Gabarito da Quarta Prova - FECD, 2018

1. **5 pontos:** Num estudo de seguro agrícola, acredita-se que a produção de trigo X_i da área i é normalmente distribuída com média θz_i, onde z_i é a quantidade *CONHECIDA* de fertilizante utilizado na área. Assumindo que as produções em diferentes áreas são independentes, e que a variância é conhecida e igual a 1 (ou seja, que $X_i \sim N(\theta z_i, 1)$, para $i = 1, \ldots, n$:

   - **1 ponto:** Encontre o MLE de θ
   - **1 ponto:** Mostre que o MLE é não viciado para θ. Lembre-se que os valores de z_i são constantes.
   - **1 ponto:** O MLE é baseado numa estatística suficiente para θ?
   - **2 pontos:** Obtenha um intervalo de aproximadamente 95% de confiança para θ baseado na distribuição assintótica do MLE.

   Solução: A verossimilhança de θ é a função

$$L(\theta) = \prod_i \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}(x_i - \theta z_i)^2\right) = \frac{1}{(2\pi)^{n/2}} \exp\left(-\frac{1}{2}\sum_i (x_i - \theta z_i)^2\right)$$

   e portanto com log-verossimilhança dada por

$$\ell(\theta) = -n/2 \log(2\pi) - \frac{1}{2}\sum_i (x_i - \theta z_i)^2$$

   A derivada com respeito com θ é igual a

$$\frac{\partial \ell(\theta)}{\partial \theta} = -\frac{1}{2}\sum_i 2(-z_i)(x_i - \theta z_i) = \sum_i (z_i x_i - \theta z_i^2) = \sum_i z_i x_i - \theta \sum_i z_i^2$$

   Igualando a derivada a zero é possível isolar θ num dos lados da equação e assim obter o MLE:

$$\hat{\theta} = \frac{\sum_i z_i x_i}{\sum_i z_i^2} = \sum_i \left(\frac{z_i}{\sum_k z_k^2}\right) x_i \,.$$

   Note que $\hat{\theta}$ é uma soma ponderada dos x_i, com pesos $z_i / \sum_k z_k^2$ que não somam 1.

   O MLE é não viciado para θ pois, escrevendo o MLE com as v.a.'s X_i, temos

$$\mathbb{E}(\hat{\theta}) = \mathbb{E}\left(\frac{\sum_i z_i X_i}{\sum_i z_i^2}\right) = \frac{\sum_i z_i \mathbb{E}(X_i)}{\sum_i z_i^2} = \frac{\sum_i z_i (\theta z_i)}{\sum_i z_i^2} = \theta$$

   O MLE é baseado numa estatística suficiente para θ pois

$$\begin{aligned} f(\mathbf{x};\theta) &= (2\pi)^{-n/2} \exp\left(-\frac{1}{2}\sum_i (x_i - \theta z_i)^2\right) = (2\pi)^{-n/2} \exp\left(-\frac{1}{2}\sum_i (x_i^2 - 2\theta z_i x_i + \theta^2 z_i^2)\right) \\ &= \underbrace{\exp\left(\theta \sum_i (z_i x_i)\right) \exp\left(-\frac{1}{2}\theta^2 \sum_i z_i^2\right)}_{g(T(\mathbf{x},\theta)} \underbrace{(2\pi)^{-n/2} \exp\left(-\frac{1}{2}\sum_i (x_i^2)\right)}_{h(\mathbf{x})} \end{aligned}$$

onde $T(\mathbf{x}) = \sum_i (z_i x_i)$. Assim, pelo teorema da fatoração $T(\mathbf{X}) = \sum_i z_i X_i$ é estatística suficiente e o MLE é função dessa estatística: $\hat{\theta} = T(\mathbf{x}) / \sum_i z_i^2$ (lembre-se que asquantidades de fertilizantes z_i's são consideradas constantes conhecidas).

