(0) - \rho'K^{-1}\rho - \eta^2[\mu'K^{-1}\mu]$$

$$\sigma_K^2 = C(0) - \rho' K^{-1} \rho - \eta.$$

Conclui-se, então que:

$$\sigma_K^2 = C(0) - \rho' K^{-1} \rho - [(\mu_0 - \lambda_0) - \rho' K^{-1} \mu] (\mu' K^{-1} \mu)^{-1}.$$

Se além da hipótese de esperanças conhecidas o processo for estacionário, isto é, $E[Z(\mathbf{x}_\alpha)] = m$, $\forall \mathbf{x} \in D$, então $Z_{SK}^* = \Lambda' Z(\mathbf{x})$, ou seja, $Z_{SK}^* = \rho' K^{-1} Z(\mathbf{x}) + (1 - \rho' K^{-1} 1)m \in D$; onde $1': [1]_{1 \times n}$, e a restrição torna-se $\Lambda' 1 = 1$. Neste caso, o preditor é dito de *Kriging Simple*, e a variância de predição é dada por:

$$\sigma_K^2 = C(0) - \rho' K^{-1} \rho.$$

2.3 Kriging Ordinário

Se as esperanças e covariâncias do processo estocástico, dado por $\{Z(\mathbf{x}); \mathbf{x} \in D \subseteq \mathcal{R}^d\}$, são conhecidas, então como visto anteriormente o problema de predição está resolvido. Mas, na prática, não se dispõe do conhecimento da função de covariância, surge daí a necessidade da introdução da hipótese de estacionaridade de segunda ordem, com o fim de tornar possível a predição. A hipótese de estacionaridade não fornece a forma da função de covariância e apenas permite estimá-la a partir da amostra. Mesmo com a restrição de que a classe de covariância do processo seja estacionária, se faz necessário estimar o semi-variograma.

Se $\{Z(\mathbf{x}); \mathbf{x} \in D \subseteq \mathcal{R}^d\}$ é um processo estocástico estacionário de segunda ordem, então por definição:

$$E[Z(\mathbf{x})] = m; \forall \mathbf{x} \in D; (m \text{ é desconhecido})$$

$$E[Z(\mathbf{x}+h)Z(\mathbf{x})] = C(h) + m^2, \text{ se a covariância for finita;}$$

$$\text{Var}[Z(\mathbf{x}+h) - Z(\mathbf{x})] = 2\gamma(h), \mathbf{x}, h \in D.$$

O preditor de *Kriging Ordinário* (OK) de $Z_v(\mathbf{x}_0)$, denotado por

Z_{OK}^* , é dado por:

$$Z_{OK}^* = Z_v^*(x_0) = \lambda_0 + \sum_{\alpha=1}^n \lambda_{\alpha} Z_{\alpha}.$$

A esperança é estacionária, porém desconhecida, então a condição de não-viciabilidade é satisfeita se, e somente se, $\lambda_0 = 0$ e $\sum_{\alpha=1}^n \lambda_{\alpha} = 1$.

A condição $\lambda_0 = 0$, significa a restrição do conjunto de possíveis preditores ao espaço vetorial $\xi_n \subset \xi_{n+1}$. O subespaço ξ_n de dimensão n é gerado pela combinação linear, $\sum_{\alpha=1}^n \lambda_{\alpha} Z_{\alpha}$, dos n dados. Já a condição $\sum_{\alpha=1}^n \lambda_{\alpha} = 1$, restringe ξ_n à variedade linear ζ definida por esta condição. Assim, no caso estacionário, a não-viciabilidade fornece um preditor linear $Z_{OK}^* \in \zeta \subset \xi_n \subset \xi_{n+1}$.

Então, o preditor de OK é definido como a projeção da variável desconhecida $Z(x_0)$ na variedade linear ζ . Esta projeção é única e caracterizada pelas condições:

- (i) $Z_{OK}^* \in \zeta$, isto é, $\sum_{\alpha=1}^n \lambda_{\alpha} = 1$;
- (ii) o vetor $Z_{OK}^* - Z(x_0)$ é ortogonal a qualquer vetor diferença $Z_{\alpha} - Z_c$, com $Z_{\alpha}, Z_c \in \zeta$, sendo Z_c um elemento arbitrário de ζ .

A condição de ortogonalidade, dada acima, é formalizada por:

$$\langle Z_{OK}^* - Z(x_0), Z_{\beta} - Z_c \rangle = 0 \Rightarrow \langle Z_{OK}^*, Z_{\beta} \rangle - \langle Z_{OK}^*, Z_c \rangle - \langle Z(x_0), Z_{\beta} \rangle + \langle Z(x_0), Z_c \rangle = 0,$$

fazendo-se $\langle Z_{OK}^*, Z_c \rangle - \langle Z(x_0), Z_c \rangle = \nu$, então, $\langle Z_{OK}^*, Z_c \rangle - \nu = \langle Z(x_0), Z_c \rangle$;

$$\langle Z(x_0), Z_{\beta} \rangle = \sigma_{x_0 \beta};$$

$$\langle Z_{OK}^*, Z_{\beta} \rangle = \lambda_0 m + \sum_{\alpha=1}^n \lambda_{\alpha} \sigma_{\alpha \beta}; \quad \forall \beta = 1, \dots, n.$$

Pela não-viciabilidade:

$$\langle Z_{OK}^* - Z(x_0), 1 \rangle = 0 \Rightarrow m - \lambda_0 + \sum_{\alpha=1}^n \lambda_{\alpha} m = 0 \Leftrightarrow \sum_{\alpha=1}^n \lambda_{\alpha} = 1 \text{ e } \lambda_0 = 0.$$

Então o sistema de OK é dado pelas (n+1) equações lineares com (n+1) incógnitas:

$$(OK): \begin{cases} \sum_{\alpha=1}^n \lambda_{\alpha} = 1 \\ \sigma_{x_0\beta} = \sum_{\alpha=1}^n \lambda_{\alpha} \sigma_{\alpha\beta} - \nu; \forall \beta=1, \dots, n; \text{ onde } \sigma_{\alpha\beta} \text{ é a covariância não centrada.} \end{cases}$$

Escrevendo-se o sistema em termos da covariância centrada $\sigma_{\alpha\beta}^c = \sigma_{\alpha\beta} - m^2$, obtém-se o sistema (OK) reduzido:

$$\sigma_{x_0\beta}^c = \sum_{\alpha=1}^n \lambda_{\alpha} \sigma_{\alpha\beta}^c - \nu; \forall \beta=1, \dots, n,$$

ou seja, em termos matriciais:

$$\rho = \Lambda' K - N.$$

Variância de predição

$$\sigma_E^2 = Z_{OK}^* - Z(x_0) \quad \sigma_E^2 = Z(x_0)^2 + Z_{OK}^{*2} - 2\langle Z(x_0), Z_{OK}^* \rangle, \text{ com } Z(x_0)^2 = \sigma_{x_0 x_0};$$

$$Z_{OK}^{*2} = \left[\sum_{\alpha=1}^n \lambda_{\alpha} Z_{\alpha} \right]^2 = \left\langle \sum_{\alpha=1}^n \lambda_{\alpha} Z_{\alpha}, \sum_{\beta=1}^n \lambda_{\beta} Z_{\beta} \right\rangle = \sum_{\alpha=1}^n \lambda_{\alpha} \sum_{\beta=1}^n \lambda_{\beta} \langle Z_{\alpha}, Z_{\beta} \rangle = \sum_{\alpha=1}^n \lambda_{\alpha} \sum_{\beta=1}^n \lambda_{\beta} \sigma_{\alpha\beta}$$

$$Z_{OK}^{*2} = \sum_{\alpha=1}^n \lambda_{\alpha} (\nu + \sigma_{x_0\alpha}) = \nu + \sum_{\alpha=1}^n \lambda_{\alpha} \sigma_{x_0\alpha};$$

$$\langle Z(x_0), Z_{OK}^* \rangle = \langle Z(x_0), \sum_{\alpha=1}^n \lambda_{\alpha} Z_{\alpha} \rangle = \sum_{\alpha=1}^n \lambda_{\alpha} \sigma_{x_0\alpha}; \quad \forall \alpha, \beta=1, \dots, n;$$

logo a variância de extensão é dada por:

$$\sigma_E^2 = \sigma_{x_0 x_0} - 2 \sum_{\alpha=1}^n \lambda_{\alpha} \sigma_{x_0\alpha} + \nu + \sum_{\alpha=1}^n \lambda_{\alpha} \sigma_{x_0\alpha}$$

$$\sigma_E^2 = \sigma_{x_0 x_0} - \sum_{\alpha=1}^n \lambda_{\alpha} \sigma_{x_0\alpha} + \nu.$$

Escrevendo-se em termos da covariância centrada:

$$\sigma_E^2 = \sigma_{x_0 x_0}^C - \sum_1^n \lambda_\alpha [\sigma_{x_0 \alpha}^C - \rho] = \sigma_{x_0 x_0}^C - 2 \sum_1^n \lambda_\alpha \sigma_{x_0 \alpha}^C + \sum_1^n \lambda_\alpha \sum_1^n \lambda_\beta \sigma_{\alpha \beta}^C,$$

ou em termos matriciais:

$$\sigma_E^2 = C(0) - 2\Lambda' \rho + \Lambda' K \Lambda.$$

Problema: Encontrar o vetor Λ que minimize σ_E^2 , sujeito à $1'\Lambda = 1$, onde $1' = [1]_{1 \times n}$.

O método utilizado é o dos multiplicadores de Lagrange.

$$\sigma_E^2 = C(0) - 2\Lambda' \rho + \Lambda' K \Lambda - 2\delta(1'\Lambda - 1), \quad \delta: \text{Lagrangeano.}$$

$$\frac{\partial \sigma_E^2}{\partial \Lambda'} = 2K\Lambda - 2\rho - 2\delta 1; \quad \frac{\partial^2 \sigma_E^2}{\partial \Lambda'^2} = 2K > 0$$

($K \equiv 0$ obtém-se uma otimalidade estatística irreal, isto é, um absurdo!)

O preditor de *Kriging Ordinário* é dado por:

$$Z_{OK}^* = \rho' K^{-1} \underline{Z}(x) + (1 - 1' K^{-1} \rho)(1' K^{-1} 1)^{-1} 1' K^{-1} \underline{Z}(x).$$

A variância de *Kriging Ordinário* :

$$\sigma_K^2 = C(0) - \rho' K^{-1} \rho + (1' K^{-1} 1)^{-1} (1 - \rho' K^{-1} 1)^2.$$

Como visto anteriormente, um processo estocástico de covariância estacionária é também intrínseco. Na prática a inferência da função semi-variograma é mais facilmente obtida que a inferência da correspondente função de covariância. Portanto é de grande valia substituir, no sistema de *Kriging*, a função de covariância pela função semi-variograma, definida por $\gamma_{\alpha\beta} = \frac{1}{2}[\sigma_{\alpha\alpha}^C + \sigma_{\beta\beta}^C - 2\sigma_{\alpha\beta}^C]$, ou seja $\gamma_{\alpha\beta} = \frac{1}{2}E[Z(x_\alpha) - Z(x_\beta)]^2$.

Utilizando-se da relação formal, σ_E^2 pode ser reescrito como:

$$\sigma_E^2 = -\gamma_{x_0 x_0} + 2 \sum_1^n \lambda_\alpha \gamma_{x_0 \alpha} - \sum_1^n \lambda_\alpha \sum_1^n \lambda_\beta \gamma_{\alpha \beta}.$$

Sejam

$$K_0: [\gamma_{\alpha\beta}], \quad \alpha, \beta = 1, \dots, n.$$

$$\rho_0: [\gamma_{x_0 \alpha}], \quad \alpha = 1, \dots, n;$$

é fácil verificar que $\rho = C(0)1 - \rho_0$ e $K = C(0)11' - K_0$; então $\sigma_E^2 = 2\Lambda' \rho_0 - \Lambda' K_0 \Lambda$.

O preditor de *Kriging* em termos do semi-variograma é dado por:

$$Z_{OK}^* = \rho_0' K_0^{-1} \underline{Z}(x) + (1 - 1' K_0^{-1} \rho_0) (1' K_0^{-1} 1)^{-1} 1' K_0^{-1} \underline{Z}(x);$$

e a variância de *Kriging*:

$$\sigma_{OK}^2 = C(0) - \rho_0' K_0^{-1} \rho_0 - (1' K_0^{-1} 1)^{-1} (1 - \rho_0' K_0^{-1} 1)^2.$$

Importante

O estimador de mínimos quadrados ponderados de "m" que também é o BLUP, é dado por:

$$\hat{m} = (1' K^{-1} 1)^{-1} 1' K^{-1} \underline{Z}(x) \quad \text{e}$$

$$\text{Var}(\hat{m}) = (1' K^{-1} 1)^{-1} 1' K^{-1} K K^{-1} 1 (1' K^{-1} 1)^{-1} = (1' K^{-1} 1)^{-1}.$$

Sabe-se que:

$$Z_{OK}^* = \rho' K^{-1} \underline{Z}(x) + (1 - 1' K^{-1} \rho) (1' K^{-1} 1)^{-1} 1' K^{-1} \underline{Z}(x), \quad \text{e}$$

$$\sigma_{OK}^2 = C(0) - \rho' K^{-1} \rho + (1' K^{-1} 1)^{-1} (1 - \rho' K^{-1} 1)^2$$

$$\sigma_K^2 = C(0) - \rho' K^{-1} \rho + \eta^2 (\mu' K^{-1} \mu) - \eta^2 (\mu' K^{-1} \mu) + (1 - \rho' K^{-1} \rho)^2 \text{Var}(\hat{m}), \quad \text{isto é,}$$

$\sigma_K^2 = \sigma_{OK}^2$ (quando m_α é conhecido) + Δ , onde Δ é a perda de precisão envolvida ao se estimar "m".

2.4 Kriging Universal

Frequentemente ocorre que o processo estocástico $\{Z(x): x \in D \subseteq \mathcal{R}^d\}$ não seja estacionário de segunda ordem, devendo ocorrer a presença de uma tendência ou de um comportamento sistemático da função sob estudo. Neste caso deve-se encontrar uma expressão para a tendência e subtraí-la de $Z(x)$, afim de se obter um resíduo ou termo de flutuação que seja estacionário. O semi-variograma da variável resíduo será, então, utilizado indiretamente para predizer o valor de interesse.

A tendência é por definição, $E[Z(x)] = m(x)$, $x \in D$ e o resíduo ou flutuação $Y(x) = Z(x) - m(x)$. Supõe-se que $\{Y(x): x \in D \subseteq \mathcal{R}^d\}$ seja um processo estocástico intrinsecamente estacionário, e ainda que a tendência tenha forma conhecida e seja dada pela expressão analítica $m(x) = \sum_{k=1}^p \alpha_k f_k(x)$, onde os p parâmetros α_k são desconhecidos.

A condição de não-viciabilidade é expressa da forma:

$$\lambda_0 + \sum_{\alpha=1}^n \lambda_{\alpha} \sum_{k=1}^p \alpha_k f_k(x_{\alpha}) = \sum_{k=1}^p \alpha_k f_k(x_0);$$

$$\lambda_0 + \sum_{k=1}^p \alpha_k \sum_{\alpha=1}^n \lambda_{\alpha} f_k(x_{\alpha}) = \sum_{k=1}^p \alpha_k f_k(x_0);$$

esta relação é satisfeita se, e somente se, $\lambda_0 = 0$, o que resulta na restrição $\xi_n \subset \xi_{n+1}$ e $\sum_{\alpha=1}^n \lambda_{\alpha} f_k(x_{\alpha}) = f_k(x_0) \quad \forall k=1, \dots, p$; estas últimas p equações resultam na restrição de ξ_n à variedade linear ζ_p de dimensão $(n-p)$ tal que $\zeta_p \subset \zeta \subset \xi_n \subset \xi_{n+1}$.

Quando p cresce, a forma da tendência é assumida ser mais complexa, e com isso ganha-se em proximidade do estimador com o valor real.

O preditor de UK, dado por $Z_{UK}^* = \sum_{\alpha=1}^n \lambda_{\alpha} Z_{\alpha}$, é então definido como a única projeção de $Z(x_0)$ na variedade linear $\zeta_p \subset \xi_{n+1}$. Para determinar o BLUP de $Z(x_0)$, deve-se determinar as ponderações λ_{α} 's. Então, no problema de otimização, as ponderações devem ser tais que:

- (i) $\sum_{\alpha=1}^n \lambda_{\alpha} f_k(x_{\alpha}) = f_k(x_0)$, $\forall k=1, \dots, p$, isto é $Z_{UK}^* \in \zeta_p$;
 (ii) a distância quadrática $Z(x_0) - Z_{UK}^*$ seja mínima.

$$Z_{UK}^* - Z(x_0) = Z(x_0) - Z_{UK}^* - 2\langle Z(x_0), Z_{UK}^* \rangle, \text{ onde}$$

$$Z(x_0) = \sigma_{x_0 x_0};$$

$$Z_{UK}^* = \sum_{\alpha=1}^n \lambda_{\alpha} Z_{\alpha} = \langle \sum_{\alpha=1}^n \lambda_{\alpha} Z_{\alpha}, \sum_{\beta=1}^n \lambda_{\beta} Z_{\beta} \rangle = \sum_{\alpha=1}^n \lambda_{\alpha} \sum_{\beta=1}^n \lambda_{\beta} \langle Z_{\alpha}, Z_{\beta} \rangle = \sum_{\alpha=1}^n \lambda_{\alpha} \sum_{\beta=1}^n \lambda_{\beta} \sigma_{\alpha\beta};$$

$$\langle Z(x_0), Z_{UK}^* \rangle = \langle Z(x_0), \sum_{\alpha=1}^n \lambda_{\alpha} Z_{\alpha} \rangle = \sum_{\alpha=1}^n \lambda_{\alpha} \sigma_{\alpha x_0};$$

$$\text{então } \sigma_E^2 = \sigma_{x_0 x_0} - 2 \sum_{\alpha=1}^n \lambda_{\alpha} \sigma_{\alpha x_0} + \sum_{\alpha=1}^n \lambda_{\alpha} \sum_{\beta=1}^n \lambda_{\beta} \sigma_{\alpha\beta}.$$

Utilizando-se o método de Multiplicadores de Lagrange, para a otimização:

$$\sigma_E^2 = \sigma_{x_0 x_0} - 2 \sum_{\alpha=1}^n \lambda_{\alpha} \sigma_{\alpha x_0} + \sum_{\alpha=1}^n \lambda_{\alpha} \sum_{\beta=1}^n \lambda_{\beta} \sigma_{\alpha\beta} - 2 \sum_{k=1}^p \mu_k \left(\sum_{\alpha=1}^n \lambda_{\alpha} f_k(x_{\alpha}) - f_k(x_0) \right).$$

ϑ_k : lagrangeano;

$$\frac{\partial \sigma_E^2}{\partial \lambda_{\alpha}} = -2\sigma_{\alpha x_0} + 2 \sum_{\beta=1}^n \lambda_{\beta} \sigma_{\alpha\beta} - 2 \sum_{k=1}^p \vartheta_k f_k(x_{\alpha}), \forall \alpha=1, \dots, n.$$

Então as $(n+p)$ equações lineares nas incógnitas λ_{α} 's e μ_k 's, fornecem o sistema de *Kriging Universal* dado por:

$$(UK): \begin{cases} \sum_{\beta=1}^n \lambda_{\beta} f_k(x_{\beta}) = f_k(x_0), \forall k=1, \dots, p \\ \sum_{\beta=1}^n \lambda_{\beta} \sigma_{\alpha\beta} - \sum_{k=1}^p \vartheta_k f_k(x_{\alpha}) = \sigma_{\alpha x_0}; \forall \alpha=1, \dots, n. \end{cases}$$

Variância de predição

$$\sigma_E^2 = Z_{UK}^* - Z(x_0)$$

$$Z_{UK}^* = \sum_{\alpha=1}^n \lambda_{\alpha} \sum_{\beta=1}^n \lambda_{\beta} \sigma_{\alpha\beta} = \sum_{\alpha=1}^n \sum_{\beta=1}^n \lambda_{\alpha} \lambda_{\beta} \sigma_{\alpha\beta} = \sum_{\alpha=1}^n \sum_{\beta=1}^p \lambda_{\alpha} \vartheta_k f_k(x_{\alpha}) + \sum_{\alpha=1}^n \lambda_{\alpha} \sigma_{\alpha x_0}$$

$$Z_{UK}^* = \sum_{k=1}^p \theta_k f_k(x_0) + \sum_{\alpha=1}^n \lambda_{\alpha} \sigma_{\alpha x_0};$$

$$\text{como } Z(x_0) = \sigma_{x_0 x_0} \text{ e } \langle Z(x_0), Z_{UK}^* \rangle = \sum_{\alpha=1}^n \lambda_{\alpha} \sigma_{\alpha x_0},$$

a variância de *Kriging* é, então, dada por:

$$\sigma_E^2 = \sigma_{x_0 x_0} - \sum_{\alpha=1}^n \lambda_{\alpha} \sigma_{\alpha x_0} + \sum_{k=1}^p \theta_k f_k(x_0).$$

Pode-se escrever o sistema (UK) através da covariância centrada $\sigma_{\alpha\beta}^c = \sigma_{\alpha\beta} - m_{\alpha} m_{\beta}$, ou por meio de semi-variogramas $\gamma_{\alpha\beta} = \frac{1}{2} E[Z(x_{\alpha}) - Z(x_{\beta})]^2$.

Na prática, quando uma simples realização do processo estocástico não estacionário $\{Z(x)\}$ é observada, surgem problemas de inferência da covariância estacionária. Para maiores detalhes consulte algo sobre "Covariância Generalizada" em A. Journel [1982].

