

# Exemplos biparamétricos: comparação de duas distribuições de Poisson

Prof. Caio Azevedo

## Exemplo 4

- Considere  $X_i|\lambda_1 \stackrel{i.i.d.}{\sim} \text{Poisson}(\lambda_1), i = 1, \dots, n$  e  $Y_j|\lambda_2 \stackrel{i.i.d.}{\sim} \text{Poisson}(\lambda_2), j = 1, \dots, n$ .
- Além disso, considere que  $X_i|\boldsymbol{\lambda} \perp Y_j|\boldsymbol{\lambda} \forall i, j, \boldsymbol{\lambda} = (\lambda_1, \lambda_2)$ .
- Como comparar, via inferência bayesiana,  $\lambda_1$  e  $\lambda_2$ ?
- Devido as suposições de independência condicional, temos que

$$\begin{aligned} p(\mathbf{x}, \mathbf{y} | (\lambda_1, \lambda_2)) &= p(\mathbf{x} | (\lambda_1, \lambda_2)) p(\mathbf{y} | (\lambda_1, \lambda_2)) \\ &= p(\mathbf{x} | \lambda_1) p(\mathbf{y} | \lambda_2) \end{aligned}$$

- Podemos, então, estudar cada verossimilhança em separado a fim de identificar a família de prioris conjugadas.

## Exemplo 4 (cont.)

- Para a primeira amostra, temos

$$p(\mathbf{x}|\lambda_1) = \frac{e^{-n\lambda_1} \lambda_1^{n\bar{x}}}{\prod_{i=1}^n x_i} \propto e^{-n\lambda_1} \lambda_1^{n\bar{x}}$$

o qual se assemelhando núcleo de uma distribuição  $\text{gama}(n\bar{x} + 1, n^{-1})$ . Logo, a família de prioris conjugada natural para o modelo Poisson é a família  $\text{gama}(a,b)$ .

- O mesmo se aplica à verossimilhança associada ao vetor  $\mathbf{Y}|\lambda_2$ .
- Por outro lado, vamos admitir que  $p(\lambda_1, \lambda_2) = p(\lambda_1)p(\lambda_2)$ .

## Exemplo 4 (cont.)

- Como as verossimilhanças associadas à  $\lambda_1$  e  $\lambda_2$  são separáveis e assumimos prioris mutuamente independentes para tais parâmetros, podemos obter as posterioris (marginais) separadamente (sem a necessidade de se obter a posteriori conjunta) uma vez que elas também serão mutuamente independentes.

- Ou seja, temos que

$$p(\lambda_1, \lambda_2 | \mathbf{x}, \mathbf{y}) = p(\lambda_1 | \mathbf{x}, \mathbf{y})p(\lambda_2 | \mathbf{x}, \mathbf{y}) = p(\lambda_1 | \mathbf{x})p(\lambda_2 | \mathbf{y})$$

- Por outro lado, assumindo que  $\lambda_1 \sim \text{gama}(a, b^{-1})$  temos que

$$p(\lambda_1 | \mathbf{x}) \propto p(\mathbf{x} | \lambda_1)p(\lambda_1) \propto e^{-n\lambda_1} \lambda_1^{n\bar{x}} e^{-b\lambda_1} \lambda_1^{a-1} = e^{-\lambda_1(n+b)} \lambda_1^{n\bar{x}+a-1}$$

- Portanto,  $\lambda_1 | \mathbf{x} \sim \text{gama}(n\bar{x} + a, (n + b)^{-1})$

## Exemplo 4 (cont.)

- Analogamente, assumindo que  $\lambda_2 \sim \text{gama}(c, b^{-1})$  temos que  $\lambda_2 \sim \text{gama}(n\bar{y} + c, (n + b)^{-1})$ .
- Para comparar  $\lambda_1$  e  $\lambda_2$ , podemos definir  $\theta = \frac{\lambda_1}{\lambda_1 + \lambda_2}$ . Quanto mais próximos estiverem  $\lambda_1$  e  $\lambda_2$  entre si, mais próximo de  $1/2$ ,  $\theta$  estará.
- Por outro lado, de resultados de cálculo de probabilidades, temos que  $\theta | (\mathbf{x}, \mathbf{y}) \sim \text{beta}(n\bar{x} + a, n\bar{y} + c)$ .
- Assim, com distribuição à posteriori de  $\theta$ , podemos comparar os parâmetros de interesse.

