LCE5700 - Geoestatística

Exercícios 3.5, 4.2, 5.2, 6.4 (DIGGLE e RIBEIRO JR, 2007)

Aluno: Everton Batista da Rocha N^o **USP:** 6686406 **Data:** 02/11/2011

Sorteio dos exercícios

```
> NRO <- 7725  ## Substitua este número por um formado pelo número de letras de cada parte do seu nome
> exercicios <- c(2.2, 2.3, 3.2, 3.4, 3.5, 4.1, 4.2, 4.4, 5.2, 5.3, 5.4, 6.4, 6.5, 7.3, 7.5, 7.7, 8.1, 8.5)
> set.seed(NRO)
> sample(exercicios, 4)
[1] 5.2 3.5 4.2 6.4
>
```

Resolução

3.5) (IZABELA E EVERTON) Consider a spatial process $S(.)$ defined by

$$S(x) = \int w(u)S^*(x-u)du$$

where $w(u) = (2\pi)^{-1}\exp(-||u||^2/2)$ and $S^*(.)$ is another stationary Gaussian process. Derive an expression for the correlation function, $\rho(u)$ say, of $S(.)$ in terms of $w(.)$ and the correlation function, $\rho^*(u)$ say, of $S^*(.)$. Give explicit expressions for $\rho(u)$ when $\rho^*(u)$ is of the form:
(a) pure nugget, $\rho^*(u) = 1$ if $u = 0$, zero otherwise;
(b) spherical;
(c) Gaussian.
(d) In each case, comment on the mean square continuity and diferentiability properties of the process $S(.)$ in relation to its corresponding $S^*(.)$.

De acordo com Diggle e Ribeiro Jr (2007), quando o suporte de cada medida estende-se por uma área ao invés de ser em um único ponto, o sinal deve ser representado por:

$$S(x) = \int w(r)S^*(x-r)dr,$$

em que $S^*(.)$ é um sinal não observado subjacente e $w(.)$ é uma função peso. Nesse caso, a forma de $w(.)$ limita a forma permitida para a função de covariância de $S(.)$. Se $\gamma(.)$ e $\gamma^*(.)$ são as funções de covariância de $S(.)$ e $S^*(.)$, segue que

$$\gamma(u) = \int\int w(r)w(s)\gamma^*(u+r-s)drds.$$

De acordo com as informações do exercício e a notação usada no mesmo, tem-se que a função de correlação $\rho(u)$ será dada por

$$\rho(u) = \int\int w(u)w(s)\rho^*(t+u-s)duds.$$

A função peso $w(.)$ de $S(x)$ é $w(u) = (2\pi)^{-1}\exp(-||u||^2/2)$ e, como foi dito que $S^*(.)$ também é processo gaussiano, tem-se que sua função peso $w(s)$ é a mesma que $w(u)$. A expressão $\rho^*(.)$ varia entre puro "nugget", esférica e gaussiana. As integrais explícitas para cada uma dessas situações foram obtidas no programa Maple 14.0 e são apresentadas no arquivo ex5_3_Maple.

4.2) (IZABELA E EVERTON) Obtain an expression for the variogram of a Poisson log-linear model in which measurements $Y_i : i = 1, \ldots, n$ at locations x_i are conditionally independent, Poisson-distributed with conditional expectations μ_i,where $\log \mu_i = \alpha + S(x_i) + Z_i$, $S(.)$ is a mean-square continuous stationary Gaussian process and $Z_i : i = 1, \ldots, n$ are mutually independent $\mathrm{N}(0, \tau^2)$. Compare your general result with the special case $\tau^2 = 0$ and comment.

Suponha que S(x) é um processo estacionário com média zero e variância σ^2, e que as observações Y_i, condicionais a $S(\cdot)$, são mutuamente independentes com esperança condicional $\mu_i = g(\alpha + S_i)$ e variância condicional $\upsilon_i = \upsilon(\mu_i)$, em que $S(\cdot)$ é a notação abreviada de $S(x_i)$ e $g(\cdot)$ é a inversa analítica da função de ligação, $h(\cdot)$. Sob estas condições o variograma é expresso por,

$$\gamma_Y(u) = E\left[\frac{1}{2}(Y_i - Y_j)^2\right] \tag{1}$$

em que $u = \|x_i - x_j\|$. Usando as propriedades de esperança, tem-se que a expressão em (1) pode ser expressa por:

$$\gamma_Y(u) = \frac{1}{2}\left(E_S\left[\{g(\alpha + S_i) - g(\alpha + S_j)\}^2\right]\right) + 2E_S\left[\upsilon(g(\alpha + S_i))\right] \tag{2}$$