Propriedades do preditor de Kriging

(1) O preditor de *Kriging* além de ser um preditor não viciado é um interpolador exato, isto é, se o suporte v a ser estimado coincide com qualquer dos suportes v_{α} dos dados avaliados, ou seja se $x_0 \equiv x_{\alpha}$, então ρ é a i -ésima coluna de K , daí $\Lambda = K^{-1}\rho$ e $\hat{Z}(x_0) = \rho' K^{-1} Z(x) = Z_{\alpha}$, $Z_{OK}^* = Z_{\alpha}$, e consequentemente $\sigma_E^2 = 0$.

(2) A variância do erro de predição depende apenas do segundo momento do processo estocástico, e não de $E[Z(x)]$, nem dos valores observados. Desta forma é possível conhecer a precisão da predição antes de se observar o processo.

(3) O sistema de *Kriging* tem solução única se, e somente se, a matriz K for estritamente positiva-definida, para isto é necessário que o modelo de covariância seja positivo-definido, e que nenhum suporte v_{α} coincida com outro.

(4) A matriz K_0 depende apenas da geometria relativa dos suportes dos dados (v_α, v_β) , e não de todo o suporte V no domínio a ser predito. Suportes diferentes a serem estimados não implicam em diferentes cálculos de K_0 , isto é, se x_0 e x_0^* têm a mesma geometria com respeito a $x_1, \dots, x_n$ é suficiente resolver as equações de *Kriging* apenas uma vez.

(5) A geometria do fenômeno sob estudo é expressa em termos do sistema e da variância de *Kriging*. O $\gamma_{x_0 x_0}$ expressa a geometria do domínio V . A geometria da configuração dos dados é expressa por $\gamma_{\alpha\beta}$ e a regionalização própria é expressa pela função intrínseca semi-variograma.

2.5 Alguns Resultados Complementares

Seja $\{Z(x): x \in D \subseteq \mathcal{R}^d\}$ um processo estocástico tal que $E[Z(x)] = m(x)$ e $m(x) = \sum_{k=1}^p a_k f_k(x)$, $x \in D$.

Faz-se $Y(x) = Z(x) - m(x)$, supõe-se que $\{Y(x)\}$ seja um processo estocástico intrinsecamente estacionário. O preditor de *Kriging* para $Z(x_0)$ é dado por $Z_{UK}^* = \sum_{\alpha=1}^n \lambda_\alpha Z_\alpha$, que pela condição de não-viciabilidade resulta em $\sum_{\alpha=1}^n \lambda_\alpha f_k(x_\alpha) = f_k(x_0)$, $\forall k=1, \dots, p$. Sendo $\sigma_E^2 = C(0) - 2 \sum_{\alpha=1}^n \lambda_\alpha \sigma_{\alpha x_0} + \sum_{\alpha=1}^n \lambda_\alpha \sum_{\beta=1}^n \lambda_\beta \sigma_{\alpha\beta}$; o problema de otimização é minimizar $\sigma_E^2 = C(0) - \Lambda' K \Lambda - 2 \Lambda' \rho$ sujeito a $\Lambda' f_k(x) = f_k(x_0)$, $k=1, \dots, p$ e $x \in D$. O vetor Λ será único posto que $f_1(x), \dots, f_p(x)$ sejam linearmente independentes.

Sejam $a' = (a_1, \dots, a_p)$, $\phi'(x) = (f_1(x), \dots, f_p(x))$ e $F' = (\phi(x_1), \dots, \phi(x_n))$, pode-se escrever $Z(x) = Y(x) + m(x)$, como $\underline{Z}(x) = \underline{Y}(x) + F a$, $x \in D$. Se a for conhecido, o BLUP de $Y(x_0)$ é dado por $Y_K^* = \rho' K^{-1} (\underline{Z}(x) - F a)$ e o BLUP de $Z(x_0)$ por $Z_K^* = \phi'(x_0) a + \rho' K^{-1} (\underline{Z}(x) - F a)$, isto é, $Z_K^* = \rho' K^{-1} \underline{Z}(x) + (\phi'(x_0) - \rho' K^{-1} F) a$.

Se a for desconhecido. Seja $\hat{a}$ o BLUE de a (que é também o estimador de mínimos quadrados ponderados), $\hat{a} = (F' K^{-1} F)^{-1} F' K^{-1} \underline{Z}(x)$, então o preditor de *Kriging* para $Z(x_0)$ será:

$$Z_K^* = \rho' K^{-1} Z(x) + (\phi'(x_0) - \rho' K^{-1} F)(F' K^{-1} F)^{-1} F' K^{-1} Z(x);$$

$$Z_K^* = [\rho' K^{-1} + (\phi'(x_0) - \rho' K^{-1} F)(F' K^{-1} F)^{-1} F' K^{-1}] Z(x).$$

Lema 2.4.1: Z_K^* é um preditor não-viciado de $Z(x_0)$.

$$\text{Dem.:} \quad E[Z_K^*] = E\{[\rho' K^{-1} + (\phi'(x_0) - \rho' K^{-1} F)(F' K^{-1} F)^{-1} F' K^{-1}] Z(x)\}$$

$$E[Z_K^*] = \rho' K^{-1} F a + E[(\phi'(x_0) - \rho' K^{-1} F) \hat{a}]$$

$$E[Z_K^*] = \rho' K^{-1} F a + \phi'(x_0) a - \rho' K^{-1} F a$$

$$E[Z_K^*] = \phi'(x_0) a = E[Z(x_0)]. \blacksquare$$

Lema 2.4.2: Seja $\psi' Z(x)$ um preditor linear de $Z(x_0)$. $\psi' Z(x)$ será um preditor não-viciado para $Z(x_0)$ se, e somente se, $\psi' F = \phi'(x_0)$.

$$\text{Dem.:} \quad E[\psi' Z(x) - Z(x_0)] = E[\psi' Z(x)] - \phi'(x_0) a = \psi' E[Z(x)] - \phi'(x_0) a$$

$$E[\psi' Z(x) - Z(x_0)] = \psi' F a - \phi'(x_0) a;$$

$$E[\psi' Z(x) - Z(x_0)] = 0 \text{ se, e somente se, } \psi' F a = \phi'(x_0) a, \text{ ou seja, } \psi' F = \phi'(x_0). \blacksquare$$

Mostra-se, agora, que Z_K^* realmente é o BLUP para $Z(x_0)$.

Teorema 2.4.1: Seja $Z(x) = \phi'(x) a + Y(x)$, $x \in D$, então o BLUP de $Z(x_0)$ é dado por $\hat{\Lambda} Z(x) = \rho' K^{-1} Z(x) + (\phi'(x_0) - \rho' K^{-1} F) \hat{a}$.

Dem.: Deve-se mostrar que a variância do erro de predição de Z_K^* é mínima, então seja $\psi' Z(x)$ um preditor linear não viciado de $Z(x_0)$.

$$\text{Var}[\psi' Z(x) - Z(x_0)] = \text{Var}[\psi' Z(x) - \hat{\Lambda}' Z(x) + \hat{\Lambda}' Z(x) - Z(x_0)]$$

$$\begin{aligned} \text{Var}[\psi' Z(x) - Z(x_0)] &= \text{Var}[\hat{\Lambda}' Z(x) - Z(x_0)] + \text{Var}[\psi' Z(x) - \hat{\Lambda}' Z(x)] + \\ &\quad + 2\text{Cov}[(\psi - \hat{\Lambda})' Z(x), \hat{\Lambda}' Z(x) - Z(x_0)] \end{aligned}$$

$$\begin{aligned}\text{Var}[\psi'Z(x)-Z(x_0)] &= \text{Var}[\Lambda'Z(x)-Z(x_0)] + \text{Var}[(\psi-\Lambda)'Z(x)] + \\ &\quad + 2[(\psi-\Lambda)'K\Lambda - (\psi-\Lambda)'\rho].\end{aligned}$$

Seja $\Lambda = K^{-1}\rho + K^{-1}F(F'K^{-1}F)^{-1}(\phi(x_0) - F'K^{-1}\rho)$, então

$$\begin{aligned}\text{Cov}[(\psi-\Lambda)'Z(x), \Lambda'Z(x)-Z(x_0)] &= \\ &= (\psi-\Lambda)' \{ K[K^{-1}\rho + K^{-1}F(F'K^{-1}F)^{-1}(\phi(x_0) - F'K^{-1}\rho)] - \rho \} \\ &= (\psi-\Lambda)' [\rho + F(F'K^{-1}F)^{-1}(\phi(x_0) - F'K^{-1}\rho)] - \rho \\ &= (\psi-\Lambda)' F(F'K^{-1}F)^{-1}(\phi(x_0) - F'K^{-1}\rho).\end{aligned}$$

Pela condição de não-viciabilidade:

$\phi(x_0) = \psi'F = \Lambda'F \Rightarrow (\psi-\Lambda)'F = 0$, logo o produto interno é zero, concluindo-se que $\text{Var}[\psi'Z(x)-Z(x_0)] \geq \text{Var}[\Lambda'Z(x)-Z(x_0)]$, onde a igualdade ocorre se, e somente se, $\psi = \Lambda$. ■

Observação: Sejam $Z_K^* = \Lambda'Z(x)$ e Λ como definida na demonstração acima,

$$\sigma_K^2 = C(0) + \rho'K^{-1}\rho + (\phi'(x_0) - \rho'K^{-1}F)'(F'K^{-1}F)^{-1}(\phi'(x_0) - F'K^{-1}\rho)$$

$$\sigma_K^2 = C(0) + \rho^2K^{-1}\rho + \tau^2,$$

onde τ^2 é a perda envolvida ao se estimar o coeficiente de tendência "a".

Importante

A função $\phi'(x_0)\hat{a}$ é a reta de regressão estimada por Mínimos Quadrados Ponderados (WLS), e um dos seus possíveis empregos é para a predição de $Z(x_0)$. O erro de predição quando se usa a reta de regressão para prever $Z(x_0)$ é dado por $[\phi'(x_0)\hat{a} - Z(x_0)]$, e tem-se que:

$$\text{Var}[\phi'(x_0)\hat{a} - Z(x_0)] = \text{Var}[\phi'(x_0)\hat{a}] - 2\text{Cov}[\phi'(x_0)\hat{a}, Z(x_0)] + \text{Var}[Z(x_0)]$$

$\text{Var}[\phi'(x_0)\hat{a} - Z(x_0)] = \phi'(x_0)[\text{Var}(\hat{a})]\phi(x_0) - 2\phi'(x_0)\text{Cov}[\hat{a}, Z(x_0)] + C(0)$, mas

$$\text{Var}(\hat{a}) = (F'K^{-1}F)^{-1} \text{ e } \text{Cov}[\hat{a}, Z(x_0)] = (F'K^{-1}F)^{-1}(F'K^{-1}\rho), \text{ então}$$

$$\text{Var}[\phi'(x_0)\hat{a}-Z(x_0)]=\phi'(x_0)(F'K^{-1}F)^{-1}\phi(x_0)-2\phi'(x_0)(F'K^{-1}F)^{-1}F'K^{-1}\rho+C(0).$$

Como $\text{Var}[Z_K^*-Z(x_0)]=C(0)-\rho'K^{-1}\rho+(\phi(x_0)-F'K^{-1}\rho)'(F'K^{-1}F)^{-1}(\phi(x_0)-F'K^{-1}\rho)$, verifica-se que $\text{Var}[\phi'Z(x_0)\hat{a}-Z(x_0)]\geq\text{Var}[Z_K^*-Z(x_0)]$; a diferença pode ser considerada como um ganho na eficiência quando se usa Z_K^* ao invés da reta de regressão para predizer $Z(x_0)$, a igualdade ocorre se Z_α e Z_β são não-correlacionados $\forall \alpha, \beta$, então $\rho \equiv 0$ para todo $x_\alpha \neq x_0$, indicando, assim, que a reta de regressão é tão boa quanto o BLUP, em termos de predição de $Z(x_0)$.

2.6 A Técnica

Geoestatística é, basicamente uma metodologia de avaliação de recursos, que há muito é utilizada por Empresas de Mineração. Esta proporciona uma ferramenta muito poderosa e flexível, quando se trata de problemas de predição local. Algumas de suas possíveis aplicações, são tais como:

- predição de reservas locais por blocos;
- predição de reservas em determinadas localidades por camadas;
- predição da reserva total em certa região;
- simulação dos limites de reservatórios, bem como das reservas;
- predição e direcionamento relativo a reservatórios futuros;
- predição de reservas recuperáveis.

Kriging Clássico é um método de predição linear local. O objetivo como visto anteriormante é predizer o valor de $Z(x)$ sob $v(x_0)$, num domínio centrado em $x_0 \in D$.

$$\hat{Z}(x_0)=\hat{Z}_v(x_0)=\frac{1}{v}\int_{v(x_0)} Z(x)dx;$$

$$\sigma_E^2=\text{Var}[Z_K^*-Z_v(x_0)]=E[Z_K^*-Z_v(x_0)]^2=E[Z_K^*]^2+E[Z_v(x_0)]^2-2E[Z_K^*Z_v(x_0)] \text{ onde,}$$

$$E[Z_v(x_0)]^2 = \frac{1}{v^2} \int_v \int_v E[Z(x)Z(x')] dx dx',$$

$$E[Z_k^* Z_v(x_0)] = E[Z_v(x_0) \sum_{\alpha=1}^n \lambda_{\alpha} Z_{\alpha}] = \sum_{\alpha=1}^n \lambda_{\alpha} E[Z_v(x_0) Z_{\alpha}]$$

$$E[Z_k^* Z_v(x_0)] = \sum_{\alpha=1}^n \lambda_{\alpha} \frac{1}{v v_{\alpha}} \int_v \int_{v_{\alpha}} E[Z(x)Z(x')] dx dx',$$

$$E[Z_k^*]^2 = \sum_{\alpha=1}^n \sum_{\beta=1}^n \lambda_{\alpha} \lambda_{\beta} \frac{1}{v v_{\alpha} v_{\beta}} \int_v \int_{v_{\alpha}} \int_{v_{\beta}} E[Z(x)Z(x')] dx dx'.$$

Como $E[Z(x)Z(x')] = C(x, x') + m(x)m(x')$, e denotando-se

$$\bar{C}(v, v') = \frac{1}{v^2} \int_{v(x_0)} \int_{v(x'_0)} [C(x, x') + m(x)m(x')] dx dx',$$

tem-se:

$$\sigma_E^2 = \bar{C}(v, v) - 2 \sum_{\alpha=1}^n \lambda_{\alpha} \bar{C}(v, v_{\alpha}) + \sum_{\alpha=1}^n \sum_{\beta=1}^n \lambda_{\alpha} \lambda_{\beta} \bar{C}(v_{\alpha}, v_{\beta}).$$

A variância de dispersão é a medida da variância de extensão quando se faz $v(x_0)$ percorrer V .

$$D(v|V)^2 = \bar{C}(v, v) - \bar{C}(V, V)$$

Os dois conceitos se completam e permitem tratar, por exemplo, o problema de avaliação de blocos de minérios a partir de sondagens de determinado volume. Desta forma o erro cometido ao se estender o teor pontual para o bloco pode ser avaliado.

Se o processo for estacionário, o que implica que este é intrinsecamente estacionário, utiliza-se a relação formal $C(x, x+h) = C(0) - \gamma(h)$, onde $x, x+h \in D$; para escrever σ_E^2 em função de

$$\text{semi-variogramas } \bar{\gamma}(v, v') = \frac{1}{v^2} \int_{v(x_0)} \int_{v(x'_0)} \gamma(x, x') dx dx';$$

$$\sigma_E^2 = 2 \sum_{\alpha=1}^n \lambda_{\alpha} \bar{\gamma}(v, v_{\alpha}) - \sum_{\alpha=1}^n \sum_{\beta=1}^n \lambda_{\alpha} \lambda_{\beta} \bar{\gamma}(v_{\alpha}, v_{\beta}) - \bar{\gamma}(v, v').$$

Se o processo for não-estacionário de segunda ordem:

$$E[Z(x)] = m(x) = \sum_{k=1}^p a_k f_k(x);$$

$$E[Z_v(x_0)] = \sum_{k=1}^p a_k \frac{1}{v} \int_{v(x_0)} f_k(x) dx;$$

$$E[Z_k^*] = \sum_{\alpha=1}^n \lambda_{\alpha} \sum_{k=1}^p a_k \frac{1}{v_{\alpha}} \int_{v_{\alpha}} f_k(x) dx;$$

$$E[Z_v(x_0)]^2 = \frac{1}{v^2} \int_v \int_v E\{[Y(x) + m(x)][Y(x') + m(x')]\} dx dx';$$

$$E[Z_v(x_0)]^2 = \frac{1}{v^2} \int_v \int_v [C(x, x') + m(x)m(x')] dx dx' = \bar{C}(v, v);$$

$$E[Z_k^*]^2 = \sum_{\alpha=1}^n \sum_{\beta=1}^n \lambda_{\alpha} \lambda_{\beta} \frac{1}{v_{\alpha} v_{\beta}} \int_{v_{\alpha}} \int_{v_{\beta}} [C(x, x') + m(x)m(x')] dx dx';$$

$$E[Z_k^*]^2 = \sum_{\alpha=1}^n \sum_{\beta=1}^n \lambda_{\alpha} \lambda_{\beta} \bar{C}(v_{\alpha}, v_{\beta});$$

$$E[Z_k^* Z_v(x_0)] = \sum_{\alpha=1}^n \lambda_{\alpha} \frac{1}{v v_{\alpha}} \int_{v_{\alpha}} \int_v [C(x, x') + m(x)m(x')] dx dx';$$

$$E[Z_k^* Z_v(x_0)] = \sum_{\alpha=1}^n \lambda_{\alpha} \bar{C}(v_{\alpha}, v); \text{ então}$$

$$\sigma_E^2 = \bar{C}(v, v') - 2 \sum_{\alpha=1}^n \lambda_{\alpha} \bar{C}(v, v_{\alpha}) + \sum_{\alpha=1}^n \sum_{\beta=1}^n \lambda_{\alpha} \lambda_{\beta} \bar{C}(v_{\alpha}, v_{\beta}),$$

que pode ser escrito em termos do semi-variograma como:

$$\sigma_E^2 = 2 \sum_{\alpha=1}^n \lambda_{\alpha} \bar{\gamma}(v, v_{\alpha}) - \sum_{\alpha=1}^n \sum_{\beta=1}^n \lambda_{\alpha} \lambda_{\beta} \bar{\gamma}(v_{\alpha}, v_{\beta}) - \bar{\gamma}(v, v').$$

Com o objetivo de modelar a variabilidade espacial do processo, os cálculos de $\bar{\gamma}(v, v_{\alpha})$, $\bar{\gamma}(v_{\alpha}, v_{\alpha})$ e $\bar{\gamma}(v, v')$ podem ser feitos de duas formas:

(A) analítica, o cálculo é feito de maneira explícita;

$$\bar{\gamma}(v, v) = \frac{1}{vv} \int_v \int_v \gamma(x, x') dx dx'.$$

(B) numérica, o cálculo é feito através da discretização do suporte geométrico e aproximação de integrais por somatórios,

$$\bar{\gamma}(v, v) \approx \frac{1}{n_v n_v} \sum_{i \in n_v} \sum_{j \in n_v} \gamma(x_i, x_j), \text{ onde } n_k \text{ é o número de pontos no suporte } k.$$

É fato que o problema inicial em predição local, utilizando-se o método de *Kriging Clássico*, é o não conhecimento da função de covariância, ou do semi-variograma, daí surge a necessidade de estimá-lo.

Se os pontos aos quais a amostra se relaciona estão distribuídos em uma malha regular, estima-se o semi-variograma na direção fixa θ , baseado no semi-variograma amostral pelo método dos momentos,

$$\hat{\gamma}_{\theta}(h) = \frac{1}{2N_{\theta}(h)} \sum_1^{N_{\theta}(h)} [Z(x_1+h) - Z(x_1)]^2, \quad x_1, h \in D \quad [\text{Matheron, 1963}],$$

onde $N_{\theta}(h)$ é o número de pares de pontos que distam entre si de h em direções paralelas às previamente fixadas (perfis paralelos).

No caso de muitos perfis, calcula-se o semi-variograma médio para cada grupo de linhas paralelas em uma direção.

A estimação do semi-variograma em várias direções tem por objetivo revelar as anisotropias que possam existir no domínio. Semi-variogramas experimentais em várias direções frequentemente revelam tal anisotropia.

Na prática os dados se encontram irregularmente distribuídos num domínio D , isto é, os pontos amostrados são alinhados ao longo de perfis, mas estes não são paralelos. Então o semi-variograma deverá ser estimado através de uma regularização (ou agrupamento) angular (perfis não são paralelos) e por classes de distâncias (os dados não mostram qualquer padrão sistemático), a qual o procedimento é dado a seguir:

- (i) considera-se um ponto qualquer $x_i \in D$,
- (ii) segundo a direção θ (pré-fixada), gera-se um cone de abertura $2\delta\theta$, onde $\delta\theta$ é dito ângulo de regularização;
- (iii) reserva-se os pontos tomados pelo volume e que estejam na classe de distância $h \pm \Delta h$.

A seleção dos pontos pertencentes a um determinado cone pode ser feita utilizando-se o método das tangentes.

Uma vez estimados os valores de $\gamma(h)$, constrói-se um gráfico pontual de $[\hat{\gamma}(h) \text{ vs } h]$ para a direção θ . Com base neste gráfico ajusta-se um modelo para a função semi-variograma.

Importante

- 1) Na prática, calcula-se o semi-variograma amostral apenas para distâncias que não excedam a um quarto das dimensões do domínio, pois quando h cresce a correlação tende a diminuir, ou seja diminui o número de pares de pontos de considerável influência.
- 2) Para o cálculo do semi-variograma considera-se, apenas, os pontos aos quais está relacionado uma variável.

Para efeito de estimação, o semi-variograma experimental não deve ser usado diretamente. Um modelo teórico deve ser ajustado. A razão para tal fato é que o semi-variograma deve satisfazer a certas condições de consistência matemática, sem as quais talvez fosse possível a obtenção de variâncias negativas.

