Estatística Aplicada a Recursos Hídricos

Docente: Rachid Muleia

(rachid.muleia@uem.mz)

Mestrado em Gestão de Recursos Hídricos - DGEO/UEM

Tema: Interpolação espacial

Ano lectivo: 2023

Interpolação espacial

- Estimar valores em locais não amostrados usando observações vizinhas
- Assume-se que o atributo de interesse é contínuo em todo domínio
- Assume-se que o atributo é espacialmente dependente, com maior similaridade para dados mais próximos, e maior dissimilaridade para dados mais afastados
- Pode-se destacar dois tipos de interpolção: determínistica e estatística

Interpolação inversa da distância ponderada (IVD)

- A IVD é simplesmente a média ponderada das observações vizinhas;

- A interpolação de um ponto não amostrado pode ser calculada usando:

$$Z(\mathrm{s}_0) = \sum_{i=1}^{n} \lambda_i Z(\mathrm{s}_i)$$

λ_i representa o peso atribuído observação amostrada. O λ_i depende da distância entre o valor amostrado e o ponto a ser estimado, e é dado por:

$$\lambda_i = \frac{d_{0,i}^{-p}}{\sum_i^n d_{0,i}^{-p}}$$

$d_{0,i}$ representa a distância entre o ponto s_0 e s_i

Cont.

- A soma dos pesos deve ser igual a 1, isto é, $\sum_i^n \lambda_i$

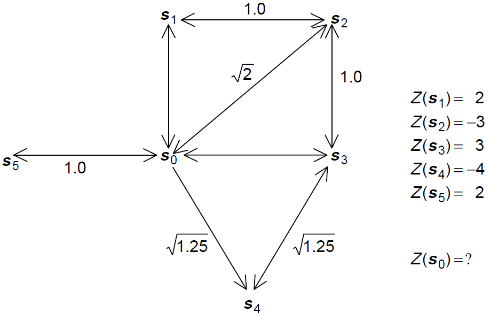
- Geralmente, para p usa-se valores entre 1 e 3

- $p = 2$ dá origem à inverso do quadrado da distância

- Assumir $p = 0$ é mesmo que usar a média aritmética para estimar o valor de $Z(s_0)$

- A medida que o p aumenta, os pesos vão diminuindo em função da distância

Exemplo - IVD



- Uma estimativa razoável para s_0 é a média aritmética;
- Usando IVD as os valores em s_2 e s_4 terão menor; contribuição
- Considerando $p = 2$, tem-se que $Z(s_0) = 0.53$

Exemplo - ilustração em R

```
library(sp)
library(gstat)
# valores amostrados
X = c(61,63,64,68,71,73,75)
Y = c(139,140,129,128,140,141,128)
Z = c(477,696,227,646,606,791,783)
# locais não amostrados. Qual é o valor de Z1
X1 = 65; Y1 = 137

obser_dt = data.frame(X,Y,Z)
coordinates(obser_dt)= ~ X + Y
nao_obser_dt = data.frame(X1,Y1)
coordinates(nao_obser_dt)=~ X1 + Y1
idwmodel = idw(Z ~1, obser_dt,nao_obser_dt, maxdist = Inf, idp = 2)
predZ= idwmodel@data$var1.pred
predZ

> 597.6204
```

Exemplo - IVD para interpolação do HIV.

```
path='C:/Users/lucp8943/Dropbox/Geostatistics Class Material/Lecture notes'
hiv.df_prop=read.csv(paste(path, 'hiv_prev.csv', sep='/'), header=TRUE); head(hiv.df_prop)
shp1<-readOGR("MZGE52FL.shp")
shp<-readOGR("MOZ-level_1.shp")
# create outer boundary
shp2 <- gUnaryUnion(shp)
plot(shp)
plot(shp2, add=T, border = 'red')

# create coordinates of outer boundary
extractCoords <- function(sp.df)
{
  results <- list()
  for(i in 1:length(sp.df@polygons[[1]]@Polygons))
  {
    results[[i]] <- sp.df@polygons[[1]]@Polygons[[i]]@coords
  }
  results <- Reduce(rbind, results)
  results
}
coord<-extractCoords(shp2)
```

