MAE5763 - Modelos Lineares Generalizados - Resolução da Prova 1

Guilherme Marthe - 8661962

6/11/2020

1 Exercício 1

1.1 Matriz X e W

O modelo descrito no enunciado consiste tem $Y_{ij} \overset{ind}{\sim} FE(\mu_i, \phi)$ com $i = 1, 2, 3$ e $j = 1, ..., m$. A parte sistemática é descrita como:

$$g(\mu_1) = \alpha$$

$$g(\mu_2) = \alpha - \Delta$$

$$g(\mu_3) = \alpha + \Delta$$

Desta forma, a matriz modelo de cada bloco $i = 1, 2, 3$ é dada por

$$X_1 = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ \vdots & \vdots \\ 1 & 0 \end{bmatrix}_{m\times 2} \quad X_2 = \begin{bmatrix} 1 & -1 \\ 1 & -1 \\ \vdots & \vdots \\ 1 & -1 \end{bmatrix}_{m\times 2} \quad X_3 = \begin{bmatrix} 1 & 1 \\ 1 & 1 \\ \vdots & \vdots \\ 1 & 1 \end{bmatrix}_{m\times 2}$$

E a matriz modelo é composta pelo subcomponentes X_i de tal forma que:

$$X = \begin{bmatrix} X_1 \\ X_2 \\ X_2 \end{bmatrix}_{3m\times 2}$$

A matriz de pesos W é composta, como de usual, por uma matriz diagonal onde elementos na diagonal principal são $w_i = \left(\frac{d\mu_i}{d\eta_i}\right)^2 V_i^{-1}$ com $\eta_i = g(\mu_i)$ e zero em todo resto. Os elementos w_i se repetem m vezes para cada $i = 1, 2, 3$ resultando numa matriz quadrada $3m \times 3m$. Ou seja:

$$W = diag\{w_1, ..., w_1, w_2, ..., w_2, w_3, ..., w_3\}_{3m\times 3m}$$

1.2 Variância assintótica

Para se obter a variância assintótica do estimador $\hat{\Delta}$ iremos utilizar o seguinte resultado de MLGs. Sendo $\theta = (\alpha, \Delta)^T$, temos que a matriz de informação de Fisher pode ser escrita em forma matricial como:

$$K_{\theta\theta} = \phi X^T W X$$

E a variância assintótica $Var(\hat{\Delta}) = K^{-1}_{\Delta\Delta}$. Seguimos com a computação de $K_{\theta\theta}$:

$$\begin{aligned}
X^T W &= \begin{bmatrix} 1 & \dots & 1 & 1 & \dots & 1 & 1 & \dots & 1 \\ 0 & \dots & 0 & -1 & \dots & -1 & 1 & \dots & 1 \end{bmatrix}_{2\times 3m} \cdot diag\{w_1, ..., w_1, w_2, ..., w_2, w_3, ..., w_3\}_{3m\times 3m} \\
&= \begin{bmatrix} w_1 & \dots & w_1 & w_2 & \dots & w_2 & w_3 & \dots & w_3 \\ 0 & \dots & 0 & -w_2 & \dots & -w_2 & w_3 & \dots & w_3 \end{bmatrix}_{2\times 3m}
\end{aligned}$$

$$\begin{aligned}
X^T W X &= \begin{bmatrix} w_1 & \dots & w_1 & w_2 & \dots & w_2 & w_3 & \dots & w_3 \\ 0 & \dots & 0 & -w_2 & \dots & -w_2 & w_3 & \dots & w_3 \end{bmatrix}_{2\times 3m} \cdot \begin{bmatrix} 1 & 0 \\ \vdots & \vdots \\ 1 & 0 \\ 1 & -1 \\ \vdots & \vdots \\ 1 & -1 \\ 1 & 1 \\ \vdots & \vdots \\ 1 & 1 \end{bmatrix}_{3m\times 2} \\
&= \begin{bmatrix} mw_1 + mw_2 + mw_3 & mw_3 - mw_2 \\ mw_3 - mw_2 & mw_2 + mw_3 \end{bmatrix} \\
&= m \begin{bmatrix} \sum_{i=1}^3 w_i & w_3 - w_2 \\ w_3 - w_2 & w_3 + w_2 \end{bmatrix}
\end{aligned}$$

Assim, a matriz de informação de Fisher fica com a seguinte forma:

$$K_{\theta\theta} = \phi m \begin{bmatrix} \sum_{i=1}^3 w_i & w_3 - w_2 \\ w_3 - w_2 & w_3 + w_2 \end{bmatrix}$$

O determinante da matriz, nomeadamente J, na expressão anterior será necessário para os próximos cálculos:

$$det(J) = \left[\sum_{i=1}^{3} w_i\right] (w_3 + w_2) - (w_3 - w_2)^2$$

Assim temos o seguinte resultado:

$$K^{-1}_{\theta\theta} = \frac{1}{\phi m \left[\sum_{i=1}^3 w_i\right] (w_3 + w_2) - (w_3 - w_2)^2} \begin{bmatrix} w_3 + w_2 & w_2 - w_3 \\ w_2 - w_3 & \sum_{i=1}^3 w_i \end{bmatrix}$$

Assim a variância assintótica de $\hat{\Delta}$ é o elemento de interesse em $K^{-1}_{\theta\theta}$, ou seja:

$$Var(\hat{\Delta}) = K_{\Delta\Delta}^{-1} = \frac{\sum_{i=1}^{3} w_i}{\phi m \left[\sum_{i=1}^{3} w_i\right](w_3 + w_2) - (w_3 - w_2)^2}$$