Seja $\{Z(x):x \in D\}$ um processo estocástico e $\sum_{\alpha=1}^n \lambda_{\alpha} Z_{\alpha}$ uma combinação linear das variáveis aleatórias envolvidas.

$$\text{"Var"} \left[\sum_{\alpha=1}^n \lambda_{\alpha} Z_{\alpha} \right] = - \sum_{\alpha=1}^n \sum_{\beta=1}^n \lambda_{\alpha} \lambda_{\beta} \gamma(x_{\alpha}, x_{\beta}) \text{ sujeito a } \sum_{\alpha=1}^n \lambda_{\alpha} = 0.$$

Como a variância deve ser não-negativa, a equação acima indica a condição a qual uma função $\gamma(h)$ deve satisfazer para ser um semi-variograma.

Note que, a condição $\sum_{\alpha=1}^n \lambda_{\alpha} = 0$, torna a classe dos semi-variogramas uma classe "pobre", embora esta contenha todos os semi-variogramas associados à função de covariância por $\gamma(h) = C(0) - C(h)$ e os semi-variogramas limitados que não são associados a uma covariância finita.

A prática mostra que a forma analítica do semi-variograma estimado depende basicamente dos seguintes itens:

- (a) inclinação na origem, que é geralmente obtida baseado nos três primeiros pontos deste;
- (b) efeito pepita (descontinuidade na origem), medidas são feitas arbitrariamente nos pontos x_{α_1} e x_{α_2} a diferença dos conteúdos relativos a estes pontos não tende a zero, no sentido de média quadrática, mas continua a flutuar com uma dispersão irredutível;
- (c) alcance, este é ajustado visualmente;
- (d) patamar, se existir este é ajustado onde o variograma se estabiliza e é comparado à variância dos dados, que podem ser iguais se o alcance for menor que a dimensão do domínio estudado;
- (e) anisotropia geométrica : considera-se o caso onde o variograma experimental em todas as direções mostram a mesma relação funcional e

patamar, porém diferentes valores para o *alcance*; anisotropia zonal: apresentam diferentes funcionais e/ou diferentes valores de *patamar* para diferentes direções. O grande problema é confundir anisotropia e tendência.

Em geral, um ajuste razoável pode ser obtido através da composição de mais de um modelo de semi-variograma. Este ajuste é sempre feito por tentativas e erros.

A Predição

O preditor de *Kriging Ordinário* de $Z_v(x_0)$ baseado nas n observações $Z(x_1), \dots, Z(x_n)$, onde $x_\alpha \in D$, $\alpha=1, \dots, n$; é obtido resolvendo-se o sistema $\Delta_0 \Lambda_0 = \Gamma_0$, onde,

$$\Delta_0 = \left[\begin{array}{c|c} \bar{\gamma}(v_\alpha, v_\beta) & 1 \\ \hline 1 & 0 \end{array} \right], \quad \Lambda_0 = \left[\begin{array}{c} \lambda_\alpha \\ \delta \end{array} \right] \text{ e } \Gamma_0 = \left[\begin{array}{c} \bar{\gamma}(v_\alpha, v) \\ 1 \end{array} \right]; \quad \forall \beta=1, \dots, n.$$

Este sistema de *Kriging* tem solução única se, e somente se, a matriz Δ_0 for estritamente positiva-definida; para isto é necessário que o modelo de semi-variograma seja estritamente negativo, e que nenhum suporte de dados v_α coincida com outro.

O preditor de *Kriging Universal* de $Z_v(x_0)$ baseado nas n observações $Z(x_1), \dots, Z(x_n)$, com $x_\alpha \in D$, $\alpha=1, \dots, n$; é obtido resolvendo-se o sistema $\Delta_U \Lambda_U = \Gamma_U$, onde,

$$\Delta_U = \left[\begin{array}{c|c} \bar{\gamma}(v_\alpha, v_\beta) & f_k(v_\alpha) \\ \hline f_k(v_\alpha) & 0 \end{array} \right], \quad \Lambda_U = \left[\begin{array}{c} \lambda_\alpha \\ \delta \end{array} \right] \text{ e } \Gamma_U = \left[\begin{array}{c} \bar{\gamma}(v_\alpha, v) \\ f_k(v_\alpha) \end{array} \right]; \quad \forall \beta=1, \dots, n.$$

Note que o sistema de *Kriging* considera, em suma, os seguintes elementos:

- (i) a posição relativa do domínio predito V e os pontos amostrais x_α , através dos termos $\bar{\gamma}(x_\alpha, V)$ ou $\gamma(x_\alpha - x_0)$;
- (ii) as distâncias entre pontos amostrais, através do termo $\gamma(x_\alpha - x_\beta)$;
- (iii) a estrutura da variável através do semi-variograma intrínseco.

Kriging se baseia na influência da vizinhança mais próxima do objeto da predição.

Em sua aplicação no campo de contorno automático e construção de modelos numéricos, *Kriging* deve aparecer como um método de interpolação dentre outros. Porém, uma propriedade muito importante de *Kriging* é que este é um interpolador exato. Isto faz com que *Kriging* seja um método apropriado para descrever mapas contorno. Ainda mais, como uma boa consequência disto, *Kriging* proporciona a variância do erro de predição, da qual um mapa de erro pode ser extraído.

Quando aplicado a mapas contorno, *Kriging* produz a cada ponto do mapa o preditor linear ótimo do valor real e seu associado desvio padrão do erro. Com a suposição da distribuição de probabilidade deste erro, pode-se construir um intervalo de confiança para o verdadeiro valor. O mapa *Kriging*, contudo, é mais suave que os dados reais, pois um preditor, mesmo sendo ótimo, não pode restituir detalhes que tenham sido suprimidos. Ainda, em algumas aplicações, o interesse está voltado ao mapa que explicita a variabilidade real.

2.7 Co-Kriging

Algumas vezes uma variável não pode ou não deve ser amostrada suficientemente (devido a alguma restrição, por exemplo técnica ou econômica)afim de proporcionar um preditor com precisão aceitável. O objetivo de *Co-Kriging* (técnica multivariada), em geral, é o de melhorar predições locais levando em consideração informações adicionais trazidas por uma variável diferente da que se deseja predizer. É evidente que a correlação espacial entre as variáveis deve ser suficientemente "forte" para que a variável adicional insira algum ganho de informação que sirva para a predição da primeira.

O uso deste método deve ser feito quando o objetivo principal é a redução da variância de predição de uma variável ou a necessidade da predição conjunta de várias variáveis, por ser este um processo onde tais variáveis podem ser conjuntamente preditas com base na intervariabilidade ou em informações da estrutura espacial.

Sob o ponto de vista teórico *Co-Kriging* é uma generalização de *Kriging*, ou melhor, *Kriging* é um *Co-Kriging* univariado..

Considera-se uma regionalização caracterizada por K funções aleatórias estacionárias de segunda ordem, espacialmente intercorrelacionadas $\{Z_k(x): k=1, \dots, K; x \in \mathcal{R}^d\}$ onde,

$$E[Z_k(x)] = m_k; \forall x;$$

$$E[Z_k(x+h)Z_k(x)] - m_k m_k = C_{k,k}(h) : \text{covariância cruzada};$$

$$E\{[Z_k(x+h) - m_k(x)][Z_k(x) - m_k(x)]\} = 2\gamma_{k,k}(h) : \text{variograma cruzado}.$$

Deseja-se prever o valor médio da função aleatória $Z_{k0}(x)$ sobre o suporte $V_{k0}(x_0)$, a partir das informações constituídas por N_k teores pontuais $z_k(x_{\alpha k})$, $\alpha k=1, \dots, N_k$ e $k=1, \dots, K$;

$$Z_{V_{k0}} = \frac{1}{V_{k0}} \int_{V_{k0}(x_0)} Z_{k0}(x) dx.$$

Os dados avaliados $\{Z_{\alpha k} : \alpha k=1, \dots, N_k, \text{ para cada } k=1, \dots, K\}$ são definidos no suporte $v_{\alpha k}$,

$$Z_{\alpha k} = \frac{1}{v_{\alpha k}} \int_{v_{\alpha k}} Z_k(x) dx.$$

O preditor de *Co-Kriging* $Z_{V_{k0}}^*$, é uma combinação linear de todos os dados avaliados de todas as K variáveis na corre regionalização:

$$Z_{V_{k0}}^*(x_{k0}) = \sum_k^K \sum_{\alpha k}^{N_k} \lambda_{\alpha k} Z_{\alpha k}(x_{\alpha k}).$$

O valor esperado do erro envolvido é:

$$E[Z_{V_{k0}} - Z_{V_{k0}}^*] = E[Z_{V_{k0}}] - E[Z_{V_{k0}}^*] = m_{k0} - \sum_{\alpha k0} \lambda_{\alpha k0} E[Z_{\alpha k0}] - \sum_{k \neq k0} \sum_{\alpha k} \lambda_{\alpha k} E[Z_{\alpha k}]$$

$$E[Z_{V_{k0}} - Z_{V_{k0}}^*] = m_{k0} \left[1 - \sum_{\alpha k0} \lambda_{\alpha k0} \right] - \sum_{k \neq k0} \sum_{\alpha k} \lambda_{\alpha k} m_k.$$

Pela condição de não-viciabilidade, $E[Z_{V_{k0}} - Z_{V_{k0}}^*] = 0$, tem-se:

$$\sum_{\alpha k0} \lambda_{\alpha k0} = 1 \quad \text{e} \quad \sum_{k \neq k0} \sum_{\alpha k} \lambda_{\alpha k} = 0.$$

A minimização da variância, $\sigma_{V_{k0}}^2 = E[Z_{V_{k0}} - Z_{V_{k0}}^*]^2$, sujeito às K restrições acima, resulta na resolução de um sistema linear de $(\sum_k N_k + K)$ equações, onde as incógnitas são as $\sum_k N_k$ ponderações $\lambda_{\alpha k}$'s e K parâmetros de Lagrange μ_k . Assim o sistema de *Co-Kriging* é dado por:

$$\sum_{k=1}^K \sum_{\alpha k}^{N_k} \lambda_{\alpha k} \bar{C}_{k'k}(v_{\beta k}, v_{\alpha k}) - \mu_k = \bar{C}_{k0k}(V_{k0}, v_{\beta k}); \quad \forall \alpha k = 1, \dots, N_k \quad \text{e} \quad k = 1, \dots, K;$$

$$\sum_{\alpha k0}^{N_{k0}} \lambda_{\alpha k0} = 1 \quad \text{e} \quad \sum_{k \neq k0} \sum_{\alpha k} \lambda_{\alpha k} = 0.$$

Escrevendo em termos do semi-variograma cruzado:

$$\sum_{k=1}^K \sum_{\alpha k}^{N_k} \lambda_{\alpha k} \bar{\gamma}_{k'k}(v_{\beta k}, v_{\alpha k}) + \mu_k = \bar{\gamma}_{k0k}(V_{k0}, v_{\beta k}); \quad \forall \alpha k = 1, \dots, N_k \quad \text{e} \quad k = 1, \dots, K;$$

isto se torna possível somente se $C_{k'k}(h) = C_{kk'}(h)$, o que implica que as covariâncias cruzadas são simétricas em $(h, -h)$, $\bar{\gamma}_{k'k}(h) = C_{k'k}(0) - C_{k'k}(h)$.

A variância de predição ou variância de *Co-Kriging* pode ser escrita como: $\sigma_{V_{k0}}^2 = \bar{C}_{k0k0}(V_{k0}, V_{k0}) + \mu_{k0} - \sum_{k=1}^K \sum_{\alpha k}^{N_k} \lambda_{\alpha k} \bar{C}_{k0k}(V_{k0}, v_{\alpha k})$.

Em termos do semi-variograma tem-se:

$$\sigma_{V_{k0}}^2 = -\bar{\gamma}_{k0k0}(V_{k0}, V_{k0}) + \mu_{k0} + \sum_{k=1}^K \sum_{\alpha k}^{N_k} \lambda_{\alpha k} \bar{\gamma}_{k0k}(V_{k0}, v_{\alpha k}).$$

A notação $\bar{C}_{k,k}(v_{\beta k}, v_{\alpha k})$ é a representação padrão do valor médio da covariância cruzada $C_{k,k}(h)$ quando uma extremidade do vetor h descreve o suporte $v_{\beta k}$, e a outra, independentemente, o suporte $v_{\alpha k}$.

Em notação matricial:

$$\underline{Z}(x)=[Z_1(x), \dots, Z_K(x)] ; E[\underline{Z}(x)]=[m_1, \dots, m_K] ; C(h)=E[Z_\alpha(x+h)Z_\beta^T(x)] ;$$

$$\bar{C}(h)=\begin{bmatrix} C_{1,1}(h) & \dots & C_{1,K}(h) \\ \vdots & & \vdots \\ C_{K,1}(h) & \dots & C_{K,K}(h) \end{bmatrix} ; \bar{C}(h)=E[\underline{Z}(x+h)\underline{Z}^T(x)] .$$

Objetivo: Predizer o valor $\underline{Z}(x_0)$ desconhecido, por:

$$\underline{Z}^*(x_0)=\sum_{\beta=1}^n \underline{Z}(x_\beta) \Lambda_\beta = [Z_1^*(x_0), \dots, Z_K^*(x_0)] , \text{ onde } \Lambda_1, \dots, \Lambda_n \text{ são matrizes } K \times K.$$

Uma condição suficiente para a não-viciabilidade, $E[\underline{Z}^*(x_0)-\underline{Z}(x_0)]=[0, \dots, 0]$, é que $\sum_{\beta=1}^n \Lambda_\beta = I$, onde n é o número de locações amostradas.

Deve-se minimizar $\sum_{\alpha=1}^n \text{Var}[Z_\alpha(x_0)-Z_\alpha^*(x_0)]$ sujeito a $\sum_{\alpha=1}^n \Lambda_\alpha = I$, originando o sistema de *Co-Kriging*, $D=U\Lambda$, onde

$$U = \begin{bmatrix} \bar{C}(x_1, x_1) & \dots & \bar{C}(x_1, x_n) & I \\ \vdots & & \vdots & \vdots \\ \bar{C}(x_n, x_1) & \dots & \bar{C}(x_n, x_n) & I \\ I & \dots & I & 0 \end{bmatrix} , \Lambda = \begin{bmatrix} \Lambda_1 \\ \vdots \\ \Lambda_n \\ \mu^n \end{bmatrix} \text{ e } D = \begin{bmatrix} \bar{C}(x_1, x_0) \\ \vdots \\ \bar{C}(x_n, x_0) \\ I \end{bmatrix} .$$

O valor mínimo de $\sum_{\alpha=1}^n \text{Var}[Z_\alpha(x_0)-Z_\alpha^*(x_0)]$ é dado por $\sigma_{CK}^2 = \text{Tr}[\bar{C}(0)] - \text{Tr}[\sum_{\beta=1}^n \bar{C}(x_\beta, x_0) \Lambda_\beta] - \text{Tr}[\mu]$, onde $\text{Tr}[A]$ denota o traço da matriz A , isto é a soma dos elementos da diagonal principal.

Em termo dos semi-variogramas cruzados: $C=V\Lambda$,

$$V = \begin{bmatrix} \bar{\gamma}(x_1, x_1) & \dots & \bar{\gamma}(x_1, x_n) & I \\ \vdots & & \vdots & \vdots \\ \bar{\gamma}(x_n, x_1) & \dots & \bar{\gamma}(x_n, x_n) & I \\ I & \dots & I & 0 \end{bmatrix} , \Lambda = \begin{bmatrix} \Lambda_1 \\ \vdots \\ \Lambda_n \\ \mu^n \end{bmatrix} \text{ e } C = \begin{bmatrix} \bar{\gamma}(x_1, x_0) \\ \vdots \\ \bar{\gamma}(x_n, x_0) \\ I \end{bmatrix} .$$

$$\sigma_{CK}^2 = -\text{Tr}[\bar{\gamma}(0)] + \text{Tr}[\sum_{\beta=1}^n \bar{\gamma}(x_\beta, x_0) \Lambda_\beta] - \text{Tr}[\mu] .$$

Pode-se mostrar que o sistema (CK) proporciona uma única solução se a matriz de covariâncias for estritamente positiva definida. Para tal é suficiente que um ponto positivo no modelo de correção seja adotado, e que nenhum valor dos dados seja redundante com respeito a outro.

CAPÍTULO 3

KRIGING NÃO-LINEAR

Em alguns problemas de predição local, não é possível, ou adequado o uso de preditores lineares, com o objetivo de predizer o valor de uma variável aleatória, baseado na realização $\{Z(x_\alpha)=z(x_\alpha):\alpha=1,\dots,n\}$ de um processo estocástico em determinado painel, ou para predizer a proporção de valores $Z(x)$ que são maiores que um dado limite, pelo motivo daquele poder ser não adequado, devido às limitações teóricas e práticas. Mas pode-se remover tal limitação de linearidade, e partir em busca de um preditor não-linear $Z^*=f(Z_1,\dots,Z_n)$, função dos n dados avaliados. Isto é, considera-se um espaço vetorial muito maior que ξ_{n+1} , com o preditor de *Kriging Não-linear* sendo deduzido através da projeção do valor desconhecido $Z(x_0)$ neste espaço maior, que por sua vez faz este preditor melhor que o preditor de *Kriging Linear*.

Seja $Z(x_0)$ uma variável aleatória a ser predita baseado na realização $\{Z(x_\alpha)=z(x_\alpha):\alpha=1,\dots,n\}$ de um processo estocástico. Sabe-se que o melhor preditor possível é, por definição, a esperança condicional, isto é, $E[Z(x_0)|Z(x_1),\dots,Z(x_n)]$, ou seja, esta é a melhor aproximação de $Z(x_0)$ por uma função mensurável $f(Z(x_1),\dots,Z(x_n))$ das n observações. O conjunto $\mathcal{H}_n$ de todas estas funções mensuráveis a n variáveis é um espaço vetorial, note que $\xi_{n+1} \subset \mathcal{H}_n$. Assim, quando $Z(x_0)$ é projetado no espaço vetorial $\mathcal{H}_n$, o correspondente preditor de *Kriging* é idêntico à esperança condicional $E_n(Z(x_0))$.

Note que $E_n(Z(x_0))$, considerado como um preditor de $Z(x_0)$ é não-viciado e, $Y=f(Z(x_1),\dots,Z(x_n))$, sendo um elemento arbitrário de $\mathcal{H}_n$, a projeção $E_n(Z(x_0))$ de $Z(x_0)$ em $\mathcal{H}_n$ é caracterizada pelas relações:

$$Z(x_0) - E_n(Z(x_0)) = \min_{Y \in \mathcal{H}_n} \|Z(x_0) - Y\|, \quad \forall Y \in \mathcal{H}_n;$$

Com o objetivo de construir a projeção $E_n(Z(x_0))$, isto é, determinar a função n -mensurável $E_n(Z(x_0)) = f_k(Z(x_1), \dots, Z(x_n)) \in \mathcal{H}_n$, é necessário a obtenção de um produto interno do tipo

$$\begin{aligned} \langle Z(x_0) - f_k(Z(x_1), \dots, Z(x_n)), f(Z(x_1), \dots, Z(x_n)) \rangle &= \\ &= E\{[Z(x_0) - f_k(Z(x_1), \dots, Z(x_n))]f(Z(x_1), \dots, Z(x_n))\}, \quad \forall f \in \mathcal{H}_n, \end{aligned}$$

para isto se realizar, deve-se conhecer a distribuição conjunta das $(n+1)$ variáveis $(Z(x_0), Z(x_1), \dots, Z(x_n))$.

Em geral, quando a informação é limitada a uma única realização do processo estocástico $\{Z(x)\}$ em cada dos n pontos $\{x_\alpha: \alpha=1, \dots, n\}$, esta lei de distribuição conjunta $(n+1)$ -dimensional é desconhecida, e portanto a esperança condicional é inacessível.

Na prática, a esperança condicional mencionada, não é facilmente obtida através dos dados, exceto possivelmente se o processo estocástico $\{Z(x)\}$ for Gaussiano e estacionário. É um resultado clássico de probabilidade que a esperança condicional de um processo estocástico Gaussiano é idêntica ao melhor preditor linear, isto é, $Z(x_0)$ tem a mesma projeção em $\mathcal{H}_n$ e ξ_{n+1} , e sua projeção pode ser determinada através de um sistema de *Kriging Linear*.

3.1 Processo de Kriging Disjuntivo

Pelo fato do espaço das funções que são combinações lineares, ξ_{n+1} , ser muito menor que o espaço das funções mensuráveis a n -variáveis $f(Z(x_1), \dots, Z(x_n)) \in \mathcal{H}_n$, a aproximação dada pelo *Kriging Linear*, o qual projeta $Z(x_0)$ no subespaço vetorial gerado por todas as combinações lineares dos n dados avaliados, fica longe de ser tão precisa quanto a esperança condicional, que é o preditor de $Z(x_0)$, utilizando-se de sua projeção no espaço $\mathcal{H}_n$, que no caso geral, não-gaussiano, é inacessível. Surgiu, então, a idéia de *Kriging Disjuntivo*, ou seja de um possível método não-linear, teoricamente mais preciso que o *Kriging Clássico*, fazendo-se a escolha de um espaço maior

que o das combinações lineares, ξ_{n+1} , porém suficientemente menor que o das funções mensuráveis, $\mathcal{H}_n$.

Na procura deste espaço vetorial, um bom ponto de partida é dado pela técnica de análise de dados conhecida como "Codificação Disjuntiva", que é descrita a seguir.

Seja $B_i, i \in I$, uma partição da reta real (em intervalos ou classes). Cada variável Z é associada a uma família de variáveis aleatórias $Z_i, i \in I$ definida por:

$$Z_i = I_{B_i}(Z) = \begin{cases} 1 & \text{se } Z \in B_i \\ 0 & \text{se } Z \notin B_i. \end{cases}$$

Note que a família das combinações lineares $\sum_{i=1}^n \lambda_i Z_i$ é idêntica à classe das funções B_i -mensuráveis, $f(Z)$. Se a partição B_i tende a ser cada vez menor (refinamento), obtém-se no limite a classe das funções Borel-mensuráveis, $f(Z)$.

