

Elementos de Computação em R para Inferência Estatística

PJ

Ultima atualização: 18 de outubro de 2019

1 Inferência para parâmetros da distribuição Gama

O objetivo deste documento é revisar conceitos de inferência estatística, com ênfase na função de verossimilhança. Tratamos aqui de um caso de modelo com dois parâmetros.

1.1 Densidade

Parametrização 1:

Esta é a parametrização utilizada nas funções `{r,p,q,d}gamma()` do R com:
 $\alpha = \text{shape} = a$ e $\beta = \text{scale} = s$ ($\text{rate} = 1/\text{scale}$).

Note que $a = 0$ define uma fdp trivial com toda massa de probabilidade em zero. Existe um alerta importante na documentação da função sobre a capacidade do computador em representar pequenos valores que ocorrem para certos valores dos parâmetros:

Note that for smallish values of shape (and moderate scale) a large parts of the mass of the Gamma distribution is on values of x so near zero that they will be represented as zero in computer arithmetic. So `rgamma` can well return values which will be represented as zero. (This will also happen for very large values of scale since the actual generation is done for `scale=1`.)

Funções e expressões: densidade, esperança, variância, (log)-verossimilhança, escore e hessiano:

$$\begin{aligned} Y &\sim G(\alpha, \beta) \\ f(y|\alpha, \beta) &= \frac{1}{\Gamma(\alpha)\beta^\alpha} y^{\alpha-1} \exp\{-y/\beta\} \quad y \geq 0 \quad \alpha \geq 0 \quad \beta > 0 \\ E[Y] &= \alpha\beta \quad \text{Var}[Y] = \alpha\beta^2 \\ L((\alpha, \beta)|y) &= \left(\frac{1}{\Gamma(\alpha)\beta^\alpha}\right)^n \prod_{i=1}^n y_i^{\alpha-1} \exp\{-y_i/\beta\} \\ l((\alpha, \beta)|y) &= n \left(-\log(\Gamma(\alpha)) - \alpha \log(\beta) + (\alpha - 1)\overline{\log(y)} - \bar{y}/\beta\right) \end{aligned} \quad (1)$$

Função escore:

$$\begin{cases} \frac{dl}{d\beta} = n \left(-\frac{\alpha}{\beta} + \frac{\bar{y}}{\beta^2}\right) \\ \frac{dl}{d\alpha} = n \left(-\psi(\alpha) - \log(\beta) + \overline{\log y}\right) \end{cases} \quad (2)$$

em que $\psi(\alpha) = \frac{d}{d\alpha} \log(\Gamma(\alpha)) = \frac{\Gamma'(\alpha)}{\Gamma(\alpha)}$ é a *função digama* (calculada no R com função `digamma()`).

Hessiano, Informação Observada e Esperada:

$$I_O(\alpha, \beta) = -H(\alpha, \beta) = - \begin{bmatrix} \frac{d^2 l}{d\beta^2} & \frac{d^2 l}{d\alpha d\beta} \\ \frac{d^2 l}{d\alpha d\beta} & \frac{d^2 l}{d\alpha^2} \end{bmatrix} = \begin{bmatrix} n \left(2 \frac{\bar{y}}{\beta^3} - \frac{\alpha}{\beta^2} \right) & \frac{n}{\beta} \\ \frac{n}{\beta} & n\psi'(\alpha) \end{bmatrix} \quad (3)$$

em que $\psi'(\alpha) = \frac{d^2}{dt^2} \log(\Gamma(\alpha))$ é a *função trigama* (calculada no R com função `trigamma()`).

$$I_E(\alpha, \beta) = E_Y[I_O(\alpha, \beta)] = \dots \quad (4)$$

Funções concentradas (perfilhadas):