1.3 Função Escore U_Δ

Para o caso dos MLGs o vetor de funções escore dos parâmtros é:

$$U_\theta = \phi X^T W^{1/2} V^{-1/2}(y - \mu)$$

Onde X é a matriz modelo, W é a matriz de pesos, apresentadas anteriormente. V é a matriz de funções variância. Porém, para a estatística de escore requisitada, só é necessário computar a função escore do parâmtro testado. Ou seja:

$$U_\Delta = \phi X_\Delta^T W^{1/2} V^{-1/2}(y - \mu)$$

Onde

$$X_\Delta = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ -1 \\ \vdots \\ -1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}_{3m \times 1}$$

Iremos a seguir calcular cada um dos fatores para se chegar em U_Δ.

$$\begin{aligned} X_\Delta^T W^{1/2} &= \begin{bmatrix} 0 & \dots & 0 & -1 & \dots & -1 & 1 & \dots & 1 \end{bmatrix}_{1 \times 3m} \cdot diag\{\sqrt{w}_1, ..., \sqrt{w}_1, \sqrt{w}_2, ..., \sqrt{w}_2, \sqrt{w}_3, ..., \sqrt{w}_3\}_{3m \times 3m} \\ &= \begin{bmatrix} 0 & ... & 0 & -\sqrt{w}_2 & ... & -\sqrt{w}_2 & \sqrt{w}_3 & ... & \sqrt{w}_3 \end{bmatrix} \end{aligned}$$

$$X_\Delta^T W^{1/2} V^{-1/2} = \begin{bmatrix} 0 & ... & 0 & \frac{-\sqrt{w}_2}{\sqrt{V}_2} & ... & \frac{-\sqrt{w}_2}{\sqrt{V}_2} & \frac{\sqrt{w}_3}{\sqrt{V}_3} & ... & \frac{\sqrt{w}_3}{\sqrt{V}_3} \end{bmatrix}$$

Por fim, temos

$$U_\Delta = \phi X_\Delta^T W^{1/2} V^{-1/2}(y - \mu)$$

$$= \phi \begin{bmatrix} 0 & ... & 0 & \frac{-\sqrt{w_2}}{\sqrt{V_2}} & ... & \frac{-\sqrt{w_2}}{\sqrt{V_2}} & \frac{\sqrt{w_3}}{\sqrt{V_3}} & ... & \frac{\sqrt{w_3}}{\sqrt{V_3}} \end{bmatrix} \begin{bmatrix} y_{11} - \mu_1 \\ \vdots \\ y_{1m} - \mu_1 \\ y_{21} - \mu_2 \\ \vdots \\ y_{2m} - \mu_2 \\ y_{31} - \mu_3 \\ \vdots \\ y_{3m} - \mu_3 \end{bmatrix}$$

$$= \phi \left[-\frac{\sqrt{w_2}}{\sqrt{V_2}} \sum_{j=1}^{m} (y_{2j} - \mu_2) + \frac{\sqrt{w_3}}{\sqrt{V_3}} \sum_{j=1}^{m} (y_{3j} - \mu_3) \right]$$

$$= \phi \left[-\frac{\sqrt{w_2}}{\sqrt{V_2}} m(\bar{y}_2 - \mu_2) + \frac{\sqrt{w_3}}{\sqrt{V_3}} m(\bar{y}_3 - \mu_3) \right]$$

$$= \phi m \left[\frac{\sqrt{w_3}}{\sqrt{V_3}} (\bar{y}_3 - \mu_3) - \frac{\sqrt{w_2}}{\sqrt{V_2}} (\bar{y}_2 - \mu_2) \right]$$

1.4 Estatística de escore

Com a função escore e a variância assintótica calculadas anteriormente, o próximo passo para calcular a estatístide escore consiste em identificar essas funções sob as suposições da hipótese nula. No nosso caso queremos testar $H_0 : \Delta = 0$ contra $H_1 : \Delta \neq 0$. Note que, sob H_0:

$$w_i \overset{H_0}{\to} w$$

$$V_i \overset{H_0}{\to} V$$

$$\mu_i = g^{-1}(\eta_i) \overset{H_0}{\to} \mu = g^{-1}(\alpha_0) = \bar{y}$$

Os componentes da estatística se escore sob H_0 são:

$$Var(\hat{\Delta}) = \frac{\sum_{i=1}^{3} w_i}{\phi m \left[\sum_{i=1}^{3} w_i \right] (w_3 + w_2) - (w_3 - w_2)^2}$$

$$\downarrow$$

$$Var_0(\hat{\Delta}) = \frac{3w}{\phi m 3w 2w - 0} = \frac{1}{2\phi m w}$$

$$U_\Delta = \phi m \left[\frac{\sqrt{w_3}}{\sqrt{V_3}} (\bar{y}_3 - \mu_3) - \frac{\sqrt{w_2}}{\sqrt{V_2}} (\bar{y}_2 - \mu_2) \right]$$

$$\downarrow$$

$$U_\Delta^0 = \phi m \left[\frac{\sqrt{w}}{\sqrt{V}} (\bar{y}_3 - \bar{y}) - \frac{\sqrt{w}}{\sqrt{V}} (\bar{y}_2 - \bar{y}) \right]$$

$$= \frac{\phi m\sqrt{w}}{\sqrt{V}}(\bar{y}_3 - \bar{y}_2)$$

Assim a estatística de escore fica com a seguinte forma:

$$\begin{aligned}
\xi_{SR} &= \left[U_\Delta^0\right]^2 \cdot \left[\text{Var}_0(\hat{\Delta})\right] \\
&= \frac{\phi^2 m^2 w}{V}(\bar{y}_3 - \bar{y}_2)^2 \frac{1}{2\phi m w} \\
&= \frac{\phi m}{2V}(\bar{y}_3 - \bar{y}_2)^2
\end{aligned}$$

Como apenas um parâmetro é testado, nomeadamente Δ, $\xi_{SR} \to \chi^2_{(1)}$ quando n cresce e prevalesce a hipótese nula.