A variância aproximada do MLE $\hat{\theta}$ é obtida mais facilmente calculando a segunda derivada:

$$\frac{\partial^2 \ell(\theta)}{\partial \theta^2} = \frac{\partial}{\partial \theta} \left(\frac{\partial \ell(\theta)}{\partial \theta} \right) = - \sum_i z_i^2$$

Como esta expressão não envolve as v.a.'s X_i mas apenas as constantes z_i's, seu valor esperado é igual à própria expressão e portanto

$$I(\theta) = -\mathbb{E} \left(- \frac{\partial^2 \ell(\theta)}{\partial \theta^2} \right) = \sum_i z_i^2 \, .$$

Portanto,

$$\mathbb{V}(\hat{\theta}) \approx I(\theta)^{-1} = \frac{1}{\sum_i z_i^2} \, .$$

Veja que $\mathbb{V}(\hat{\theta}) \to 0$ a medida que mais observações sejam coletadas (não sendo todas com $z_i = 0$).

O I.C. de (aproximadamente) 95% é igual a $\hat{\theta} \pm 1.96 \frac{1}{\sqrt{\sum_i z_i^2}}$.

---

2. **5 pontos:** O vento é uma força vetorial que está sempre mudando de direção e de intensidade (magnitude). Por causa, disso, sua medição num instante de tempo qualquer é tratada como uma variável aleatória. O modelo usualmente adotado para a intensidade X do vento (uma v.a. contínua) é a densidade $f(x) = \frac{x}{\phi} e^{-x^2/(2\phi)}$ se $x > 0$. Estamos considerando apenas a magnitude escalar da intensidade do vento, ignorando a sua direção.

   - **0.5 ponto:** Encontre o MLE de ϕ com uma amostra i.i.d. de medições $X_1, X_2, \ldots, X_n$.
   - **1 pontos:** Encontre uma estatística suficiente para estimar ϕ.
   - **0.5 ponto:** O MLE é função da estatística suficiente?
   - **0.5 ponto:** Sabe-se que $\mathbb{E}(X) = \sqrt{\phi \pi / 2}$ e $\mathbb{V}(X) = \phi(4 - \pi)/2$. Obtenha então $\mathbb{E}(X^2)$ usando a expressão da variância em termos de $\mathbb{E}(X)$ e de $\mathbb{E}(X^2)$.
   - **1 pontos:** Obtenha a informação de Fisher $\mathbb{I}(\theta)$ através da derivada segunda da log-verossimilhança.
   - **0.5 ponto:** Diga qual é a distribuição aproximada do MLE quando o tamanho da amostra é grande.
   - **1 pontos:** Use a distribuição assintótica do MLE para construir um intervalo de confiança de 95% para o parâmetro θ. OBS: se $Z \sim N(0,1)$ então $\mathbb{P}(|Z| \leq 2) \approx 0.95$

   Solução: A densidade conjunta é

$$f(\mathbf{x}, \phi) = \prod_i \frac{x_i}{\phi} \exp\left(-\frac{x_i}{2\phi} \right) = \frac{\prod_i x_i}{\phi^n} \exp\left(-\frac{1}{2\phi} \sum_i x_i^2 \right) = \underbrace{\prod_i x_i}_{h(\mathbf{x})} \; \underbrace{\phi^{-n} \exp\left(-\sum_i x_i^2 / (2\phi) \right)}_{g(T(\mathbf{x}), \phi)}$$

   onde $T(\mathbf{x}) = \sum_i x_i^2$. Pelo teorema da fatoração, $T(\mathbf{X})$ é estatística suficiente

   A log-verossimilhança é

$$\ell(\phi) = -n \log(\phi) + \sum_i \log(x_i) - \frac{1}{2\phi} \sum_i x_i^2$$

   e portanto a sua derivada (função escore) é

$$\frac{\partial \ell(\phi)}{\partial \phi} = -\frac{n}{\phi} + \frac{1}{2\phi^2} \sum_i x_i^2 \, .$$

