

# Atividade 02 - Introdução ao Aprendizado Estatístico

**Marcel Dantas de Quintela**

---

Atividade apresentada como parte das avaliações da disciplina de Introdução ao Aprendizado Estatístico, ministrada pela Profa. *Mariane Barros Alves* para o curso de Especialização em Ciência e Dados do Instituto de Matemática da Universidade Federal do Rio de Janeiro

---

## Instruções

Uma empresa de suporte técnico está fazendo um estudo sobre a duração de atendimentos a clientes e, para tal fim, registra o número de clientes atendidos, em intervalos com uma hora de duração. Uma amostra aleatória de contagens de atendimentos foi tomada sobre um longo período de tempo, com observações registradas em horas não consecutivas, de tal forma que se pode assumir independência entre as observações. O arquivo *atendimentos.csv* guarda 100 registros dessas contagens. Assuma que as contagens sigam distribuição Poisson( $\theta$ ) e que se adote um enfoque bayesiano para o processo inferencial, com descrição a priori sobre  $\theta$  dada por uma distribuição Gama( $\alpha; \lambda$ ).

---

$Y_1, \dots, Y_n \sim \text{Poisson}(\theta)$  e assume-se, a priori,  $\theta \sim \text{Gama}(\alpha, \lambda)$ , então, a posteriori, tem-se:  
 $(\theta | y_1, \dots, y_n) \sim \text{Gama}(\alpha + t, \lambda + n), t = \sum_{i=1}^n y_i$

---

## Priori 1

- a) Assuma, a princípio, que  $\alpha = \lambda = 0,001$ . Passaremos a denominar essa especificação de *priori 1*. Determine os seguintes sumários sobre a especificação subjetiva inicial sobre  $\theta$  (*priori 1*): valor esperado ( $E[\theta]$ ) e desvio-padrão ( $DP[\theta]$ ). A especificação a priori é bastante informativa ou é vaga? Justifique.

---

```
# informações da priori 1
a<-1<-0.001
```

Considerando tais especificações subjetivas a cerca de  $\theta$ , obtém-se as seguintes estimativas:

- $E(\theta) = \frac{\alpha}{\lambda} = 1$
- $DP(\theta) = \sqrt{\frac{\alpha}{\lambda^2}} = 31.6227766$

Estas informações, a priori, podem ser considerada pouco informativa ou vaga. Uma vez que a distribuição Gama com parâmetros  $\alpha$  e  $\lambda$  muito pequenos (tendendo a zero) nos leva a uma curva com decaimento extremamente rápido e cauda praticamente constante, expressando ausência de conhecimento prévio.

Assim, pode-se afirmar que a *Priori 1*,  $[Gama(0.001, 0.001)]$  contribui com pouca ou nenhuma informação nas estimativas de  $\theta$ .

---

### *Priori 2*

- b) Assuma que um segundo analista creia, a priori, que, em média, haja em torno de 5 atendimentos por hora. Perguntado sobre a probabilidade de que a média de atendimentos por hora seja superior a 10, ele responde que acredita que essa probabilidade seja próxima de 5%. Especifique valores para os parâmetros  $\alpha$ , e  $\lambda$ , de forma a traduzir essa informação a priori. Passamos a nos referir a essa segunda especificação como *priori 2*. As linhas de código abaixo podem ajudar nessa tarefa.
- 

```
#Elicitacao dos parametros da priori 2
lambda<-seq(0,5,0.1)
alpha<-5*lambda
matriz<-cbind(alpha,lambda,P=1-pgamma(10,shape=alpha,rate=lambda))

#probabilidades próximas a 0.05 entre [0.046,0.059]
matriz[round(matriz[,3],4)%in%seq(0.0460,0.0590,0.0001),]

##      alpha      lambda      P
## 3.50000000 0.70000000 0.05118135
```

Considerando uma probabilidade de cerca de 0.05 para uma média de mais de 10 atendimentos por hora, podemos assumir a *Priori 2* como  $\theta \sim Gama(\alpha = 3.5, \lambda = 0.7)$ .