2 Exercício 2

O enunciado apresenta $Z_i \overset{ind}{\sim} \text{ZAP}(\mu_i, \pi_i)$, ou seja, uma poisson ajustada em zeros, com a seguinte função de probabilidades:

$$f_Z(z_i; \mu_i, \pi_i) = \begin{cases} \pi_i & \text{se } z_i = 0 \\ (1 - \pi_i)\frac{f_Y(z_i;\mu_i)}{1 - f_Y(0;\mu_i)} & \text{se } z_i > 0 \end{cases}$$

Onde $f_Y(z_i, \mu_i) = \frac{e^{-\mu_i}\mu_i^{z_i}}{z_i!}$, $z_i = 1, 2, 3, \ldots$ e $i = 1, ..., n$. Pede-se então para encontrar a função desvio para uma estimativa dessa distribuição, assumindo que o modelo saturado é o mesmo de uma variável aleatória Poisson usual. A definição da função desvio parte de uma diferença de log-verossimilhanças do modelo saturado e do modelo ajustado:

$$D^*(\boldsymbol{z}, \hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\pi}}) = 2\{L(\boldsymbol{z}, \tilde{\boldsymbol{\mu}}, \tilde{\boldsymbol{\pi}}) - L(\boldsymbol{z}, \hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\pi}})\}$$

Quando escrita em função dos parâmestros genéricos, sem indicação de se é sob o modelo saturado ou estimado, o log da função de verossimilhanças é escrito como uma soma, onde defino um compoenente da soma como l_i, ou seja

$$L(\boldsymbol{z}, \boldsymbol{\mu}, \boldsymbol{\pi}) = \sum_{i=1}^{n} log(f_Z(z_i, \mu_i, \pi_i)) = \sum_{i=1}^{n} l_i$$

No caso da distribuição poisson ajustada em zero, o componente l_i é escrito de duas maneiras distintas. Se $z_i > 0$

$$l_i = log(1 - \pi_i) - \mu_i + z_i log\, \mu_i - log\, z_i! - log(1 - e^{-\mu_i})$$

E se $z_i = 0$

$$l_i = log(\pi_i)$$

2.1 Modelo saturado

Conforme fornecido pelo enunciado, o modelo saturado para o nosso caso é talque $\tilde{\mu}_i = z_i$. Porém precisamos pensar um pouco sobre o parâmetro π_i. Como foi definido o modelo, podemos pensar na ocorrência de zeros como uma variável Bernoulli onde o sucesso corresponde em $z_i = 0$. Heuristicamente, podemos definir a situação em que a maximização da verossimilhança seria "ideal", ou "teto da igreja", para esta porção da distribuição ZAP desta forma. Então, o parâmetro saturado para π_i é definido por nós como:

$$\tilde{\pi}_i = \begin{cases} 0 & \text{se } \; z_i > 0 \\ 1 & \text{se } \; z_i = 0 \end{cases}$$

2.2 Componente do desvio

Com a definição do modelo saturado podemos partir para o cálculo dos componente do desvio. Para o caso em que $z_i > 0$:

$$d^*(z_i, \hat{\mu}_i, \hat{\pi}_i) = 2\left(\tilde{l}_i - \hat{l}_i\right)$$

Expandindo essa diferença temos:

$$2\left(log(1-\tilde{\pi}_i) - \tilde{\mu}_i + z_i log\, \tilde{\mu}_i - log\, z_i! - log(1 - e^{-\tilde{\mu}_i}) - \;\; log(1-\hat{\pi}_i) + \hat{\mu}_i - z_i log\, \hat{\mu}_i + log\, z_i! + log(1 - e^{-\hat{\mu}_i})\right)$$

E substituindo os parâmetros saturados:

$$2\left(log(1-0) - z_i + z_i log\, z_i - log\, z_i! - log(1 - e^{-z_i}) - \;\; log(1-\hat{\pi}_i) + \hat{\mu}_i - z_i log\, \hat{\mu}_i + log\, z_i! + log(1 - e^{-\hat{\mu}_i})\right)$$

Resultando em:

$$d^*(z_i, \hat{\mu}_i, \hat{\pi}_i) = 2\left[log\left(\frac{1}{1-\hat{\pi}_i}\right) + (\hat{\mu}_i - z_i) + z_i log\left(\frac{z_i}{\hat{\mu}_i}\right) + log\left(\frac{1-e^{-z_i}}{1-e^{-\hat{\mu}_i}}\right)\right]$$

O próximo caso de interesse é se $z_i = 0$. Partindo da mesma definição, porém com uma computação mais direta, temos:

$$\begin{aligned} d^*(z_i, \hat{\mu}_i, \hat{\pi}_i) &= 2\left(\tilde{l}_i - \hat{l}_i\right) \\ &= 2(log(\tilde{\pi}_i) - log(\hat{\pi}_i)) \\ &= 2(log(1) - log(\hat{\pi}_i)) \\ &= -2log(\hat{\pi}_i) \end{aligned}$$