Igualando a derivada a zero, isolamos ϕ e obtemos o MLE de forma explícita:

$$\hat{\phi} = \frac{\sum_i x_i^2}{2n}$$

Temos

$$\mathbb{E}(X^2) = \mathbb{V}(X) + [\mathbb{E}(X)]^2 = \frac{\phi(4-\pi)}{2} + \frac{\phi\pi}{2} = 2\phi$$

Obtendo a informação de Fisher:

$$\begin{aligned}
I(\phi) &= -\mathbb{E}\left(\frac{\partial^2 \ell(\theta)}{\partial \phi^2}\right) = -\mathbb{E}\left(\frac{n}{\phi^2} - \frac{2}{2\phi^3}\sum_i X_i^2\right) \\
&= -\frac{n}{\phi^2} + \frac{2\sum_i \mathbb{E}(X_i^2)}{2\phi^3} \\
&= -\frac{n}{\phi^2} + \frac{2 \times (n2\phi)}{2\phi^3} \\
&= \frac{n}{\phi^2}
\end{aligned}$$

A distribuição assintótica do MLE é $\hat{\phi} \approx N(\phi_0, \phi_0^2/n)$ onde ϕ_0 é o verdadeiro (e desconhecido) valor do parâmetro ϕ.

O I.C. de (aproximadamente) 95% é igual a $\hat{\phi} \pm 1.96\hat{\phi}/\sqrt{n}$. Note que usamos a paorximação $\hat{\phi}$ ao calcular o desvio-padrão no I.C.

---

3. **10 pontos:** A distribuição exponencial será usada como modelo para o tempo de sobrevida de pacientes diagnosticados com certo tipo de câncer. Se $Y \sim \exp(\lambda)$ então sua densidade de probabilidade (para $y > 0$) é dada por $f(y) = \lambda \exp(-\lambda y)$. Sabe-se que $\mathbb{E}(Y) = 1/\lambda$ e que $\mathbb{V}(Y) = 1/\lambda^2$ (não precisa mostrar isto).

   Sabe-se que a distribuição de probabilidade do tempo de sobrevida depende do estágio em que o câncer foi diagnosticado e do sexo do indivíduo. Considere duas variáveis independentes x_1 e x_2 medidas em cada indivíduo. Para o i-ésimo indivíduo, $x_{i1} = 1$, se ele é homem, e $x_{i1} = 0$, se mulher. A medida x_{i2} é um valor contínuo entre 1 a 10 e mede o estágio do câncer no momento do diagnóstico, um estágio mais avançado correspondendo a valores maiores.

   Um modelo estatístico para este problema supõe que $Y_1, \ldots, Y_n$ sejam variáveis aleatórias independentes mas não identicamente distribuídas. O valor esperado $E(Y_i) = 1/\lambda_i$ do i-ésimo indivíduo depende dos valores de x_{i1} e de x_{i2}. Isto é, cada indivíduo tem uma distribuição própria para o seu tempo de vida adicional, uma exponencial com valor esperado $1/\lambda_i$ que depende de seu sexo e do estágio do seu câncer no momento do diagnóstico.

   Um modelo que é muito usado é o modelo linear generalizado que adota a seguinte relação entre o parâmetro λ e as características x_1 e x_2: $\lambda_i(\boldsymbol{\theta}) = \exp(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2})$ onde $\boldsymbol{\theta} = (\beta_0, \beta_1, \beta_2)$ é o parâmetro desconhecido. Dessa forma, a distribuição de probabilidade do tempo sobrevida Y_i de um dado indivíduo dependem apenas de seu sexo e do estágio da doença e $E(Y_i|x_{i1}, x_{i2}) = 1/\lambda_i = e^{-(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2})}$. Podemos também escrever

$$\begin{aligned}
E(Y_i|x_{i1}, x_{i2}) &= e^{-\beta_0}\ e^{-\beta_1 x_{i1}}\ e^{-\beta_1 x_{i1}} = e^{-\beta_0}\ \left(e^{-\beta_1}\right)^{x_{i1}}\ \left(e^{-\beta_2}\right)^{x_{i2}} \\
&= b_0\ b_1^{x_{i1}}\ b_2^{x_{i2}} \quad \text{onde } b_j = e^{\beta_j} \\
&= \begin{cases} b_0 b_2^{x_{i2}}, & \text{se } i \text{ é mulher} \\ b_0 b_1 b_2^{x_{i2}}, & \text{se } i \text{ é homem} \end{cases}
\end{aligned}$$