No problema de predição ao invés de se usar o preditor de *Kriging Clássico* para $Z(x_0)$, baseado nas realizações $\{Z(x_\alpha): \alpha=1, \dots, n\}$, faz-se o *Kriging* destas, através de uma família mais rica constituída pelas variáveis $Z_{i,\alpha} = I_{B_i}(Z_\alpha)$.

O correspondente preditor terá a forma $Z^* = \sum_{i=1}^n g_{i,\alpha} Z_{i,\alpha}$, onde $g_{i,\alpha} = \sum_{j \in I} \lambda_{j,\alpha} I_{B_j}$ é uma função B_i -mensurável.

A matriz de covariância não-centrada das $Z_{i,\alpha}$'s depende apenas da lei bi-dimensional $F_{\alpha\beta}$, onde $F_{\alpha\beta}(B_i, B_j) = E[Z_{i,\alpha} Z_{j,\beta}]$.

Assim, este método requer apenas o conhecimento das leis bi-dimensionais dos pares $(Z(x_0), Z_\alpha)$ e (Z_α, Z_β) .

No caso limite de uma partição infinitamente fina B_i , tem-se que a procura por uma melhor aproximação de $Z(x_0)$, é dada pela soma $\sum_{i=1}^n g_i Z(x_i)$, onde os g_i 's são funções mensuráveis a uma variável. O

correspondente preditor, então, é chamado de *Kriging Disjuntivo*, que é um *Co-Kriging* de funções indicatrizes.

Desta forma, um espaço com projeção, incluído em $\mathcal{H}_n$ e incluindo ξ_{n+1} deve ser considerado; as inclusões neste caso são, naturalmente, estritas. Uma boa escolha para este espaço intermediário é o espaço vetorial $\mathcal{D}_n$, gerado pela soma de n -funções mensuráveis a uma variável, isto é,

$$\mathcal{D}_n = \{g_1(Z(x_1)) + \dots + g_n(Z(x_n)) : g_i \in \mathcal{H}_n, i=1, \dots, n\},$$

que satisfaz a $\xi_{n+1} \subset \mathcal{D}_n \subset \mathcal{H}_n$.

O processo de *Kriging Disjuntivo* é, então, por definição, o processo de projeção Z_{DK}^* , de $Z(x_0)$ em $\mathcal{D}_n$.

Esta projeção é caracterizada pelas seguintes condições:

- (a) $Z_{DK}^* \in \mathcal{D}_n$, isto é, $Z_{DK}^* = \sum_{\alpha=1}^n g_{\alpha} Z(x_{\alpha})$;
- (b) o vetor $(Z(x_0) - Z_{DK}^*)$ é ortogonal a qualquer vetor Z de $\mathcal{D}_n$, isto é, $\langle Z(x_0) - Z_{DK}^*, Z \rangle = 0, \forall Z \in \mathcal{D}_n$.

Como $\mathcal{D}_n$ é gerado pelas funções mensuráveis $g(Z_{\beta})$, então a condição (b) pode ser escrita da forma:

- (b') $\langle Z(x_0), g(Z_{\beta}) \rangle = \langle Z_{DK}^*, g(Z_{\beta}) \rangle$, para qualquer $\beta=1, \dots, n$; e qualquer função mensurável a uma variável, g .

Teorema 4.1.1 A condição (b') é satisfeita se, e somente se, $Z(x_0)$ e Z_{DK}^* admitem a mesma esperança condicional, uma vez dado cada um dos Z_{β} 's, isto é,

$$E\{Z(x_0) | Z_{\beta}\} = E\{Z_{DK}^* | Z_{\beta}\}, \forall \beta=1, \dots, n.$$

Dem.: $\langle Z(x_0), g(Z_{\beta}) \rangle = \langle Z_{DK}^*, g(Z_{\beta}) \rangle$, $\forall \beta=1, \dots, n$ e $\forall g \in \mathcal{D}_n$, é satisfeita se, e somente se, (b') ocorre. Seja δ_{β} o espaço vetorial gerado por todas as

funções mensuráveis a uma variável de um dado particular Z_β . A esperança condicional

$$E\{Z(x_0)|Z_\beta\} = E_\beta\{Z(x_0)\}$$

é por definição a projeção de $Z(x_0)$ em δ_β e, então, o vetor $(Z(x_0) - E_\beta Z(x_0))$ é ortogonal a qualquer vetor $g(Z_\beta) \in \delta_\beta$, isto é,

$$\langle Z(x_0), g(Z_\beta) \rangle = \langle E_\beta(Z(x_0)), g(Z_\beta) \rangle, \quad \forall g(Z_\beta) \in \delta_\beta,$$

então (b') pode ser reescrita da forma:

$$\langle E_\beta(Z(x_0)), g(Z_\beta) \rangle = \langle E_\beta(Z_{DK}^*), g(Z_\beta) \rangle, \quad \forall g(Z_\beta) \in \delta_\beta,$$

que é satisfeita se, e somente se, $E_\beta(Z(x_0)) = E_\beta(Z_{DK}^*)$, $\forall \beta = 1, \dots, n$. ■

Considerando a expressão $Z_{DK}^* = \sum_{\alpha=1}^n g_\alpha Z(x_\alpha)$, dada por (a), o preditor de *Kriging Disjuntivo* é, finalmente, caracterizado pelo seguinte sistema de n equações:

$$(DK): \sum_{\alpha=1}^n E\{g_\alpha(Z_\alpha)|Z_\beta\} = E\{Z(x_0)|Z_\beta\}, \quad \forall \beta = 1, \dots, n.$$

No caso geral, este sistema de equações, proporciona como soluções funções g_α , $\alpha = 1, \dots, n$, na forma de equações integrais. Estas funções são, pelo fato de não serem inferidas facilmente, então, aproximadas por expansões limitadas em polinômios ortogonais, ver-se-á que na realidade são expansões em fatores ortogonais.

O sistema de *Kriging Disjuntivo* requer apenas cálculos de esperanças condicionais do tipo $E\{g_\alpha(Z_\alpha)|Z_\beta\}$, isto é, o conhecimento das distribuições bivariadas do processo estocástico $\{Z(x), x \in D\}$.

Note que, *Kriging Disjuntivo* pode ser usado para prever qualquer função mensurável ϕ da incógnita $Z(x_0)$. O valor $\phi(Z(x_0))$ é projetado no espaço $\mathcal{D}_n$, proporcionando o preditor $\phi_{DK}^* = \sum_{\alpha=1}^n g_\alpha \phi(Z(x_\alpha))$, caracterizado pelo sistema:

$$(DK)_\phi: \sum_1^n E\{g_\alpha \phi(Z_\alpha) | \phi(Z_\beta)\} = E\{\phi(Z(x_0)) | \phi(Z_\beta)\}, \quad \forall \beta=1, \dots, n.$$

Variância de Predição

A resolução do sistema de equações (DK), proporciona o preditor de *Kriging Disjuntivo* Z_{DK}^* e a correspondente variância de predição:

$$\sigma_{DK}^2 = Z(x_0) - Z_{DK}^* \quad {}^2 = E\{[Z(x_0) - Z_{DK}^*]^2\}.$$

Pelo fato do espaço $\mathcal{D}_n \subset \mathcal{H}_n$, a projeção de $Z(x_0)$ em $\mathcal{H}_n$, que é a esperança condicional $E_n\{Z(x_0)\}$, é mais precisa na predição de $Z(x_0)$, que a projeção deste em $\mathcal{D}_n$, que fornece o preditor Z_{DK}^* . O teorema das três perpendiculares, mostra que Z_{DK}^* é também a projeção da $E_n[Z(x_0)]$ em $\mathcal{D}_n$, e do teorema de Pitágoras obtém-se a seguinte relação:

$$\sigma_{DK}^2 = Z(x_0) - Z_{DK}^* \quad {}^2 = Z(x_0) - E_n(Z(x_0)) \quad {}^2 + E_n(Z(x_0)) - Z_{DK}^* \quad {}^2,$$

e do teorema da projeção que

$$\sigma_{DK}^2 = E[Z(x_0)]^2 + E[Z_{DK}^*]^2 = E[Z(x_0)]^2 + E[Z(x_0)Z_{DK}^*],$$

onde a última igualdade é obtida posto que para *Kriging Disjuntivo* $E[Z(x_0)] = E[Z_{DK}^*]$.

3.2 Representação Fatorial

No caso geral, não se está apto a resolver o sistema de (DK). Isto, porém só se torna possível, utilizando aproximações (uma discretização adequada), o que implica em cálculos computacionais dispendiosos. Por esta razão deve-se procurar por casos particulares de simplificação, e usar uma técnica de análise multivariada, conhecida como "representação fatorial" de uma lei bi-dimensional.

Seja $F(x, y)$ a lei conjunta das variáveis aleatórias $X=Z(x)$ e $Y=Z(x+h)$, e ainda, $H_1=L^2(\mathcal{R}, F(x))$ e $H_2=L^2(\mathcal{R}, F(y))$ os dois espaços de Hilbert associados a $F(x)$ e $F(y)$. Para cada função $f \in H_2$, deve-se

associar sua esperança condicional dado X , diga-se $E_x f$; define-se $g = E_x f$, ou seja $g(x) = E[f(Y) | X=x]$.

Este operador $E_x: H_2 \rightarrow H_1$ é linear, pois

$$E[f(Y) + f'(Y) | X=x] = E[f(Y) | X=x] + E[f'(Y) | X=x] \quad \forall f, f' \in H_2;$$

contínuo, uma vez que o operador linear E_x é limitado, isto é, $E_x < \infty$; e de norma unitária, posto que

$$E_x = \sup \{ E_x f : f \in H_2, \|f\| = 1 \}.$$

Da mesma forma, a esperança condicional dado Y , proporciona um operador $E_y: H_1 \rightarrow H_2$, com as mesmas características.

Para algum $f \in H_2$ e $g \in H_1$, tem-se:

$$\langle E_x f, g \rangle = E[E(f(Y) | X=x)g] = gE[f(Y)] = E[gf(Y)] = \langle f, g \rangle = \langle f, E_y g \rangle,$$

tal que E_x e E_y são adjuntos, isto é, $E_x = E_y^*$; ou seja, sendo $E_x: H_2 \rightarrow H_1$, onde H_2 e H_1 são espaços lineares normados, o adjunto de E_x é um operador $E_x^*: H_2^* \rightarrow H_1^*$, definido assim: se $g \in H_2^*$ faça-se $(E_x^* g)(f) = g(E_x f)$; $f \in H_2$. Segue, então, que o operador $E_x E_y = E_y E_x^*$ de H_1 em si é Hermitiano.

Como $\langle E_x E_y f, f \rangle = \langle E_y E_x^* f, f \rangle = \langle E_y f, E_y f \rangle = \|E_y f\|^2 \geq 0$, e $E_x E_y$ é de norma unitária, então, seus autovalores λ são reais e satisfazem a $0 \leq \lambda \leq 1$.

A função constante 1 é uma autofunção associada ao autovalor $\lambda=1$ (esta é única exceto se existem funções mensuráveis f e g tais que $f(Y)=g(X)$ q.c., isto é, com probabilidade um).

Seja λ um autovalor não nulo de $E_x E_y$ ($0 < \lambda \leq 1$), e f_λ uma autofunção associada a λ . Deve-se supor $f_\lambda = 1$. Nota-se que $E_x E_y f_\lambda = \lambda f_\lambda$ implica em $E_y E_x (E_y f_\lambda) = \lambda E_y f_\lambda$, e tem-se que $g = E_y f_\lambda$ é uma autofunção de $E_y E_x$, associada pelo mesmo autovalor λ . Ainda mais, $E_y f_\lambda^2 = \langle E_x E_y f_\lambda, f_\lambda \rangle = \lambda \geq 0$.

Assim, $g_\lambda = (E_Y f_\lambda) / \sqrt{\lambda}$ é de norma unitária e ainda, $f_\lambda = (\sqrt{\lambda})^{-1} E_X g_\lambda$, definem uma correspondência 1-1 entre a autofunção de norma unitária do operador $E_X E_Y$ e $E_Y E_X$, associado pelo mesmo autovalor $\lambda > 0$. Em termos de análise multivariada, as autofunções f_λ e g_λ são chamadas fatores.

Supõe-se, agora, que a lei $F(x, y)$ admite uma densidade mensurável ϕ com respeito ao produto, $F_X(x)F_Y(y)$, das leis marginais, isto é,

$$F(x, y) = \phi(x, y) F_X(x) F_Y(y).$$

Em particular $\phi(x, y) F_Y(y)$ é a lei condicional de Y dado $X=x$, e neste caso o operador E_X é definido por:

$$(E_X f)(x) = \int \phi(x, y) f(y) dF_Y(y).$$

Suponha, que a densidade ϕ tenha quadrado integrável, isto é,

$$\iint |\phi(x, y)|^2 dF_X(x) dF_Y(y) < \infty.$$

Então, pode ser mostrado que os operadores $E_X E_Y$ e $E_Y E_X$ são compactos. Seus autovalores formam uma sequência não-crescente $1 = \lambda_0 \geq \lambda_1 \geq \lambda_2 \geq \dots$, tais que $\sum_{k=1}^{\infty} \lambda_k^2 < \infty$. Onde cada autovalor $\lambda_k > 0$ é associado pela autofunção f_k de $E_X E_Y$ (g_k de $E_Y E_X$), e é possível escolher estas autofunções tais que sejam ortonormadas. Então a densidade ϕ da lei $F(x, y)$, admite a seguinte expansão:

$$\phi(x, y) = \sum_{k=0}^{\infty} \sqrt{\lambda_k} f_k(y) g_k(x),$$

chamada representação fatorial da lei $F(x, y)$.

As correspondentes representações dos operadores esperança condicional E_X e E_Y , são dadas por:

$$E_Y f = \sum_{k=0}^{\infty} \sqrt{\lambda_k} \langle f, f_k \rangle g_k$$

e

$$E_x g = \sum_{k=0}^{\infty} \sqrt{\lambda_k} \langle g, g_k \rangle f_k.$$

3.3 Modelo Isofatorial

Sejam Y_i , $i=1, \dots, n$, variáveis aleatórias, F_i a lei marginal de Y_i e F_{ij} a lei bi-dimensional do par (Y_i, Y_j) . Diz-se que os Y_i 's formam um modelo isofatorial, se:

- (i) todos os Y_i 's têm a mesma lei marginal, $F_i = F$;
- (ii) a lei bi-dimensional F_{ij} admite representação da forma fatorial, com os mesmos fatores $\chi_k = \pm f_k = \pm g_k$, para todos os pares (i, j) ; isto é,

$$F_{ij}(x, y) = \sum_{k=0}^{\infty} T_k(i, j) \chi_k(x) \chi_k(y) F(x) F(y).$$

A lei F e os fatores χ_k são os mesmos para todos os pares (i, j) . Os coeficientes $T_k(i, j) = E[\chi_k(Y_i) \chi_k(Y_j)]$, são coeficientes de correlação, $-1 \leq T_k(i, j) \leq 1$, e em particular $T_0(i, j) = 1$.

Segue que, para qualquer função mensurável f tal que $E[f^2(Y_j)] < \infty$, tem-se:

$$(S): E[f(Y_j) | Y_i] = \sum_{k=0}^{\infty} T_k(i, j) \langle f, \chi_k \rangle \chi_k(Y_i),$$

$$\text{onde, } \langle f, \chi_k \rangle = \int f(y) \chi_k(y) dF(y).$$

Agora, considera-se o problema de *Kriging Disjuntivo* de $Z(x_0)$, baseado em Z_α , $\alpha=1, \dots, n$, e suponha que cada uma destas variáveis seja dada pela transformação da correspondente variável aleatória Y_i de um modelo isofatorial, isto é, $Z(x_0) = f(Y_0)$, com segundo momento finito e $Z(x_\alpha) = g_\alpha(Y_\alpha)$, $\alpha=1, \dots, n$.

Deve-se encontrar o preditor de *Kriging Disjuntivo* de $Z(x_0)=f(Y_0)$, baseado em Y_α , $\alpha=1,\dots,n$.

Fazendo-se uso de (S), o sistema (DK) pode ser reescrito como:

$$(DK_R): \sum_{k=0}^{\infty} T_k(0,\alpha) \langle f, \chi_k \rangle \chi_k(Y_\alpha) = \sum_{k=0}^{\infty} \sum_{l=1}^n T_k(\alpha,\beta) \langle \chi_k, f_\beta \rangle \chi_k(Y_\alpha).$$

Os fatores χ_n 's sendo ortogonais, o sistema (DK_R) torna-se equivalente a:

$$\sum_{l=1}^n T_k(\alpha,\beta) \langle \chi_k, f_\beta \rangle = T_k(0,\alpha) \langle f, \chi_k \rangle; \quad \forall \alpha=1,\dots,n \quad \text{e} \quad k=0,1,2,\dots$$

Por conveniência faz-se $f_k = \langle f, \chi_k \rangle$ e $f_k^\alpha = \langle f_\alpha, \chi_k \rangle$, tais que a função a ser predita e seu preditor de *Kriging Disjuntivo* $Z_{DK}^* = \sum_{l=1}^n f_l(Y_\alpha)$, admitam, respectivamente, as expansões:

$$f(Y_0) = \sum_{k=0}^{\infty} f_k \chi_k(Y_0),$$

e

$$Z_{DK}^* = \sum_{k=1}^n \sum_{l=0}^{\infty} f_k^\alpha \chi_k(Y_\alpha).$$

Então, para cada $k=1,2,\dots$ os coeficientes f_k^α são obtidos resolvendo-se o sistema:

$$(CK): \sum_{l=1}^n f_k^\beta T_k(\alpha,\beta) = f_k T_k(0,\alpha).$$

De (S) segue que $(T_k(i,j))$ é a matriz de covariâncias de $\chi_k(Y_\alpha)$, tal que o último sistema acima representa o *Kriging Clássico* de $\chi_k(Y_0)$ baseado em $\chi_k(Y_\alpha)$. Ou seja, na estrutura de um modelo isolado, o problema é largamente simplificado, deve-se submeter a *Kriging* cada fator separadamente. Na prática usa-se expansões limitadas a uma dada ordem K , e daí deve-se apenas resolver, separadamente, K sistemas $n \times n$.

Note que, outra consequência da ortogonalidade dos fatores χ_k é que a variância de predição σ_{DK}^2 é dada pela soma das variâncias de *Kriging* dos fatores, isto é,

$$\sigma_{DK}^2 = \sum_{k=1}^{\infty} \sigma_{K,k}^2; \text{ onde } \sigma_{K,0}^2 = 0 \text{ e}$$

$$\sigma_{K,k}^2 = f_k^2 + f_k \sum_{\alpha=1}^n T_k(0, \alpha) f_k^{\alpha}.$$

CAPÍTULO 4

ANAMORFISMO GAUSSIANO

Pelo fato de que no processo Gaussiano estacionário o melhor preditor de $Z(\mathbf{x}_0)$ pode ser obtido, seria conveniente, se possível, transformar a variável aleatória $Z(\mathbf{x})$ com distribuição estacionária (estacionaridade, aqui, se refere à estacionaridade estrita) arbitrária, em uma variável aleatória gaussiana, centrada, que origine um processo Gaussiano estacionário $\{Y(\mathbf{x})\}$.

Seja ϕ a transformação $Z(\mathbf{x})=\phi[Y(\mathbf{x})]$; esta transformada é dita anamorfismo gaussiano, no apêndice 4, mostra-se que ela existe é única e invertível (se a variável for contínua). Note que isto deve ocorrer para $Z(\mathbf{x}_\alpha) \neq Z(\mathbf{x}_\beta) \forall \alpha \neq \beta$.

Ainda mais, esta transformada gaussiana produz uma variável aleatória que é Normal univariada com média zero e variância unitária, obtida através de uma variável com distribuição arbitrária.

Uma outra hipótese verificada é que todas as distribuições multivariadas do processo estocástico $\{Y(\mathbf{x})\}$ são também gaussianas. Sob estas hipóteses a esperança condicional $E_n\{Y(\mathbf{x}_0)\}$ é idêntica ao preditor de *Kriging Simples* $Y_{SK}^* = \sum_{\alpha=1}^n \lambda_\alpha Y_\alpha$, obtido pelo sistema de equações (SK).

Quando os n dados transformados Y_α são fixados, a lei condicional de $Y(\mathbf{x}_0)$ é também gaussiana, com valor esperado $Y_{SK}^* = E_n\{Y(\mathbf{x}_0)\}$ e variância de *Kriging* σ_{DK}^2 . Conhecendo-se esta lei é simples obter a esperança condicional da variável original

$$Z(\mathbf{x}_0) = \phi[Y(\mathbf{x}_0)], \text{ isto é,}$$

$$E_n \{Z(x_0)\} = E \{ \phi[Y(x_0)] | \phi[Y(x_1)], \dots, \phi[Y(x_n)] \}.$$

Note que, pelo fato da transformada ϕ ser geralmente não-linear, a esperança condicional $E_n \{Z(x_0)\}$ é não-linear com respeito aos n dados originais $Z(x_1), \dots, Z(x_n)$.

Esta transformação é aplicada a $Z(x)$ por três razões básicas:

- (i) é conveniente trabalhar com a distribuição Normal devido à facilidade de definir relações tal como a esperança condicional;
- ii) como esta transformação sempre existe, o método de predição é generalizado no sentido de que apenas uma formulação seja necessária para todas as possíveis distribuições da variável aleatória $Z(x)$;
- (iii) elimina-se a necessidade de que a distribuição de $Z(x)$ seja especificada.

Este anamorfismo gaussiano ϕ , na prática, deve ser conhecido, o qual requer a estacionaridade do processo estocástico $\{Z(x)\}$, e o conhecimento da esperança estacionária $E\{Z(x)\} = m$. Então, o preditor de *Kriging Linear* Y_{SK}^* pode ser construído, onde ainda se requer o conhecimento da função de covariância estacionária $\langle Y(x_\alpha), Y(x_\beta) \rangle$ do processo estocástico transformado, $\{Y(x)\}$. A inferência desta covariância é geralmente feita baseada nos dados transformados. Contudo o preditor de *Kriging Linear*, Y_{SK}^* , pode ser assumido idêntico à esperança condicional $E_n \{Y(x_0)\}$, a qual requer a suposição de que todas as distribuições multivariadas do processo estocástico transformado sejam gaussianas.