Em conclusão, o componente do desvio para a variável é:

$$d^*(z_i, \hat{\mu}_i, \hat{\pi}_i) = \begin{cases} -2log(\hat{\pi}_i) & \text{se } \; z_i = 0 \\ 2\left[log\left(\frac{1}{1-\hat{\pi}_i}\right) + (\hat{\mu}_i - z_i) + z_i log\left(\frac{z_i}{\hat{\mu}_i}\right) + log\left(\frac{1-e^{-z_i}}{1-e^{-\hat{\mu}_i}}\right)\right] & \text{se } \; z_i > 0 \end{cases}$$

3 Exercício 3

Uma variável aleatória modelada por uma distribuição logarítimica $Y_i \overset{iid}{\sim} LG(\rho)$ tem a seguinte função de probabilidades:

$$f(y_i;\rho) = \frac{\rho^{y_i}}{-y_i \log(1-\rho)}$$

Com $y_i = 1, 2, ...,\ i = 1, ..., n$ e $0 < \rho < 1$

Vamos iniciar a resolução do problema mostrando que Y_i pertencen à família exponencial, pois os resultados nos ajudarão a montar a estatística de razão de verossimilhanças.

3.1 Família exponencial

A forma da função de probabilidades pertencente à família exponencial é a seguinte:

$$f_Y(y;\theta,\phi) = exp\{\phi[y\theta - b(\theta)]\ +\ c(y;\phi)\}$$

Se manipularmos a função de probabilidades de Y_i de tal forma que (omitindo o subscrito i)

$$f(y;\rho) = \frac{\rho^{y}}{-y \log(1-\rho)} = \exp\left[y\log\rho - \log(-\log(1-\rho)) - \log y\right]$$

Podemos definir $\phi = 1$, $c(y;\phi) = -\log y$, o parâmetro canônico $\theta = \log\rho$ e

$$b(\theta) = \log(-\log(1-\rho)) = \log(-\log(1-e^{\theta}))$$

O valor esperado de Y_i pode ser obtido derivando $b(\theta)$

$$\begin{aligned}
E(Y) &= \mu \\
&= b'(\theta) \\
&= \frac{d}{d\theta}\log(-\log(1-e^{\theta})) \\
&= \frac{1}{-\log(1-e^{\theta})}(-1)\frac{1}{1-e^{\theta}}(-1)e^{\theta} \\
&= \frac{-1}{\log(1-e^{\theta})}\frac{e^{\theta}}{1-e^{\theta}} \\
&= \frac{-1}{\log(1-\rho)}\frac{\rho}{1-\rho}
\end{aligned}$$

No enunciado, é sugerida a seguinte parametrização

$$\rho = \frac{e^{\alpha}}{1+e^{\alpha}}$$

Assim, o valor esperado μ em termos dessa parametrização é

$$\mu = \frac{-1}{\log(1-\rho)}\frac{\rho}{1-\rho}$$

$$
\begin{aligned}
&= \frac{ae^{\alpha}(1+e^{\alpha})^{-1}}{\log\left(1-\frac{e^{\alpha}}{1+e^{\alpha}}\right)\left(1-\frac{e^{\alpha}}{1+e^{\alpha}}\right)} \\
&= \frac{-e^{\alpha}(1+e^{\alpha})^{-1}}{-\log(1+e^{\alpha})(1+e^{\alpha})^{-1}} \\
&= \frac{e^{\alpha}}{\log(1+e^{\alpha})}
\end{aligned}
$$

3.2 Variância assintótica

O cálculo da variância assintótica de $\hat{\alpha}$ passa pela determinação do log da função de verossimilhança, em seguida pela função escore e por fim da informação de Fisher associada à esse parâmetro. Iniciaremos pela verossimilhança:

$$
LL(y;\rho) = \sum_{i=1}^{n} y_i \log\rho - \log(-y_i \log(1-\rho))
$$

Porém, é necessário deixar a verossimilhança em função de α. Para tanto, é necessário mostrar algumas identidades da parametrização:

$$
\begin{aligned}
\rho &= \frac{e^{\alpha}}{1+e^{\alpha}} \leftrightarrow \\
\log\rho &= \log(e^{\alpha}) - \log(1+e^{\alpha}) \leftrightarrow \\
\log\rho &= \alpha - \log(1+e^{\alpha})
\end{aligned}
$$

E

$$
\begin{aligned}
\rho &= \frac{e^{\alpha}}{1+e^{\alpha}} \leftrightarrow \\
-\rho &= -\frac{e^{\alpha}}{1+e^{\alpha}} \leftrightarrow \\
1-\rho &= 1-\frac{e^{\alpha}}{1+e^{\alpha}} = \frac{1}{1+e^{\alpha}} \leftrightarrow \\
\log(1-\rho) &= \log(1) - \log\left(1+e^{\alpha}\right) \leftrightarrow \\
\log(1-\rho) &= -\log\left(1+e^{\alpha}\right)
\end{aligned}
$$