   (a) **3 pontos:** Responda às questões abaixo:

- Considere dois indivíduos no mesmo estágio x_2 da doença mas de sexos diferentes. Para obter o tempo esperado de sobrevida do homem basta **(somar ou multiplicar?)** $b_1 = \exp(-\beta_1)$ ao tempo esperado de sobrevida da mulher.
  Solução: Para uma mulher, temos o tempo esperado $b_0 b_2^{x_{i2}}$ enquanto que, para um homem, temos $b_0 b_1 b_2^{x_{i2}}$. Assim, a diferença entre os sexos (para um mesmo valor do estágio inicial x_{i2} do câncer) é multiplicativa: para obter o tempo de vida esperado do homem, multiplique o tempo da mulher por b_1.

- V ou F: O efeito em $E(Y_i)$ de mudar o valor da variável sexo depende do valor estágio x_2 da doença: o efeito é maior nas pessoas com estágio mais avançado.
  Solução: Falso:
  $$E(Y_i|x_{1i}=1, x_{2i}) = b_0 b_1 b_2^{x_{2i}} = b_1 \times E(Y_i|x_{1i}=0, x_{2i})$$
  Assim, o efeito no valor tempo de vida esperado ao mudar a variável sexo (variável x_1) de valor (de 0 para 1) é multiplicar o tempo esperado pelo mesmo valor b_1, não importa o valor da variável x_2.

- V ou F: Suponha que $\beta_1 = 0.2$. Então o fato de ser homem reduz o tempo esperado de sobrevida pelo fator $\exp(-0.2) \approx 0.82$ (isto é, reduz em aproximadamente 18%) em relação ao tempo esperado de uma mulher com o mesmo valor de x_2.
  Solução: Verdadeiro. Como $b_1 = e^{-\beta_1} = e^{-0.2} \approx 0.82$, temos
  $$E(Y_i|x_{1i}=1, x_{2i}) = 0.82 \times E(Y_i|x_{1i}=0, x_{2i}) \, .$$
  Assim, o tempo de vida masculio é 18% menor que o tempo esperado feminino no mesmo nível inicial de câncer, qualquer que seja este nível.

- V ou F: Sejam Y_1 e Y_2 os tempos esperados de sobrevida de um homem e uma mulher, respectivamente, ambos com $x_2 = 5$. Então $E(Y_1) = e^{-\beta_1} E(Y_2)$. Entretanto, se x_2 não for igual a 5 essa relação entre $E(Y_1)$ e $E(Y_2)$ não é válida.
  Solução: A questão está um tanto ambígua e o sinal de β_1 deveria ser negativo (como acima, depois que corrigi). Se você respondeu Falso por causa deste sinal errado, OK.
  Considerando a resposta para o que deveria ter sido escrito, num sentido, a resposta é *Falso*. A relação entre os dois é esta para todos os valores de x_2, desde que este valor seja o mesmo para os dois sexos. Se a variável x_2 para o homem tiver um valor diferente do seu valor para a mulher, então a relação entre $E(Y_1)$ e $E(Y_2)$ terá de ser ajustada também para esta diferença. Isto é, se $x_2(M)$ e $x_2(H)$ são os valores diferentes de x_2 para a mulher e o homem, respectivamente, então
  $$E(Y_1|x_1=1, x_2=x_2(H)) = e^{-\beta_1} e^{-\beta_2(x_2(H)-x_2(M))} E(Y_2|x_1=0, x_2=x_2(M))$$
  e neste caso a resposta é *Verdadeiro.*