## Exemplo 4 (cont.)

- Uma outra forma de obter as posteriores marginais é através da posteriori conjunta, ou seja

$$\begin{aligned} p(\lambda_1, \lambda_2 | \mathbf{x}, \mathbf{y}) &\propto p(\mathbf{x}, \mathbf{y} | \lambda_1, \lambda_2) p(\lambda_1, \lambda_2) \\ &= p(\mathbf{x} | \lambda_1) p(\mathbf{y} | \lambda_2) p(\lambda_1) p(\lambda_2) \\ &= p(\mathbf{x} | \lambda_1) p(\lambda_1) p(\mathbf{y} | \lambda_2) p(\lambda_2) \end{aligned}$$

- Assim, a posteriori de  $\lambda_1$  é obtida através de

$$\begin{aligned} p(\lambda_1 | \mathbf{x}, \mathbf{y}) &= \int p(\lambda_1, \lambda_2 | \mathbf{x}, \mathbf{y}) d\lambda_2 \propto \int p(\mathbf{x} | \lambda_1) p(\lambda_1) p(\mathbf{y} | \lambda_2) p(\lambda_2) d\lambda_2 \\ &= p(\mathbf{x} | \lambda_1) p(\lambda_1) \int p(\mathbf{y} | \lambda_2) p(\lambda_2) d\lambda_2 \propto p(\mathbf{x} | \lambda_1) p(\lambda_1) \\ &\propto p(\lambda_1 | \mathbf{x}) \end{aligned}$$

- Análogamente para  $\lambda_2$ .

# Dados reais: comparação do número de acidentes

- Descrição: número de acidentes (com algum tipo de trauma para as pessoas envolvidas) em 92 dias (correspondentes) em dois anos distintos (1961 e 1962), medidos em algumas regiões da Suécia.
- Considerou-se apenas 43 dias, correspondendo a dias de 1961 em que não havia limite de velocidade e de 1962 em que havia limites de velocidade (90 ou 100 km/h).
- Vamos assumir que

$$X_i | \lambda_1 \stackrel{i.i.d.}{\sim} \text{Poisson}(\lambda_1), i = 1, \dots, 43 \text{ (1961)}.$$

$$Y_i | \lambda_2 \stackrel{i.i.d.}{\sim} \text{Poisson}(\lambda_2), i = 1, \dots, 43 \text{ (1962)}.$$

# Dados reais: comparação do número de acidentes

- Assumiremos também que  $p(\lambda_1, \lambda_2) \propto \mathbb{1}_{(0,\infty)}(\lambda_1)\mathbb{1}_{(0,\infty)}(\lambda_2)$ .
- Dessa forma, teremos que  
 $\lambda_1|\mathbf{x} \sim \text{gama}(n\bar{x} + 1, n^{-1}) \perp \lambda_2|\mathbf{y} \sim \text{gama}(n\bar{y} + 1, n^{-1})$  (exercício).
- Assim,  $\theta = \frac{\lambda_1}{\lambda_1 + \lambda_2} | (\mathbf{x}, \mathbf{y}) \sim \text{beta}(n\bar{x} + 1, n\bar{y} + 1)$ .

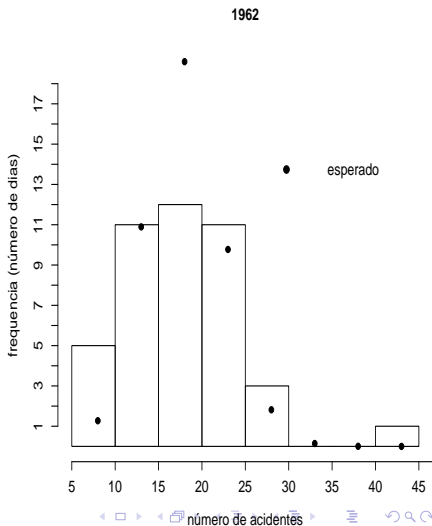
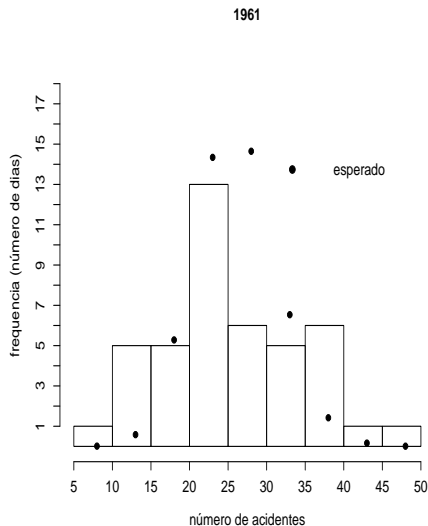