Sob a condição de que a distribuição uni-variada de $\{Z(x)\}$ seja conhecida, ou a transformação ϕ bem determinada, a prática tem mostrado que o preditor não-linear Z_{DK}^* é geralmente melhor que Z_{SK}^* ou Z_{OK}^* obtidos por métodos de *Kriging Linear* aplicados diretamente aos dados Z_α , $\alpha=1, \dots, n$, isto é, pela projeção de $Z(x_0)$ em ξ_{n+1} ou na variedade linear ζ .

4.1 Modelo Hermitiano

Todo interesse em polinômios de Hermite vem destes constituírem uma base ortonormal de $L^2(\mathcal{R}, G)$.

Polinômios de Hermite são funções ortogonais em relação à função de ponderação $\exp[-y^2/2]$ no intervalo $(-\infty, +\infty)$. Muitas funções podem ser representadas por uma série de polinômios de Hermite da forma $\phi(y) = \sum_{k=1}^{\infty} \phi_k H_k(y)$, onde os ϕ_k 's podem ser determinados usando as propriedades de ortogonalidade, e $H_k(y)$ é um polinômio de Hermite de ordem k que é definido como:

$$H_k(y) = (-1)^k \exp[-y^2/2] \frac{\partial^k}{\partial y^k} (\exp[-y^2/2]) .$$

$H_k(y)$ pode ser avaliado pela relação de recursividade:

$$H_{k+1}(y) = yH_k(y) - kH_{k-1}(y),$$

onde $H_0(y) = 1$ e $H_1(y) = y$.

Uma condição suficiente para que a função $\phi(y)$ tenha uma representação de Hermite é que esta convirja em média quadrática, isto é:

$$\int_{-\infty}^{\infty} (\phi(y))^2 \exp[-y^2/2] dy < \infty .$$

É a condição acima que permite o cálculo dos coeficientes ϕ_k .

A relação de ortogonalidade para polinômios de Hermite (não-normalizados), é dada por:

$$\int_{-\infty}^{\infty} H_k(y) H_{k'}(y) \exp[-y^2/2] dy = 0 \text{ se } k \neq k' ;$$

$$\int_{-\infty}^{\infty} H_k(y) H_{k'}(y) \exp[-y^2/2] dy = k! \sqrt{2\pi} \text{ se } k = k' .$$

Usando a relação de ortogonalidade:

$$\varphi_k = \left\{ \int_{-\infty}^{\infty} \phi(y) H_k(y) \exp[-y^2/2] dy \right\} / (k! \sqrt{2\pi}) .$$

Posto que para *Kriging Disjuntivo* $\phi(y)$ é geralmente desconhecido, não se pode, em geral, avaliar esta integral analiticamente. A integração de Hermite propicia um método alternativo de técnica numérica que usa apenas poucos pontos y_i 's:

$$\varphi_k = \left\{ \sum_{i=1}^J \omega_i \phi(y_i) H_k(y_i) \exp[-y_i^2/2] \right\} / (k! \sqrt{2\pi}) ;$$

onde y_i e ω_i podem ser obtidos através de consultas a Abramowitz e Stegun [1965].

Caso Gaussiano

O caso gaussiano proporciona um simples exemplo de modelo isofatorial, onde os fatores são polinômios de Hermite normalizados

$$\eta_k = H_k / \sqrt{k!}, \text{ com } H_k(y) = (-1)^k \exp[-y^2/2] \frac{\partial^k}{\partial y^k} (\exp[-y^2/2]).$$

Sabe-se que estes polinômios formam uma base ortogonal de $L^2(\mathcal{R}, G)$, onde G é a lei Normal com esperança zero e variância unitária. Seja $G_\rho(x, y)$ a lei Normal bi-dimensional com esperança zero, variância unitária e coeficiente de correlação ρ . Sua correspondente função característica é:

$$\psi_\rho(u, v) = \exp \left\{ -\frac{(u^2 + v^2 + 2\rho uv)}{2} \right\} = \int_0^\infty \frac{\rho^k (iu)^k (iv)^k}{k!} \exp \left\{ -\frac{(u^2 + v^2)}{2} \right\} .$$

Se $|\rho| < 1$, segue do teorema da convergência dominada que a transformada de Fourier pode ser obtida termo a termo, tal que a densidade gaussiana g_ρ admita a expansão:

$$g_{\rho}(x,y) = \frac{1}{2\pi} \sum_{k=0}^{\infty} \frac{\rho^k}{k!} \frac{\partial^k}{\partial y^k} (\exp[-y^2/2]) \frac{\partial^k}{\partial x^k} (\exp[-x^2/2]).$$

Utilizando-se $H_k(y)$, obtém-se a seguinte representação fatorial da lei Normal G_{ρ} :

$$(G): G_{\rho}(x,y) = \sum_{k=0}^{\infty} \rho^k \eta_k(y) \eta_k(x) G(y) G(x).$$

Posto que $E[H_k(X)|Y] = \rho^k H_k(Y)$, segue que qualquer modelo gaussiano é isofatorial.

Substituindo-se em (G) ρ^k por T_k , generaliza-se o modelo gaussiano e obtém-se a lei bi-dimensional da forma:

$$F(x,y) = \phi(x,y) G(x) G(y); \text{ onde } \phi(x,y) = \sum_{k=0}^{\infty} T_k \eta_k(x) \eta_k(y).$$

Uma condição necessária e suficiente que T_k deve satisfazer, pois ϕ deve ser positiva, é $T_k = E(P^k)$; onde P é uma variável aleatória, tal que $-1 \leq P \leq 1$ q.c., ou seja, T_k 's são os momentos de uma lei de probabilidade concentrada em $[-1,1]$.

Posto que o caso $T_k = \rho^k$ corresponde à lei Normal G_{ρ} , a lei F surge como uma mistura de leis Normais com vários coeficientes de correlação. Os fatores associados a F são os polinômios de Hermite η_k , e dir-se-á que a lei F é Hermitiana.

Isto sugere o seguinte modelo isofatorial chamado modelo Hermitiano. As variáveis aleatórias Y_i 's são Normais $(0,1)$, isto é, $F_1 = G$, e as leis bidimensionais são da forma:

$$F_{ij}(x,y) = \sum_{k=0}^{\infty} T_k(i,j) \eta_k(y) \eta_k(x) G(y) G(x),$$

com coeficientes $T_k(i,j) = E[\eta_k(Y_i) \eta_k(Y_j)]$. Em particular, se $T_k(i,j) = \rho_{ij}^k$, as leis F_{ij} são Normais bivariadas. Caso contrário tem-se $E[\eta_k(Y_i) | Y_j] = T_k(i,j) \eta_k(Y_j)$, a relação (CK) ocorre.

4.2 Sistema de DK Hermitiano

Baseado nas seguintes propriedades:

(i) $\eta_k(u)$ são polinômios ortonormais, isto é:

$$\langle \eta_k(u), \eta_{k'}(u) \rangle = E\{\eta_k(u)\eta_{k'}(u)\} = \begin{cases} 1 & \text{se } k=k' \\ 0 & \text{c.c.} \end{cases};$$

(ii) $\langle \eta_k(Y_1), \eta_k(Y_j) \rangle = E\{\eta_k(Y_1)\eta_k(Y_j)\} = \rho_{1j}^k$;

(iii) qualquer função mensurável $f(u)$ com variância finita, tal que:

$$\int_{-\infty}^{\infty} f^2(u)G(du) < \infty,$$

pode ser expandida como uma série polinomial de Hermite:

$$f(u) = \sum_{k=0}^{\infty} \varphi_k \eta_k(u),$$

onde

$$\varphi_k = \langle f(u), \eta_k(u) \rangle = \int_{-\infty}^{\infty} f(u)\eta_k(u)G(du), \quad \forall k;$$

(iv) a esperança condicional de qualquer função mensurável $f(u)$ tem a seguinte expansão:

$$E\{f(Y_1) | Y_j\} = \sum_{k=0}^{\infty} \rho_{1j}^k \varphi_k \eta_k(Y_j),$$

e sob a hipótese de normalidade univariada e bivariada as esperanças condicionais que aparecem no sistema de (DK) podem ser calculadas.

A expansão da função ϕ (anamorfismo gaussiano) em polinômios de Hermite, truncada no termo de ordem K , é dada por:

$$\phi(u) = \sum_{k=0}^K \varphi_k \eta_k(u).$$

Analogamente, considera-se a expansão de Hermite de ordem K relativo ao preditor de *Kriging Disjuntivo*:

$$Z_{DK}^* = \sum_{\alpha=1}^n g_{\alpha}(Z_{\alpha}) = \sum_{\alpha=1}^n g_{\alpha}(\phi(Y_{\alpha})) = \sum_{\alpha=1}^n g_{\alpha} \sum_{k=1}^K \varphi_k \eta_k(Y_{\alpha})$$

$$Z_{DK}^* = \sum_{\alpha=1}^n \sum_{k=1}^K g_{\alpha} \varphi_k \eta_k(Y_{\alpha}) = \sum_{\alpha=1}^n \sum_{k=1}^K f_{\alpha,k} \eta_k(Y_{\alpha});$$

onde $f_{\alpha,k}$ é constante dependente apenas de α e k .

As duas condições impostas para a obtenção do melhor preditor são não-viciabilidade e mínima variância do erro de predição, ou seja,

$$E[Z(x_0) - Z_{DK}^*] = 0$$

e

$$(D): \sigma_{DK}^2 = E[Z(x_0) - Z_{DK}^*]^2 \text{ mínima.}$$

A condição (D) ocorre quando $[Z(x_0) - Z_{DK}^*]$ é perpendicular a qualquer função g_{α} no hiperplano definido pelas funções mensuráveis $g_{\alpha}(Z_{\alpha})$. Portanto os vetores $E[Z(x_0) - Z_{DK}^*]$ e $g_{\alpha}(Z_{\alpha})$ são ortogonais. Então pode-se escrever:

$$E\{[Z(x_0) - Z_{DK}^*] g_{\alpha}(Z_{\alpha})\} = 0$$

ou

$$(F): E\{[Z(x_0) g_{\alpha}(Z_{\alpha})]\} = E\{[Z_{DK}^* g_{\alpha}(Z_{\alpha})]\}.$$

Em termos de esperança condicional pode-se escrever (F) da seguinte forma:

$$E\{[Z(x_0) | (Z_{\beta})]\} = E\{[Z_{DK}^* | (Z_{\beta})]\}, \quad \forall \beta = 1, \dots, n.$$

Posto que nenhuma afirmação fora feita sobre a distribuição de

$Z(x)$, se faz necessária a transformação ϕ , ou seja, a aplicação de um anamorfismo gaussiano. Portanto escreve-se esta função em termo de polinomiais de Hermite:

$$Z(x) = \phi[Y(x)] = \sum_{k=0}^{\infty} \varphi_k \eta_k[Y(x)]$$

onde os φ_k 's são determinados utilizando-se a relação de ortogonalidade e integrais de Hermite.

$$\text{Tem-se: } E\{\phi[Y(x_0)] | Y(x_\beta)\} = \sum_{\alpha=1}^n E\{g_\alpha(\phi[Y(x_\alpha)]) | Y(x_\beta)\}, \quad \forall \beta=1, \dots, n.$$

As incógnitas são determinadas utilizando-se o fato: "Se (X, Y) é Normal bivariado com média zero e variância unitária, então a densidade conjunta é dada por:

$$f(x, y) = \left[\sum_{k=0}^{\infty} (\rho_{xy})^k H_k(y) H_k(x) g(y) g(x) \right] / (k!),$$

onde g é a densidade da Normal padronizada".

Desta forma, tem-se:

$$E\{\phi(X) | Y\} = \sum_{k=0}^{\infty} (\rho_{xy})^k \varphi_k H_k(Y),$$

$$\sum_{k=0}^K \varphi_k H_k[Y(x_\beta)] \left\{ (\rho_{0\beta})^k - \sum_{\alpha=1}^n f_{\alpha,k} (\rho_{\alpha\beta})^k / \varphi_k \right\} = 0, \quad \forall \beta=1, \dots, n.$$

Definindo $b_{\alpha,k} = f_{\alpha,k} / \varphi_k$, então $Z_{DK}^* = \sum_{k=1}^K \varphi_k H_k^*(Y(x_0))$, onde $H_k^*(Y(x_0))$ é o valor predito de $H_k(Y(x_0))$, que pode ser escrito como:

$$H_k^*(Y(x_0)) = \sum_{\alpha=1}^n b_{\alpha,k} H_k(Y(x_\alpha)),$$

e os $b_{\alpha,k}$'s são determinados resolvendo-se o seguinte sistema de *Kriging Linear*:

$$(\rho_{0\beta})^k = \sum_{\alpha=1}^n b_{\alpha,k} (\rho_{\alpha\beta})^k, \quad \beta=1, \dots, n.$$

Para $k=0$, o sistema acima reduz-se à condição de não-viciabilidade $\sum_1^n b_{\alpha,0} = 1$.

Variância de Kriging Disjuntivo

Se $\phi(Y)$ for uma função de uma variável aleatória Normal $(0,1)$, então a média e a variância de $Z=\phi(Y)$, onde $\phi(Y)$ é expandida da forma $\sum_0^\infty \phi_k \eta_k(Y)$, podem ser obtidas através dos coeficientes ϕ_k 's como:

$$\mu_Z = E[\phi(Y)] = \phi_0 \quad \text{e} \quad \sigma_Z^2 = \text{Var}[\phi(Y)] = \sum_1^\infty k! \phi_k^2.$$

A variância de *Kriging Disjuntivo* é dada por:

$$\text{Var}[Z(\mathbf{x}_0) - Z_{DK}^*(\mathbf{x}_0)] = E[Z(\mathbf{x}_0)]^2 + E[Z_{DK}^*(\mathbf{x}_0)Z(\mathbf{x}_0)];$$

que utilizando-se do fato relacionado à esperança e variância, anteriormente citado, e aquela decomposição da lei bidimensional, fornece:

$$\sigma_{DK}^2 = \sum_1^K k! \phi_k^2 \left[1 + \sum_1^n b_{\alpha,k} (\rho_{0\alpha})^k \right].$$

4.6 Cálculo Aproximado

Sejam $z(\mathbf{x}_\alpha)$'s os valores conhecidos experimentalmente que são realizações de uma função aleatória $Z(\mathbf{x})$ com função de probabilidade $F(z)=P[Z(\mathbf{x}) \leq z]$, contínua. Deseja-se determinar a função anamorfismo (bijetiva) ϕ , relativo à transformação de $Z(\mathbf{x})$ em uma gaussiana $Y(\mathbf{x})$ cuja relação é $Z(\mathbf{x})=\phi[Y(\mathbf{x})]$.

Obtém-se a identidade:

$$G(y)=P[Y(\mathbf{x}) \leq y]=P[\phi[Y(\mathbf{x})] \leq \phi(y)]=P[Z(\mathbf{x}) \leq \phi(y)]=F[\phi(y)].$$

Transformar a função aleatória original em uma que é suposta ser Normal uni e bivariada (para cada par de locação) não é simples. Experimentalmente um método para aproximar esta função é

primeiramente ordenar os dados em ordem crescente; a distribuição de frequência acumulada $F(z_\alpha)$ e uma aproximação para esta é utilizada para obter um valor transformado $Y(x_\alpha)$. Uma possibilidade para tal cálculo é o uso da relação $P[Z(x_\alpha) \leq z_0] \approx (\alpha - 0.5)/n$, onde α é o número total de z_α 's menores ou iguais a z_0 e n o número de pontos amostrados (malha regular no espaço implica em ponderar por $1/n$).

Associando o valor y_α de uma gaussiana Y , correspondente à mesma frequência acumulada $G(y_\alpha) = F(z_\alpha)$, sendo $y_\alpha = G^{-1}[F(z_\alpha)]$, no caso da relação anteriormente citada, o $Y(x_\alpha)$ associado é calculado invertendo a função de probabilidade

$$y_\alpha = P^{-1}[(\alpha - 0.5)/n].$$

Torna-se conhecida, assim, a função anamorfismo para cada um destes pontos (y_α, z_α) . Os valores intermediários para a relação $z_\alpha = \phi(y_\alpha)$ são calculados por interpolação linear ou algum outro método equivalente, ajustando um polinômio de ordem n aos pares de dados (y_α, z_α) . Isto se faz necessário pois o valor da abscissa na integração de Hermite necessariamente não corresponde a um z_α contido no conjunto de dados.

Na prática é necessário modelar tal função ϕ que será determinada pelos coeficientes η_k 's do desenvolvimento em polinômios de Hermite truncados na ordem K .

$$\phi(y) = \sum_{k=0}^K \frac{\eta_k}{k!} H_k(y).$$

Os coeficientes são teoricamente definidos por:

$$\eta_k = E[\phi(Y)H_k(Y)],$$

e os H_k 's podem ser obtidos pela recursividade:

$$H_{k+1}(y) = yH_k(y) - kH_{k-1}(y), \text{ onde } H_0(y) = 1 \text{ e } H_1(y) = y.$$

Indica-se uma possibilidade para a determinação de η_k por interpolação entre os pontos experimentais (y_α, z_α) , η_k é obtido por integração de $\phi(y)H_k(y)g(y)$.

Para verificar se valores apropriados de η_k 's foram obtidos, calcula-se a média e a variância do processo. Determina-se assim o número K do truncamento "plotando" a função $\phi(Y(x))$ "versus" os dados originais para ajudar na determinação do número de η_k 's necessários.

CAPÍTULO 5

PREDIÇÃO DA PROBABILIDADE CONDICIONAL

Pode-se desejar, por questões de gerenciamento, prever a proporção de valores de uma função aleatória que sejam maiores que certo limite, baseado nos valores avaliados em determinado suporte.

Predizer uma probabilidade condicional em dada locação é possível, posto que o preditor de *Kriging Disjuntivo* é não-linear. O método em si consiste de duas etapas: a primeira é encontrar a probabilidade condicional de que o valor de $Z(x)$, em um ponto aleatoriamente localizado, x_0 , em um bloco V , seja em conteúdo, superior a um dado ponto-de-corte z_c . A seguinte é obter a probabilidade condicional de que o valor do bloco esteja acima de dado ponto de corte. A função de probabilidade é então integrada em todo bloco.

Em termos de predição pontual o problema se apresenta da seguinte forma: é conhecida a realização de um processo estocástico $\{Z(x): x \in D\}$ em certo número de pontos x_α , $\alpha=1, \dots, n$. Procura-se prever a quantidade de valores $f[Z(x_0)]$ a partir das informações disponíveis.

Mais frequentemente $f[Z(x_0)]$ representa a tonelagem de mineral $T[Z(x_0)] = \mathcal{F}[Z(x_0) \geq z_c]$, ou uma quantidade de metal $Q[Z(x_0)] = Z(x_0) \mathcal{F}[Z(x_0) \geq z_c]$, que são chamadas "funções de recuperação pontual".

5.1 Funções Indicatrizes

Em termos da predição pontual,

$$P[Z(x_0) \geq z_c | Z(x_1), \dots, Z(x_n)] = P[\phi(Y(x_0)) \geq \phi(y_c) | Z(x_1), \dots, Z(x_n)];$$

onde $Z=\phi[Y]$ e y_c é a transformada associada ao valor-de-corte. ϕ é o anamorfismo de uma função multigaussiana Y .

A probabilidade média sobre o bloco é dada por:

$$P[\bar{Z}(x_0) \geq z_c | Z(x_1), \dots, Z(x_n)] = \frac{1}{V} \int_V P[Z(x_0) \geq z_c | Z(x_1), \dots, Z(x_n)] dx.$$

De fato, para se determinar a probabilidade condicional usando *Kriging Disjuntivo* o problema deve ser remodelado tal que o preditor de DK, que prediz a esperança condicional, também possa predizer a probabilidade condicional. Assim define-se uma função indicatriz $f_{y_c}(Y)$, baseada no ponto-de-corte transformado y_c como:

$$f_{y_c}(Y) = \begin{cases} 1 & \text{se } Y \geq y_c \\ 0 & \text{c.c.} \end{cases}$$

então,

$$E[f_{y_c}(Y)=1] = P[(Y(x) \geq y_c)]$$

o que permite escrever:

$$E[f_{y_c}[Y(x_0)]=1 | Y(x_1), \dots, Y(x_n)] = P[[Y(x_0) \geq y_c] | Y(x_1), \dots, Y(x_n)],$$

posto que

$$f_{y_c}(Y) = \begin{cases} 1 & \text{se } y_c \leq Y < \infty \\ 0 & \text{se } -\infty < Y < y_c \end{cases}.$$

Portanto a probabilidade condicional de que $Y(x)$ seja maior que y_c é dada como a esperança condicional da função indicatriz $f_{y_c}(Y)$.

Citando em particular as funções de recuperação:

$$f[Y(x_0)] = f[Y(x_0) \geq y_c] = f_{y_c}[Y(x_0)];$$

$$f[Y(x_0)] = Y(x_0) \mathcal{F}[Y(x_0) \geq y_c] = Y(x_0) f_{y_c}[Y(x_0)].$$

Frequentemente, o método de predição de $\mathcal{F}[Y(x_0) \geq y_c]$, é denotado por $f^*[Y(x_0)]$ e é deduzido como uma combinação linear dos $\{\mathcal{F}[Y(x_0) \geq y_c]\}^*$.