Cont

```
# make grid within area
plot.x = seq(30.21786,40.84447,0.1)
plot.y = seq(-26.86867,-10.47188,0.1)
x.pred = expand.grid(plot.x, plot.y)[,1]
y.pred = expand.grid(plot.x, plot.y)[,2]
pred.loc = data.frame(cbind(x.pred,y.pred))
pred.loc2<-SpatialPoints(pred.loc)
point.in.polygon = over(pred.loc2,shp2)
plot(shp2)
points(pred.loc2[!is.na(point.in.polygon),],cex=0.2)
coordinates(hiv.df_prop)= ~ long +lat
coordinates(pred.loc)=~x.pred+y.pred
idwmodel = idw(prev ~1, hiv.df_prop,pred.loc,
               maxdist = Inf, idp = 2)
predZ= idwmodel@data$var1.pred
predZ
plot.index = matrix(point.in.polygon, nrow=length(plot.x), ncol=length(plot.y))
plot.mean.idw = matrix(predZ, nrow=length(plot.x), ncol=length(plot.y))
plot.mean.idw[is.na(plot.index)] = NA
image.plot(x=plot.x,y=plot.y,z=plot.mean.idw,xlab="",ylab="")
contour(x=plot.x,y=plot.y,z=plot.mean.idw,xlab="",ylab="",add=TRUE, drawlabels=FALSE)
plot(shp2,add=T)
```

Cont.

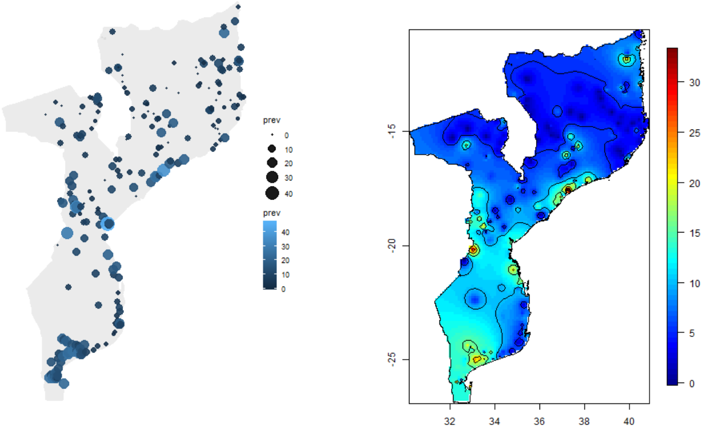


Figura: Do lado esquerdo temos valores observados e do lado direito temos valores previsto para todo o domínio

Interpolação por Krigagem

- Krigagem- Técnica geoestatística para interpolação espacial;
- Igual ao IVD, a Krigagem usa valores vizinhos para prever locais não amostrados ou estimar a média sobre um determinado bloco;

$$Z(\mathsf{s}_0) = \sum_{i=1}^{n} \lambda_i Z(\mathsf{s}_i)$$

- A Krigagem é conhecida como um interpolador exacto;

Cont.

Salientar que, a qualidade da interpolação depende de vários factores, nomeadamente:

- [i)] tamanho da amostra e a qualidade dos dados;
- [ii)] localização das observações (uniforme ou agrupados);
- [iii)] distância entre os pontos observados e o ponto a ser previsto;
- [iv)] continuidade espacial da variável aleatória;

Os métodos de Krigagem levam uma vantagem sobre os outros métodos de interpolação, pois para além de considerar a característica geométrica do processo, consideram também a estrutura espacial do fenómeno.

Noção de vizinhança

Krigagem usa valores circunvizinhos para estimação/previsão de uma local não amostrado.

