Trabalho de Teoria do Risco

Margarida Biscaia Caleiras
Mestrado em Matemática, Universidade de Coimbra

Conteúdo

Exercício 1. **2**
- Distribuição de X_1 2
- Intervalo de confiança para os parâmetros de X_1 4
- Distribuição de X_2 6
- Intervalo de confiança para os parâmetros de X_2 7

Exercício 2. **8**
- Distribuição de Y_1 8
- Distribuição de Y_2 15

Exercício 3. **18**

Exercício 4. **19**

Exercício 5. **20**

Exercício 6. **21**

Uma seguradora vende dois tipos de apólices, cobrindo riscos de natureza distinta. Cada cliente adquire uma apólice de cada tipo. Cada uma das apólices, independentemente do risco seguro submete um quantidade variável de pedidos de indemnização, cujos valores são, naturalmente aleatórios, descritos cada um por uma distribuição de probabilidade conveniente. Os dados disponíveis descrevem, para cada tipo de apólice, o número de pedidos de indemnização submetidos por apólice e, caso existam, os seu valores.

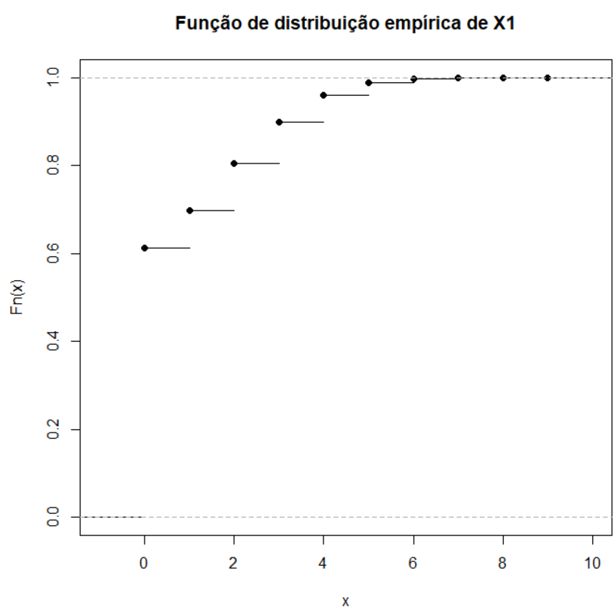
Consideremos X_1, X_2 as variáveis aleatórias que descrevem o número de pedidos de indemnização submetidos pelas apólices 1 e 2, respetivamente, (correspondentes a **b1$num.ind** e **b2$num.ind** no ficheiroR.txt) e Y_1, Y_2 as variáveis aleatórias que descrevem o valor desses pedidos (correspondentes a **b1$inds** e **b2$inds** no ficheiro leitura.txt).

Exercício 1.

Queremos uma aproximação para as distribuições de X_1, X_2 e intervalos de confiança para os respetivos parâmetros. Como as variáveis X_1 e X_2 descrevem números de pedidos de indemnização, é de esperar que tenham distribuições discretas.

Distribuição de X_1

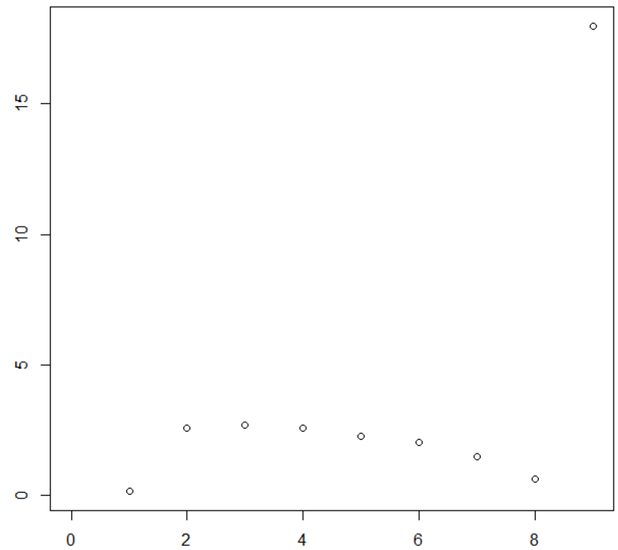
Comecemos por analisar a função de distribuição empírica de X_1.