Constrói-se uma família de pontos-de-corte y_i , $i=1, \dots, N$; tal que:

$$m_i = E\{\mathcal{F}[Y(x) \geq y_i]\} = P[Y(x) \geq y_i].$$

O preditor de *Kriging Disjuntivo* (*Co-Kriging* das indicatrizes) à média conhecida de $\mathcal{F}[Y(x_0) \geq y_i]$ é da forma:

$$\{\mathcal{F}[Y(x_0) \geq y_c]\}^* = m_i + \sum_{j=1}^N \sum_{\alpha=1}^n \lambda_{j\alpha} [\mathcal{F}(Y_{\alpha} \geq y_j) - m_j];$$

que conduz ao sistema de *Kriging Disjuntivo*:

$$(C-K): \sum_{j=1}^N \sum_{\alpha=1}^n \lambda_{j\alpha} K_{jk}^{\alpha\beta} = K_{ik}^{\alpha 0}, \quad j=1, \dots, N \text{ e } \alpha=1, \dots, n;$$

$$\text{com, } K_{jk}^{\alpha\beta} = \text{Cov}[\mathcal{F}(Y_{\alpha} \geq y_j), \mathcal{F}(Y_{\beta} \geq y_k)].$$

O que se sabe é que este método é dispendioso, pois deve-se tornar conhecidas as N funções de covariância das indicatrizes. Com o objetivo de evitar o conhecimento de tais funções de covariância A. Journel [1982], propôs o *Kriging* de cada indicatriz independentemente uma da outra. Porém não cabe neste trabalho sua discussão e avaliação.

Sabe-se que teoricamente a melhor predição possível de $\mathcal{F}[Y(x_0) \geq y_c]$ a partir dos n dados avaliados, Y_{α} 's, é a esperança condicional $E[\mathcal{F}[Y(x_0) \geq y_c] | Y_1, \dots, Y_n]$, isto é, a projeção de $Y(x_0)$ sobre o espaço das funções mensuráveis da forma $f(Y_1, \dots, Y_n)$. Já é notório também que para se obter esta esperança condicional faz-se necessário recorrer a um modelo. Considera-se que Z seja um anamorfismo de uma função multigaussiana Y , e ainda que existe uma função bijetiva ϕ tal que $Y(x) = \phi^{-1}[Z(x)]$, de forma que

$$E[\mathcal{I}[Y(\mathbf{x}_0) \geq y_c | Y_1, \dots, Y_n] = E[\mathcal{I}[\phi^{-1}(Z(\mathbf{x}_0)) \geq \phi^{-1}(z_c) | Z_1, \dots, Z_n].$$

Sob este ponto de vista, deduz-se que a média e a variância são respectivamente as de *Kriging Linear* para a lei gaussiana e obtém-se:

$$\begin{aligned} E[\mathcal{I}[Y(\mathbf{x}_0) \geq y_c | Y_\alpha = \phi^{-1}(Z_\alpha); \alpha=1, \dots, n] = \\ = \int_{-\infty}^{\infty} [\phi^{-1}(Z(\mathbf{x}_0)) \geq y_c] \left[\frac{1}{(2\pi\sigma_K^2)} \right]^{1/2} \exp\left\{ -\frac{1}{2} \left(\frac{z - z_K}{\sigma_K} \right)^2 \right\} dz. \end{aligned}$$

Em particular:

$$\{\mathcal{I}[Y(\mathbf{x}_0) \geq y_c]\}^* = 1 - G\left\{ \frac{\phi^{-1}(z_c) - y_K}{\sigma_K} \right\};$$

$$\{Y(\mathbf{x}_0)\mathcal{I}[Y(\mathbf{x}_0) \geq y_c]\}^* = \int_{\phi^{-1}(z_c)}^{\infty} \phi[y] \left[\frac{1}{(2\pi\sigma_K^2)} \right]^{1/2} \exp\left\{ -\frac{1}{2} \left(\frac{y - y_K}{\sigma_K} \right)^2 \right\} dy.$$

5.2 Kriging Disjuntivo Bigaussiano

A função esperança condicional,

$$E[f_{y_c}[Y(\mathbf{x}_0)] | Y(\mathbf{x}_1), \dots, Y(\mathbf{x}_n)],$$

é por definição a projecção da função $f_{y_c}[Y(\mathbf{x})]$ no espaço das funções mensuráveis das $(n+1)$ variáveis $Y(\mathbf{x}_0), Y(\mathbf{x}_1), \dots, Y(\mathbf{x}_n)$, para $\mathbf{x}_\alpha \in D$.

O preditor de *Kriging Disjuntivo* P_{DK}^* da função $f_{y_c}[Y(\mathbf{x}_0)]$ é então a projecção desta função no subespaço vetorial $\mathcal{D}_n \subset \mathcal{H}_n$ da soma das funções mensuráveis de cada dos n dados Y_α separadamente. Do teorema clássico das três perpendiculares, resulta que P_{DK}^* é a projecção de T em $\mathcal{D}_n$.

Modelo Isofatorial Bigaussiano

Sendo U e V duas variáveis aleatórias com mesma lei Normal padrão g e coeficiente de correlação ρ , a densidade conjunta de (U,V), é dada por:

$$f(u,v) = (2\pi\sqrt{1-\rho^2})^{-1} \exp\left\{-\frac{1}{2} \frac{u^2 - 2\rho uv + v^2}{1-\rho^2}\right\}.$$

Introduzindo polinômios normalizados de Hermite:

$$\eta(u) = (\sqrt{k!})^{-1} \frac{1}{g(u)} \frac{\partial^k}{\partial u^k} g(u),$$

pode-se reescrever a densidade como:

$$f(u,v) = \sum_{k=0}^{\infty} \rho^k \eta_k(u) \eta_k(v) g(u) g(v).$$

Portanto, isto fornece a seguinte fórmula fundamental:

$$(FF): E[\eta_k(U) | V] = \rho^k \eta_k(V).$$

$$E[\eta_k(U) | V=v] = \int_{-\infty}^{\infty} \eta_k(u) f(u|v) dv$$

$$E[\eta_k(U) | V=v] = \sum_{l=0}^{\infty} \rho^l \eta_l(v) \int_{-\infty}^{\infty} \eta_k(u) \eta_l(u) g(u) du$$

$$E[\eta_k(U) | V=v] = \rho^k \eta_k(v),$$

obtido através da ortogonalidade dos η_k 's, isto é,

$$\text{Cov}[\eta_k(U), \eta_k(V)] = \begin{cases} 0 & \text{se } k \neq 0 \\ \rho^k & \text{se } k = 0. \end{cases}$$

Como visto no capítulo anterior, se ϕ é uma função de quadrado integrável, ϕ pode ser representada em termos dos η_k 's, o que permite concluir que:

$$E[\phi(U) | V] = E\left[\sum_0^\infty \varphi_k \eta_k(U) | V\right] = \sum_0^\infty \varphi_k \rho^k \eta_k(V).$$

Mantida a condição de que $\{Y(x)\}$ seja um processo estocástico estacionário, diz-se que $\{Y(x)\}$ é bigaussiano se este verifica as seguintes propriedades:

- i) $Y(x)$ é Normal padrão;
- ii) a densidade conjunta de $(Y(x), Y(y))$ é bigaussiana.

Esta densidade deve ser representada, como:

$$f_{xy}(u, v) = \sum_0^\infty \rho_{xy}^k \eta_k(u) \eta_k(v) g(u) g(v);$$

onde $g(\cdot)$ é a função de densidade de uma Normal (0,1).

O preditor de *Kriging Disjuntivo* de $f_{zc}[Z(x_0)]$, baseado nos Z_α 's é caracterizado pelas relações:

$$E[f_{zc}^*[Z(x_0)] | Z_\beta] = E[f_{zc}[Z(x_0)] | Z_\beta], \quad \forall \beta = 1, \dots, n.$$

Suponha ϕ um anamorfismo gaussiano de um processo estocástico bigaussiano $\{Y(x)\}$, $Z(x) = \phi[Y(x)]$. Então pode-se exprimir o sistema de *Kriging Disjuntivo*, como:

$$E[f_{zc}^*[Z(x_0)] | Y_\beta] = E[f_{zc} \phi[Y(x_0)] | Y_\beta], \quad \forall \beta = 1, \dots, n.$$

Através da expansão de ϕ em uma série polinomial de Hermite obtém-se:

$$E[f_{zc} \phi[Y(x_0)] | Y_\beta] = \sum_0^\infty \varphi_k E[\eta_k[Y(x_0)] | Y_\beta]$$

$$E[f_{zc}^*[Z(x_0)] | Y_\beta] = \sum_0^\infty \varphi_k E[\eta_k^*[Y(x_0)] | Y_\beta], \quad \forall \beta = 1, \dots, n;$$

que pela unicidade do preditor de DK, fornece:

$$f_{yc}^*[Y(x_0)] = \sum_0^{\infty} \varphi_k \left[\eta_k^*[Y(x_0)] \right];$$

onde os $\eta_k^*[Y(x_0)]$'s são os preditores de DK dos $\eta_k[Y(x_0)]$'s, obtidos através dos Y_α 's.

Voltando ao problema original, expandindo $f_{yc}(Y)$ em termos de uma série polinomial de Hermite truncada no termo de ordem K tem-se,

$$f_{yc}^*[Y(x_0)] = \sum_0^K \varphi_k \eta_k^*[Y(x_0)],$$

que é o preditor da probabilidade condicional, P_{DK}^* .

Aplicando a ortogonalidade:

$$\varphi_k = \langle f_{yc}[Y(x)], \eta_k[Y(x)] \rangle = \int_{-\infty}^{\infty} f_{yc}(u) \eta_k(u) g(u) du = \int_{yc}^{\infty} \eta_k(u) g(u) du.$$

Usando a definição de η_k e a de $g(u)$, os parâmetros (ou coeficientes da expansão) φ_k 's, podem ser expressos por:

$$\varphi_0 = (2\pi)^{-1/2} \int_{yc}^{\infty} e^{-u^2/2} du = 1 - G(y_c),$$

onde $G(\cdot)$ é a correspondente função de distribuição de $g(\cdot)$;

$$\varphi_k = (2\pi)^{-1/2} \int_{yc}^{\infty} \eta_k(u) g(u) du = [g(y_c) H_{k-1}(y_c)] / \sqrt{k!}$$

$$\varphi_k = \sqrt{\frac{(k-1)!}{k!}} [g(y_c) \eta_{k-1}(y_c)];$$

então,

$$f_{yc}^*[Y(x_0)] = \sum_0^K \varphi_k \eta_k^*[Y(x_0)] = 1 - G(y_c) + g(y_c) \sum_{k=1}^K \frac{1}{\sqrt{k!}} H_{k-1}(y_c) \eta_k^*[Y(x_0)],$$

onde a única incógnita é $\eta_k^*[Y(x_0)]$.

Fixando-se k , pode-se avaliar $\eta_k^*[Y(x_0)]$. Este preditor é caracterizado por:

$$E[\eta_k^*[Y(x_0)]|Y_\beta] = E[\eta_k[Y(x_0)]|Y_\beta], \quad \forall \beta=1, \dots, n.$$

Se $k=0 \Rightarrow \eta_k[Y(x_0)]=1$, então $\eta_k^*[Y(x_0)]=1$. Suponha, então $k \neq 0$.

De (FF) o segundo membro da equação acima é $\rho_{\alpha 0}^k \eta_k(Y_\alpha)$. Quanto ao primeiro membro sabe-se que $\eta_k^*[Y(x_0)] = \sum_{l=1}^n \lambda_{kl}^\beta f_{k\beta}(Y_\beta)$. Note que os λ_{kl}^β são os coeficientes de $f_{k\beta}$. Então reduz-se o sistema de DK a:

$$\sum_{l=1}^n \left(\sum_{k=1}^n \lambda_{kl}^\beta \rho_{\alpha\beta}^1 \right) \eta_l(Y_\alpha) = \rho_{\alpha 0}^k \eta_k(Y_\alpha); \quad \forall \alpha=1, \dots, n.$$

Um estudo polinomial fornece:

$$\sum_{l=1}^n \lambda_{kl}^\beta \rho_{\alpha\beta}^1 = \begin{cases} 0 & \text{se } l \neq k \\ \rho_{\alpha 0}^k & \text{se } l = k; \end{cases} \quad \alpha=1, \dots, n.$$

Considere $l \neq k$. Se $l \neq 0$, a matriz do termo genérico, $\rho_{\alpha\beta}^1 = \text{Cov}[\eta_l(Y_\alpha), \eta_l(Y_\beta)]_{1 \leq \alpha, \beta \leq n}$ é regular. Então $\lambda_{kl}^\beta = 0$ para $\beta=1, \dots, n$.

Se $l=0$, $\sum_{l=1}^n \lambda_{kl}^\beta = 0$. Porém estes termos não entram no cálculo de $\eta_k^*[Y(x_0)]$, posto que $\eta_0(Y_\beta)=1$, $\forall \beta=1, \dots, n$.

Finalmente, os termos a se considerar são os λ_{kk}^β . O preditor de $\eta_k[Y(x_0)]$ é da forma:

$$\eta_k^*[Y(x_0)] = \sum_{l=1}^n \lambda_{kk}^\beta \eta_k(Y_\beta)$$

[reconheça que esta última equação é a de *Kriging Simples* de $\eta_k[Y(x_0)]$, baseado nos $\eta_k(Y_\alpha)$]. Os λ_{kk}^β 's são obtidos através da resolução do sistema:

$$\sum_{l=1}^n \lambda_{kk}^\beta \rho_{\alpha\beta}^1 = \rho_{\alpha 0}^k; \quad \alpha=1, \dots, n.$$

5.3 Função de recuperação

Usando o preditor $H_k^*[Y(x_0)]$, para $H_k[Y(x_0)]$, obtém-se o preditor pontual para a probabilidade condicional:

Quantidade de Mineral: $\{T[Z(x_0)] = \mathcal{I}[Z(x_0) \geq z_c]\}$

$$P_{DK}^* = 1 - G(y_c) + g(y_c) \sum_{k=1}^K \frac{1}{\sqrt{k!}} H_{k-1}(y_c) \eta_k^*[Y(x_0)].$$

Pode-se então escrever o preditor da probabilidade condicional da forma alternativa:

$$P_{DK}^*[Y(x_0)] = f_{yc}^*[Y(x_0)] = \int_{y_c}^{\infty} f_{DK}(u) du, \text{ onde}$$

$$f_{DK}(u) = g(u) \left[1 + \sum_{k=1}^K \frac{1}{k!} H_{k-1}(u) \eta_k^*[Y(x_0)] \right].$$

Afim de se obter um preditor da probabilidade condicional para o grau médio de um determinado bloco, $\bar{P}$, procede-se da seguinte forma:

$$\bar{P}[Y(x_0) \geq y_c | Y(x_1), \dots, Y(x_n)] = \frac{1}{V} \int_V P_{DK}^*[Y(x_0)] dx.$$

Dai obtém-se o preditor para a probabilidade condicional do grau médio do bloco, $\bar{P}_{DK}^*[Y(x_0)]$,

$$\bar{P}_{DK}^*[Y(x_0)] = \frac{1}{V} \int_V \left\{ 1 - G(y_c) + g(y_c) \sum_{k=1}^K \frac{1}{\sqrt{k!}} H_{k-1}(y_c) \eta_k^*[Y(x_0)] \right\} dx.$$

Posto que apenas os $b_{\alpha k}$'s dependem da locação (x), pode-se reescrever a equação acima como:

$$\bar{P}_{DK}^*[Y(x_0)] = 1 - G(y_c) + g(y_c) \sum_{k=1}^K \frac{1}{\sqrt{k!}} H_{k-1}(y_c) \sum_{\alpha=1}^n b_{\alpha k} \eta_k[Y(x_{\alpha})].$$

Quantidade de Metal: $\{Q[Z(x_0)] = Z(x_0) \mathcal{I}[Z(x_0) \geq z_c]\}$

Escreve-se $Q[Z(x_0)] = \phi[Y(x_0)] \mathcal{I}[\phi(Y(x_0)) \geq \phi y_c]$, onde $y_c = \phi^{-1}(z_c)$, donde se obtém os coeficientes φ_k 's por:

$$\varphi_k = \int_{-\infty}^{\infty} \phi(y) \mathcal{I}(y \geq y_c) \eta_k(y) g(y) dy,$$

que é equivalente a:

$$\varphi_k = \int_{y_c}^{\infty} \phi(y) \eta_k(y) g(y) dy.$$

Desenvolvendo-se ϕ em expansões polinomiais de Hermite, da forma:

$$\phi = \sum_{l=0}^{\infty} \varphi_l \eta_l;$$

obtém-se, então,

$$\varphi_k = \sum_{l=0}^{\infty} \varphi_l \int_{y_c}^{\infty} \eta_l(y) \eta_k(y) g(y) dy.$$

Os coeficientes $\int_{y_c}^{\infty} \eta_l(y) \eta_k(y) g(y) dy$, que serão denotados por U_{lk} , são deduzidos facilmente uns dos outros pelas relações:

$$U_{kl} = U_{lk};$$

$$U_{00} = 1 - G(y_c);$$

$$U_{k0} = -\eta_{k-1}(y_c) g(y_c) \quad (k > 0);$$

$$U_{kl} = -\eta_l(y_c) \eta_{k-1}(y_c) g(y_c) + l U_{k-1, l-1} \quad (k, l > 0).$$

5.4 Predição de Recuperação de Blocos (KD)

O sistema de *Kriging* de $Y(v)$ por Y_α 's é:

$$\sum_{\alpha=1}^n \lambda_\alpha \text{Cov}(Y_\alpha, Y_\beta) = \text{Cov}(Y(v), Y_\beta), \quad \forall \beta = 1, \dots, n;$$

isto é,

$$\lambda_{\alpha} + \sum_{\beta=1}^n \lambda_{\beta} \rho_{\alpha\beta}^2 = \rho_{\alpha\alpha}, \quad \forall \alpha=1, \dots, n,$$

onde $\text{Cov}(Y_{\alpha}, Y_{\alpha})=1$.

O preditor de uma função de $Y(v)$, diga-se $\psi[Y(v)]$, é

$$E[\psi(Y(v)) | Y_{\alpha's}] = \int_{-\infty}^{\infty} \psi(y) (2\pi\sigma_K^2)^{-1/2} \exp\left\{-\frac{1}{2} \left(\frac{y-Y_v^*}{\sigma_K}\right)^2\right\} dy.$$

Em particular, para as funções de recuperação T e Q de um bloco, segue a regra:

$$E[T | Y_{\alpha's}] = 1 - G\left[\frac{y_c - Y_v^*}{\sigma_K}\right];$$

$$E[Q | Y_{\alpha's}] = \int_{y_c}^{\infty} \phi_{\rho}(y) (2\pi\sigma_K^2)^{-1/2} \exp\left\{-\frac{1}{2} \left(\frac{y-Y_v^*}{\sigma_K}\right)^2\right\} dy.$$

Desenvolvendo-se Q em uma polinomial de Hermite truncada:

$$Q = \sum_{k=0}^K \phi_k \eta_k(Y(v)).$$

Seja $s^2 = 1 - \sigma_K^2 = \text{Var}(Y^*)$. Assim como em *Kriging* à média conhecida:

$$s = \text{Cov}(Y(v), Y_v^*/s),$$

deduz-se, então:

$$E[\eta_k[Y(v)] | Y_{\alpha's}] = s^k \eta_k(Y_v^*/s),$$

$$E[Q | Y_{\alpha's}] = \sum_{k=0}^K \phi_k s^k \eta_k(Y_v^*/s).$$

Sabe-se que: $\frac{1}{N} \sum_{i=1}^N \psi(Y(v_i)) = \sum_{k=0}^K \psi_k \left[\frac{1}{N} \sum_{i=1}^N \eta_k(Y(v_i)) \right]$.

Deseja-se prever $\frac{1}{N} \sum_{i=1}^N \eta_k(Y(v_i))$. O preditor é da forma:

$$\sum_{\alpha=1}^n \lambda_{\alpha,k} \eta_k(Y_{\alpha}),$$

com o sistema

$$\sum_{\alpha=1}^n \lambda_{\alpha,k} [\text{Cov}(Y_{\alpha}, Y_{\beta})]^k = \frac{1}{N} \sum_{i=1}^N [\text{Cov}(Y_{\alpha}, Y(v_i))]^k; \quad \forall \alpha=1, \dots, n,$$

ou

$$\lambda_{\alpha,k} + \sum_{\beta=1}^n \lambda_{\beta,k} [\rho^2_{\alpha\beta}]^k = \frac{1}{N} \sum_{i=1}^N [\rho_{\alpha, v_i}]^k; \quad \forall \alpha=1, \dots, n.$$

Donde se deduz diretamente a predição de $\frac{1}{N} \sum_{i=1}^N \psi(Y(v_i))$, é então, em particular a função de recuperação.

Para N blocos v_i 's de um painel, tem-se a forma geral:

$$E \left[\frac{1}{N} \sum_{i=1}^N \psi(Y(v_i)) \mid Y_{\alpha} \right] = \frac{1}{N} \sum_{i=1}^N E \left[\psi(Y(v_i)) \mid Y_{\alpha} \right].$$

É necessário o *Kriging* de cada $Y(v_i)$ com as informações Y_{α} 's.

De forma análoga os preditores de P e Q, para todos os blocos do painel são obtidos.

Através de uma expansão em polinômios de Hermite, tem-se:

$$\frac{1}{N} \sum_{i=1}^N \psi(Y(v_i)) = \frac{1}{N} \sum_{i=1}^N \sum_{k=0}^K \phi_k \eta_k(Y(v_i)) = \sum_{k=0}^K \phi_k \left[\frac{1}{N} \sum_{i=1}^N \eta_k(Y(v_i)) \right].$$

É neste caso mais simples prever $\frac{1}{N} \sum_{i=1}^N \eta_k(Y(v_i))$ por:

$$\frac{1}{N} \sum_{i=1}^N s_i^{\rho} \eta_{\rho}(Y(v_i)/s_i),$$

com o objetivo de se obter diretamente o preditor de $\frac{1}{N} \sum_{i=1}^N \psi(Y(v_i))$.

5.5 Predição por Condicionamento Uniforme

Sabe-se que a lei de um bloco $Y(v_1)$ condicionado às informações Y_α 's, não depende do *Kriging* $Y^*(v_1)$ de tal bloco, o que significa que um bloco pode ser condicionado por seu valor de *Kriging*.