Para este caso não existem expressões fechadas para o estimador de máxima verossimilhança. As funções dadas anteriormente são funções de dois parâmetros (α e β) mas, neste caso, é possível *reduzir a dimensionalidade* e escrevê-las como funções de apenas um parâmetro. Para isto considera-se o comportamento da função no estimador de máxima verossimilhança. Igualando a zero primeira função escore em 2 obtemos uma expressão fechada para um dos parâmetros dada por:

$$\hat{\alpha}\hat{\beta} = \bar{y}. \quad (5)$$

Pode-se então isolar um dos parâmetros e reescrever as funções de verossimilhança, escore e hessiana como funções de apenas um parâmetro. Substituindo β por $\hat{\beta} = \bar{y}/\hat{\alpha}$ temos que a log-verossimilhança (1) pode ser escrita na forma concentrada (perfilhada) como função apenas do parâmetro α :

$$pl(\alpha|y) = l_{\hat{\beta}}(\alpha|y) = n \left(-\log(\Gamma(\alpha)) - \alpha(\log(\bar{y}) - \log(\alpha)) + (\alpha - 1)\overline{\log(y)} - \alpha \right). \quad (6)$$

A função escore fica:

$$U_{\hat{\beta}}(\alpha) = n \left(-\psi(\alpha) - \log \bar{y} + \log(\alpha) + \overline{\log y} \right). \quad (7)$$

O hessiano para α é:

$$H_{\hat{\beta}}(\alpha) = n \left(-\psi'(\alpha) + \frac{1}{\alpha} \right). \quad (8)$$

E, neste caso, como a expressão não depende da v.a. Y as informações observadas e esperadas são iguais:

$$I_E(\alpha) = I_O(\alpha) = -H(\alpha)$$

Alternativamente, substituindo $\alpha = \hat{\alpha} = \bar{y}/\hat{\beta}$ pode-se obter expressões para o parâmetro β :

$$pl(\beta|y) = l_{\hat{\alpha}}(\beta|y) = \dots \quad (9)$$

$$U_{\hat{\beta}}(\alpha) = \dots H_{\hat{\beta}}(\alpha) = \dots I_O(\beta) = \dots I_E(\beta) = \dots \quad (10)$$

Obtenção das estimativas de máxima verossimilhança

O *EMV* é solução conjunta de

$$\begin{cases} \overline{\log y} - \log \beta + & = \psi(\bar{y}/\beta) \\ \hat{\alpha}\hat{\beta} & = \bar{y} \end{cases}$$

que neste caso devem ser obtidas por algoritmos numéricos. Diferentes algoritmos podem ser utilizados conforme será ilustrado mais adiante. Quatro opções serão consideradas:

1. algoritmos de maximização:

- (a) da log-verossimilhança de dois parâmetros,
 - (b) da log-verossimilhança concentrada (de um parâmetro) com posterior substituição em (5) para obtenção da estimativa do outro parâmetro
2. algoritmos de solução de equações (obtidas igualando a zero as funções escore);
- (a) do sistema de equações de dois parâmetros,
 - (b) do sistema concentrado (equação de um parâmetro) com posterior substituição em (5) para obtenção da estimativa do outro parâmetro.

Parametrização 2:

Rizzo (2008) e outros autores utilizam a parametrização com $r = \alpha$, $\lambda = 1/\beta$ em que λ é o parâmetro de taxa (*rate*). Esta é também a parametrização original do **S** e **S-Plus**.

Densidade, esperança, variância, (log)verossimilhança, escore e hessiano:

$$\begin{aligned}
Y &\sim G(r, \lambda) \\
f(y|r, \lambda) &= \frac{\lambda^r}{\Gamma(r)} y^{r-1} \exp\{-\lambda y\} \quad y \geq 0 \quad r \geq 0 \quad \lambda > 0 \\
E[Y] &= r/\lambda \quad Var[Y] = r/\lambda^2 \\
L((r, \lambda)|y) &= \left(\frac{\lambda^r}{\Gamma(r)}\right)^n \prod_{i=1}^n y_i^{r-1} \exp\{-\lambda y_i\} \\
l((r, \lambda)|y) &= n \left(r \log(\lambda) - \log(\Gamma(r)) + (r-1)\overline{\log(y)} - \lambda \bar{y} \right)
\end{aligned} \tag{11}$$

Função escore:

$$\begin{cases} \frac{dl}{d\lambda} = n \left(\frac{r}{\lambda} - \bar{y} \right) \\ \frac{dl}{dr} = n \left(\log(\lambda) - \frac{\Gamma'(r)}{\Gamma(r)} + \overline{\log y} \right) \end{cases}$$