O segundo termo do lado direito da igualdade em 2 é uma constante que escreve-se como $2\tau^2$. Isto mostra que τ^2 é obtido pela média da variância condicional da distribuição de $S(\cdot)$, que é análoga a variância *nugget* de um processo gaussiano estacionário. Para aproximar o primeiro termo do lado direito, considera-se uma aproximação de série de Taylor de primeira ordem $g(\alpha + S) \approx g(\alpha) + Sg^{'}(\alpha)$, que resulta em,

$$\gamma_Y(u) \approx g^{'}(\alpha)^2\gamma_S(u) + \tau^2. \tag{3}$$

Agora, considere que as observações $Y_i : \quad i = 1, \cdots, n$ nas localizações x_i são condicionalmente independentes, distribuídas de acordo com um modelo de Poisson com esperança condicional μ_i, em que $\log(\mu_i) = \alpha + S(x) + Z_i$, em que $S(\cdot)$ é uma média quadrática de um processo gaussiano e $Z_i : \quad i = 1, \cdots, n$ são mutuamente independentes $N(0, \tau^2)$. Para este caso, note que a função de ligação $h(\cdot)$ é dada por $h(\cdot) = \log(\mu_i) = \alpha + S(x) + Z_i$, assim,

$$\begin{aligned} g(\cdot) &= h(\cdot)^{-1} \Rightarrow \\ g(\cdot) &= \log(\mu_i)^{-1}. \end{aligned}$$

Desta forma,

$$\begin{aligned} g^{'}(\cdot) &= -\frac{1}{\mu_i \log(\mu_i)^2} \Rightarrow \\ g^{'}(\cdot)^2 &= \frac{1}{\mu_i^2 \log(\mu_i)^4}. \end{aligned}$$

Então,

$$\begin{aligned} \gamma_Y(u) &\approx g^{'}(\alpha)^2 \gamma_S(u) + \tau^2 \Rightarrow \\ \gamma_Y(u) &\approx \frac{1}{\mu_i^2 \log(\mu_i)^4} \gamma_S(u) + \tau^2. \end{aligned}$$

Mas,

$$\begin{aligned} \log(\mu_i) &= \alpha + S(x) + Z_i \Rightarrow \\ \mu_i &= \exp\{\alpha + S(x) + Z_i\} \end{aligned}$$

logo,

$$\begin{aligned} \gamma_Y(u) &\approx \frac{1}{\exp\{\alpha + S(x) + Z_i\}^2 \log(\exp\{\alpha + S(x) + Z_i\})^4} \gamma_S(u) + \tau^2 \Rightarrow \\ \gamma_Y(u) &\approx \frac{1}{\exp\{\alpha + S(x) + Z_i\}^2 \left[\log(\exp\{\alpha + S(x) + Z_i\})\right]^4} \gamma_S(u) + \tau^2 \Rightarrow \\ \gamma_Y(u) &\approx -\frac{1}{\exp\{\alpha + S(x) + Z_i\}^2 (\alpha + S(x) + Z_i)^4} \gamma_S(u) + \tau^2. \end{aligned}$$

Note que quando $\tau^2 \neq 0$ a interpretação é dada em termos da variância *nugget*, de forma aditiva, o que pode ser inadequado, uma vez que claramente nota-se que para este modelo há uma incorporação não-linear da relação entre Y e $S(x)$. Ao se considerar $\tau^2 = 0$ nota-se que o variograma de Y_i necessariamente depende dos momentos de alta ordem dos processo latente $S(x)$.

5.2) (IZABELA e EVERTON) Write a programme to simulate data from a stationary Gaussian model with exponential covariance function, $\rho(u) = \sigma^2 \exp(-u/\phi)$. Apply the method of maximum likelihood estimation to replicate simulations of this model, and investigate the joint sampling distribution of σ^2 and ϕ. Compare this with the joint sampling distribution for Zhang's re-parameterisation to $\theta_1 = \log(\sigma^2/\phi)$ and $\theta_2 = \log(\phi)$.

```
#Everton Batista da Rocha
#Izabela Regina Cardoso de Oliveira


#----------------------------------------------------------------------------
# Carregando os pacotes
#----------------------------------------------------------------------------
require(MASS)
require(geoR)
rm(list=ls())

#----------------------------------------------------------------------------
```

```
# Parâmetros reais
#-------------------------------------------------------------------------------

phi    <- 0.1
sigma2 <- 1
tau2   <- 1
mu     <- 0
theta1 <- log(sigma2/phi)
theta2 <- log(sqrt(sigma2))
np     <- 125

#-------------------------------------------------------------------------------
# Coordenadas, matriz de distância e funções de covariância
#-------------------------------------------------------------------------------
set.seed(341134)
X    <- cbind(runif(np),runif(np))                              # Coordenadas
D    <- as.matrix(dist(X, diag=T, upper=T))                     # Matriz de distâncias
Cov  <- sigma2 * exp(- D/phi)    + diag(sqrt(tau2), nrow(X))      # Função de covariância exponenencial
Zhg  <- theta2 * exp(- D/theta1) + diag(sqrt(tau2), nrow(X))      # Função de covariância com a reparametrização de Zhang


#-------------------------------------------------------------------------------
# Simulações
#-------------------------------------------------------------------------------
ns   <- 100
sims <- mu + t(chol(Cov)) %*% matrix(rnorm(nrow(X)*ns), ncol=ns)  # dados das simulações
simZ <- mu + t(chol(Zhg)) %*% matrix(rnorm(nrow(X)*ns), ncol=ns)

#-------------------------------------------------------------------------------
# Ajustes
#-------------------------------------------------------------------------------
resS <- matrix(0,ns,2)  # coluna 1: sigma2 estimado; coluna2: phi estimado
resZ <- matrix(0,ns,2)
i    <- 0



for(i in (i+1):ns){
resS[i,]<-likfit(as.geodata(cbind(X,sims[,i])),ini=c(1,0.1),messages=F)$cov.pars
modZ    <-likfit(as.geodata(cbind(X,simZ[,i])),ini=c(1,0.1),messages=F)
resZ[i,]<-cbind(exp(modZ$cov.pars[1])*exp(modZ$cov.pars[2]),exp(modZ$cov.pars[2]))
}



hist(resS)
hist(resZ)
```