Uma solução mais simples para a predição condicional consiste no *Condicionamento Uniforme* (K-UNI) dos $Y(v_1)$'s (lei do bloco) pelo valor de *Kriging* Y_K^* de $\frac{1}{N} \sum_{i=1}^N Y(v_i)$ pelos Y_α 's:

$$Y_K^* = \sum_{\alpha=1}^n \lambda_\alpha Y_\alpha,$$

onde se obtém o sistema:

$$\sum_{\alpha=1}^n \lambda_\alpha \text{Cov}(Y_\alpha, Y_\beta) = \frac{1}{N} \sum_{i=1}^N \text{Cov}(Y(v_i), Y_\beta); \quad \forall \beta=1, \dots, n$$

ou

$$\lambda_\alpha + \sum_{\beta \neq \alpha} \lambda_\beta \text{Cov}(Y_\alpha, Y_\beta) = \frac{1}{N} \sum_{i=1}^N \text{Cov}(Y(v_i), Y_\alpha); \quad \forall \alpha=1, \dots, n.$$

Sendo $s_K^2 = \text{Var}(Y_K)$, o coeficiente de correlação de Y_K/s_K e $Y(v_1)$ é:

$$\rho_1 = \rho \sum_{\alpha=1}^n \lambda_\alpha \text{Cov}(Y_\alpha, Y(v_1)) / s_K.$$

Donde se deduz a lei de $Y(v_1)$ dado Y_K :

$$f(Y(v_1) | Y_K) = \frac{1}{(1-\rho_1^2)^{1/2}} \mathcal{G} \left(\frac{Y(v_1) - \rho_1 Y_K / s_K}{(1-\rho_1^2)^{1/2}} \right).$$

A predição que é de interesse se escreve:

$$\begin{aligned} E \left[\frac{1}{N} \sum_{i=1}^N \psi(Y(v_i)) | Y_K \right] &= \frac{1}{N} \sum_{i=1}^N \left[\psi(Y(v_i)) | Y_K \right] = \\ &= \frac{1}{N} \sum_{i=1}^N \int_{-\infty}^{\infty} (2\pi(1-\rho_1^2))^{-1/2} \psi(y) \exp \left\{ -\frac{1}{2} \left(\frac{Y(v_1) - \rho_1 Y_K / s_K}{(1-\rho_1^2)^{1/2}} \right)^2 \right\} dy. \end{aligned}$$

Em particular para T e Q:

$$E[T|Y_K] = \frac{1}{N} \sum_1^N \left(1 - G \left(\frac{Y(v_1) - \rho_1 Y_K / s_K}{(1 - \rho_1^2)^{1/2}} \right) \right);$$

$$E[Q|Y_K] = \frac{1}{N} \sum_1^N \int_{y_c}^{\infty} \varphi_k(y) (2\pi(1 - \rho_1^2))^{-1/2} \exp \left\{ - \frac{1}{2} \left(\frac{Y(v_1) - \rho_1 Y_K / s_K}{(1 - \rho_1^2)^{1/2}} \right)^2 \right\} dy.$$

Em termos de polinômios de Hermite:

$$E[\eta_k(Y(v_1)) | Y_K / s_K] = \rho_1^k \eta_k(Y_K / s_K);$$

em consequência obtém-se o preditor K-UNI dos coeficientes,

$$[\eta_k^*]_{K-UNI} = E \left[\frac{1}{N} \sum_1^N \eta_k(Y(v_1)) | Y_K / s_K \right] = \frac{1}{N} \sum_1^N \rho_1^k \eta_k(Y_K / s_K),$$

originando o preditor final de $\frac{1}{N} \sum_1^N \psi(Y(v_1)) : \sum_0^K \psi_k \eta_k^*$.

CAPÍTULO 6

PREDIÇÃO DE RESERVAS RECUPERÁVEIS

Um dos principais eixos de pesquisa em *geoestatística* atualmente é o desenvolvimento de métodos para a predição de reservas recuperáveis.

Decisões críticas de gerenciamento sobre a viabilidade de uma proposta operação de mineração, caem inteiramente sobre a predição feita para a reserva global. Tal predição é feita com o pleno conhecimento de que com alguma extensão existe um erro. O analista deve conhecer a extensão deste erro, com o objetivo de tomar decisões corretas.

O conjunto dos recursos "*in situ*" de um jazimento, apresenta em muitos poucos casos, características suficientes para permitir um exploração não seletiva. Para julgar a rentabilidade de um jazimento ou de uma zona, ou para decidir o melhor método de exploração, dever-se-á que estudar diversas possíveis seleções. O interesse está, então concentrado na predição destas reservas selecionadas. Para abordar o estudo das seleções, é necessário fixar uma primeira idéia, que é essencial: "A realidade é desconhecida, e um preditor não se comporta como a realidade".

6.1 Reservas Recuperadas

Quando se busca selecionar um mineral num jazimento J , para recuperar uma parte $j \subset J$, objetiva-se um controle de qualidade do minério recuperado, ou um controle das condições de exploração, mais geralmente, a otimização de uma certa função das características do minério extraído. Tais características dependem, dentre outros fatores, de um certo número de parâmetros de corte. A cada n -upla destes

parâmetros de corte corresponde um certo minério recuperado de características associadas a estes parâmetros. Surge, então, o problema de definir o conjunto ótimo que maximiza uma função associada ao benefício sob certas condições eventuais. Chama-se critério de seleção a esta função das características do minério eventualmente recuperado. Vê-se que este critério variará segundo o método de exploração e as condições econômicas.

Uma vez definido tal critério de seleção, mencionado acima, chama-se parâmetro ótimo de seleção ao conjunto dos parâmetros de corte que otimizam tal função das características do minério. A este parâmetro ótimo de seleção corresponde um certo conjunto j^0 de minério recuperado, de características associadas a ele.

Em geral, há que se estudar vários critérios de seleção e seus impactos sobre a recuperação de recursos. Fixa-se um critério. Antes de definir o parâmetro ótimo de seleção, trata-se de observar como variam as diferentes características inerentes do minério recuperado em função do parâmetro de corte.

Posto que a realidade é inacessível, ou seja não se conhece as verdadeiras características do minério recuperável, os estudos são feitos baseados apenas em preditores destas.

É clara a distinção entre recurso "*in situ*" e reserva recuperável, o recurso "*in situ*" presente em um dado corpo mineral não pode ser extraída completamente, por duas razões básicas: (i) técnica; equipamento, material e problemas organizacionais; (ii) econômica; o custo da extração.

É fato que o recurso "*in situ*" de um depósito não pode, em geral, ser extraída completamente. Restrições técnicas assim como econômicas, fazem com que seja necessário selecionar apenas uma fração desta reserva. Isto define reserva recuperável. A seletividade é fortemente relacionada ao tamanho do suporte, que é o volume mínimo de material que pode ser selecionado.

O termo predição recuperável, adotado em *geoestatística*, se refere a uma situação muito particular:

— o meio técnico da exploração pode ser sumarizado na definição de um volume mínimo para seleção;

— o meio econômico pode ser resumido na definição de uma seleção de corte aplicado à seleção do volume médio do conteúdo.

A predição recuperável, como citada anteriormente, é utilizada com o objetivo de se aproximar exatamente das condições reais de mineração. Identificar o problema de predição recuperável e melhorar as capacidades de solução trabalháveis têm sido o maior incentivo no campo de reservas minerais.

A proposta da predição recuperável é a avaliação da tonelagem e conteúdo médio recuperável por seleção de um dado corte em um dado corpo mineral.

O passo básico de seleção aplica-se à unidade de seleção do volume V , chamada seleção por bloco. O critério de seleção é um corte z_c aplicado a um conteúdo médio predito Z_v^* do bloco v selecionado. Neste caso a seleção é dita indireta. Se acontece que o verdadeiro teor do bloco Z_v seja conhecido, ou se sua predição é considerada ser próxima do valor exato (real), a seleção é dita direta. Se o bloco for selecionado sem qualquer restrição, a seleção é dita livre.

Na avaliação de recursos o corpo mineral é dividido em painéis regulares V . Na predição recuperável o corpo mineral será também dividido em painéis regulares V , e assume o mineral a ser selecionado:

- (i) em unidades de blocos de tamanho v ;
- (ii) sem restrições geométricas;
- (iii) aplicando um critério de corte em um conteúdo predito Z_v^* , em um bloco.

Consequentemente, partes estritas de cada painel V , serão selecionadas como "minério", e uma fração de tonelagem e metal será

recuperável. Para qualquer painel V define-se as duas funções de recuperação:

- (a) $T(z_c)$; proporção de tonelagem de minério no valor de corte z_c , predita no painel V ;
- (b) $Q(z_c)$; quantidade de metal recuperado por unidade de tonelagem em V , a um ponto de corte z_c .

A função recuperável será determinada direta ou indiretamente. Na prática o resultado será apresentado como segue, para um painel V de tonelagem TV :

- (i) tonelagem recuperável predita em $V \Rightarrow T = TV[T(z_c)]$;
- (ii) quantidade de metal recuperável predito em $V \Rightarrow Q = TV[Q(z_c)]$;
- (iii) teor médio predito do minério recuperável em $V \Rightarrow M = Q(z_c)/T(z_c)$.

Note que as duas funções $T(z_c)$ e $Q(z_c)$ para um dado painel dependem:

- (a) do tamanho da unidade, v , de seleção do bloco; se v é reduzido a um ponto, $T(z_c)$ e $Q(z_c)$ serão ditas funções de recuperação pontual;
- (b) da quantidade de informação avaliável quando se faz a predição e incorporados aos preditores de T e Q ;
- (c) da quantidade de informações avaliadas quando da mineração e incorporados em cada preditor de bloco, Z_v^* .

Quando reservas recuperáveis são preditas a partir de um sistemático conhecimento dos dados, estas avaliações correspondem a suportes muito pequenos, que são frequentemente muito pobres ou muito ricos, e podem fornecer a ilusão de fácil separação entre minério e estéril. Posto que o objeto da exploração é um bloco e não um corpo, a distribuição do conteúdo de seleção dos blocos deve ser estimada baseado no conteúdo de tais corpos. Esta estimativa é dita ser global quando se refere ao total do depósito. Um modelo para mudança de suporte na distribuição do conteúdo deve ser usado. Aquela é dita ser local, quando se refere a uma parte do depósito, como os painéis, neste caso o modelo de distribuição de blocos deve ser condicionado pela amostra na vizinhança do painel.

6.2 Relação de Cartier

O modelo para mudança de suporte, que permite deduzir a distribuição do teor do bloco V a partir da distribuição do conteúdo $Z(v)$, de pontos amostrais v , que é conhecido, deve satisfazer a relação de Cartier:

$$E[Z(v)|Z(V)] = Z(V),$$

isto é, dado um bloco V com conteúdo $Z(V)$, o conteúdo de uma amostra v , aleatoriamente localizada em V , deve ser igual em média ao teor de V .

Com isto, tem-se que a diferença entre o modelo Gaussiano e outro qualquer, deve-se à hipótese de bivariedade que seja feita com relação aos pares $(Z(v), Z(V))$. No modelo Gaussiano, estes pares são considerados Normais bivariados, ou após serem submetidos a um anamorfismo gaussiano, esta consideração é feita.

A distribuição do conteúdo da amostra $Z(v)$, é então perfeitamente definida por esta função anamorfismo gaussiano ϕ , que satisfaz a: $Z(v) = \phi(Y(v))$, onde $Y(v)$ é uma variável Normal $(0,1)$, associada à amostra.

De forma análoga, a distribuição dos conteúdos dos blocos $Z(V)$, será completamente determinada por sua função anamorfismo ϕ_v , $\phi_v(Y(V)) = Z(V)$, onde $Y(V)$ é a variável Normal padronizada associada aos blocos.

Como os pares $(Y(v), Y(V))$ são supostos serem Normais bivariados, ϕ_v é então dada pela relação de Cartier:

$$\phi_v(y) = E[\phi(Y(v)) | Y(V)].$$

Se g denota a densidade Normal padrão, então a relação acima pode ser escrita como:

$$\phi_v(y) = \int_{-\infty}^{\infty} \phi(\rho y + (1-\rho^2)^{-1/2} u) g(u) du.$$

A correlação ρ , entre $Y(v)$ e $Y(V)$ é escolhida de forma que a variância de $Z(V)$ satisfaça a:

$$\text{Var}(Z(V)) = \text{Var}(Z(v)) - \bar{\gamma}_v(V, V).$$

6.3 Função de Recuperação Pontual

Seja $\{Z(x): x \in D \subseteq \mathcal{R}^d\}$, um processo estocástico estacionário e com variância finita. Se x pertence a um depósito D , define-se as seguintes funções aleatórias:

(a) indicatriz para o grau de corte z_c :

$$g_{z_c}(Z) = \begin{cases} 1 & \text{se } Z(x) < z_c \\ 0 & \text{c.c.} \end{cases};$$

(b) indicatriz inversa:

$$g'_{z_c}(Z) = 1 - g_{z_c}(Z) = \begin{cases} 1 & \text{se } Z(x) \geq z_c \\ 0 & \text{c.c.} \end{cases};$$

(c) indicatriz Z -inversa:

$$Zg'_{z_c}(Z) = Z(x)g'_{z_c}(Z) = \begin{cases} Z(x) & \text{se } Z(x) \geq z_c \\ 0 & \text{c.c.} \end{cases}.$$

Os parâmetros associados a estas são:

(i) o teor médio de minério;

$$m = E[Z(x)] = \frac{1}{D} \int_D Z(u) du;$$

(ii) proporção de núcleos abaixo do ponto de corte;

$$F(z_{zc}) = P[Z(x) < z_{zc}] = E[\mathcal{F}_{zc}(Z(x))] = \frac{1}{D} \int_D t_{zc}(u) du \quad .$$

A função de recuperação pontual é então definida em termos das funções aleatórias acima citadas.

Tonelagem Recuperável:

$$T(z_c) = E[\mathcal{F}'_{zc}(Z(x))] = \frac{1}{D} \int_D t'_{zc}(u) du \quad ;$$

Quantidade de Metal Recuperável:

$$Q(z_c) = E[Z \mathcal{F}'_{zc}(Z(x))] = \frac{1}{D} \int_D z(u) t'_{zc}(u) du \quad ;$$

Teor Médio Recuperável:

$$m(z_c) = Q(z_c) / T(z_c) \quad .$$

6.4 Função de Recuperação de Blocos

Suponha que a seleção possa ser feita sem a desclassificação do valor real do conteúdo de minério no bloco. Define-se o conjunto de funções aleatórias de blocos e parâmetros, análogos aos do item anterior, incluindo o subscrito V (suporte do bloco) naquelas expressões. Ainda mais, defina o conteúdo minério do bloco como:

$$Z_V(x) = \frac{1}{V} \int_{V(x)} z(u) du \quad .$$

Também será assumido que a distribuição estacionária, dos conteúdos dos blocos, pode ser obtida diretamente da distribuição do conteúdo nuclear por uma correção afim da variância.

Dada a hipótese de correção afim, funções de recuperação de

blocos podem ser obtidas diretamente das funções de recuperação pontual.

Na prática mineira, a verdadeira exploração ou lavra se faz sobre unidades de seleção, que ultrapassam largamente o suporte da amostra. O efeito de suporte não é negligível. O problema é saber como aproximar a distribuição dos blocos em um painel V a partir de uma informação pontual.

De forma geral, o problema é chamado "mudança de suporte", que é um dos principais problemas em *geoestatística*.

Se v designa o suporte da unidade de seleção, define-se a função indicatriz de blocos:

$$f(Z(v) \geq z_c) = \begin{cases} 1 & \text{se } Z(v) \geq z_c \\ 0 & \text{c.c.} \end{cases}.$$

Define-se, desta forma, o preditor das funções de recuperação de um painel V contendo N bloco v_i 's:

(i) Mineral recuperável

$$T = \frac{1}{N} \sum_{i=1}^N f(Z(v_i) \geq z_c);$$

(ii) Metal recuperável

$$Q = \frac{1}{N} \sum_{i=1}^N Z(v_i) f(Z(v_i) \geq z_c).$$

Os elementos de resposta são extremamente reduzidos. Sabe-se que a média dos teores dos blocos será a mesma dos corpos de prova. Sabe-se, também, que a estrutura média pontual sobre v é próxima à variância do bloco. Por outro lado, nada se sabe sobre a forma de F_v . Porém, na prática, pode-se tirar informações úteis como a distribuição do bloco condicionada à pontual (do corpo), trata-se evidentemente de uma aproximação.

Outra informação utilizada para construir o modelo de mudança de suporte é a relação de Cartier:

$$E[Z(x)|Z(v)] = Z(v).$$

Utiliza-se aqui, um modelo gaussiano, já que o anamorfismo gaussiano é possível e conduz a pares de bigaussianas (por hipótese).

Conhecendo-se a distribuição dos teores pontuais $Z(x)$, graças ao anamorfismo pontual ϕ :

$$Z(x) = \phi(Y(x)) = \sum_{k=0}^n \varphi_k \eta_k(Y(x)),$$

é possível determinar a distribuição dos teores dos blocos $Z(v)$, devido ao anamorfismo dos blocos. Seja $Y(v)$, o anamorfismo dos conteúdos de blocos $Z(v)$. Então, no modelo Gaussiano (discreto) o espaço é particionado em blocos.

Seja x , um ponto aleatório no bloco v . Desta forma obtém-se a relação de Cartier:

$$Z(v) = E[Z(x)|Z(v)] = \sum_{k=0}^n \varphi_k E[\eta_k(Y(x))|Y(v)] = \sum_{k=0}^n \varphi_k \rho^k \eta_k(Y(v)),$$

supondo que o par seja bigaussiano com correlação ρ .

Uma vez os coeficientes $\varphi_k \rho^k$, determinados, desenvolve-se ϕ , de maneira a respeitar a variância de $Z(v)$, dada por:

$$\text{Var}(Z(v)) = \sum_{k=0}^n (\varphi_k \rho^k)^2.$$

Sejam v_i 's os blocos de anamorfismo $Y(v_i)$, e x_i 's pontos aleatoriamente escolhidos em cada um de tais blocos, que correspondem aos anamorfismos $Y(x_i)$, pela multigaussianidade a lei é perfeitamente determinada pela covariância dois a dois.

(A) Covariância Bloco-Bloco: $\rho_{ij} = \text{Cov}(Y(v_i), Y(v_j))$, que é obtida a partir de $\text{Cov}(Z(v_i), Z(v_j))$, determinada através de um sistema de *Kriging Linear*, por:

$$\text{Cov}(Z(v_i), Z(v_j)) = \sum_{k=0}^n (\varphi_k \rho^k)^2 (\rho_{ij})^k.$$

(B) Covariância Ponto-Bloco: $\rho_{\alpha j} = \text{Cov}(Y(x_\alpha), Y(v_j)) = \rho \text{Cov}(Y(v_1), Y(v_j))$.

(C) Covariância Ponto-Ponto: $\rho_{\alpha\beta} = \text{Cov}(Y(x_\alpha), Y(x_\beta)) = \rho^2 \text{Cov}(Y(v_1), Y(v_j))$.

Deseja-se, portanto, predizer uma função $\psi(Y(v))$, ou até mesmo uma ponderação de tal função pelos N blocos v_1 's de um painel V . O formalismo da predição de $\psi(Y(v))$, é exatamente o mesmo envolvido na predição pontual de $\psi(Y_0)$. É suficiente, então, recriar os sistema de *Kriging Linear*, fazendo uso das covariâncias ponto-ponto e ponto-bloco do modelo Gaussiano.

6.5 Efeito de Suporte

Considere todos os blocos v_1 a serem selecionados em um dado corpo mineral. Z_{v_1} seu conteúdo, e $Z_{v_1}^*$ seu conteúdo predito baseado em algum conjunto de informações. Sejam, ainda, $F(z)$ e $F^*(z)$ suas respectivas funções de distribuição de probabilidades (FDP).

Se a seleção fora feita com o verdadeiro conteúdo Z_{v_1} (informação perfeita), a tonelagem de minério, a quantidade de metal e o benefício acima de um ponto de corte z_c , serão respectivamente:

$$T(z_c) = T_M (1 - F(z_c)) \quad ;$$

$$Q(z_c) = T_M \int_{z_c} z \, dF(z) \quad ;$$

$$B(z_c) = Q(z_c) - z_c T(z_c) \quad ;$$

onde T_M é a tonelagem do corpo mineral.

Estas três funções dependem da distribuição do conteúdo do bloco. É conhecido da *geoestatística* que, quanto maior o bloco v_1 , mais suave (no sentido de menor variância, ou menor coeficiente de variação) é tal distribuição do conteúdo. Consequentemente para alto valor de corte, é maior a seleção de blocos e menor a tonelagem recuperada. A

influência do bloco selecionado é chamado efeito de suporte. G. Matheron [1983] mostrou matematicamente que para qualquer distribuição do teor, é maior a seleção do suporte e menor a operação de beneficiamento.

No caso de seleção indireta do bloco v_1 , a tonelagem recuperada e quantidade de metal, serão:

$$T(z_c) = T_M (1 - F^*(z_c)) \quad ;$$

$$Q(z_c) = T_M \int_{z_c}^{\infty} z^* dF^*(z) \quad .$$

A tonelagem recuperada será dependente diretamente da FDP do conteúdo predito do bloco, Z_v^* .

6.6 Método para Predição de Recuperação Pontual

Considera-se, primeiramente, o método de predição que permite a análise de dois importantes problemas:

- (i) predição da FDP local;
- (ii) mudança de suporte.

O primeiro problema tem várias soluções matemáticas exatas, importantes, enquanto o segundo tem apenas soluções aproximadas. Do ponto de vista prático, predição de pontos recuperáveis é importante no estudo de minas já em operação, desta forma se torna possível comparar a distribuição local estimada com as de cada painel.

Como visto anteriormente, prever a função de recuperação pontual em dado painel V , é equivalente a estimar a distribuição local dos conteúdos de pontos x , $Z(x)$ em V , que se sabe é equivalente a prever a função indicatriz de conteúdos pontuais para todos os pontos do painel V .