- Todos valores amostrados são incluídos na interpolação;
  - ➜ O impacto das observações distantes é insignificante;
  - ➜ Inclusão de todas observações exige que o processo seja estacionário em todo o domínio;
- Apenas os pontos mais próximos do local a ser interpolado é que são considerados;
  - ➜ Apenas exige-se que o processo seja estacionário de segunda ordem ou quase intrínsecamente estacionário

Cont.

Não existe uma regra clara para definir a dimensão da vizinhança, contudo Webster Oliver (2001) sugerem algumas directrizes:

- Se os dados forem densos e o semivariograma tiver efeito pepita menor, então o raio da vizinhança pode ser igual a amplitude ou amplitude prática;
- Se os o efeito pepita for maior, observações distantes do ponto a ser previsto irão ter um impacto significativo na interpolação e por essa razão devem ser incluídos na vizinhança;
- Por outra, pode-se definir a vizinhança em termos de número mínimo e máximo de observações próximos do ponto a ser interpolado. Geralmente recomenda-se um mínimo de $n \approx 7$ e um máximo $n \approx 20$.

Krigagem ordinária

Considere-se que uma função aleatória seja estacionária de segunda ordem, então

$$E[Z(\mathrm{s})] = \mu$$

com covariância definida por:

$$E[Z(\mathrm{s})Z(\mathrm{s}+\mathrm{h})] - \mu^2$$

e variograma dado por

$$E\left([Z(\mathrm{s}+\mathrm{h}) - Z(\mathrm{s})]^2\right)$$

Cont.

A interpolação por Krigagem é dada por

$$Z^*(s_0) = \sum_{i=1}^{n} \lambda_i Z(s_i)$$

A interpolação depende dos valores dos λ_i

- Estimador não enviesado

$$E[Z^*(s_0)] = \mu \sum_{i=1}^{n} \lambda_i = E[Z(s)]$$

- Variância do erro de estimação (erro quadrático médio) deve ser mínima. O que significa minimizar $E\left[(Z^*(s_0) - Z(s_0))^2\right]$ sujeito a restrição $\sum_{i=1}^{n} \lambda_i = 1$

Cont.

Usando o método dos multiplicadores de Lagrange os pesos (λ_i) podem ser calculados usando a seguinte expressão:

$$\boldsymbol{\lambda} = \boldsymbol{\Gamma}^{-1}\boldsymbol{\gamma}_0,$$

onde

$$\boldsymbol{\lambda} = \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_n \\ m \end{bmatrix}, \boldsymbol{\Gamma} = \begin{bmatrix} \gamma(s_1 - s_1) & \cdots & \gamma(s_1 - s_n) & 1 \\ \gamma(s_2 - s_1) & \cdots & \gamma(s_2 - s_n) & 1 \\ \vdots & \ddots & \vdots & \vdots \\ \gamma(s_n - s_1 & \ldots & \gamma(s_n - s_n) & 1 \\ 1 & \cdots & 1 & 0 \end{bmatrix}, \boldsymbol{\gamma}_0 = \begin{bmatrix} \gamma(s_0 - s_1) \\ \gamma(s_0 - s_2) \\ \vdots \\ \gamma(s_0 - s_n) \\ 1 \end{bmatrix}$$

Cont.

A variância de estimativa resultante da Krigagem é dada por:

$$\sigma^2(\mathsf{s}_0) = \boldsymbol{\lambda}^T\boldsymbol{\gamma}_0 = \sum_{i=1}^{n} \lambda_i \gamma(\mathsf{s}_0 - \mathsf{s}_i) + m$$

Assumindo-se que o erro de previsão $Z^*(\mathsf{s}_0) - Z(\mathsf{s}_0)$ segue distribuição normal, então o intervalo de previsão a 95

$$(Z^*(\mathsf{s}_0 \pm 1.96\sigma(\mathsf{s}_0))$$

Exemplo

$\rightarrow$ Considere exercício no slide 9.