Igualando as funções a zero, da primeira equação temos $\hat{\lambda} = \hat{r}/\bar{y}$. Substituindo λ por $\hat{\lambda}$ a segunda expressão é escrita como:

$$n \left(\log \frac{\hat{r}}{\bar{y}} + \overline{\log y} - \frac{\Gamma'(\hat{r})}{\Gamma(\hat{r})} \right) = 0$$

O *MLE* é solução conjunta de:

$$\begin{cases} \log \lambda + \overline{\log y} = \psi(\lambda \bar{y}) \\ \bar{y} = r/\lambda \end{cases} \tag{12}$$

em que $\psi(t) = \frac{d}{dt} \log(\Gamma(t)) = \frac{\Gamma'(t)}{\Gamma(t)}$ (função **digamma()**).

(Incluir Hessiano, informação e perfiladas como na parametrização anterior)

Parametrização 3:

Aitkin (2010) menciona duas parametrizações sendo a primeira, a mesma implementada no R, apontada como a mais usual, e uma segunda parametrizada por $r = \alpha$ e $\mu = \alpha\beta$.

Esta última parametrização tem propriedades interessantes para inferência. A primeira é a ortogonalidade entre r e μ na matriz de informação. Além disto em geral μ é o usualmente o parâmetro de interesse para inferências e r é um parâmetro *nuisance*. Finalmente a parametrização é adequada

para modelagem estatística na qual usualmente se propõe um modelo de regressão para média μ , como por exemplo em modelos lineares generalizados (*GLM*).

Densidade, esperança, variância e verossimilhança:

$$\begin{aligned}
 Y &\sim G(r, \mu) \\
 f(y|r, \mu) &= \frac{r^r}{\Gamma(r)\mu^r} y^{r-1} \exp\{-ry/\mu\} \quad y \geq 0 \quad r \geq 0 \quad \mu \geq 0 \\
 E[Y] &= \mu \quad Var[Y] = \mu^2/r \\
 L((r, \mu)|y) &= \left(\frac{r^r}{\Gamma(r)\mu^r}\right)^n \prod_{i=1}^n y_i^{r-1} \exp\{-ry_i/\mu\} \\
 l((r, \mu)|y) &= n \left(r(\log(r) - \log(\mu)) - \log(\Gamma(r)) + (r-1)\overline{\log(y)} - \frac{r}{\mu}\bar{y} \right)
 \end{aligned}$$

Função escore:

$$\begin{cases} \frac{dl}{d\mu} = n \left(-\frac{r}{\mu} + \frac{r\bar{y}}{\mu^2} \right) \\ \frac{dl}{dr} = n \left(\log(r) + 1 - \log(\mu) - \frac{\Gamma'(r)}{\Gamma(r)} + \overline{\log y} - \frac{\bar{y}}{\mu} \right) \end{cases}$$

Igualando as funções a zero, da primeira equação temos $\hat{\mu} = \bar{y}$. Substituindo μ por $\hat{\mu}$ a segunda expressão é escrita como:

$$n \left(\log(\hat{r}) + 1 - \log(\bar{y}) - \frac{\Gamma'(\hat{r})}{\Gamma(\hat{r})} + \overline{\log y} - 1 \right) = 0$$

O *MLE* é solução conjunta de:

$$\begin{cases} \log \hat{r} - \psi(\hat{r}) = \log \bar{y} - \overline{\log y} \\ \hat{\mu} = \bar{y} \end{cases}$$

em que $\psi(t) = \frac{d}{dt} \log(\Gamma(t)) = \frac{\Gamma'(t)}{\Gamma(t)}$ (calculado no R pela função `digamma()`).

