Métodos de estimação para modelo CAR

Fernando G. Moro

Universidade Federal do Paraná

8 de junho de 2014

1. Estatística espacial e dados de área;

1. Estatística espacial e dados de área;
2. Vizinhança e matriz de vizinhança;

1. Estatística espacial e dados de área;
2. Vizinhança e matriz de vizinhança;
3. Dados;

1. Estatística espacial e dados de área;
2. Vizinhança e matriz de vizinhança;
3. Dados;
4. O modelo CAR;

1. Estatística espacial e dados de área;
2. Vizinhança e matriz de vizinhança;
3. Dados;
4. O modelo CAR;
5. Estimação por MV;

1. Estatística espacial e dados de área;
2. Vizinhança e matriz de vizinhança;
3. Dados;
4. O modelo CAR;
5. Estimação por MV;
6. MCMC da versão bayesiana do modelo CAR;

1. Estatística espacial e dados de área;
2. Vizinhança e matriz de vizinhança;
3. Dados;
4. O modelo CAR;
5. Estimação por MV;
6. MCMC da versão bayesiana do modelo CAR;
7. Versão INLA;

1. Estatística espacial e dados de área;
2. Vizinhança e matriz de vizinhança;
3. Dados;
4. O modelo CAR;
5. Estimação por MV;
6. MCMC da versão bayesiana do modelo CAR;
7. Versão INLA;

Estatística espacial e dados de área

Estatística espacial: É um ramo da estatística que analisa dados distribuídos em uma área. Onde geralmente procura-se identificar padrões neste espaço que possam explicar parte da variabilidade dos dados através de um modelo estatístico.

Dados de área: Ela é um caso particular da estatística espacial. Onde dentro da área abrangente existem pequenas subáreas, sendo uma observação em cada uma.

Vizinhança e matriz de vizinhança

Vizinhança: É um agrupamento das subáreas de uma forma que seja condizente com o contexto do problema e/ou explique a disperção dos dados no espaço.

Matriz de vizinhança: É uma matriz de n por n, onde n é o número de subáreas, e seja a forma mais simples dada por:

$$w_{ij} = \begin{cases} 1 & \text{se i vizinho de j} \\ 0 & \text{se i é não vizinho de j} \end{cases} \tag{1}$$

para i,j$= 1, 2, ..., n$

Vizinhança e matriz de vizinhança

```
> require(sp)
> require(maptools)
> PRmun <- readShapePoly("C:/Users/PR/munpr.shp")
> plot(PRmun)
```

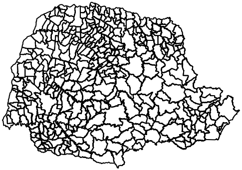


Figura: Polígono das 399 cidades do Paraná

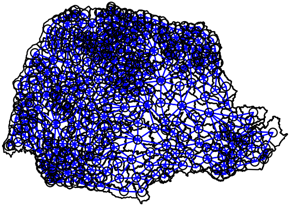


Figura: Grafo sobreposto ao polígono

Vizinhança e matriz de vizinhança

```
> require(spdep)
> Estneig<-poly2nb(PRmun)
> estbin<-nb2listw(Estneig,style='B')
> summary(estbin)
```

```
Characteristics of weights list object:
Neighbour list object:
Number of regions: 399
Number of nonzero links: 2230
Percentage nonzero weights: 1.400745
Average number of links: 5.588972
Link number distribution:

 2  3  4  5  6  7  8  9 10 11
 7 40 71 83 85 60 26 14 10  3
7 least connected regions:
0 103 134 231 293 316 398 with 2 links
3 most connected regions:
158 345 385 with 11 links

Weights style: B
Weights constants summary:
    n     nn   S0   S1    S2
B 399 159201 2230 4460 55040
```

Número médio de filhos por mulher

```
> filhos <- read.table("C:/Users//mfilhos.txt",
+                      sep=";", dec=",", quote="\"",h=T)
> head(filhos[,c(1,3)])

          Localidade X2010
1             Abatiá  2.24
2        Adrianópolis  2.37
3       Agudos do Sul  2.14
4 Almirante Tamandaré  2.13
5  Altamira do Paraná  2.38
6        Alto Paraíso  2.14

>
>
```

