Geoestatística - Exercícios

Natalia da Silva Martins[1], Paulo Ribeiro Justiniano Jr.[2]

1 Exercícios

Exercício 2.2

Consider the following two models for a set of responses, $Y_i : i = 1, ..., n$ associated with a sequence of positions $x_i : i = 1, ..., n$ along a onedimensional spatial axis x.

(a) $Y_i = \alpha + \beta x_i + Z_i$, where α and β are parameters and the Z_i are mutually independent with mean zero and variance σ_Z^2.

Calculando a esperança de $Y_i = \alpha + \beta x_i + Z_i$

$$\begin{aligned} E(Y_i) &= E(\alpha + \beta x_i + Z_i) \\ &= E(\alpha) + E(\beta x_i) + E(Z_i) \\ &= \alpha + \beta x_i + 0 \\ &= \alpha + \beta x_i \end{aligned} \tag{1}$$

$$\begin{aligned} V(Y_i) &= V(\alpha + \beta x_i + Z_i) \\ &= V(\alpha) + V(\beta x_i) + V(Z_i) \\ &= V(\alpha) + (x_i)^2 V(\beta) + V(Z_i) \\ &= 0 + (x_1^2)0 + \sigma_z^2 \\ &= \sigma_z^2 \end{aligned} \tag{2}$$

[1]ESALQ-USP, naty-martins@hotmail.com
[2]ESALQ-USP, paulojus@ufpr.br

(b) $Y_i = A + Bx_i + Z_i$ where the Z_i are as in (a) but A and B are now random variables, independent of each other and of the Z_i, each with mean zero and respective variances σ_A^2 e σ_B^2.

Calculando a esperança de $Y_i = A + Bx_i + Z_i$

$$\begin{aligned} E(Y_i) &= E(A + Bx_i + Z_i) \\ &= E(A) + E(Bx_i) + E(Z_i) \\ &= 0 + 0 + 0 \\ &= 0 \end{aligned} \tag{3}$$

$$\begin{aligned} V(Y_i) &= V(A + Bx_i + Z_i) \\ &= V(A) + V(Bx_i) + V(Z_i) \\ &= \sigma_A^2 + (x_i)^2\sigma_B^2 + \sigma_z^2 \end{aligned} \tag{4}$$

For each of these models, find the mean and variance of Y_i and the covariance between Y_i and Y_j for any $j \neq i$. Given a single realisation of either model, would it be possible to distinguish between them?

Seria sim possível distinguir os modelos, uma vez que estes tem diferentes medidas de posição (média) e diferentes medidas de dispersão (variância).

Exercício 7.7

Reproduce the simulated binomial data shown in Figure 4.6. Use geoRglm in conjunction with priors of your choice to obtain predictive distributions for the signal S(x) S at locations x = $(0.6, 0.6)$ and x = $(0.9, 0.5)$.

A programação encontra-se no Anexo.

Exercício 7.7

Compare the predictive inferences which you obtained in Exercise 7.6 with those obtained by fitting a linear Gaussian model to the empirical logit transformed data, $log(y + 0.5)/(n - y + 0.5)$

A programação encontra-se no Anexo.

Por meio da Figura 1 pode-se observar os dados originais e os dados trasformados.

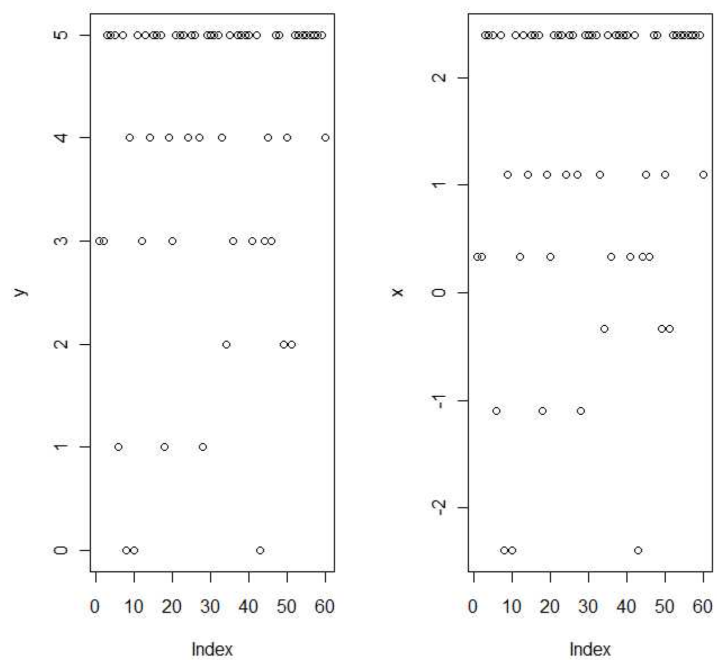

Figura 1: Dados originais e dados transformados, respectivamente.