ToDo:(Incluir Hessiano, informação e perfiladas como na parametrização anterior)

1.2 Códigos R para obtenção das estimativas

Não é possível obter *MLE* em forma analítica fechada para ambos parâmetros da distribuição Gama sendo então necessário o uso de métodos numéricos para obter estimativas. Algumas abordagens numéricas possíveis são:

1. *via* solução de equações (função escore)
 - a. conjunta para dois parâmetros: resolução numérica do sistema de duas equações
 - b. por substituição: resolução numérica de uma equação para encontrar a estimativa de um parâmetro que é substituída na equação conhecida para obter a estimativa do outro parâmetro.
2. *via* maximização de função de log-verossimilhança
 - a. maximização bidimensional (encontra numericamente estimativas dos 2 parâmetros)

```

jfig=Tj= set.seed(201107) dadosG j- rgamma(20, shape = 4.5, rate=2) hist(dadosG, prob=T,
xlim=c(0,5), main=) lines(density(dadosG));rug(dadosG) curve(dgamma(x, shape = 4.5, rate=2),
0, 5, add=T, lwd=2) @

```

Figura 1: Histograma, densidade empírica e *rug* dos dados originais e densidade da distribuição simulada (linha grossa)

- b. unidimensional (verossimilhança concentrada/perfilhada para 1 parâmetro, seguida de substituição para o outro parâmetro)

```

jjecho=Fj= rm(list=ls()) @

```

Gerando dados (simulados).

Estatísticas amostrais que são as quantidades calculadas com os dados utilizadas nas avaliações das funções. `jkeep.source=Tj= am j- list(media=mean(dadosG), media.logs = mean(log(dadosG)), n=length(dadosG)) @`

Estimação utilizando métodos numéricos utilizando a **Parametrização 2**.

(1.a.) Equações de estimação Definimos uma função em R com sistema de equações definido em 12. As estimativas são dadas pela solução deste sistema que é obtida utilizando a função `multiroot()` do pacote **rootSolve**. `jij= Umat j- function(par, amostra) lambda j- par[1] ; r = par[2] UlamUr j- local(c(n * ((r/lambda) - media), n*(log(lambda) - digamma(lambda*media) + media.logs)) , env=amostra) return(UlamUr) rootSolve::multiroot(f=Umat, start=c(1,1), amostra=am)[["root"]] @`

(1.b.) Equações de estimação - substituição

$$\left(\log \frac{\hat{r}}{\bar{y}} + \overline{\log y} - \frac{\Gamma'(r)}{\Gamma(r)} \right) = 0 \quad (\text{numericamente})$$

$$\hat{\lambda} = \hat{r}/\bar{y} \quad (\text{substituição})$$

A função `uniroot()` pode ser usada para resolver numericamente uma equação. `jij= Ulam j- function(lambda, amostra) with(amostra, digamma(lambda*media) - media.logs - log(lambda)) (lambda.est j- uniroot(Ulam, lower=1e-3, upper=1e5, amostra=am)root)(r.est < -lambda.est * ammedia) @`

(2.a.) Maximização da função de log-verossimilhança (2 parâmetros) A função de (negativo) log-verossimilhança pode ser definida no R como uma função a 2 parâmetros. `jij= negLik j- function(par, amostra, modelo=2) if(modelo == 1) alpha j- par[1]; beta j- par[2] ll j- with(amostra, n*(-alpha*log(beta) - log(gamma(alpha)) + (alpha-1) * media.logs - media/beta)) if(modelo == 2) r j- par[1]; lambda j- par[2] ll j- with(amostra, n*(r*log(lambda) - log(gamma(r)) + (r-1) * media.logs - lambda * media)) if(modelo == 3) r j- par[1]; mu j- par[2] ll j- with(amostra, n*(r*(log(r) - log(mu)) - log(gamma(r)) + (r-1) * media.logs - (r/mu) * media)) return(-ll) @`

A seguir é mostrada a obtenção das estimativas em cada parametrização. Para impor limites no espaço paramétrico utiliza-se os argumentos `lower` e `upper` e `method="L-BFGS-B"`. `jij= (mod1 j- optim(c(1,1), negLik, amostra=am, method="L-BFGS-B", modelo=1, lower=c(0,0), upper=c(Inf, Inf)))[["par"]] (mod2 j- optim(c(1,1), negLik, amostra=am, method="L-BFGS-B", modelo=2, lower=c(0,0), upper=c(Inf, Inf)))[["par"]] (mod3 j- optim(c(1,1), negLik, amostra=am, method="L-BFGS-B", modelo=3, lower=c(0,0), upper=c(Inf, Inf)))[["par"]] @`

Obtenção de estimativas em uma parametrização a partir das estimativas de outra parametrização é obtida diretamente. `jij= 1/mod1[2] prod(mod1) @`