- V ou F: Sejam Y_i e Y_j os tempos de sobrevida de dois homens com $x_{i2} = 5$ e $x_{j2} = 5 + x$. O efeito de passar do estágio $x_2 = 5$ para o estágio $x_2 = 5 + x$ pode ser explicado como: multiplique o tempo esperado de vida de Y_i por $b_2^x = \exp(-\beta_2 x)$.
  Solução: Verdadeiro:
  $$E(Y_1|x_1=1, x_2=x+5) = b_0 b_1 b_2^{5+x} = b_2^x \ (b_0 b_1 b_2^5) = b_2^x \ E(Y_2|x_1=1, x_2=5)$$

- V ou F: O efeito em $E(Y)$ de aumentar em x unidades o estágio x_2 da doença entre os homens é diferente do efeito desse mesmo aumento entre as mulheres.
  Solução: Falso: para mulheres, temos $x_1 = 0$ e
  $$E(Y_1|x_1=0, x_2=x+5) = b_0 b_2^{5+x} = b_2^x \ (b_0 b_2^5) = b_2^x \ E(Y_2|x_1=0, x_2=5)$$

(b) **2 pontos:** Obtenha a densidade conjunta das observações e obtenha a a estatística suficiente para estimar $\boldsymbol{\theta}$.

Solução: Como as observações são v.a.'s exponenciais independentes, temos

$$
\begin{aligned}
f(\mathbf{y}|\boldsymbol{\theta}) &= \prod_i \lambda_i \exp(-\lambda_i y_i) \\
&= \prod_i \exp\left(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2}\right)\ e^{-y_i \exp(\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2})} \\
&= \exp\left(n\beta_0 + \beta_1 \sum_i x_{i1} + \beta_2 \sum_i x_{i2}\right)\ e^{-\sum_i (y_i \exp(\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}))}
\end{aligned}
$$

Se você chegou até este ponto, terá todos os pontos neste item. Para a parte da suficiência, a questão deveria ter sido trabalhada em listas e em sala de aula. Passou batido (copiei do meu livro de exercícios), esta pergunta adicional neste item não deveria ter entrado na prova e lamento por isto. O restante da solução vai abaixo:

Lembrando que os valores das features x_1 e x_2 devem ser consideradas constantes, a única parte da densidade conjunta que envolve os dados aleatórios $y_1, \ldots, y_n$ e os parâmetros $\boldsymbol{\theta} = (\beta_0, \beta_1, \beta_2)$ é $\sum_i (y_i \exp(\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}))$, o expoente da exponencial. Entretanto, esta função *não pode* ser escrita da forma $g(T(\mathbf{y}), \boldsymbol{\theta})$ onde $T(\mathbf{y})$ depende *apenas dos dados aleatórios* $\mathbf{y}$ (e possivelmente de constantes como $\mathbf{x}_i$), sem envolver os parâmetros desconhecidos $\boldsymbol{\theta}$. Assim, não podemos usar o teorema da fatoração para encontrar uma estatística suficiente. A estatística suficiente é $T(\mathbf{Y}) = \mathbf{Y}$, o próprio vetor de dados, sem haver redução de dimensionalidade. Este modelo probabilístico não cai na classe da família exponencial (embora usea distribuição exponencial).

(c) **1 ponto:** (INVERTI OS SINAIS NA PROVA. Solução da lista 13 enviada pela Karen corrigiu isto) Para obter o estimador de máxima verossimilhança de $\boldsymbol{\theta}$, encontre a equação de verossimilhança mostrando que

$$
D\ell(\boldsymbol{\theta}) = \begin{bmatrix} \partial\ell/\partial\beta_0 \\ \partial\ell/\partial\beta_1 \\ \partial\ell/\partial\beta_2 \end{bmatrix} = \begin{bmatrix} n - \sum_i y_i \lambda_i(\boldsymbol{\theta}) \\ n\bar{x}_1 - \sum_i y_i x_{i1} \lambda_i(\boldsymbol{\theta}) \\ n\bar{x}_2 - \sum_i y_i x_{i2} \lambda_i(\boldsymbol{\theta}) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}
$$

onde $\bar{x}_j = \sum_i x_{ij}/n$.