Transformação Box-Cox:

```
> require(MASS)
> x<-filhos$X2010
> par(cex.lab=0.7,cex.axis=0.7)
> Vero<-boxcox(x~1,plotit=T)
> title('Verossimilhança Box-Cox')
```

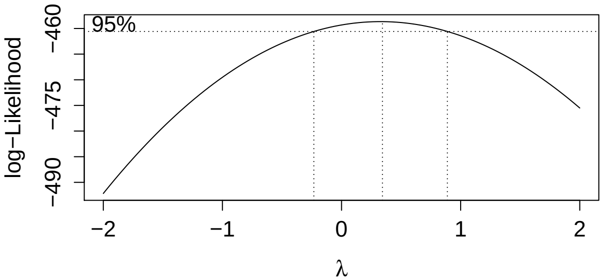
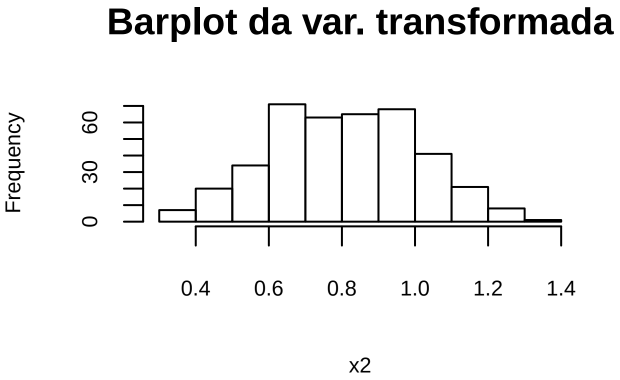


Figura: Gráfico do Box-Cox

Dados

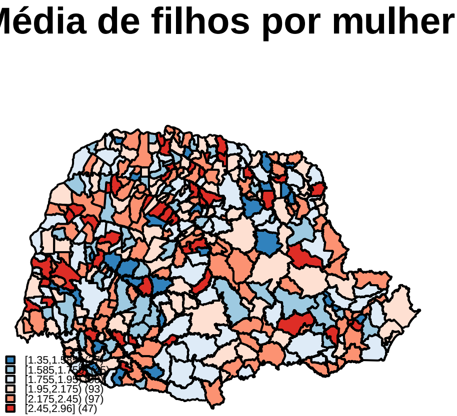
```
> x2<-((x^0.35)-1)/0.35
> par(cex.lab=0.7,cex.axis=0.7)
> hist(x2,main='Barplot da var. transformada')
>
```



Comandos para gráfico com legenda de cores

```
> require(classInt)
> require(RColorBrewer)
> require(maptools)
> require(rgeos)
> pal_raw <- c(rev(brewer.pal(3, "Blues")), brewer.pal(3, "Reds"))
> rkmeans7_BEL <- classIntervals(round(x,2), style="kmeans",n=6)
> par(cex.main=1,1,cex=1.2)
> plot(PRmun, col=findColours(rkmeans7_BEL, pal_raw),main='')
> title(main="Média de filhos por mulher")
> cols <- findColours(rkmeans7_BEL, pal_raw)
> table <- attr(cols, "table")
> legtext <- paste(names(table), " (", table, ")", sep="")
> legend("bottomleft", fill=attr(cols, "palette"), legend=legtext, cex=0.3,
+        y.inter=0.7,bty='n')
>
```

Gráfico com legenda de cores:



O modelo CAR

Seja Z o vetor de observações, então a forma marginal é dada por:

$$Z \sim NM(\mathbf{X}\beta, \Sigma_{car}) \quad (2)$$

E seja e Σ_{car} dados por:

$$\Sigma_{car} = (\mathbf{I} - \rho\mathbf{W})^{-1}\sigma^2\mathbf{I}$$

(3)

O modelo CAR:

Seja Z_i uma observação qualquer e Z_{-i} o conjunto de todas as observações fora i, então a distribuição condicional é da forma:

$$Z_i|Z_{-i} \sim N(\mu_i + \sum_j W_{ij}(Z_{-i} - \mu_{-i}), \tau^2)$$

(4)

Observação:Diferentemente da distribuição marginal, a distribuição condicional do modelo é independente.