$\rightarrow$ Considere, igualmente, um semivariograma esférico com $c_0 = 0$, $c_s = 1$ e $a = 1.5$

$$\begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \\ \lambda_4 \\ \lambda_5 \\ m \end{bmatrix} = \begin{bmatrix} 0 & 0.852 & 0.995 & 1.00 & 0.995 & 1 \\ 0.852 & 0 & 0.852 & 1.00 & 1.00 & 1 \\ 0.995 & 0.852 & 0 & 0.911 & 1.00 & 1 \\ 1.00 & 1.00 & 0.911 & 0 & 1.00 & 1 \\ 0.995 & 1.00 & 1.00 & 1.00 & 0 & 1 \\ 1 & 1 & 1 & 1 & 1 & 0 \end{bmatrix}^{-1} \begin{bmatrix} 0.852 \\ 0.995 \\ 0.852 \\ 0.911 \\ 0.852 \\ 1 \end{bmatrix} = \begin{bmatrix} 0.256809 \\ 0.049996 \\ 0.241484 \\ 0.186251 \\ 0.265459 \\ 0.118743 \end{bmatrix}$$

$$Z^*(s_0) = 0.257 \times 62 + 0.050 \times (-3) + 0.241 \times 3 + 0.186 \times (-4) + 0.265 \times 2 = 0.88$$

Exemplo - Smooky Mountain

```
ph_geo=as.geodata(ph_df,coords.col=c(2,1),data.col=3)
variogram_ph=variog(ph_geo,uvec = seq(0,120,l=10),messages = FALSE)
fit3=variofit(variogram_ph,cov.model="sph",ini.cov.pars=c(0.23,100),
              fix.nugget=FALSE,nugget=0,weights='npairs',messages=FALSE)
x.range=as.integer(range(ph_df[,1]))
y.range=as.integer(range(ph_df[,2]))
grd=expand.grid(x=seq(from=x.range[1], to=x.range[2], length.out=50),
                      y=seq(from=y.range[1], to=y.range[2], length.out  =50))
kg_ph=krige.conv(ph_geo,locations = grd, krige = krige.control(obj.m = fit3))
kg_ph$predict
```

Cont.

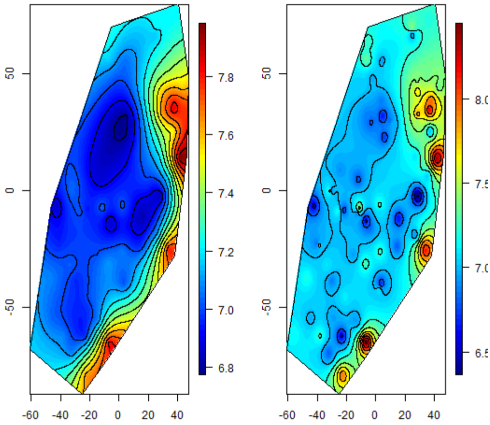


Figura: A figura mostra as interpolações, do lado esquerdo por Krigagem e do lado direito por Inverso da distância

Cont

Durante o processo do estudo da continuidade espacial observou-se que, este processo apresenta um comportamento anisotrópico.

```
plot.x = seq(-60.93033 , 47.48567,0.1)
plot.y = seq(-87.08972,79.58228,0.1)
x.pred = expand.grid(plot.x, plot.y)[,1]
y.pred = expand.grid(plot.x, plot.y)[,2]
pred.loc = data.frame(cbind(x.pred,y.pred))
pred.loc2<-SpatialPoints(pred.loc)
point.in.polygon = over(pred.loc2,sps)
plot(sps)
points(pred.loc2[!is.na(point.in.polygon),],cex=0.2)
plot.index = matrix(point.in.polygon, nrow=length(plot.x), ncol=length(plot.y))

# Krigagem com correcção  da anisotropia
kg_ph=krige.conv(ph_geo,locations = pred.loc,
                krige = krige.control(cov.model ='exponential',nugget =0,
    cov.pars =c(0.2725,36.25),
    aniso.pars = c(7*pi/18,36.25/16.93) ))

pred.kg=kg_ph$predict

# visualização das intgerpolações por Krigagem
plot.mean.kg= matrix(pred.kg, nrow=length(plot.x), ncol=length(plot.y))
plot.mean.kg[is.na(plot.index)] = NA
image.plot(x=plot.x,y=plot.y,z=plot.mean.kg,xlab="",ylab="")
contour(x=plot.x,y=plot.y,z=plot.mean.kg,xlab="",ylab="",add=TRUE, drawlabels=FALSE)
plot(sps,add=T)
```

