Terceira Prova - FECD

23/06/2018

1. Suponha que $X_1, \ldots, X_n$ forme uma amostra aleatória de v.a.'s i.i.d. com densidade de probabilidade Rayleigh dada por
$$f(x;\theta) = (x/\theta)^2 \exp\left(-x^2/(2\theta^2)\right)$$
para $x > 0$ e com $\theta \in (0, \infty)$ sendo um parâmetro desconhecido controlando a forma da densidade. Obtenha o MLE de θ.

   Solução: A log-verossimilhança de θ baseada numa amostra é igual a
$$\begin{aligned} \ell(\theta) &= \log\left[\frac{\prod_i x_i^2}{\theta^{2n}} \exp\left(-\frac{\sum_i x_i^2}{2\theta^2}\right)\right] \\ &= -2n\log(\theta) + \sum_i \log(x_i^2) - \frac{1}{2\theta^2}\sum_i x_i^2 \,. \end{aligned}$$
   Derivando com respeito a θ e igualando a zero, temos
$$0 = \frac{\partial \ell(\theta)}{\partial \theta} = -\frac{2n}{\theta} + \frac{\sum_i x_i^2}{\theta^3}$$
   o que implica no MLE
$$\hat{\theta} = \sqrt{\frac{\sum_i x_i^2}{2n}}$$

---

2. Suponha que $X_1, \ldots, X_n$ forme uma amostra aleatória de v.a.'s i.i.d. com densidade de probabilidade Weibull dada por
$$f(x;\alpha,\beta) = \alpha\beta x^{\beta-1} \exp\left(-\alpha x^\beta\right)$$
para $x > 0$. Você vai usar o método de Newton para obter o MLE de $\boldsymbol{\theta} = (\alpha, \beta)$. Como valor inicial você vai usar $\boldsymbol{\theta}^{(0)} = (1/\bar{x}, 1)$.

   Obtenha explicitamente a equação de iteração do algoritmo de Newton com a densidade acima (mas não precisa inverter a matriz).

   DICA: A derivada em β da função x^β é igual a
$$\frac{d}{d\beta}\left(x^\beta\right) = \log(x)\ x^\beta$$
   onde log é o logaritmo na base e.

   Solução: A log-verossimilhança de $\boldsymbol{\theta} = (\alpha, \beta)$ baseada numa amostra é igual a
$$\begin{aligned} \ell(\alpha,\beta) &= \log\left[\alpha^n \beta^n (\prod_i x_i^{\beta-1}) \exp\left(-\alpha \sum_i x_i^\beta\right)\right] \\ &= n\log(\alpha) + n\log(\beta) + (\beta-1)\sum_i \log(x_i) - \alpha\sum_i x_i^\beta \,. \end{aligned}$$
   Derivando com respeito a α e depois com respeito a β encontramos o vetor gradiente:
$$\nabla \ell(\boldsymbol{\theta}) = \frac{\partial \ell}{\partial \boldsymbol{\theta}} = \begin{bmatrix} \frac{\partial \ell(\alpha,\beta)}{\partial \alpha} \\ \frac{\partial \ell(\alpha,\beta)}{\partial \beta} \end{bmatrix} = \begin{bmatrix} \frac{n}{\alpha} - \sum_i x_i^\beta \\ \frac{n}{\beta} + \sum_i \log(x_i) - \alpha\sum_i \left(\log(x_i) x_i^\beta\right) \end{bmatrix}$$

Não é possível obter o MLE de forma explícita. Devemos usar métodos numéricos. Para usar o método de Newton precisamos da matriz de derivadas parciais de segunda ordem (o Hessiano da função $\ell(\boldsymbol{\theta})$):

$$\begin{aligned}
D^2(\boldsymbol{\theta}) &= \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^2} \\
&= \begin{bmatrix} \frac{\partial^2 \ell(\alpha,\beta)}{\partial \alpha^2} & \frac{\partial^2 \ell(\alpha,\beta)}{\partial \alpha \partial \beta} \\ \frac{\partial^2 \ell(\alpha,\beta)}{\partial \beta \partial \alpha} & \frac{\partial^2 \ell(\alpha,\beta)}{\partial \beta^2} \end{bmatrix} \\
&= \begin{bmatrix} -\frac{n}{\alpha^2} & -\sum_i \left(\log(x_i) x_i^\beta\right) \\ -\sum_i \left(\log(x_i) x_i^\beta\right) & -\frac{n}{\beta^2} - \alpha \sum_i \left(\log^2(x_i) x_i^\beta\right) \end{bmatrix}
\end{aligned}$$

O MLE é encontrado aplicando iterativamente a seguinte atualização:

$$\boldsymbol{\theta}^{(\boldsymbol{k}+1)} = \boldsymbol{\theta}^{(\boldsymbol{k})} - \left[D^2(\boldsymbol{\theta}^{(\boldsymbol{k})})\right]^{-1} \nabla \ell(\boldsymbol{\theta}^{(\boldsymbol{k})})$$

---

3. No modelo de regressão logística com uma única feature numérica x, temos variáveis aleatórias binárias $Y_1, \ldots, Y_n$ em que a probabilidade de sucesso do item i depende de sua feature x_i da seguinte forma:

$$p_i = p(x_i) = \frac{1}{1 + \exp(-\beta_0 - \beta_1 x_i)}$$

- Mostre que $p_i = \frac{\exp(\beta_0 + \beta_1 x_i)}{1 + \exp(\beta_0 + \beta_1 x_i)}$
- Obtenha a função log-verossmilhança de $\boldsymbol{\theta} = (\beta_0, \beta_1)$ mostrando que ela é igual a

$$\ell(\boldsymbol{\beta}) = \beta_0 \sum_i y_i + \beta_1 \sum_i x_i y_i - \sum_i \log\left(1 + e^{\beta_0 + \beta_1 x_i}\right)$$

(Não precisa obter a derivada da log-verossimilhança).