Solução: A log-verossimilhança de $\boldsymbol{\theta}$ é

$$
\begin{aligned}
\ell(\boldsymbol{\theta}) &= n\beta_0 + \beta_1 \sum_i x_{i1} + \beta_2 \sum_i x_{i2} - \sum_i (y_i \exp(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2}) \\
&= n\beta_0 + \beta_1 n\bar{x}_1 + \beta_2 n\bar{x}_2 - \sum_i (y_i \exp(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2})
\end{aligned}
$$

Encontrando cada derivada parcial separadamente:

$$
\begin{aligned}
\frac{\partial\ell}{\partial\beta_0} &= n - \sum_i y_i \exp(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2}) = n - \sum_i y_i \lambda_i(\boldsymbol{\theta}) \\
\frac{\partial\ell}{\partial\beta_1} &= n\bar{x}_1 - \sum_i y_i x_{i1} \exp(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2}) = n\bar{x}_1 - \sum_i y_i x_{i1} \lambda_i(\boldsymbol{\theta}) \\
\frac{\partial\ell}{\partial\beta_2} &= n\bar{x}_2 - \sum_i y_i x_{i2} \exp(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2}) = n\bar{x}_2 - \sum_i y_i x_{i2} \lambda_i(\boldsymbol{\theta})
\end{aligned}
$$

(d) **1 ponto:** Mostre que a matriz com as derivadas parciais de segunda ordem é dada por

$$\begin{aligned}
D^2\ell(\boldsymbol{\theta}) &= \begin{bmatrix} \partial^2\ell/\partial\beta_0^2 & \partial^2\ell/\partial\beta_1\beta_0 & \partial^2\ell/\partial\beta_2\beta_0 \\ \partial^2\ell/\partial\beta_0\beta_1 & \partial^2\ell/\partial\beta_1^2 & \partial^2\ell/\partial\beta_2\beta_1 \\ \partial^2\ell/\partial\beta_0\beta_2 & \partial^2\ell/\partial\beta_1\beta_2 & \partial^2\ell/\partial\beta_2^2 \end{bmatrix} \\
&= -\begin{bmatrix} \sum_i y_i\lambda_i(\boldsymbol{\theta}) & \sum_i y_i x_{i1}\lambda_i(\boldsymbol{\theta}) & \sum_i y_i x_{i2}\lambda_i(\boldsymbol{\theta}) \\ \sum_i y_i x_{i1}\lambda_i(\boldsymbol{\theta}) & \sum_i y_i x_{i1}^2\lambda_i(\boldsymbol{\theta}) & \sum_i y_i x_{i1}x_{i2}\lambda_i(\boldsymbol{\theta}) \\ \sum_i y_i x_{i2}\lambda_i(\boldsymbol{\theta}) & \sum_i y_i x_{i1}x_{i2}\lambda_i(\boldsymbol{\theta}) & \sum_i y_i x_{i2}^2\lambda_i(\boldsymbol{\theta}) \end{bmatrix}
\end{aligned}$$

Solução: Vou mostrar como encontrar apenas um dos elementos dessa matriz pois os cálculos são análogos em todas as entradas:

$$\begin{aligned}
\frac{\partial^2\ell}{\partial\beta_0\beta_1} &= \frac{\partial}{\partial\beta_1}\frac{\partial\ell}{\partial\beta_0} \\
&= \frac{\partial}{\partial\beta_1}\left(n - \sum_i y_i\lambda_i(\boldsymbol{\theta})\right) \\
&= -\sum_i y_i \frac{\partial\lambda_i(\boldsymbol{\theta})}{\partial\beta_1} \\
&= -\sum_i y_i x_{i1}\lambda_i(\boldsymbol{\theta})
\end{aligned}$$

Veja que a matriz3×3 de derivadas parciais de segunda ordem pode ser escrita matricialmente como $\mathbf{X}'\,\mathbf{W}(\boldsymbol{\theta})\,\mathbf{X}$ onde $\mathbf{X}$ é a usual matriz $n\times(2+1)$ dos modelos de regressão onde as colunas são as features (com a primeira coluna composta apenas da constante 1) e com $\mathbf{W}(\boldsymbol{\theta})$ sendo uma matriz diagonal $n\times n$ com o i-ésimo elemento sendo $y_i\lambda_i(\boldsymbol{\theta})$.

(e) **1 ponto:** Se $\beta_1 = \beta_2 = 0$ caímos no caso usual de variáveis i.i.d. exponenciais com parâmetro comum $\lambda = \exp(\beta_0)$. A estimativa de máxima verossimilhança de λ é $1/\overline{y}$ e portanto β_0 pode ser estimado como $-\log(\overline{y})$. Você vai usar $\boldsymbol{\theta}^{(0)} = (\overline{y}, 0, 0)$ como valor inicial para $\boldsymbol{\theta}$ num procedimento de Newton-Raphson para obter a estimativa de máxima verossimilhança de $\boldsymbol{\theta}$. Mostre como fazer isto usando a equação de iteração de Newton. Deixe as contas apenas indicadas.

Solução: Basta usar a equação de iteração até convergência de $\boldsymbol{\theta}^{(k)}$:

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

(f) **1 ponto:** Obtenha a matriz 3×3 de informação de Fisher $I(\boldsymbol{\theta}) = -\mathbb{E}(D^2\ell(\boldsymbol{\theta}))$:

$$I(\boldsymbol{\theta}) = -\mathbb{E}\left[D^2\ell(\boldsymbol{\theta})\right] = n\begin{bmatrix} 1 & \overline{x_1} & \overline{x_2} \\ \overline{x_1} & \overline{x_1^2} & \overline{x_1x_2} \\ \overline{x_2} & \overline{x_1x_2} & \overline{x_2^2} \end{bmatrix}$$

onde $\overline{x_j^2} = \sum_i x_{ij}^2/n$ e $\overline{x_1x_2} = \sum_i x_{i1}x_{i2}/n$.

Solução: Ao tomar a esperança na matriz $D^2\ell(\boldsymbol{\theta})$ todos os elementos são considerados aleatórios exceto as v.a.'s Y_i. Temos $\mathbb{E}(Y_i) = 1/\lambda_i(\boldsymbol{\theta})$. Considerando por exemplo a entrada

$(2,3)$ da matriz $D^2\ell(\boldsymbol{\theta})$, temos

$$\begin{aligned}
-\mathbb{E}\left[D^2\ell(\boldsymbol{\theta})\right]_{(2,3)} &= -\mathbb{E}\left(-\sum_i Y_i x_{i1}x_{i2}\lambda_i(\boldsymbol{\theta})\right) \\
&= \sum_i \mathbb{E}(Y_i)x_{i1}x_{i2}\lambda_i(\boldsymbol{\theta}) \\
&= \sum_i \frac{1}{\lambda_i(\boldsymbol{\theta})}x_{i1}x_{i2}\lambda_i(\boldsymbol{\theta}) \\
&= \sum_i x_{i1}x_{i2} \\
&= n\overline{x_1x_2}
\end{aligned}$$

Os demais elementos são obtidos de forma análoga.

(g) **1 ponto:** Como você obteria intervalos de confiança de 95% para β_1 e β_2 se a estimativa de MLE com $n = 30$ pacientes resultar em $\hat{\boldsymbol{\theta}} = (-2, -0.4, -0.005)$ e numa matriz de informação de Fisher cuja **inversa** seja dada por

$$\begin{bmatrix} 0.32 & -0.10 & -0.04 \\ -0.10 & 0.14 & 0.01 \\ -0.04 & 0.01 & 0.01 \end{bmatrix}$$

Solução: Os I.C. de (aproximadamente) 95% para β_1 e β_2 são da forma $\hat{\beta}_j \pm 1.96\sqrt{I^{-1}_{jj}}$. Assim,

- para β_1: I.C. é $-0.10 \pm 1.96\sqrt{0.14}$;
- para β_2: I.C. é $-0.04 \pm 1.96\sqrt{0.01}$.