Seja $\{Z(x)\}$, um processo estocástico estacionário com FDP $F(z)$, e FDP bivariada entre dois pontos x e $x+h$, $F_h(z, z)$, sua função de covariância. O processo é realizado em n pontos amostrais. A função

indicatriz para o conteúdo de corte z_c é associada à função numérica $f_{z_c}(Z)$. Nestas condições, $f_{z_c}(Z)$ define um novo processo estocástico estacionário com as seguintes condições:

$$E[f_{z_c}(Z)] = 1 - F(z_c);$$

$$\text{Var}[f_{z_c}(Z)] = F(z_c)[1 - F(z_c)].$$

A $f_{z_c}(Z)$ tem função de covariância estacionária:

$$C_{z_c}(h) = \int_{z_c}^{\infty} \int_{z_c}^{\infty} F_h(u, u') du du' - [1 - F(z_c)]^2.$$

A idéia do uso de função indicatriz é baseada em *Kriging Disjuntivo*, que é proposto aqui como um método prático para a predição de recuperação pontual.

Como qualquer função contínua pode ser representada por uma combinação de funções indicatrizes, tal preditor de (DK), pode ser aproximado por uma combinação linear de funções indicatrizes baseadas na amostra:

$$Z = \sum_{i=1}^n \psi_i f_{z_i}(Z_{\alpha});$$

onde os z_i 's formam uma sequência de valores de corte, permitindo uma boa representação da variável Z .

Suponha que a variável $Z(x)$ seja representada de uma forma contínua. No caso de uma variável positiva $Z(x)$, tem-se:

$$Z(x) = \int_0^{\infty} f_z(x) dz.$$

Seja $f_z^*(x)$, o *Kriging Disjuntivo* (Co-Kriging completo) da função $f_z(x)$, por linearidade $Z^*(x)$ será:

$$Z^*(x) = \int_0^\infty f_z^*(x) dz \quad ;$$

onde

$$f_z^*(x) = \sum_{\alpha=1}^n \int_0^\infty \lambda_\alpha^*(z, z') f_{z, (x_\alpha)} dz' \quad ;$$

então, obtém-se:

$$Z^*(x) = \sum_{\alpha=1}^n \int_0^\infty \int_0^\infty \lambda_\alpha^*(z, z') f_{z, (x_\alpha)} dz dz' \quad ;$$

definindo-se

$$\psi_\alpha^*(z') = \int_0^\infty \lambda_\alpha^*(z, z') dz \quad ;$$

isto resulta em:

$$Z^*(x) = \sum_{\alpha=1}^n \int_0^\infty \psi_\alpha^*(z') f_{z, (x_\alpha)} dz' \quad ;$$

Cada termo

$$\int_0^\infty \psi_\alpha^*(z') f_{z, (x_\alpha)} dz'$$

é a definição de uma função $g_\alpha(Z_\alpha)$ através de funções indicatrizes.

Voltando à usual representação de *Kriging Disjuntivo*, $\sum g_\alpha(Z_\alpha)$. Apresentando o preditor de (DK), como um *Co-Kriging Indicador*, permite um fácil entendimento das especificações necessárias para se determinar um preditor de DK em uma forma contínua.

6.7 Predição por Distribuição Condicional

Considere a realização do processo estocástico em n pontos amostrais x_α , $\alpha=1, \dots, n$, e um ponto alvo x_0 , define-se o processo vetorial $\{Z(x_0), Z_\alpha: \alpha=1, \dots, n\}$ por sua FDP $F(z_0, z_1, \dots, z_n)$.

Pode-se derivar desta FDP:

— a densidade marginal de $Z(x_0)$ e $Z_\alpha \Rightarrow f_0(z)$ e $f_\alpha(z)$, $\forall \alpha$;

— a densidade condicional de $Z(x_0)$ quando cada Z_α é fixado $\Rightarrow$
 $\Rightarrow f(Z(x_0)|Z_1, \dots, Z_n)$.

Para prever $\phi[Z(x_0)]$, sendo ϕ qualquer função mensurável, o preditor deve ser uma função $g(Z_1, \dots, Z_n)$ dos valores amostrados $(Z_1, \dots, Z_n)$ que minimiza a variância do erro é a esperança condicional de $\phi[Z(x_0)]$ dado $Z_1, \dots, Z_n$.

$\phi^*[Z(x_0)] = g(Z_1, \dots, Z_n)$ tal que $E(\phi^* - \phi)^2$ seja mínima;

$$\phi^*[Z(x_0)] = \int \phi(z) f_0(z|z_1, \dots, z_n) dz$$

Em particular se a função a ser predita for a função indicatriz $\mathcal{F}[Z(x_0)]$, o preditor ótimo é a FDP:

$$\mathcal{F}^*[Z(x_0)] = \int_z^\infty f_0(z'|z_1, \dots, z_n) dz' = 1 - F_0(z|z_1, \dots, z_n).$$

Na aplicação em mineração é frequente o uso acima de 10 pontos amostrais para avaliar um preditor, o correspondente modelo $F(z_0, z_1, \dots, z_n)$ será muito complexo no caso geral, a não ser quando gaussiano multivariado.

Sendo que para qualquer variável aleatória contínua, existe uma transformação desta em uma gaussiana, aplicando tal transformada a $Z_0, Z_1, \dots, Z_n$, cada variável originada $(Y_0, Y_1, \dots, Y_n)$ assim pode ser considerada individualmente gaussiana "padrão" e tal suposição consiste de que o par seja bigaussiano com média 0, variância 1 e matriz de covariância $P: [\rho_{\alpha\beta}]$.

A maior vantagem da distribuição gaussiana multivariada é a grande simplicidade do cálculo da distribuição condicional, pois esta ainda é gaussiana. De fato, a distribuição condicional de $Y_0|Y_1, \dots, Y_n$ é $\sum_{\alpha=1}^n \lambda_\alpha Y_\alpha + \sigma(x_0) \nu \sim N(0,1)$, onde λ_α é solução de $\sum_{\alpha=1}^n \lambda_\alpha \rho_{\alpha\beta} = \rho_{\beta x_0}$ e $\sigma(x_0) = 1 - \sum_{\alpha=1}^n \lambda_\alpha \rho_{\alpha x_0}$.

Consequentemente, para a função indicadora, tem-se:

$$f^*(Y(x_0)) = 1 - G \left(\frac{y - \sum_{\alpha=1}^n \lambda_{\alpha} Y_{\alpha}}{\sigma(x_0)} \right).$$

Por construção o preditor $f^*(Z(x_0)) = g(Z_1, \dots, Z_n)$ tem a propriedade de $1 - f^*(Z(x_0))$ ser uma FDP. Não é necessário que $\{Z(x)\}$ seja estacionário, quando se pode especificar uma $(n+1)$ FDP conjunta para $(Z_0, Z_1, \dots, Z_n)$ a esperança condicional é implicitamente definida.

CONSIDERAÇÕES FINAIS

As propriedades teóricas do preditor de *Kriging Disjuntivo* são as mesmas de qualquer preditor de *Co-Kriging*, e depende apenas da variante de *Kriging* utilizada.

(i) $\hat{F}^*(Z(x_0)) = 1 - F(z) + \sum_{\alpha} \sum_{\beta} \lambda_{\alpha}(Z_0, Z_{\beta}) [\hat{F}(Z_{\beta}(x_{\alpha})) - 1 + F(Z_{\beta})]$, onde $\lambda_{\alpha}(Z_0, Z_{\beta})$ é o coeficiente de *Co-Kriging* na predição de $\hat{F}(Z_{\beta}(x_0))$, $\hat{F}^*(Z(x_0))$ é o preditor de *Kriging Simples* baseado no modelo da FDP de $Z(x)$, $F(z)$. Como qualquer SK à média conhecida $1 - F(z)$, a validade deste preditor é baseada na estacionaridade estrita sobre o campo amostrado, em particular quando a informação for esparsa, a predição será igual à média $1 - F(z)$.

(ii) $\hat{F}^*(Z(x_0)) = \sum_{\alpha} \sum_{\beta} \lambda_{\alpha}(Z_0, Z_{\beta}) [\hat{F}(Z(x_{\alpha}))]$, este é o preditor de *Kriging Ordinário*, principalmente usado em uma vizinhança local. É sabido que o *KO* requer apenas uma estacionaridade local da covariância.

(iii) $\hat{F}^*(Z(x_0))$ pode ser definido por um sistema de *Kriging Universal*, embora não de total coerência, baseado em um conjunto de funções de tendência espacial para $\hat{F}(Z(x_0))$. De fato, crescendo o número de condições de universalidade, cresce o número de ponderações negativas de *Kriging*, o que indica uma maior adequação para *Co-Kriging Indicador*. Desta forma valores negativos, nulos ou um para a ponderação, devem proporcionar um preditor negativo para $\hat{F}(Z(x_0))$, ou $\hat{F}^*(Z(x_{\alpha+1})) > \hat{F}^*(Z(x_{\alpha}))$ com $Z(x_{\alpha+1}) > Z(x_{\alpha})$. Na prática o *Co-Kriging Simples* é recomendado.

Os passos essenciais para um *KD* são:

(1) transformar os valores amostrais (originais) usando uma

transformada matemática ou uma "transformação gráfica". A hipótese fundamental comum a todos os métodos de predição apresentados aqui é a possibilidade de se fazer um anamorfismo gaussiano bijetivo dos conteúdos, conduzindo a pares bigaussianos, o que em geral ocorre em depósitos de Cobre, Chumbo, Zinco, Alumínio, Ferro, etc.

(2) Calcular o semi-variograma e ajustar um modelo teórico a este, e assim o ajuste de $\rho(h)$ (correlograma).

(3) Resolver K sistemas de *Kriging* com covariância ρ^k . Estes sistemas são usualmente num montante de 10, a maior ordem sendo facilmente resolvida pelo fato destes serem diagonal dominante. De fato, estes sistemas referem-se diretamente à função média indicadora $\overline{F}(V)$, num painel completo V, através da densidade média predita $\overline{f}_V^*$. Como visto a tonelagem pontual recuperada e o metal pontualmente recuperado são ligados a esta distribuição:

$$T_V^*(Z) = \int_{y_c}^{\infty} \overline{f}_V^*(u) du;$$

$$Q_V^*(Z) = \int_{y_c}^{\infty} \phi(u) \overline{f}_V^*(u) du, \text{ onde } y_c = \phi^{-1}(z_c).$$

É possível, desta forma, calcular o grau de co-produtos recuperados quando o grau principal é selecionado a um ponto de corte z_c . O grau do co-produto recuperado no painel é derivado do grau do co-produto predito do painel modificado por uma expansão de termos incluindo os H_k^* s do painel e os coeficientes de correlação entre o principal transformado e o grau do co-produto.

A seleção minério/estéril é feita quando se dá a mineração pela aplicação do critério de corte a um preditor Z_V^* de cada unidade de seleção do bloco. Tem sido mostrado que isto propicia um preditor condicionalmente não-viciado, as funções de recuperação T_Z e Q_Z podem ser derivadas da distribuição condicional (ou um preditor desta) do preditor Z_V^* : $F_V^* = P(Z_V^* > z_c | z_1, \dots, z_n)$. Como exemplo:

$$f_v^*(Z_c) = \int_{z_c}^{\infty} f(z_v^* | z_1, \dots, z_n) dz_v^*.$$

A exata determinação de $f(z_v^* | z_1, \dots, z_n)$ é impossível na prática, pelo fato do preditor Z_v^* a ser calculado no estágio da mineração não ser conhecido no estágio da exploração. Uma aproximação tem sido feita considerando tal preditor como de *Kriging* com dada variância de predição. Em suma deve-se antecipar a quantidade de informação avaliável no estágio de mineração.

Teoricamente no *Co-Kriging Indicador* de bloco, $F_v(z)$ pode ser derivada de uma lei espacial de $Z(x)$ para todos os pontos no bloco v . Assim em geral, $F_v(z)$ não pode ser ligada a uma FDP pontual $F(z)$ apenas, e derivando $F_v(z)$ requer um modelo probabilístico completo. Em casos excepcionais é permitido uma determinação matemática para este modelo, por meras razões de cálculo.

Na prática, a mudança de suporte irá consistir em definir uma relação entre o grau pontual $F(z)$ e o grau do bloco $F_v(z)$. Estas operações devem satisfazer a um determinado número de restrições devido à hipótese de estacionaridade e a relação de Cartier para FDP's.

- $F_v(z)$ e $F_x(z)$ têm a mesma média, m ;
- a variância associada a $F_v(z)$ deve ser igual ao valor dado pela integração do variograma de $Z(x)$, isto é $\sigma_v^2 = \sigma_v^2 - \gamma(V, V)$;
- o intervalo de definição de Z_v deve ser o mesmo que $Z(x)$;
- definindo o operador beneficiamento gerado pela distribuição do grau selecionado a um ponto de corte ao nível z_c como:

$$B(z_c) = \int_{z_c}^{\infty} (z - z_c) dF(z),$$

o benefício produzido por uma seleção no bloco deve ser menor ou igual ao produzido por uma seleção no grau pontual, para todo z_c ;

$$B_v(z_c) \leq B_x(z_c), \quad \forall z_c.$$

Há duas possíveis visões para se determinar a predição local de blocos recuperados: ou se determina a distribuição condicional do teor do bloco através de um modelo probabilístico completo ou predizendo tal distribuição por um *Co-Kriging* completo. Determinando a distribuição condicional do grau do bloco presume-se que a fórmula de mudança de suporte fora usada para definir o modelo de grau do bloco e para definir as probabilidades pontuais $(Z_v, Z_1, \dots, Z_n)$. O resultado final será para um dado bloco, a FDP condicional:

$$F_v(z|z_1, \dots, z_n) = P(Z_v \leq z | z_1, \dots, z_n).$$

Predizer reservas recuperáveis é de grande importância sob um aspecto econômico, pois este se torna o único caminho para avaliar os dois maiores fatores que influenciam a seletividade: o efeito de suporte e o efeito de informação. Apenas pelo conhecimento destes efeitos é que a Geoestatística está apta a resolver corretamente o problema de "Tonelagens Faltantes". Em muitos corpos minerais cada grau médio predito não é muito preciso, pois esta se faz, geralmente, utilizando-se de histogramas do grau pontual

Resulta que uma efetiva predição recuperável depende de dois pontos primordiais que fazem com que esta seja ou não razoável:

- necessidade do uso de algum método sofisticado baseado em um modelo probabilístico bem definido;
- necessidade de maior despenho para ajustar e checar os modelos, sabe-se que mais completo o modelo mais sofisticado é o preditor.

Há uma necessidade no estágio da exploração de obter globalmente para grandes unidades, o resultado da seletividade de mineração a diferentes graus de pontos de corte. *Kriging Indicador* baseado no *Kriging* do painel, irá proporcionar uma predição inicial da reserva recuperada, o método já é bastante utilizado a um nível global para determinação de histogramas mais sensíveis a "clusters". Um importante sub-produto desta prática é proporcionar informações para ajustar modelos no caso de não estacionaridade. A este primeiro nível de uma predição global, quando métodos simples são requeridos, a

mudança de suporte pode ser equivalente a usar correção afin ou um método gaussiano discreto. Finalmente, não se deve esquecer que alguns métodos de seleção são muito complexos e não podem ser idealizados como uma seleção direta livre, não há alternativa senão simular o método de mineração em uma simulação da distribuição do grau. Existe uma proximidade muito grande entre os modelos utilizados na predição recuperável e na simulação condicional.

APÊNDICE A

Sejam X variável aleatória com função de distribuição F e $y=F(x)$ e defina F^{-1} como segue:

$$F^{-1}(y)=\inf\{x\in\mathcal{R}: F(x)\geq y\}. \quad (A)$$

Desta definição fica claro que quando F for estritamente crescente, para cada $x\in\mathcal{R}$, existe exatamente um $y\in(0,1)$ tal que $F(x)=y$. Fica claro também que se F for contínua então a definição acima vem a ser:

$$F^{-1}(y)=\inf\{x\in\mathcal{R}: F(x)=y\}$$

Lema: Seja F^{-1} definida por (A). Então $F^{-1}(y)\leq z$ se, e somente se, $y\leq F(z)$.

Dem: Tem-se $F^{-1}(y)=\inf\{x\in\mathcal{R}: F(x)\geq y\}$, assim existe $x_n\in\{x\in\mathcal{R}, F(x)\geq y\}$, tal que $x_n\downarrow F^{-1}(y)$, então $F(x_n)\rightarrow F(F^{-1}(y))$, pela continuidade a direita de F e $F(F^{-1}(y))\geq y$. Assume-se que $F^{-1}(y)\leq z$, então $F(F^{-1}(y))\leq F(z)$, posto que F seja não decrescente, daí obtém-se $F(z)\geq y$. Agora, assumamos que $y\leq F(z)$, isto significa que $z\in\{x\in\mathcal{R}: F(x)\geq y\}$ e assim $F^{-1}(y)\leq z$.

Teorema: Seja Y uma variável aleatória com distribuição $U(0,1)$ e seja F uma função de distribuição. Defina a variável aleatória X por $X=F^{-1}(Y)$, onde F^{-1} é definida por (A). Então a função de distribuição de X é F .

Dem: Tem-se $P(X\leq x)=P[F^{-1}(Y)\leq x]=P[Y\leq F(x)]=F(x)$, onde a última igualdade segue do fato que Y seja distribuída uniformemente em $(0,1)$ e a anterior pelo Lema acima. ■

Fato 1: Seja X uma variável aleatória com distribuição acumulada F (contínua e invertível), então $Y=F(X) \sim U[0,1]$.

Fato 2: Sejam Y uma variável aleatória tal que $Y \sim U[0,1]$ e F uma distribuição acumulada qualquer com inversa F^{-1} , então $X=F^{-1}(Y)$ tem distribuição acumulada F .

Teorema: Seja X uma variável aleatória com distribuição F (invertível), então existe uma transformada gaussiana (única e invertível) ϕ^* .

Dem: Se $Z \sim N(0,1)$ com distribuição acumulada ϕ , então $\phi(Z) \sim U(0,1)$, pois ϕ é invertível. Se X tem distribuição acumulada F , então $X=F^{-1}(\phi(Z))$, isto é $X=(F^{-1} \circ \phi)(Z) \Rightarrow Z=\phi^{*-1}(X)$. Esta transformada é única posto que a função de distribuição acumulada de uma dada variável aleatória é única. ■

REFERÊNCIAS BIBLIOGRÁFICAS

Abramowitz M. e Stegun A. (1965), "*Handbook of Mathematical Functions*". Dover, New York.

Armstrong M. e Diamond P. (1984), *Testing Variograms for Positive-Definiteness*. Mathematical Geology, v16, n04.

Armstrong M. e Matheron G. (1986), *Disjunctive Kriging Revisited - Part I/ Part II*. Mathematical Geology, v18.

Braga L. (1982), *Mapping Technics using Kriging*. U. des Montreal, publication #453.

Cressie N. (1987), *A Nonparametric View of Generalized Covariances for Kriging*. Mathematical Geology, v19, n05.

————— (1989), *Geostatistics*. The American Statistician, v43, n07.

David M. (1977), "*Geostatistics Ore Reserve Estimation*". Elsevier, Amsterdam.

Demange C. et ali (1987), *Global Recoverable Reserves: Testing Various Changes of Support Models on Uranium Data*. Geostatistics Case Studies, D. Reidel Publishing Company.

Graybill L. (1976), "*Theory and Application of the Linear Model*".

Guibal D. e Remacre A. (1984), *Local Estimation of Recoverable Reserves: Comparing Various Methods with the Reality on a Porphyry Copper Deposit*. Geostatistics for Natural Resources Characterization, D. Reidel Publishing Company.

Journel A. (1977), *Kriging in Terms of Projections*. Mathematical Geology, v09, n06.

———— (1982), *Indicator Approach to Spatial Distributions*. Proceedings of 17th APCOM, Denver.

Journel A. e Huijbregts Ch. (1978), *"Mining Geostatistics"*. Academic Press, London.

Marechal A. (1984), *Recovery Estimation: A Review of Models and Methods*. Geostatistics for Natural Resources Characterization, D. Reidel Publishing Company.

Matheron G. (1965), *"Les Variables Regionalisées et leur Estimation"*. Masson & Cia, Paris.

———— (1973), *The Intrinsec Random Functions and their Applications*. Adv. Appl. Probability, v05.

———— (1976), *A Simple Substitute for Condicional Expectation: The Disjunctive Kriging*. Advanced Geostatistics in the Mining Industry, D. Reidel Publishing Company.

———— (1983), *The Selectivity of the Distributions and the Second Principle of Geostatistics*. NATO ASI, D. Reidel Publishing Company.

———— (1984), *Isofactorial Models and Change of Support*. Geostatistics for Natural Resources Characterization, D. Reidel Publishing Company.

Myers D. (1982), *Matrix Formulation of Co-Kriging*. Mathematical Geology, v14, n03.

Nelson G. e Guibal D. (1986), *Recoverable Reserves on an Australian Gold Deposit*. Geostatistical Case Studies, D. Reidel Publishing Company.

Petenate, A. (1983), "*Optimal Allocation and others aspects of Kriging on a Line*". Ph.D. Thesis, Iowa State Univerty, Iowa

Remacre A. (1984), "*L'estimation du Récupérable Local: Le Conditionnement Uniform*". These de Docteurat Présentée à L'École Nationale Supérieure des Mines de Paris – Option Géostatistique.

————— (1989), *Uniform Conditioning versus Indicator Kriging: A Comparative Study with the Actual Data*. Geostatistics, v02, Kluwer Academic Publishers.

Ripley B. (1981), "*Spatial Statistics*". John Wiley, New York.

Valente J (1982), "*Lições de Geoestatística*". Fundação Gorceix, Ouro Preto.

Yates S., Warrick A. e Myers D. (1986), *Disjunctive Kriging - 1. Overview of Estimation and Conditional Probability*. Water Resources Research, v22, n05