Estimação por MV

Log verossimilhança do modelo marginal:

$$-\frac{n}{2} log(2\Pi\sigma^2) + \frac{1}{2} \left|\mathbf{I} - \rho\mathbf{W}\right| - \frac{1}{2\sigma^2}(\mathbf{Z} - \mathbf{X}\beta)'(\mathbf{I} - \rho\mathbf{W})(\mathbf{Z} - \mathbf{X}\beta) \tag{5}$$

Derivando em relação a ρ, a σ^2, e ao vetor $\boldsymbol{\beta}$ e igualando a zero, encontramos os estimadores de Máxima verossimilhança do modelo CAR.

Estimação por MV

Rodando o modelo:

```
> Model1<-spautolm(x2~1,listw=estbin,family='CAR')
> names(Model1)
```

```
 [1] "fit"          "lambda"      "LL"         "LL0"        "call"
 [6] "parameters"   "aliased"     "method"     "family"     "zero.policy"
[11] "weights"      "interval"    "trs"        "timings"    "LLNullLlm"
[16] "fdHess"       "lambda.se"   "X"          "Y"          "weights"
```

```
>
```

Saída do modelo:

```
> options(width=60)
> summary.spautolm(Model1)

Call:
spautolm(formula = x2 ~ 1, listw = estbin, family = "CAR")

Residuals:
        Min          1Q      Median          3Q         Max
-0.49446370 -0.13852100  0.00035551  0.14651915  0.50908468

Coefficients:
             Estimate Std. Error z value  Pr(>|z|)
(Intercept) 0.8057949  0.0088107  91.456 < 2.2e-16

Lambda: -0.05405 LR test value: 2.9168 p-value: 0.087661
Numerical Hessian standard error of lambda: 0.03198

Log likelihood: 72.82389
ML residual variance (sigma squared): 0.040331, (sigma: 0.20082)
Number of observations: 399
Number of parameters estimated: 3
AIC: -139.65
```

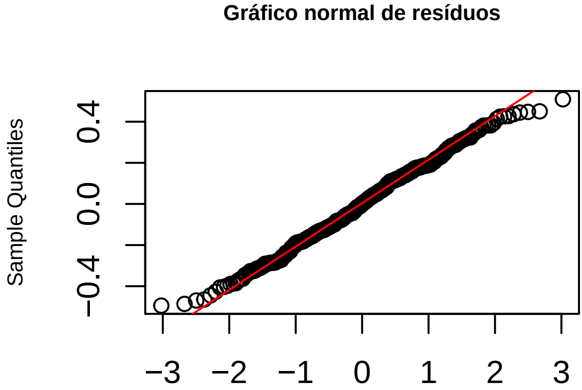
Valores preditos:

```
> head(Model1$fit$fitted)

        1         2         3         4         5         6
0.7916669 0.8051600 0.8097974 0.7575523 0.8080695 0.8208979
```

Gráfico de resíduos:

```
> res<-Model1$fit$residuals
> par(cex.main=0.7,cex.lab=0.7)
> qqnorm(res,main='Gráfico normal de resíduos')
> qqline(res,col='red')
```



Relembramos primeiramente o modelo:

$$Z \sim NM(\mathbf{X}\beta, (\mathbf{I} - \rho\mathbf{W})^{-1}\sigma^2 I) \tag{6}$$