Parâmetros obtidos por meio da probabilidade que mais se aproxima de  $P(\theta > 10) \approx 0.05$  obtido por testes sequenciais para  $\lambda [0 : 5, \text{ by } 0.1]$  e  $\alpha = 5 \times \lambda$  (média de 5 atendimentos por hora).

---

- c) Determine valor esperado ( $E[\theta]$ ) e desvio-padrão ( $DP[\theta]$ ) de  $\theta$  sob a *priori 2* e esboce, em um mesmo gráfico, as duas densidades a priori (use as linhas de código abaixo). A *priori 2* é mais vaga ou mais informativa que a *priori 1*?
-

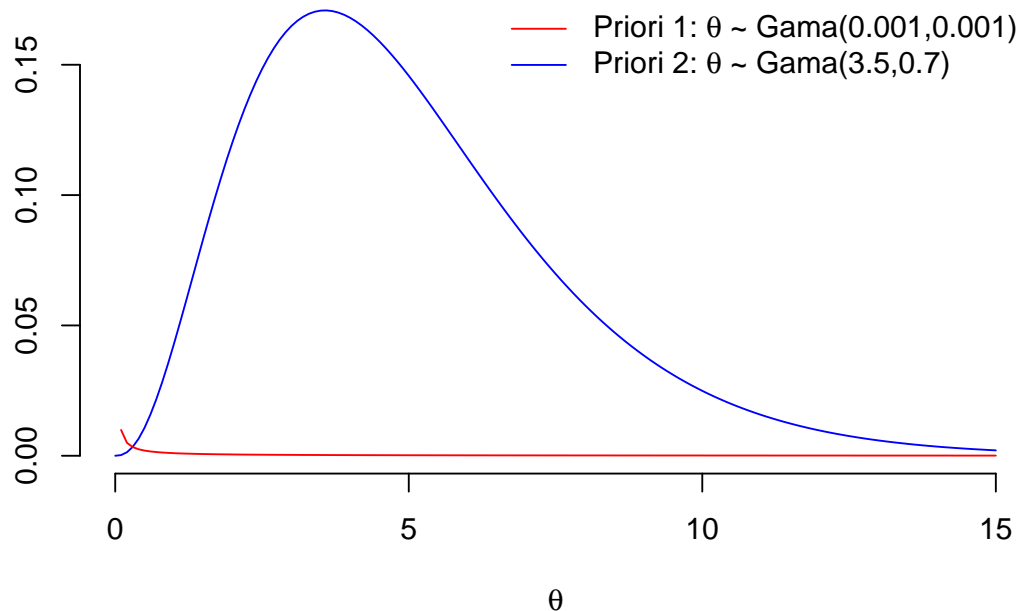
```
# informações da priori 2
l<-0.7; a<-5*l
```

Considerando as informações da *Priori 2*, temos:

- $E(\theta) = \frac{\alpha}{\lambda} = 5$
- $DP(\theta) = \sqrt{\frac{\alpha}{\lambda^2}} = 2.6726124$

```
#Grafico das prioris para uma grade de valores de theta entre 0 e 15
theta<-seq(0,15,0.1)
priori1<-dgamma(theta,shape=0.001, rate=0.001)
#prioriJ<-dgamma(theta,shape=1/2, rate=0.001) #Priori não informativa de Jeffreys
priori2<-dgamma(theta,shape=a, rate=l)

plot(theta, priori2, type = "l",lty=1, col="blue",
      ylab="", xlab=expression(theta), bty = "n")
lines(theta, priori1, col="red")
#lines(theta, prioriJ, col="green", lty=2)
legend("topright",
      legend=c(
        expression(paste("Priori 1: ",theta," ~ Gama(0.001,0.001)")),
        expression(paste("Priori 2: ",theta," ~ Gama(3.5,0.7)")))
      ,lty=c(1,1),col=c("red","blue"),bty = "n")
```



A *Priori 2* se concentra em torno de  $E(\theta) = 3.5$ , se espalhando lentamente para esquerda, ampliando assim uma região de maior probabilidades possíveis para  $\theta$ . Sendo assim, mais informativa que a *Priori 1*.