6.4) (EVERTON E ELISÂNGELA) Consider a stationary trans-Gaussian model with known transformation function $h(\cdot)$, let x be an arbitrary location within the study region and define $T = h^{-1}\{S(x)\}$. Find explicit expressions for $P(T > c|Y)$ where $Y = (Y_1, \cdots, Y_n)$ denotes the observed measurements on the untransformed scale and:
(a) $h(u) = u$;
(b) $h(u) = \log(u)$;
(c) $h(u) = sqrt(u)$.
Solução:

Assim como em outras áreas da estatística, há pelo menos três motivos diferentes para se utilizar transformação de dados. Em primeiro lugar, uma transformação particular pode ser sugerida por argumentos qualitativos ou mesmo por convenção. Em segundo lugar, a transformação pode ser usada como um dispositivo de estabilização da variância de uma distribuição amostral conhecida não gaussiana. Em terceiro e último lugar, pode-se introduzir uma família paramétrica de transformações simplismente como a generalização empírica do modelo gaussiano, o que é o considerado neste exercício. Considere um modelo estacionário trans-Gaussiano com observações $Y_i \quad : i = 1, \cdots, n$, $T = S(x)$ é o valor do *signal*, nas localizações x_i. Considere ainda uma transformação conhecida $h(\cdot)$ e defina então $T = h^{-1}\{S(x)\}$.

(a)
Sendo,

$$h(u) = u \quad \Rightarrow \quad h^{-1}(u) = \frac{1}{u} = u^{-1},$$

tem-se que,

$$T(x) = S^{-1}(x).$$

Segue então que $P(T > c|Y)$ é expressa por,

$$P(T > c|Y) = \int_c^{\infty} S^{-1}(x|y)dx.$$

em que $S(\cdot)$ é um processo estacionário trans-gaussiano.

(b)
De modo análogo ao anterior, considerando,

$$h(u) = \log(u) \quad \Rightarrow \quad h^{-1}(u) = \exp(u),$$

segue que,

$$T(x) = \exp\{S(x)\}$$

Daí tem-se que $P(T > c|Y)$ é expressa por,

$$P(T > c|Y) = \int_c^{\infty} \exp\{S(x|y)\}dx.$$

em que $S(\cdot)$ é um processo estacionário trans-gaussiano.

(c)
Sendo,

$$h(u) = \sqrt{u} \quad \Rightarrow \quad h^{-1}(u) = u^2,$$

tem-se que,

$$T(x) = \{S(x)\}^2.$$

Segue então que $P(T > c|Y)$ é expressa por,

$$P(T > c|Y) = \int_c^{\infty} \{S(x|y)\}^2 dx.$$

em que $S(\cdot)$ é um processo estacionário trans-gaussiano.

Referências

[1] DIGGLE, P. J.; RIBEIRO JR, P. J., *Model-based Geostatistics*, Londres: Springer Series in Statistics, 2007, 230 p.

[2] GELFAND, A. E.; DIGGLE, P. J.; FUENTES, M.; GUTTORP, P., *Handbook of Spatial Statistics*, Boca Raton: Chapman & Hall/CRC handbooks of modern statistical methods. 2010, 575 p.

[3] R Development Core Team (2011). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org/.

[4] RIBEIRO JR, P. J.; DIGGLE, P. J. , geoR: A package for geoestatistical analysis. *R-NEWS*, v. 1, n.2, p. 15-18, 2001