Cont.

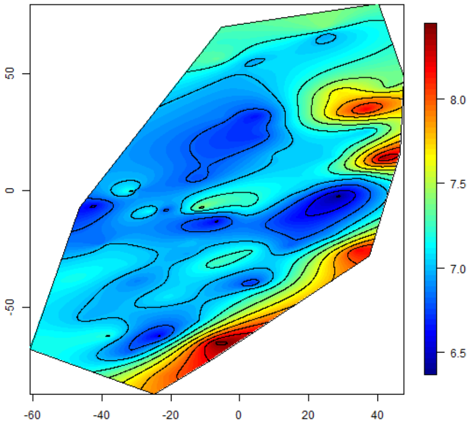


Figura: Distribuição espacial do pH. Valores interpolados com correcção da anisotropia geométrica

Krigagem simples

- Este método não difere tanto da Krigagem ordinária
- Assume que a média do processo é conhecida
- O processo de interpolação leva em consideração o conhecimento da média

$$Z^*(\mathsf{s}_0) = \sum_{i=1}^{n} \lambda_i Z(\mathsf{s}_i) + (1 - \sum_{i=1}^{n} \lambda_i)\mu$$

Cont.

Visto que a média é conhecida o cáculo dos pesos será feito apenas com base na função de covariância.

$$\boldsymbol{\lambda} = \mathsf{C}^{-1}\mathsf{c}_0,$$

onde

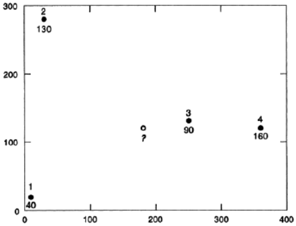
$$\boldsymbol{\lambda} = \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_n \end{bmatrix}, \mathsf{C} = \begin{bmatrix} C(\mathsf{s}_1 - \mathsf{s}_1) & \cdots & C(\mathsf{s}_1 - \mathsf{s}_n) \\ C(\mathsf{s}_2 - \mathsf{s}_1) & \cdots & C(\mathsf{s}_2 - \mathsf{s}_n) \\ \vdots & \ddots & \vdots \\ C(\mathsf{s}_n - \mathsf{s}_1) & \dots & C(\mathsf{s}_n - \mathsf{s}_n) \end{bmatrix}, \mathsf{c}_0 = \begin{bmatrix} C(\mathsf{s}_0 - \mathsf{s}_1) \\ C(\mathsf{s}_0 - \mathsf{s}_2) \\ \vdots \\ C(\mathsf{s}_0 - \mathsf{s}_n) \end{bmatrix}$$

A variância da estimativa para krigagem simples é dada por:

$$\sigma^2_{KS} = C(0) - \boldsymbol{\lambda}^T \mathsf{c}_0$$

Exemplo

Considere a figura (lado esquerdo) e a tabela abaixo (lado direito)



índice	X	Y	Medições
1	10	20	40
2	30	280	130
3	250	130	90
4	360	120	160

Assuma que o atributo em estudo tem uma média de 110 e uma função de covariância $C(h) = 2000 \exp\left(-\frac{h}{250}\right)$. Se o ponto a ser estimado é $x_0 = (180, 120)$. Calcule os peso e o valor do atributo para o ponto não amostrado.