---

Agora, vamos estabelecer dois cenários de disponibilidade de dados: no primeiro cenário, assumamos amostra grande, composta pelas 100 contagens de atendimentos disponíveis no arquivo *atendimento.csv*. Para a amostra pequena, tome um subconjunto aleatório contendo 10 dessas observações:

```
set.seed(100)
amostra.gr<-scan(file="atendimentos.csv")
amostra.pq<-sample(amostra.gr,10)
```

---

Resumo das medidas amostrais:

- Média:
    - Amostra grande: 1
    - Amostra pequena: 1.6
  - Desvio Padrão:
    - Amostra grande: 1.0730867
    - Amostra pequena: 1.3498971
- 

*Verossimilhança*

- d) Esboce, em um mesmo gráfico, a função de verossimilhança, sob as duas amostras (ou seja, sob o cenário de dados abundantes e escassos). Comente o comportamento da função de verossimilhança sob as duas amostras. (Especifique  $n_1, n_2, t_1$  e  $t_2$  adequadamente e use o código abaixo).
- 

```
theta<-seq(0,5,0.01)

n1<-length(amostra.gr)
n2<-length(amostra.pq)
t1<-sum(amostra.gr)
t2<-sum(amostra.pq)
```

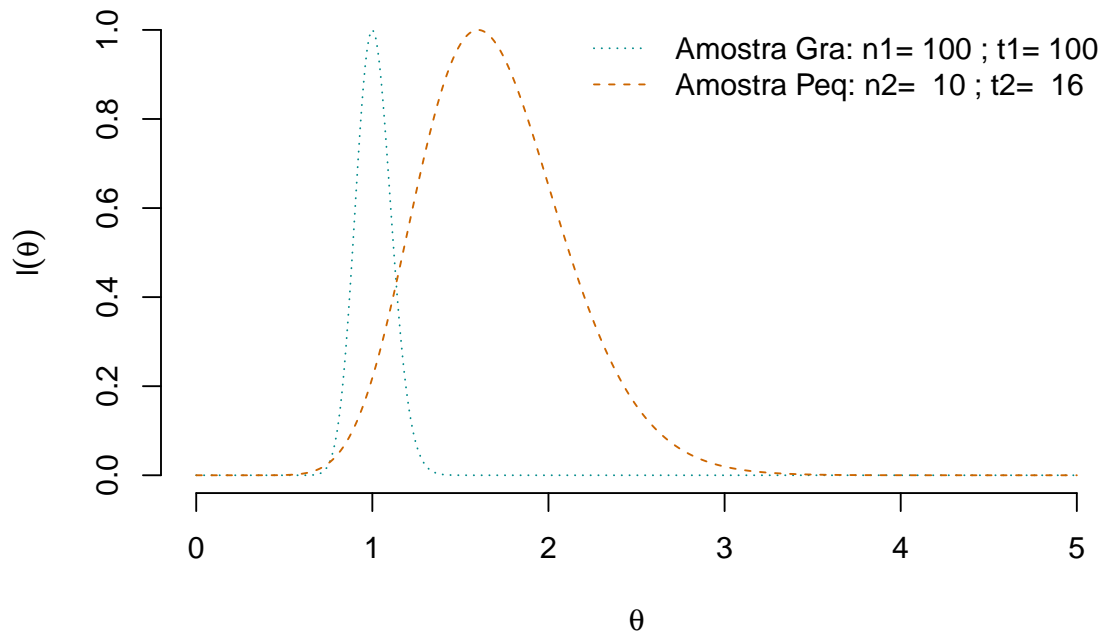
```

vero.gr<-exp(-n1*theta)*theta^t1
vero.pq<-exp(-n2*theta)*theta^t2

plot(theta,vero.gr/max(vero.gr),lty=3, col="cyan4",type="l",bty = "n",
      ylab = expression(l(theta)),
      xlab=expression(theta))
lines(theta,vero.pq/max(vero.pq), col="darkorange3", lty=2)

legend("topright",
      legend=c(paste("Amostra Gra: n1=",n1, "; t1=",t1),
                paste("Amostra Peq: n2= ",n2, "; t2= ",t2)),
      lty=c(3,2),col=c("cyan4","darkorange3"),bty = "n")

```



As funções de verossimilhança resumizam os dados coletados, trazendo informações de quais são os possíveis valores do parâmetro  $\theta$ .

O gráfico acima, mostra que quanto maior o número de informações disponíveis na amostra, maior será o caráter elucidativo da curva de verossimilhança. Ou seja, quanto maior o tamanho da amostra, menor será a variação da função de verossimilhança em torno do parâmetro investigado.

## Posteriori

e) Passaremos agora a estudar o comportamento da distribuição a posteriori sob as duas especificações a priori, considerando os dois cenários: amostra grande e amostra pequena. Obtenha a distribuição a posteriori:

- sob priori 1 e amostra pequena;
- sob priori 2 e amostra pequena;
- sob priori 1 e amostra grande;
- sob priori 2 e amostra grande.

Esboce gráficos das distribuições a posteriori obtidas sob as quatro especificações acima e teça comentários sobre a variação em seu comportamento. (O código R abaixo pode ajudar na confecção dos gráficos. Cabe a você definir quem é a curva a posteriori, em cada caso).

```
# grade de thetas para gerar posteriors
theta<-seq(0,5,0.01)

# posteriors
poster.1.pq<-dgamma(theta,shape=0.001+t2, rate=0.001+n2)
poster.2.pq<-dgamma(theta,shape= 3.5+t2, rate= 0.7+n2)
poster.1.gr<-dgamma(theta,shape=0.001+t1, rate=0.001+n1)
poster.2.gr<-dgamma(theta,shape= 3.5+t1, rate= 0.7+n1)

# poster.J.pq<-dgamma(theta,shape= 0.5+t2, rate=0.001+n2)
# poster.J.gr<-dgamma(theta,shape= 0.5+t1, rate=0.001+n1)

par(mfrow=c(2,1), mar=c(2,6,2,6))

#Posteriori para theta sob amostra pequena e priori 1:
plot(theta,poster.1.pq,type="l",ylab="n=10", xlab=expression(theta),
      col="red",bty="n")
#Posteriori para theta sob amostra pequena e priori 2:
lines(theta,poster.2.pq,type="l",ylab="n=10",col="blue", lty=1)

#Verossimilhança n=10
# lines(theta,vero.pq/max(vero.pq), col="green", lty=2)
# #Posteriori para theta sob amostra pequena e priori não informativade Jeffreys:
# lines(theta,poster.J.pq,type="l",ylab="n=10",col="green", lty=2)

legend("topright",cex=.6,
      legend=c(
        expression(paste("Post 1: ",theta," ~ Gama(0.001+16, 0.001+10)")),
        expression(paste("Post 2: ",theta," ~ Gama(3.5+16,0.7+10)")),
        lty=c(1,1,2),col=c("red","blue", "green"),bty = "n")

#Posteriori para theta sob amostra grande e priori 1:
plot(theta,poster.1.gr,type="l",ylab="n=100",xlab=expression(theta),
```

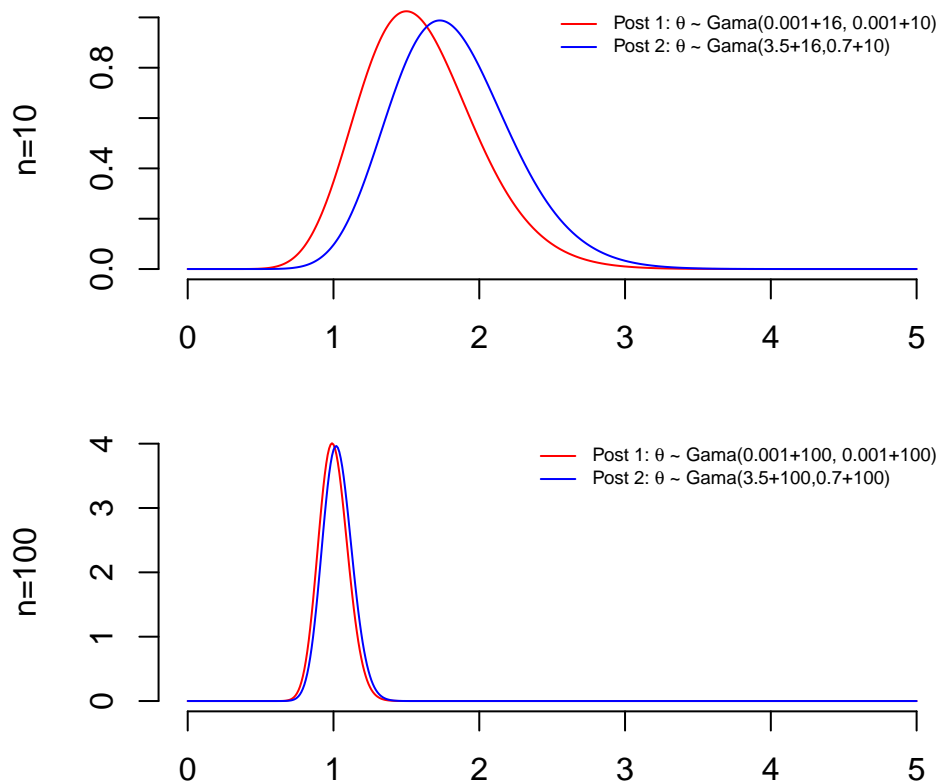
```

col="red",bty="n")
#Posteriori para theta sob amostra grande e priori 2:
lines(theta,poster.2.gr,type="l",ylab="n=100",col="blue", lty=1)

#Verossimilhança n=10
# lines(theta,vero.gr/max(vero.gr), col="green", lty=2)
# #Posteriori para theta sob amostra Grande e priori não informativa de Jeffreys:
# lines(theta,poster.J.gr,type="l",ylab="n=10",col="green", lty=2)

legend("topright",cex=.6,
      legend=c(
        expression(paste("Post 1: ",theta," ~ Gama(0.001+100, 0.001+10)")),
        expression(paste("Post 2: ",theta," ~ Gama(3.5+100,0.7+10)"))),
      lty=c(1,1,2),col=c("red","blue","green"),bty = "n")

```



É possível notar o impacto das informações a priori à luz das funções de verossimilhança considerando os diferentes tamanhos de amostra.

Quando existe escassez de informações, a função de verossimilhança carrega maior incerteza, ampliando o intervalo de possibilidades para o parâmetro investigado, incorrendo em posteriores carregadas de crenças viesadas ou não-informativas.

Considerando a amostra pequena ( $n = 10$ ), observa-se na *Posteriori 1* predominância de informações vindas da função de verossimilhança. Isso ocorre pela não informação, acima constatada, da *Priori 1*. Na *Posteriori 2*, nota-se um leve efeito da *Priori 2*, tendo um sutil deslocamento da curva para a direção dos parâmetros informados previamente.

Quando a amostra aumenta ( $n = 100$ ), a função de verossimilhança incide praticamente em toda a posteriori, reduzindo a quase nada a influência das informações a priori.