Solução: Multiplicando o numerador e o denominador de p_i por $\exp(\beta_0 + \beta_1 x_i)$ temos

$$p_i = \frac{\exp(\beta_0 + \beta_1 x_i)}{\exp(\beta_0 + \beta_1 x_i) + \exp(\beta_0 + \beta_1 x_i)\exp(-\beta_0 - \beta_1 x_i)} = \frac{\exp(\beta_0 + \beta_1 x_i)}{\exp(\beta_0 + \beta_1 x_i) + 1}$$

Para a função log-verossmilhança temos

$$\begin{aligned}
\ell(\boldsymbol{\beta}) &= \log \prod_i \left[\mathbb{P}(Y_i = y_i | \boldsymbol{\beta})\right] \\
&= \sum_i \log\left[p_i^{y_i} (1 - p_i)^{1 - p_i}\right] \\
&= \sum_i \left[y_i \log(p_i) + (1 - y_i) \log(1 - p_i)\right] \\
&= \sum_i \left[y_i \log(p_i) + (1 - y_i) \log(1 - p_i)\right] \\
&= \sum_i \left[y_i \log\left(\frac{\exp(\beta_0 + \beta_1 x_i)}{1 + \exp(\beta_0 + \beta_1 x_i)}\right) + (1 - y_i) \log\left(\frac{1}{1 + \exp(\beta_0 + \beta_1 x_i)}\right)\right] \\
&= \sum_i \left[y_i(\beta_0 + \beta_1 x_i) - y_i \log(1 + \exp(\beta_0 + \beta_1 x_i)) - (1 - y_i) \log(1 + \exp(\beta_0 + \beta_1 x_i))\right] \\
&= \sum_i \left[y_i(\beta_0 + \beta_1 x_i) - \log(1 + \exp(\beta_0 + \beta_1 x_i))\right] \\
&= \beta_0 \sum_i y_i + \beta_1 \sum_i (y_i x_i) - \sum_i \log(1 + \exp(\beta_0 + \beta_1 x_i))
\end{aligned}$$

---

4. Considere o modelo de regressão usual com matriz de desenho $\mathbf{X}$ de dimensão $n \times p$ cuja primeira coluna é o vetor $\mathbf{1}$. As colunas são linearmente independentes. O estimador de mínimos quadrados dos coeficientes $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_p)$ coincide o MLE e é igual a $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$.

   - Para avaliar se o modelo de regressão linear é adequado, costuma-se trablhar com o vetor de resíduos $\mathbf{r}$. Tendo $\mathbf{X}$ e $\mathbf{Y}$, quais as contas matricias você precisa fazer para obter o vetor $\mathbf{r}$? Isto é, qual a expressão matricial do vetor $\mathbf{r}$ em função de $\mathbf{X}$ e $\mathbf{Y}$?
   - Mostre que o vetor de resíduos $\mathbf{r}$ é ortogonal ao vetor de valores preditos $\hat{\mathbf{Y}}$.
   - A partir do item acima, conclua que a soma dos resíduos $\sum_i r_i$ é igual a zero quando a primeira coluna de $\mathbf{X}$ é o vetor $\mathbf{1}$.

   Solução: Como o resíduo é o vetor diferença entre as observações $\mathbf{Y}$ e os vetor de predições $\mathbf{X}\hat{\boldsymbol{\beta}}$, temos

   $$\mathbf{r} = \mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{Y} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$$

   O vetor $\mathbf{r}$ é ortogonal ao vetor de valores preditos $\hat{\mathbf{Y}}$ pois o seu produto interno é zero. Existem muitas maneiras de provar isto, uma delas sendo:

   $$\hat{\mathbf{Y}}'\mathbf{r} = \left(\mathbf{X}\hat{\boldsymbol{\beta}}\right)'\left(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}\right) = \hat{\boldsymbol{\beta}}'\mathbf{X}'\left(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}\right) = \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{Y} - \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} \tag{1}$$

   Use agora a expressão fornecida $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$ para obter

   $$\hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} = \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{Y}$$

   Portanto, o produto interno em (1) é zero.

   Novamente, existem muitas maneira distintas de provar este último resultado. Aqui está uma delas. Como $\mathbf{r}$ está no espaço ortogonal das colunas de $\mathbf{X}$, o produto interno de $\mathbf{r}$ com estas coluna é zero. Como $\mathbf{1}$ é uma das colunas de $\mathbf{X}$, temos $0 = \mathbf{1}'\mathbf{r} = \sum_i r_i$.

   Outra maneira, apenas manipulando as fórmulas dos itens acima:

   $$\mathbf{X}'\mathbf{r} = \mathbf{X}'\left(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}\right) = \mathbf{X}'\mathbf{Y} - \mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}'\mathbf{Y} - \mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} = \mathbf{X}'\mathbf{Y} - \mathbf{X}'\mathbf{Y} = \mathbf{0}$$

   Assim, cada linha de $\mathbf{X}'$ é ortogonal a $\mathbf{r}$. Isto é, cada coluna de $\mathbf{X}$ é ortogonal a $\mathbf{r}$ e, como $\mathbf{1}$ é uma das colunas de $\mathbf{X}$, temos $0 = \mathbf{1}'\mathbf{r} = \sum_i r_i$.