Sejam as prioris dadas por:

$$[\beta] \sim N(0, 1000) \tag{7}$$

$$[\rho] \sim U[-1, 1] \tag{8}$$

$$[\sigma] \sim Gamma(0.001, 0.001) \tag{9}$$

Declarando o modelo:

```
> N<-399
> W<-nb2mat(Estneig,style='B')
> I<-diag(1,399)
> cat('model{
+     for(i in 1:N){
+       m[i]<-eps
+       for(j in 1:N){
+         V[i,j]<- (I[i,j] - rho*W[i,j])*tau*I[i,j]
+     }}
+     z ~ dmnorm(m[ ],V[ , ])
+     rho ~ dunif(-1,1)
+     eps ~ dnorm(0,0.001)
+     tau<-pow(sigma,-2)
+     sigma ~ dgamma(0.001,0.001)
+ } ',file='CAR.model')
```

Sumário das estimativas dos parâmetros:

```
1. Empirical mean and standard deviation for each variable,
   plus standard error of the mean:

           Mean       SD  Naive SE Time-series SE
eps    0.806790 0.010260 1.026e-04      1.026e-04
rho   -0.009246 0.578309 5.783e-03      7.144e-03
sigma  0.203032 0.007239 7.239e-05      9.234e-05

2. Quantiles for each variable:

         2.5%     25%      50%    75%  97.5%
eps    0.7868  0.7999  0.80667 0.8138 0.8274
rho   -0.9526 -0.5085 -0.01485 0.4932 0.9506
sigma  0.1894  0.1980  0.20280 0.2078 0.2176
```

Gráficos das posterioris:

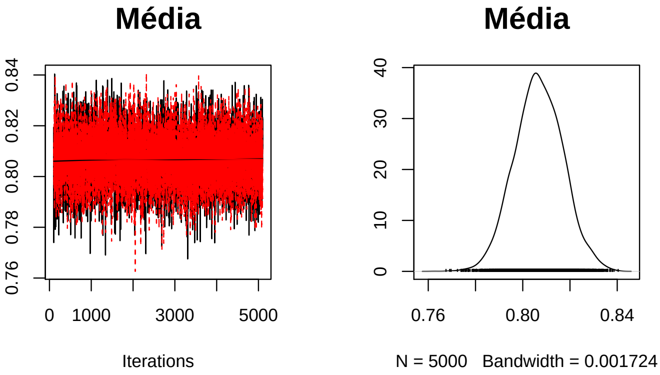


Figura: Densidade da posteriori da Média

Gráficos das posterioris:

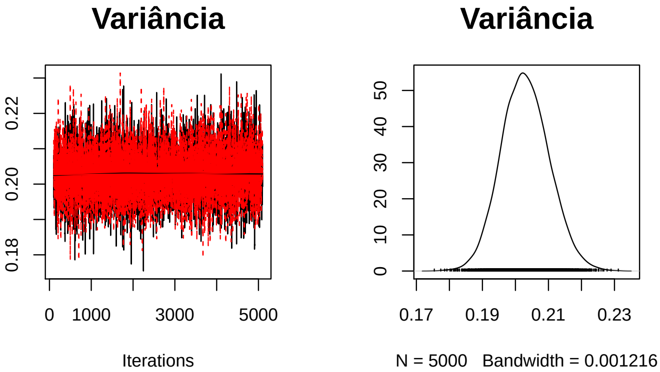


Figura: Densidade da posteriori da Variância

Gráficos das posterioris:

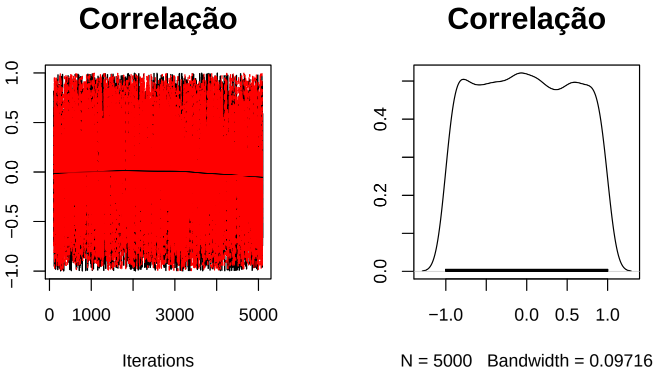


Figura: Densidade da posteriori da Correlação

Intervalo HPD:

```
[[1]]
           lower     upper
eps    0.7862131 0.8268225
rho   -0.9887493 0.9999985
sigma  0.1891857 0.2169789
attr(,"Probability")
[1] 0.95
```