# Dados reais: comparação do número de acidentes (con.t)

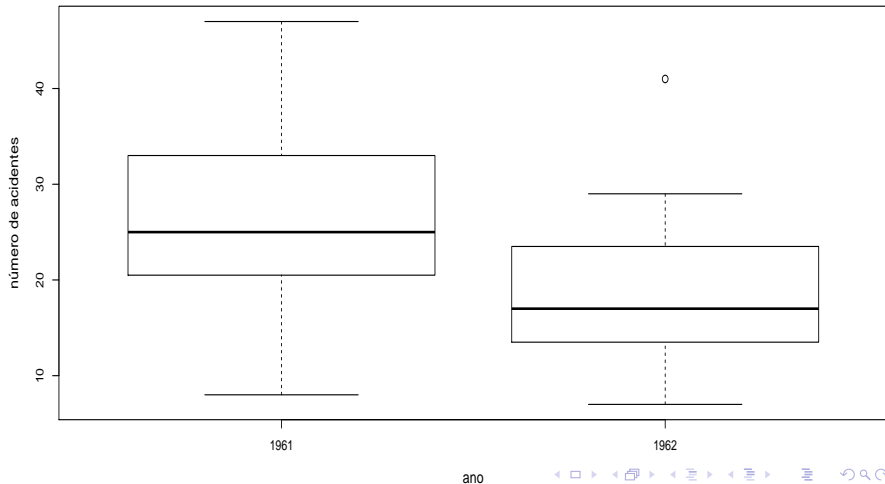
## ■ Medidas resumo

Ano	média	var.	dp	cv(%)	1º quartil	med.	3º quartil
1961	26,05	82,66	9,09	0,35	8,00	25,00	47,00
1962	18,05	44,71	6,69	0,37	7,00	17,00	41,00

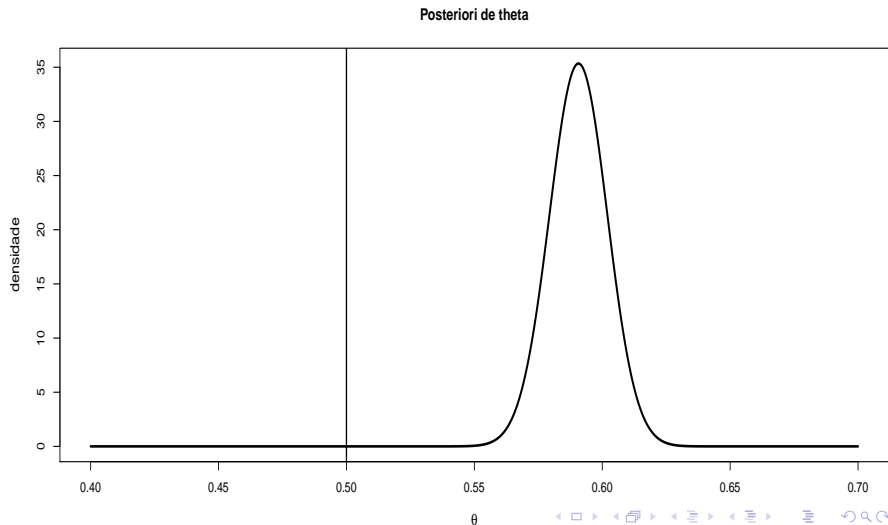
# Dados reais: comparação do número de acidentes (con.t)



# Dados reais: comparação do número de acidentes (con.t)



# Dados reais: comparação do número de acidentes (con.t)



## Exemplo 4 (cont.)

- Vamos supor agora que :

$$X_i | \lambda_1 \stackrel{ind.}{\sim} \text{Poisson}(\lambda_1), i = 1, \dots, n \text{ e}$$

$$Y_j | \lambda_2 \stackrel{ind.}{\sim} \text{Poisson}(\lambda_2), j = 1, \dots, m.$$

- Além disso,  $\lambda_1 \sim \text{gama}(a, b^{-1})$ ,  $\lambda_2 \sim \text{gama}(c, d^{-1})$  e  $\lambda_1 \perp \lambda_2$ .

- Analogamente ao caso anterior, temos que

$$\lambda_1 | \mathbf{x} \sim \text{gama}(n\bar{x} + a, (n + b)^{-1}), \lambda_2 | \mathbf{y} \sim \text{gama}(n\bar{y} + c, (m + d)^{-1}) \text{ e}$$

$$\lambda_1 | (\mathbf{x}, \mathbf{y}) \perp \lambda_2 | (\mathbf{x}, \mathbf{y}).$$

- Contudo,  $\theta | (\mathbf{x}, \mathbf{y})$  não tem mais distribuição beta.
- Neste caso, como obter a distribuição de  $\theta | (\mathbf{x}, \mathbf{y})$ ?

## Exemplo 4 (cont.)

- Uma vez que  $\lambda_1|\mathbf{x} \sim \text{gama}(n\bar{x} + a, (n + b)^{-1})$ ,  
 $\lambda_2|\mathbf{y} \sim \text{gama}(n\bar{y} + c, (m + d)^{-1})$  e  $\lambda_1|\mathbf{x}, \mathbf{y} \perp \lambda_2|\mathbf{x}, \mathbf{y}$ , podemos simular “R” variáveis aleatórias, mutuamente independentes, com distribuições gama específicas, e calcular  $\theta$  para cada par, ou seja:
- Simular  $(\lambda_1^{(r)}, \lambda_2^{(r)})$ ,  $r = 1, \dots, R$  (das respectivas distribuições) e calcular  $\theta^{(r)} = \frac{\lambda_1^{(r)}}{\lambda_1^{(r)} + \lambda_2^{(r)}}$ .
- Dessa forma, teremos uma aproximação numérica da posteriori de interesse.

# Dados reais: comparação do número de acidentes (con.t)

Densidade verdadeira e histograma dos valores simulados