Assim, a expressão de $LL(y_i;\rho)$ pode ser escrita em função de α tal que:

$$
\begin{aligned}
LL(y_i;\alpha) &= \sum_{i=1}^{n} y_i\left(\alpha - \log(1+e^{\alpha})\right) - \log(-y_i(-\log(1+e^{\alpha}))) \\
&= \sum_{i=1}^{n} y_i\alpha - y_i\log(1+e^{\alpha}) - \log(y_i\log(1+e^{\alpha})) \\
&= \sum_{i=1}^{n} y_i\alpha - y_i\log(1+e^{\alpha}) - \log(y_i) - \log(\log(1+e^{\alpha}))
\end{aligned}
$$

Assim, a função escore é obtida ao derivar a expressão anterior:

$$
\begin{aligned}
U_\alpha &= \frac{dLL}{d\alpha} \\
&= \sum_{i=1}^{n} \left[y_i - \frac{y_i e^\alpha}{1+e^\alpha} - \frac{e^\alpha}{\log(1+e^\alpha)(1+e^\alpha)} \right] \\
&= \sum_{i=1}^{n} y_i - \frac{e^\alpha}{1+e^\alpha} \sum_{i=1}^{n} y_i - \frac{ne^\alpha}{\log(1+e^\alpha)(1+e^\alpha)} \\
&= \left[1 - \frac{e^\alpha}{1+e^\alpha} \right] \sum_{i=1}^{n} y_i - \frac{ne^\alpha}{\log(1+e^\alpha)(1+e^\alpha)} \\
&= \frac{1}{1+e^\alpha} \sum_{i=1}^{n} y_i - \frac{ne^\alpha}{\log(1+e^\alpha)(1+e^\alpha)} \\
&= \frac{n}{1+e^\alpha} \left[\bar{y} - \frac{e^\alpha}{\log(1+e^\alpha)} \right]
\end{aligned}
$$

Em seguida, partimos para o cálculo da segunda derivada de $LL(y_i;\alpha)$:

$$
\frac{d^2LL}{d\alpha^2} = \frac{d}{d\alpha} \left\{ \frac{n}{1+e^\alpha} \left[\bar{y} - \frac{e^\alpha}{\log(1+e^\alpha)} \right] \right\}
$$

Onde, ao aplicarmos a regra da derivada do produto de funções, resulta em:

$$
\frac{d^2LL}{d\alpha^2} = -\frac{n}{(1+e^\alpha)^2} \left[\bar{y} - \frac{e^\alpha}{\log(1+e^\alpha)} \right] + \frac{n}{1+e^\alpha} \frac{e^\alpha \log(1+e^\alpha) - e^\alpha(1+e^\alpha)^{-1}e^\alpha}{\log(1+e^\alpha)^2}
$$

Com o negativo do valor esperado dessa expressão chegaremos na informação de Fisher para α.

$$
\begin{aligned}
K_{\alpha\alpha} &= E\left(-\frac{d^2LL}{d\alpha^2} \right) \\
&= -\frac{n}{(1+e^\alpha)^2} \left[\mu - \frac{e^\alpha}{\log(1+e^\alpha)} \right] + \frac{n}{1+e^\alpha} \left[\frac{e^\alpha \log(1+e^\alpha) - e^\alpha(1+e^\alpha)^{-1}e^\alpha}{\log(1+e^\alpha)^2} \right]
\end{aligned}
$$

Porém, note que μ em termos do parâmetro α foi calculado no etapa de prova da família exponencial. E sua expressão é:

$$
\mu = \frac{e^\alpha}{\log(1+e^\alpha)}
$$

Por consequência a primeira parcela de $K_{\alpha\alpha}$ é anulada, sendo

$$
K_{\alpha\alpha} = \frac{n}{1+e^\alpha} \left[\frac{e^\alpha \log(1+e^\alpha) - e^\alpha(1+e^\alpha)^{-1}e^\alpha}{\log(1+e^\alpha)^2} \right]
$$

No enunciado nos é fornecido a função $\tau(\alpha) = (1+e^\alpha)\log(1+e^\alpha)$. Manipulando expressão anterior, podemos evidenciar essa função e simplificar ainda mais a informação de Fisher:

$$
\begin{aligned}
K_{\alpha\alpha} &= \frac{n}{1+e^{\alpha}}\left[\frac{e^{\alpha}\log(1+e^{\alpha}) - e^{\alpha}(1+e^{\alpha})^{-1}e^{\alpha}}{\log(1+e^{\alpha})^2}\right] \\
&= \frac{ne^{\alpha}}{1+e^{\alpha}}\left[\frac{\log(1+e^{\alpha}) - (1+e^{\alpha})^{-1}e^{\alpha}}{\log(1+e^{\alpha})^2}\right] \\
&= \frac{ne^{\alpha}}{(1+e^{\alpha})\log(1+e^{\alpha})^2}\left[\log(1+e^{\alpha}) - (1+e^{\alpha})^{-1}e^{\alpha}\right] \\
&= \frac{ne^{\alpha}(1+e^{\alpha})}{(1+e^{\alpha})^2\log(1+e^{\alpha})^2}\left[\log(1+e^{\alpha}) - (1+e^{\alpha})^{-1}e^{\alpha}\right] \\
&= \frac{ne^{\alpha}}{(1+e^{\alpha})^2\log(1+e^{\alpha})^2}\left[(1+e^{\alpha})\log(1+e^{\alpha}) - (1+e^{\alpha})(1+e^{\alpha})^{-1}e^{\alpha}\right] \\
&= \frac{ne^{\alpha}}{(1+e^{\alpha})^2\log(1+e^{\alpha})^2}\left[(1+e^{\alpha})\log(1+e^{\alpha}) - e^{\alpha}\right] \\
&= \frac{ne^{\alpha}}{\tau(\alpha)^2}\left[\tau(\alpha) - e^{\alpha}\right]
\end{aligned}
$$

Finalmente, variância assintótica de $\hat{\alpha}$ é obtida a partir da inversa da informação de Fisher $K_{\alpha\alpha}$, ou seja:

$$
\mathrm{Var}(\hat{\alpha}) = K_{\alpha\alpha}^{-1} = \frac{\tau(\alpha)^2}{ne^{\alpha}\left[\tau(\alpha) - e^{\alpha}\right]}
$$

3.3 Teste de razão de verossimilhanças

Em seguida iremos buscar o teste de razão de verossimilhanças para testar $H_0 : \alpha = 0$ contra $H_1 : \alpha \neq 0$. O teste pode ser escrito para o nosso caso como:

$$
\xi_{RV} = 2[LL(y_i; \hat{\alpha}) - LL(y_i; 0)]
$$

Como já escrevemos anteriormente, o logarítimo da função verossimilhança para a distribuição logarítimica na parametrização que estamos estudando é

$$
LL(y_i; \alpha) = \sum_{i=1}^{n} y_i\alpha - y_i\log(1+e^{\alpha}) - \log(y_i) - \log(\log(1+e^{\alpha}))
$$

Assim, a estimativa sob H_0 é:

$$
\begin{aligned}
LL(y_i; 0) &= \sum_{i=1}^{n} y_i 0 - y_i\log(1+1) - \log(y_i) - \log(\log(1+1)) \\
&= \sum_{i=1}^{n} -y_i\log(2) - \log(y_i) - \log(\log(2)) \\
&= -\log(2)n\bar{y} - n\log(\log(2)) - \sum_{i=1}^{n}\log(y_i)
\end{aligned}
$$

E sob a estimativa

$$
\begin{aligned}
LL(y_i;\hat{\alpha}) &= \sum_{i=1}^{n} y_i\hat{\alpha} - y_i \log(1+e^{\hat{\alpha}}) - \log(y_i) - \log(\log(1+e^{\hat{\alpha}}))\\
&= \hat{\alpha}\sum_{i=1}^{n} y_i - \log(1+e^{\hat{\alpha}})\sum_{i=1}^{n} y_i - n\log(\log(1+e^{\hat{\alpha}})) - \sum_{i=1}^{n}\log(y_i)\\
&= \hat{\alpha}n\bar{y} - \log(1+e^{\hat{\alpha}})n\bar{y} - n\log(\log(1+e^{\hat{\alpha}})) - \sum_{i=1}^{n}\log(y_i)
\end{aligned}
$$

A estatística do teste de razão de verossimilhanças é:

$$
\begin{aligned}
\xi_{RV} &= 2[LL(y_i;\hat{\alpha}) - LL(y_i;0)]\\
&= \hat{\alpha}n\bar{y} - \log(1+e^{\hat{\alpha}})n\bar{y} - n\log(\log(1+e^{\hat{\alpha}})) - \sum_{i=1}^{n}\log(y_i) - \left(-\log(2)n\bar{y} - n\log(\log(2)) - \sum_{i=1}^{n}\log(y_i)\right)\\
&= \hat{\alpha}n\bar{y} - \log(1+e^{\hat{\alpha}})n\bar{y} - n\log(\log(1+e^{\hat{\alpha}})) - \sum_{i=1}^{n}\log(y_i) + \log(2)n\bar{y} + n\log(\log(2)) + \sum_{i=1}^{n}\log(y_i)\\
&= \hat{\alpha}n\bar{y} - \log(1+e^{\hat{\alpha}})n\bar{y} - n\log(\log(1+e^{\hat{\alpha}})) + \log(2)n\bar{y} + n\log(\log(2))\\
&= n\left(\hat{\alpha}\bar{y} - \log(1+e^{\hat{\alpha}})\bar{y} - \log(\log(1+e^{\hat{\alpha}})) + \log(2)\bar{y} + \log(\log(2))\right)\\
&= n\left[\bar{y}\left(\hat{\alpha} + \log(2) - \log(1+e^{\hat{\alpha}})\right) - \log(\log(1+e^{\hat{\alpha}})) + \log(\log(2))\right]\\
&= n\left[\bar{y}\left(\hat{\alpha} + \log(2) - \log(1+e^{\hat{\alpha}})\right) - \log(\log(1+e^{\hat{\alpha}})\log(2))\right]
\end{aligned}
$$

A estatística ξ_{RV} segue, assintoticamente e sob H_0 uma distribuição qui-quadrado com um grau de liberdade, uma vez que só um parâmetro, α, é testado.

4 Exercício 4

Seja $Y_i \overset{iid}{\sim} ZTP(\lambda)$, ou seja uma variável Poisson truncada em zeros. Sua função de probabilidades é

$$
f(y_i;\lambda) = \frac{e^{-\lambda}\lambda^{y_i}}{y_i!(1-e^{-\lambda})}
$$

Onde $i = 1, ..., n$ e $y_i = 1, 2, ...$ Calcularemos primeiro o valor esperado de Y_i e em seguida sua variância.

4.1 Valor esperado

Para o cálculo do valor esperado, iremos omitir o subscrito i, para manter uma manipulação mais limpa.

$$
\begin{aligned}
E(Y) &= \sum_{y=1}^{\infty} y f(y;\lambda)\\
&= \sum_{y=1}^{\infty} y\frac{e^{-\lambda}\lambda^{y}}{y!(1-e^{\lambda})}\\
&= \frac{e^{-\lambda}}{(1-e^{-\lambda})}\sum_{y=1}^{\infty}\frac{\lambda^{y}}{(y-1)!}
\end{aligned}
$$

$$= \frac{\lambda e^{-\lambda}}{(1-e^{-\lambda})}\sum_{y=1}^{\infty}\frac{\lambda^{y-1}}{(y-1)!}$$

Fazendo a transformação de índices tal que $z = y - 1$ e notando que se $y = 1 \to z = 0$, temos uma expansão de taylor exponencial na série com relação à variável z. Com esse fato e um pouco mais de manipulação algébrica, chegamos na expressão buscada:

$$\begin{aligned}
E(Y) &= \frac{\lambda e^{-\lambda}}{(1-e^{-\lambda})}\sum_{y=1}^{\infty}\frac{\lambda^{y-1}}{(y-1)!}\\
&= \frac{\lambda e^{-\lambda}}{(1-e^{-\lambda})}\sum_{z=0}^{\infty}\frac{\lambda^{z}}{(z)!}\\
&= \frac{\lambda e^{-\lambda}}{(1-e^{-\lambda})}e^{\lambda}\\
&= \frac{\lambda}{(1-e^{-\lambda})}\\
&= \frac{\lambda}{(1-e^{-\lambda})}\frac{e^{\lambda}}{e^{\lambda}}\\
&= \frac{\lambda e^{\lambda}}{(e^{\lambda}-1)} = \mu
\end{aligned}$$

4.2 Variância

O cálculo variância será algebricamente similar ao do valor esperado. Porém utilizaremos adicionalmente a seguinte identidade:

$$E(Y) = E(Y^2) - [E(Y)]^2$$

Iniciaremos pelo cálculo do segundo momento de Y:

$$\begin{aligned}
E(Y^2) &= \sum_{y=1}^{\infty} y^2 f(y;\lambda)\\
&= \sum_{y=1}^{\infty} y^2 \frac{e^{-\lambda}\lambda^y}{y!(1-e^{\lambda})}\\
&= \frac{e^{-\lambda}}{(1-e^{-\lambda})}\sum_{y=1}^{\infty}\frac{y\lambda^{y}}{(y-1)!}\\
&= \frac{\lambda e^{-\lambda}}{(1-e^{-\lambda})}\sum_{y=1}^{\infty}\frac{y\lambda^{y-1}}{(y-1)!}\\
&= \frac{\lambda e^{-\lambda}}{(1-e^{-\lambda})}\sum_{z=0}^{\infty}\frac{(z+1)\lambda^{z}}{(z)!} \qquad \text{(tomando} \quad z = y-1)\\
&= \frac{\lambda e^{-\lambda}}{(1-e^{-\lambda})}\left[\sum_{z=0}^{\infty}\frac{\lambda^{z}}{(z)!} + \sum_{z=0}^{\infty}\frac{z\lambda^{z}}{(z)!}\right]
\end{aligned}$$

$$
\begin{aligned}
&= \frac{\lambda e^{-\lambda}}{(1-e^{-\lambda})}\left[e^{\lambda} + \sum_{z=1}^{\infty} \frac{z\lambda^z}{(z)!}\right] \qquad \text{o primeiro termo da soma é nulo} \\
&= \frac{\lambda e^{-\lambda}}{(1-e^{-\lambda})}\left[e^{\lambda} + \lambda\sum_{z=1}^{\infty} \frac{\lambda^{z-1}}{(z-1)!}\right] \\
&= \frac{\lambda e^{-\lambda}}{(1-e^{-\lambda})}\left[e^{\lambda} + \lambda e^{\lambda}\right]
\end{aligned}
$$

Onde a última igualdade parte de um outra mudança de variável que culmina novamente numa expansão de Taylor exponencial ao redor de λ. Avançando com as manipulações para uma versão mais amigável do segundo momento, temos

$$
\begin{aligned}
E(Y^2) &= \frac{\lambda}{(1-e^{-\lambda})}\left[1+\lambda\right] \\
&= \mu[1+\lambda]
\end{aligned}
$$

Por fim

$$
\begin{aligned}
E(Y) &= E(Y^2) - [E(Y)]^2 \\
&= \mu[1+\lambda] + \mu^2 \\
&= \mu\left[1+\lambda-\mu\right]
\end{aligned}
$$

Que é a expressão no enunciado.

4.3 Função escore

Para o cálculo da informação de Fisher do parâmetro λ e da estatística de escore, é necessário partir do função escore do parâmetro.

O logaritmo da verossimilhança para a amostra $i = 1, ..., n$ de $Y_i \overset{iid}{\sim} ZTP(\mu)$ é:

$$
\begin{aligned}
LL(y_i; \lambda) &= \sum_{i=1}^{n} \log \frac{e^{-\lambda}\lambda^{y_i}}{y_i!(1-e^{\lambda})} \\
&= \sum_{i=1}^{n} \log e^{-\lambda} + \log \lambda^{y_i} - \log y_i! - \log(1-e^{-\lambda}) \\
&= \sum_{i=1}^{n} -\lambda + y_i \log \lambda - \log y_i! - \log(1-e^{-\lambda}) \\
&= -n\lambda + \log \lambda \sum_{i=1}^{n} y_i - \sum_{i=1}^{n} \log y_i! - n\log(1-e^{-\lambda})
\end{aligned}
$$

A função escore é obtida ao derivar o logaritmo da verossimilhança para com relação ao parâmetro de interesse (que no caso é apenas o λ):

$$
U_\lambda = \frac{dLL}{d\lambda}
$$

$$
\begin{aligned}
&= -n + \lambda^{-1} \sum_{i=1}^{n} y_i - n \frac{e^{-\lambda}}{1 - e^{-\lambda}} \\
&= n \left(\bar{y}\lambda^{-1} - \frac{e^{-\lambda}}{1 - e^{-\lambda}} - 1 \right) \\
&= n \left(\bar{y}\lambda^{-1} - \frac{1}{1 - e^{-\lambda}} \right) \\
&= n \left(\bar{y}\lambda^{-1} - \frac{e^{\lambda}}{e^{\lambda} - 1} \right)
\end{aligned}
$$

4.4 Informação de Fisher

Para o cálculo da informação de Fisher é necessário primeiro obter a segunda derivada do logaritmo função de verossimilhança:

$$
\begin{aligned}
\frac{d^2 LL}{d\lambda^2} &= \frac{dU_\lambda}{d\lambda} \\
&= n \left(-\bar{y}\lambda^{-2} - \frac{d}{d\lambda} \frac{e^{\lambda}}{e^{\lambda} - 1} \right)
\end{aligned}
$$

A derivada restante é calculada separadamente por razões de clareza no desenvolvimento

$$
\frac{d}{d\lambda} \frac{e^{\lambda}}{e^{\lambda} - 1} = \frac{e^{\lambda}(e^{\lambda} - 1) - e^{\lambda} e^{\lambda}}{(e^{\lambda} - 1)^2}
$$

Resultando na segunda derivada

$$
\frac{d^2 LL}{d\lambda^2} = n \left(-\bar{y}\lambda^{-2} - \frac{e^{\lambda}(e^{\lambda} - 1) - e^{\lambda} e^{\lambda}}{(e^{\lambda} - 1)^2} \right)
$$

A informação de Fisher é resultando do valor esperado do negativo da segunda derivada da função de verossimilhança, ou seja:

$$
\begin{aligned}
K_{\lambda\lambda} &= E\left(-\frac{d^2 LL}{d\lambda^2} \right) \\
&= n \left(\frac{\lambda^{-2}}{n} \sum_{i=1}^{n} E(y_i) + \frac{e^{\lambda}(e^{\lambda} - 1) - e^{\lambda} e^{\lambda}}{(e^{\lambda} - 1)^2} \right) \\
&= n \left(\lambda^{-2}\mu + \frac{e^{\lambda}(e^{\lambda} - 1) - e^{2\lambda}}{(e^{\lambda} - 1)^2} \right) \\
&= n \left(\frac{\lambda^{-2}\lambda e^{\lambda}}{e^{\lambda} - 1} + \frac{e^{2\lambda} + e^{\lambda} - e^{2\lambda}}{(e^{\lambda} - 1)^2} \right) \\
&= n \left(\frac{e^{\lambda}}{\lambda(e^{\lambda} - 1)} + \frac{e^{\lambda}}{(e^{\lambda} - 1)^2} \right) \\
&= n \left(\frac{e^{\lambda}(e^{\lambda} - 1)}{\lambda(e^{\lambda} - 1)^2} + \frac{\lambda e^{\lambda}}{\lambda(e^{\lambda} - 1)^2} \right)
\end{aligned}
$$

$$
\begin{aligned}
&= n\left(\frac{e^{2\lambda} - e^{\lambda} - \lambda e^{\lambda}}{\lambda(e^{\lambda}-1)^2}\right) \\
&= ne^{\lambda}\left(\frac{e^{\lambda} - 1 - \lambda}{\lambda(e^{\lambda}-1)^2}\right) \\
&= \frac{ne^{\lambda}(e^{\lambda} - 1 - \lambda)}{\lambda(e^{\lambda}-1)^2}
\end{aligned}
$$

4.5 Estatística de Escore

O próximo e último passo é calcular a estatística escore para testar $H_0 : \lambda = 1$ contra $H_1 : \lambda \neq 1$. Primeiro vamos obter as funções de escore e a informação de Fisher sob H_0:

$$
U_{\lambda} = n\left(\bar{y}\lambda^{-1} - \frac{e^{\lambda}}{e^{\lambda}-1}\right) \overset{H_0}{\rightarrow} U_{\lambda}^{0} = n\left(\bar{y} - \frac{e}{e-1}\right)
$$

$$
K_{\lambda\lambda} = \frac{ne^{\lambda}(e^{\lambda} - 1 - \lambda)}{\lambda(e^{\lambda}-1)^2} \overset{H_0}{\rightarrow} K_{\lambda\lambda}^{0} = \frac{ne(e-2)}{(e-1)^2}
$$

Assim, a estatística de escore é

$$
\xi_{SR} = \frac{\left[U_{\lambda}^{0}\right]^2}{K_{\lambda\lambda}^{0}} = \frac{n^2\left(\bar{y} - \frac{e}{e-1}\right)^2 (e-1)^2}{ne(e-2)}
$$

Ou

$$
\xi_{SR} = \frac{n\left(\bar{y} - \frac{e}{e-1}\right)^2 (e-1)^2}{e(e-2)}
$$

Por fim, como apenas um parâmetro é testado, a distribuição assintótica de ξ_{SR} sob H_0 é uma qui-quadrado com 1 grau de liberdade.