Valores preditos:

```
1. Empirical mean and standard deviation for each variable,
   plus standard error of the mean:
         Mean SD Naive SE Time-series SE
z[1] 0.9318  0        0              0
z[2] 1.0074  0        0              0
z[3] 0.8717  0        0              0
z[4] 0.8656  0        0              0
z[5] 1.0131  0        0              0
z[6] 0.8717  0        0              0

2. Quantiles for each variable:
         2.5%    25%    50%    75%  97.5%
z[1] 0.9318 0.9318 0.9318 0.9318 0.9318
z[2] 1.0074 1.0074 1.0074 1.0074 1.0074
z[3] 0.8717 0.8717 0.8717 0.8717 0.8717
z[4] 0.8656 0.8656 0.8656 0.8656 0.8656
z[5] 1.0131 1.0131 1.0131 1.0131 1.0131
z[6] 0.8717 0.8717 0.8717 0.8717 0.8717
```

Modelando com o INLA:

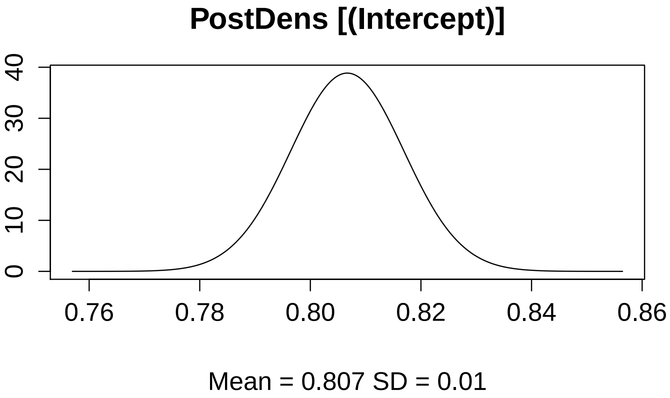
```
> require(INLA)
> N<-399
> data<-data.frame(x2,W,N)
> nb2INLA('ex.inla.adj',Estneig)
> data$ID<-1:399
> modelo = x2 ~ 1 + f(data$ID,model='besagproper',
+                     graph='ex.inla.adj')
> Model3 <- inla(modelo, data=data, family="gaussian",Ntrials=N,
+                control.predictor=list(compute=TRUE))
```

Sumário dos efeitos fixos:

```
> Model3$summary.fixed
```

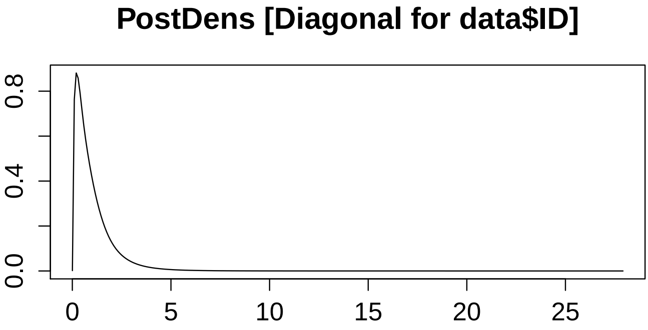
```
                 mean         sd 0.025quant  0.5quant
(Intercept) 0.8066776 0.01029232  0.7864562 0.8066773
            0.975quant      mode          kld
(Intercept)  0.8268805 0.8066776 2.620132e-13
```

Densidade posteriori da média:



Versão INLA

Densidade a posteriori de σ^2:



Sumário dos valores preditos:

> *head(Model3$summary.fitted.values[1])*

```
                          mean
fitted.predictor.001 0.8086180
fitted.predictor.002 0.8083125
fitted.predictor.003 0.8068430
fitted.predictor.004 0.8081304
fitted.predictor.005 0.8073888
fitted.predictor.006 0.8066468
```

Comportamento dos preditos: