Trabalho 2

Rodrigo Malta Esteves

10/03/2021

1. a)
2. b)
3. c)
4. d)
5. e)

```
library(readr)
library(dplyr)
```

```
atend <- read.csv("C:\\Users\\malta\\Desktop\\Pós Graduação\\Introdução ao Aprendizado Estatístico\\Tra
```

Como descrito na introdução, temos que $X1$ é a variável aleatória do número de clientes atendidos por hora e θ é a média do número de clientes atendidos. Essas variáveis seguem as distribuições de probabilidade: $X1 \sim Poisson(\theta)$ e $\theta \sim Gamma(\alpha, \lambda)$. Além disso, temos $n = 100$ registros de contagem de atendimentos.

a)

Assumimos uma *priori 1* em que nosso parâmetro $\theta \sim Gamma(0.001, 0.001)$.

Calculando as médidas de variância, desvio padrão e média desse parâmetro, podemos observar que as medidas de dispersão são muito maiores que a média. Logo, é possível afirmar que, nesse caso, θ tem uma distribuição vaga a priori.

```
#priori 1
theta1 <- rgamma(100000,.001,.001)
theta1 <- as.data.frame(theta1)
theta1 %>%
  summarize(MediaPriori = mean(theta1),
            VarianciaPriori = var(theta1),
            DesvioPriori = sd(theta1))
```

```
##   MediaPriori VarianciaPriori DesvioPriori
## 1   0.9637529        913.9191     30.23109
```

b)

O analista informa que a média de atendimento gira em torno de 5 por hora. Como temos que o parâmetro θ tem distribuição *gama*, sabemos que $E[\theta] = \frac{\alpha}{\lambda}$. Logo, podemos descrever uma relação entre os parâmetro: $\frac{\alpha}{\lambda} = 5 \implies \alpha = 5\lambda$. Também é informado que a $Prob(\theta > 10) \approx 5\%$.

Associando essas duas informações é possível encontrar uma aproximação para os parâmetro da *priori 2*. Para isso, criamos uma grade de valores de 0 até 5 para λ, uma variável α, respeitando a relação descrita anteriormente, e geramos uma matriz com as probabilidades associadas às distribuições *gama* com esses parâmetros.

Observamos que os parâmetros que mais se aproximam dos 5% de probabilidade são $\alpha = 3.5$ e $\lambda = 0.7$.

```r
#Elicitação dos par^ametros da priori 2
lambda<-seq(0,5,0.1)
alpha<-5*lambda
matriz<-cbind(alpha,lambda,1-pgamma(10,shape=alpha,rate=lambda))
matriz #alpha = 3.5 e lambda = 0.7
```

```
##       alpha lambda
##  [1,]   0.0    0.0 1.000000e+00
##  [2,]   0.5    0.1 1.572992e-01
##  [3,]   1.0    0.2 1.353353e-01
##  [4,]   1.5    0.3 1.116102e-01
##  [5,]   2.0    0.4 9.157819e-02
##  [6,]   2.5    0.5 7.523525e-02
##  [7,]   3.0    0.6 6.196880e-02
##  [8,]   3.5    0.7 5.118135e-02
##  [9,]   4.0    0.8 4.238011e-02
## [10,]   4.5    0.9 3.517354e-02
## [11,]   5.0    1.0 2.925269e-02
## [12,]   5.5    1.1 2.437324e-02
## [13,]   6.0    1.2 2.034103e-02
## [14,]   6.5    1.3 1.700084e-02
## [15,]   7.0    1.4 1.422792e-02
## [16,]   7.5    1.5 1.192150e-02
## [17,]   8.0    1.6 9.999781e-03
## [18,]   8.5    1.7 8.396123e-03
## [19,]   9.0    1.8 7.056009e-03
## [20,]   9.5    1.9 5.934709e-03
## [21,]  10.0    2.0 4.995412e-03
## [22,]  10.5    2.1 4.207748e-03
## [23,]  11.0    2.2 3.546600e-03
## [24,]  11.5    2.3 2.991155e-03
## [25,]  12.0    2.4 2.524130e-03
## [26,]  12.5    2.5 2.131152e-03
## [27,]  13.0    2.6 1.800249e-03
## [28,]  13.5    2.7 1.521432e-03
## [29,]  14.0    2.8 1.286361e-03
## [30,]  14.5    2.9 1.088058e-03
## [31,]  15.0    3.0 9.206824e-04
## [32,]  15.5    3.1 7.793409e-04
## [33,]  16.0    3.2 6.599276e-04
## [34,]  16.5    3.3 5.589956e-04
## [35,]  17.0    3.4 4.736489e-04
## [36,]  17.5    3.5 4.014524e-04
## [37,]  18.0    3.6 3.403570e-04
## [38,]  18.5    3.7 2.886376e-04
## [39,]  19.0    3.8 2.448403e-04
## [40,]  19.5    3.9 2.077400e-04
## [41,]  20.0    4.0 1.763029e-04
## [42,]  20.5    4.1 1.496569e-04
## [43,]  21.0    4.2 1.270656e-04
## [44,]  21.5    4.3 1.079069e-04
## [45,]  22.0    4.4 9.165515e-05
## [46,]  22.5    4.5 7.786593e-05
```

```
## [47,]  23.0    4.6 6.616341e-05
## [48,]  23.5    4.7 5.622960e-05
## [49,]  24.0    4.8 4.779540e-05
## [50,]  24.5    4.9 4.063296e-05
## [51,]  25.0    5.0 3.454931e-05
```

c)

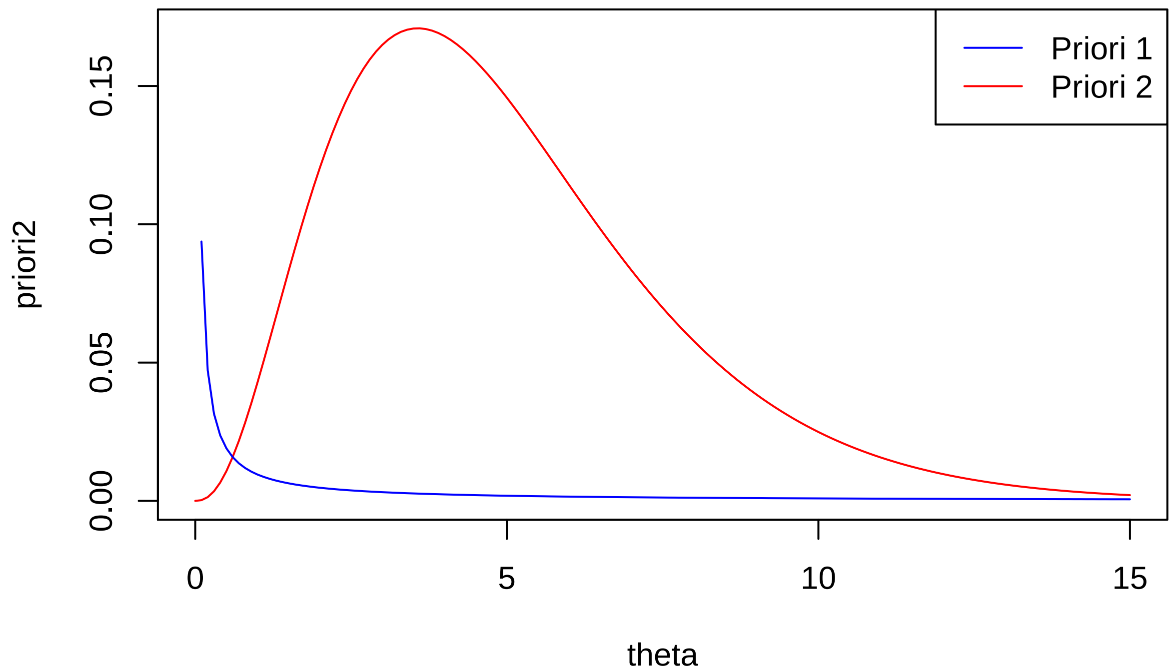
Comparando as distribuições das duas prioris, podemos ver claramente como a distribuição a *priori 1* reflete uma falta de conhecimento sobre o parâmetro ao ter quase todos os seus valores próximos ao zero para toda grade de valores de θ. Já a *priori 2* explica muito mais sobre o parâmetro, uma vez que atribui probabilidades de forma mais heterogênea entre os valores de θ.

```r
#Esperança e Variância da Priori 2
#Teórico
MediaTeoPriori2 <- 3.5/.7 #média teórica
VarianciaTeoPriori2 <- 3.5/(.7^2) #variância teórica

#Amostral
theta2 <-rgamma(100000,3.5,0.7)
theta2 <- as.data.frame(theta2)
theta2 %>%
  summarize(MediaPriori2 = mean(theta2),
            VarianciaPriori2 = var(theta2),
            MediaTeoPriori2,
            VarianciaTeoPriori2)
```

```
##   MediaPriori2 VarianciaPriori2 MediaTeoPriori2 VarianciaTeoPriori2
## 1     5.014959         7.230478               5            7.142857
```

```r
#Gráfico das prioris para uma grade de valores de theta entre 0 e 15
theta<-seq(0,15,0.1)
priori1<-dgamma(theta,shape=0.01, rate=0.01)
priori2<-dgamma(theta,shape=3.5, rate=0.7)
plot(theta, priori2,t="l",col="red") #Priori 2 muito mais informativa
lines(theta, priori1,col="blue")
legend("topright",legend=c("Priori 1","Priori 2"),lty=c(1,1),col=c(4,2))
```



d)

Função de verossimilhança para uma Poisson:

$l(\theta; \underline{x1}) = \prod_{i=1}^{n} \frac{e^{-\theta}\theta^{(x1)_i}}{(x1)!} = \frac{e^{-n\theta}\theta^{\sum (x1)_i}}{\prod (x1_i!)}$.

Estimador de máxima verossimilhança para θ:

$L(\theta; \underline{x1}) = log(l(\theta; X1)) = \sum (x1)_i log(\theta) - n\theta - log(\prod (x1)_i)$.

$\frac{dL(\theta;\underline{x1})}{d\theta} = \frac{\sum (x1)_i}{\theta} - n = 0$.

$\hat{\theta} = \frac{\sum (x1)_i}{n}$

As distribuições apresentam um resultado esperado. A amostra grande tende a concentrar seus valores mais próximo da máxima verossímilhança, enquanto a amostra pequena apresenta uma dispersão maior.

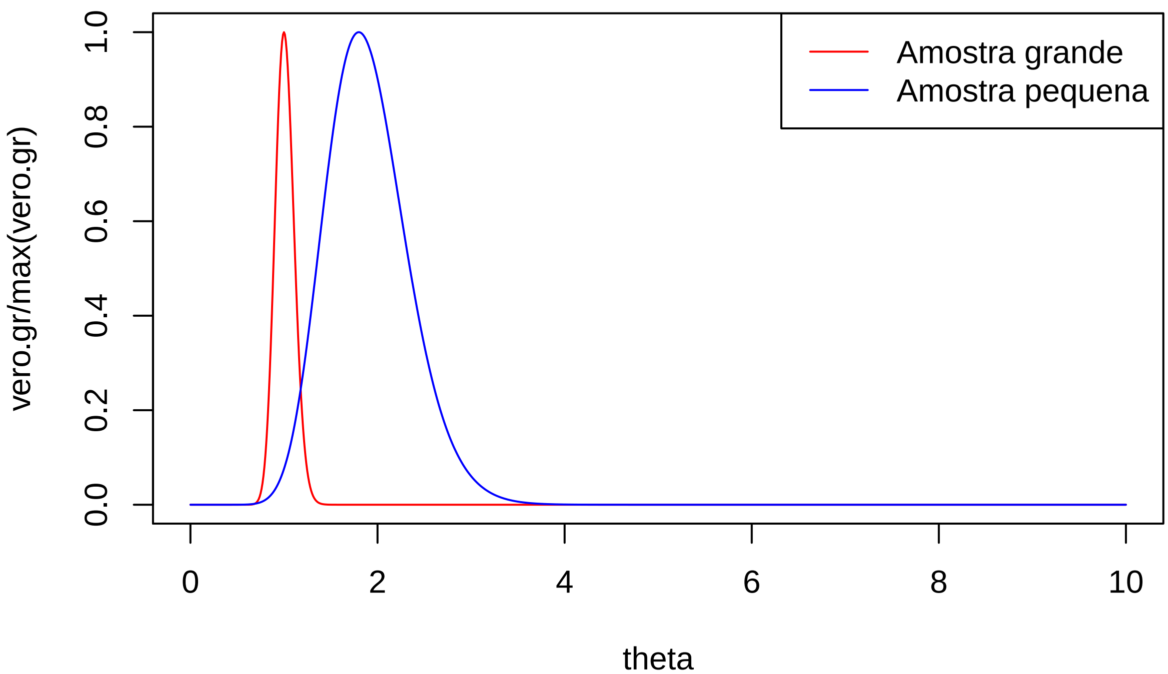
```
#set.seed(1234)
#amostras
amostra.gr<- atend
amostra.pq<-sample(amostra.gr$X1,10)

#parâmetros
t1 <- sum(amostra.gr$X1)
n1 <- length(amostra.gr$X1)
t2 <- sum(amostra.pq)
```

```
n2 <- length(amostra.pq)

#estimadores de máxima verossimilhança
mv1 <- t1/n1
mv2 <- t2/n2

theta<-seq(0,10,0.01)
vero.gr<-exp(-n1*theta)*theta^t1
vero.pq<-exp(-n2*theta)*theta^t2
plot(theta,vero.gr/max(vero.gr),t="l",col="red")
lines(theta,vero.pq/max(vero.pq),col="blue")
legend("topright",legend=c("Amostra grande","Amostra pequena"),lty=c(1,1),col=c(2,4))
```



e)

Sabemos que se $Y_1, ..., Y_n$ independentes e identicamente distribuidos seguem uma distribuição $Poisson(\theta)$ e $\theta \sim Gama(\alpha, \beta)$, então, a posteriori, temos que $\theta|\underline{y} \sim Gama(\alpha + \sum y_i, \beta + n)$.

Observamos aqui 4 posterioris possíveis (consideramos $t1$ como o somatório dos valores da amostra grande, $t2$ como o somatório dos valores da amostra pequena,n_1 o tamanho da amostra grande e n_2 o tamanho da amostra pequena):

a) Para priori 1 e amostra pequena: $\theta|\underline{y} \sim Gama(0.001 + t2, 0.001 + n_2)$
b) Para priori 2 e amostra pequena: $\theta|\underline{y} \sim Gama(3.5 + t2, 0.7 + n_2)$

c) Para priori 1 e amostra grande: $\theta|\underline{y} \sim Gama(0.001 + t1, 0.001 + n_1)$

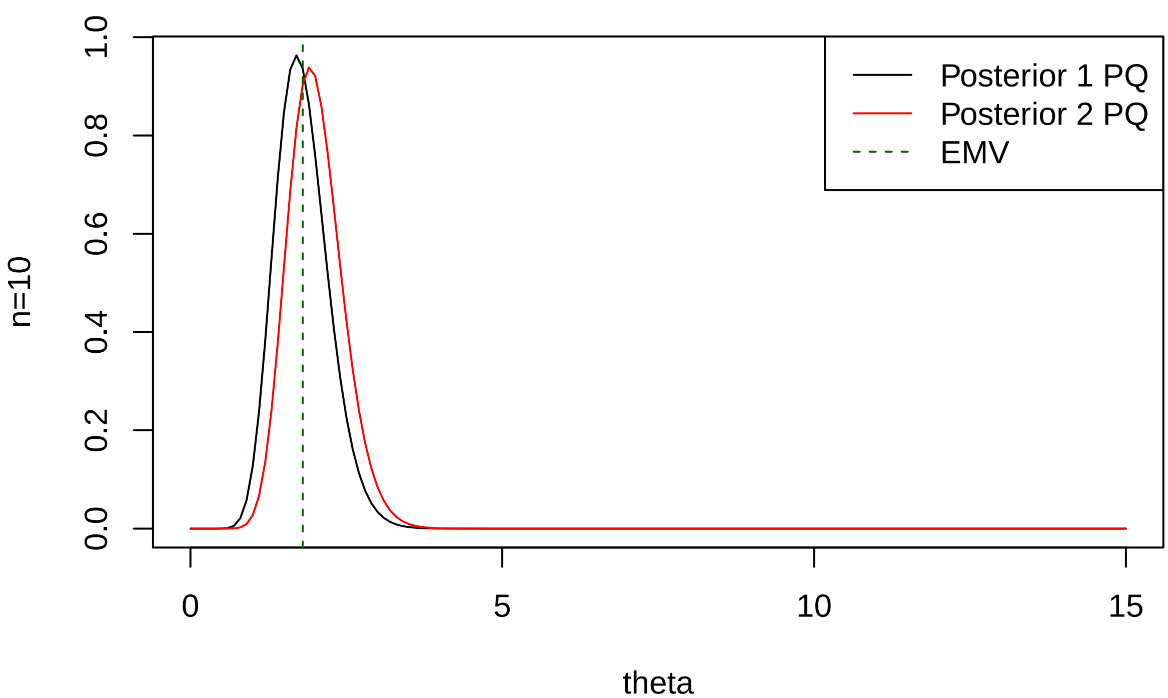
d) Para priori 2 e amostra grande: $\theta|\underline{y} \sim Gama(3.5 + t1, 0.7 + n_1)$

Podemos perceber que para as posterioris com a uma amostra grande a informação vinda dos dados, representada pela verossimilhança, dominará a distribuição. Essa característica fica bem clara dada a concentração de ambas as posterioris em torno do estimador de máxima verossimilhança.

Para as posterioris com uma amostra pequena, observamos uma dispersão maior para os valores de θ. Percemos também que como a priori 1 é pouco informativa a sua posteriori será muito semelhante à distribuição dos dados, ou seja, a própria curva de verossimilhança. Já a segunda posteriori terá os seus valores levados um pouco mais para a direita, uma vez que a priori 2 é mais informativa e tem uma média maior.

```
theta<-seq(0,15,0.1)
poster.1.pq <- dgamma(theta,0.001+t2, 0.001+n2)
poster.2.pq <- dgamma(theta,3.5+t2, 0.7 + n2)
poster.1.gr <- dgamma(theta,0.001+t1, 0.001+n1)
poster.2.gr <- dgamma(theta,3.5+t1, 0.7 + n1)

##Preparação de uma janela gráfica particionada em duas linhas e duas colunas, para acomodar os 4 gráfi
#par(mfrow=c(1,2))
#Posteriori para theta sob amostra pequena e priori 1:
plot(theta,poster.1.pq,type="l",ylab="n=10")
#Posteriori para theta sob amostra pequena e priori 2:
lines(theta,poster.2.pq,type="l",ylab="n=10",col=2)
abline(v=mv2,lty=2,col="darkgreen")
legend("topright",legend=c("Posterior 1 PQ","Posterior 2 PQ","EMV"),lty=c(1,1,2),col=c(1,2,"darkgreen"))
```



```r
#Posteriori para theta sob amostra grande e priori 1:
plot(theta,poster.1.gr,type="l",ylab="n=100")
#Posteriori para theta sob amostra grande e priori 2:
lines(theta,poster.2.gr,type="l",ylab="n=100",col=2)
abline(v=mv1,lty=2,col="darkgreen")
legend("topright",legend=c("Posterior 1 GR","Posterior 2 GR","EMV"),lty=c(1,1,2),col=c(1,2,"darkgreen"))
```