Cont.

```
x=c(10,30,250,360,180)
> y=c(20,280,130,120,120)
> XY=as.matrix(cbind(x,y))
> dist=matrix(data=NA, nrow=5,ncol=5)
>
> for(i in 1:5){
+   for(j in 1:5){
+     dist[i,j]=sqrt((XY[i,1]-XY[j,1])^2+(XY[i,2]-XY[j,2])^2)
+   }
+ }
>
> Cov=matrix(data=NA, nrow=5,ncol=5)
> for(i in 1:5){
+   for(j in 1:5){
+     Cov[i,j]=2000*exp(-dist[i,j]/250)
+   }
+ }
>
> Cov1=Cov[1:4,1:4]
> Cov0=Cov[1:4,5]
> Cov0=t(Cov0)
> w=solve(Cov1)%*%t(Cov0)
> w
             [,1]
[1,]  0.184679065
[2,]  0.128482048
[3,]  0.645838236
[4,] -0.001128155
>
```

krigagem Universal

$\rightarrow$ A Krigagem Ordinária, assim como, a Krigagem simples assumem que a média do processo no campo aleatório é contaste.
$\rightarrow$ Na prática, campos ambientais e geológicos muitas vezes apresentam valores médios não constantes (a média do processo não é constante em todo espaço aleatório).

$$Z(\mathrm{x}) = \mu(\mathrm{s}) + \epsilon(\mathrm{s})$$

onde, μ(s) é uma função que depende da localização (s), e ϵ(s) é um processo estacionário de segunda ordem (com média zero).

krigagem Universal

A componente $\mu(\mathrm{s})$ caracteriza a tendência do processo (e designa-se por drift). Suponha que $\mu(\mathrm{s})$ pode ser representado como uma combinação linear de funções conhecidas $\left\{f_l(s), l = 1, \cdots, k\right\}$, com coeficientes desconhecidos $\{a_l\}$

$$\mu(\mathrm{s}) = \sum_{l=1}^{k} a_l f_l(s)$$

A média do processo bem como a covariância podem ser expressas da seguinte forma:

$$E\left[Z(s)\right] = \sum_{l=1}^{k} a_l f_l(s)$$

$$E\left\{[Z(s_1) - \mu(s_1)][(Z(s_2) - \mu(s_2)]\right\} = E\left[\epsilon(s_1)\epsilon(s_2)\right] = C(s_1 - s_2)$$

krigagem Universal

Tal como a Krigagem Ordinária, a Krigagem Universal também é um interpolador que resulta da combinação linear das observações circundantes ao local a ser estimado,

$$Z^*(s_0) = \sum_{i=1}^{n} \lambda_i Z(s_i)$$

onde λ_i é escolhido de tal maneira que o estimador seja não enviesado e erro de estimação seja mínimo. O estimador será não enviesado, se e somente se,

$E[Z^*(s_0)] = E[Z(s_0)]$, ou

$$\sum_{i=1}^{n} \lambda_i \mu(s_i) = \mu(s_0) \Rightarrow \sum_{l=0}^{k} a_l \sum_{i=1}^{n} \lambda_i f_l(s_i) = \sum_{l=0}^{k} a_l f_l(s_0)$$

O interpolador só será não-enviesado, se e somente se, $\sum_{i=1}^{n} \lambda_i f_l(s_i) = f_l(s_0)$.

krigagem Universal

A variância para este interpolador pode ser dada por:

$$\sigma_{KU}^2(x_0) = E\left\{[Z^*(s_0) - Z(s_0)]^2\right\} = \sum_{i=1}^{n}\sum_{j=1}^{n}\lambda_i\lambda_j C(s_i, s_j) - 2\sum_{i=1}^{n}\lambda_i C(s_i, s_0) + C(0) \quad (1)$$