Exercício 8.1

Consider a stationary Gaussian process in one spatial dimension, in which the design consists of n equally spaced locations along the unit interval with $x_i = (.1{+}2i)/(2n)$: $i = 1, ..., n$. Suppose that the process has unknown mean μ but known variance $\sigma^2 = 1$ and correlation function $\rho(u) = exp(\frac{-u}{\phi})$ with known $\phi = 0.2$. Investigate, using simulation if necessary, the impact of n on the efficiency of the maximum likelihood estimator for μ. Does the variance of $\hat{\mu}$ approach zero in the limit as $n \to \infty$? If not, why not?

Observa-se que quanto maior o número de simulações mais a estimativa de máxima verossimilhança da média se aproxima do verdadeiro valor, logo sua variação é menor, como é apresentado na Tabela 1.

Tabela 1: Resultado da estimativa da média para diferentes valores de n.

tamanho do n	Média
50	0,4240
100	0,5752
150	0,6001
200	0,5332
300	0,5417
500	0,4849
700	0,4960
1000	0,5102

Ainda para a visualização da convergência da variância temos uma simulação das médias no histograma, como pode ser observado na Figura 2.

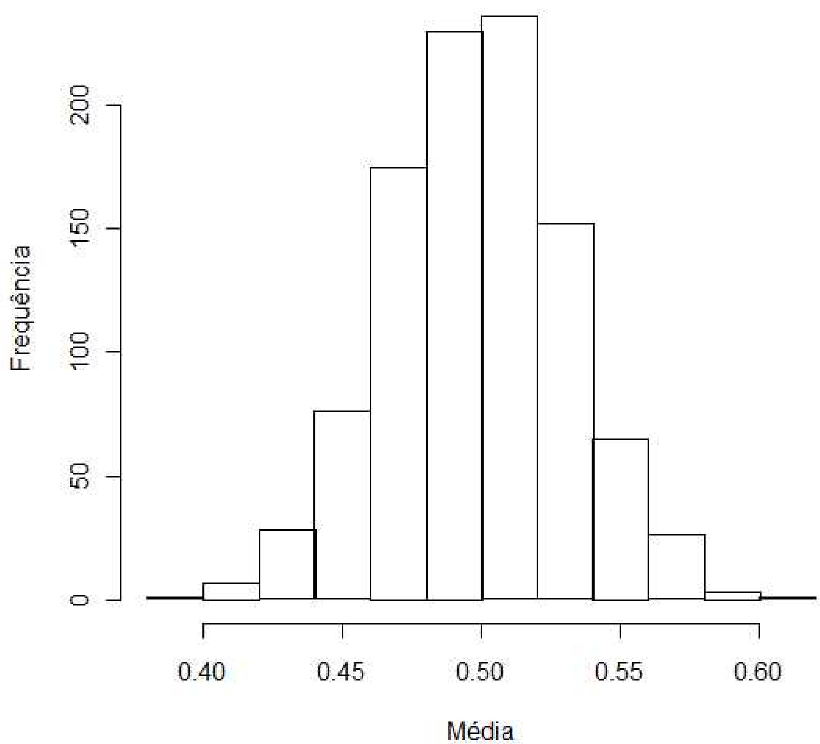

Figura 2: Histograma da distribuição das médias.

Exercício 8.5

An existing design on the unit square A consists of four locations, one at each corner of A. Suppose that the underlying model is a stationary Gaussian process with mean μ, signal variance σ^2, correlation function $\rho(u) = exp(\frac{-u}{\phi})$ and nugget variance τ^2. Suppose also that the objective is to add a fifth location, x, to the design in order to predict the spatial average of the signal process $S(x)$ with the smallest possible prediction mean square error, assuming that the model parameter values are known.

(a) Guess the optimal location for the fifth point.

(b) Suppose that we use the naive predictor $\overline{y}$. Compare the mean square prediction errors for the original four-point and the augmented five-point design.

(c) Repeat, but using the simple kriging predictor.

ANEXO

```
# Exercício 8.1
require(geoR)
sigma2<- 1
phi<-0.2
mi=0.5
n<-1000

X<-seq(n/2,(n^2-n/2),n)
Y<-rep(0,length(X))
M<-cbind(X,Y)
View(M)

set.seed(171)
dados<- grf(nrow(M),grid=M,cov.pars = c(sigma2,phi)
            ,nsim=1000, cov.model="exp", mean=mi)

# Abaixo comandos para usar método da máxima verossimilhança

ml1 <- likfit(dados, ini=c(sigma2, phi), cov.model="exp")
```

```
ml1$beta

# Comandos para visualizar a variação da média por histograma
hist(apply(dados$data,2,mean), xlab="Média", ylab="Frequência")

#exercicio 7.6
#simulando os dados da Figura 4.6
require(geoR)
n <- 5
set.seed(23) #semente para a simulação
s <- grf(60, cov.pars = c(5, 0.25))
p <- exp(2 + s$data)/(1 + exp(2 + s$data))
#simulando dados da binomial
y <- rbinom(length(p), size = 5, prob = p)
points(s)
text(s$coords, label = y, pos = 3, offset = 0.3)

#exercicio 7.7
#transformar os dados em log{(y + 0.5)/(n - y + 0.5)}
x <- log ((y + 0.5)/(n - y + 0.5))
plot(x)
par(mfrow=c(1,2))
plot(y)
plot(x)
```