Esta expressão deve ser minimizada sob a condição de não-enviesamento $\sum_{i=1}^{n}\lambda_i f_l(s_i) = f_l(s_0)$. isto pode ser feito usando multiplicador de Lagrange, que irá resultar no seguinte sistema de equações:

$$\begin{cases} \sum_{j=1}^{n}\lambda_j C(\mathrm{s}_i - \mathrm{s}_j) - \sum_{l=0}^{k}\alpha_l f_l(s_i) = C(\mathrm{s}_i - \mathrm{s}_0), & i = 1, 2, ..., n \\ \sum_{i=1}^{n}\lambda_i f_l(s_i) = f_l(s_0), & l = 0, 1, 2, \cdots, k \end{cases} \quad (2)$$

Krigagem Indicatriz

A krigagem indicatriz consiste na aplicação da krigagem ordinária numa variável transformada, isto é, uma variável resultante da aplicação da função não linear:

$$I(\mathrm{s}, z_T) = \begin{cases} 1, & \text{se} \quad Z(\mathrm{s}) \leq z_T \\ 0, & \text{se} \quad Z(\mathrm{s}) > z_T, \end{cases}$$

tal que

$$P(I(\mathrm{s}, z_T) = 1) = P(Z(\mathrm{s}) \leq z_T) = F_{Z(\mathrm{s})}(z_T)$$
$$P(I(\mathrm{s}, z_T) = 0) = P(Z(\mathrm{s}) > z_T) = 1 - F_{Z(\mathrm{s})}(z_T)$$

Cont.

Se assumirmos que a variável regionalizada em estudo é estacionária de segunda ordem, então, tem-se:

$$E(I(\mathrm{s}, z_T)) = F_{Z(\mathrm{s})}(z_T) \tag{3}$$

$$V(I(\mathrm{s}, z_T)) = F_{Z(\mathrm{s})}(z_T)(1 -_{Z(\mathrm{s})} (z_T)) \tag{4}$$

$$\begin{aligned} C_{z_T}(\mathrm{h}) &= E(I(\mathrm{s}, z_T)I(\mathrm{s}+\mathrm{h}, z_T)) - E(I(\mathrm{s}, z_T))E(I(\mathrm{s}+\mathrm{h}, z_T)) \\ &= P(I(\mathrm{s}, z_T)I(\mathrm{s}+\mathrm{h}, z_T) = 1) - P(I(\mathrm{s}, z_T) = 1)P(I(\mathrm{s}+\mathrm{h}, z_T) = 1) \\ &= F_{Z(\mathrm{s}),Z(\mathrm{s}+\mathrm{h})}(z_T) - (F_{Z(\mathrm{s})}(z_T))^2 \end{aligned} \tag{5}$$

Cont.

A relação entre o variograma e covariograma fica definido da seguinte forma

$$\begin{aligned} \gamma(\mathrm{h}) &= C_{z_T}(\mathbf{0}) - C_{z_T}(\mathrm{h}) \\ &= V(I(\mathrm{s}_0, z_T)) - C_{z_T}(\mathrm{h}, z_T) \\ &= F_{Z(\mathrm{s})}(z_T) - F_{Z(\mathrm{s}),Z(\mathrm{s}+\mathrm{h})}(z_T) \end{aligned} \tag{6}$$

A interpolação por krigagem ordinário para uma variável regionalizada indicadora é dada por :

$$I(\mathrm{s}_0, z_T) = \sum_{i=1}^{n} \lambda_i I(\mathrm{s}_i, z_T)$$

Os λ_i são obtidos seguindo o mesmo procedimento da krigagem ordinária

A krigagem indicatriz requer variograma da variável indicadora para cada teor de corte

Krigagem indicatriz - variancia da estimativa

$$V(I^*(\mathrm{s}, z_T)) = I^*(\mathrm{s}, z_T)(1 - I^*(\mathrm{s}, z_T))$$