A tabela seguinte indica o número de apólices em função dos pedidos de indemnização que a apresentaram:

No. pedidos (k)	0	1	2	3	4	5	6	7	8	9	Total
No. apólices (n_k)	4027	551	705	627	404	182	61	13	1	2	6573
$k\frac{n_k}{n_{k-1}}$	–	0.14	2.56	2.67	2.58	2.25	2.01	1.49	0.62	18	

A representação dos valores registados na última linha é



Esta representação é aproximadamente linear pelo que é aceitável que a distribuição subjacente seja de classe $(a, b, 0)$. O último ponto estraga a linearidade da figura. Contudo, este último ponto não é da forma p_n pelo que é aceitável ignorá-lo. A disposição dos pontos indica que o declive da reta subjacente, isto é, a aproximação para a, possa ser negativo ou 0. Trata-se assim de uma distribuição que aparenta ser da família das Binomiais ou Poisson. No entanto, como podemos observar na tabela apresentada acima, n_0 é muito superior aos restantes n_k. Isto sugere que consideremos as distribuições "zero-inflanted Poisson"e "zero-inflanted Binomial". Para estimar os parâmetros de cada distribuição podemos utilizar o método dos momentos.

– **Distribuição zero-inflated Poisson** (λ, ϕ)

Estes modelos são um mixing de um distribuição de Poisson e o valor 0; assume valor 0 com probabilidade ϕ, caso contrário tem distribuição Poisson(λ). Assim, vem

$$P(X_1 = 0) = \phi + (1 - \phi) \exp(-\lambda)$$

e para $k = 1, 2, \ldots$

$$P(X_1 = k) = (1 - \phi) \exp(-\lambda) \frac{\lambda^k}{k!}$$

onde ϕ satisfaz $0 < \phi < 1$.

Tem-se

$$\mathrm{E}(X_1) = (1 - \phi)\lambda, \quad \mathrm{Var}(X_1) = (1 - \phi)\lambda(1 + \phi\lambda).$$

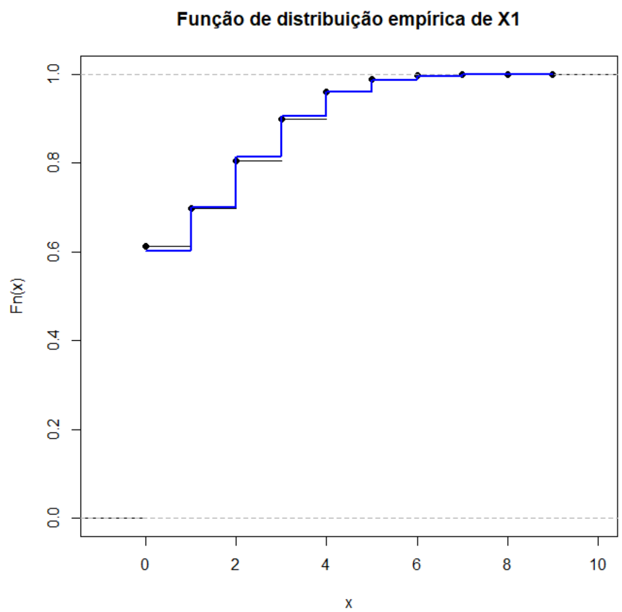
Assim, pelo método dos momentos vem

$$\begin{cases} (1 - \phi)\lambda = \overline{X_1} \\ (1 - \phi)\lambda(1 + \phi\lambda) = \mathrm{Var}(X_1) \end{cases}$$

pelo que obtemos os estimadores

$$\widehat{\lambda} = \frac{\text{Var}(X_1)}{\overline{X_1}} + \overline{X_1} - 1, \quad \widehat{\phi} = -\frac{\widehat{\lambda} - \overline{X_1}}{\widehat{\lambda}}.$$

Aplicando aos dados obtemos $\widehat{\lambda} = 2.370227$, $\widehat{\phi} = 0.5602556$. Uma comparação entre a função de distribuição empírica e a função de distribuição zero-inflated Poisson correspondente aos parâmetros obtidos é descrita na figura abaixo:



Como podemos observar, a distribuição zero-inflated Poisson parece aproximar bem a distribuição de X_1. No entanto, o teste de Anderson-Darling leva à rejeição desta distribuição (pois o p-valor obtido é inferior aos níveis de significância usuais).

Observação: Fiz várias tentativas para aproximar a distribuição de X_1 à distribuição zero-inflated Binomial mas não consegui determinar os estimadores pelo que optei por aceitar a distribuição zero-inflated Poisson para continuar o exercício.

Intervalo de confiança para os parâmetros de X_1

X_1 tem distribuição zero-inflated Poisson (λ, ϕ) pelo que a função de verosimilhança é dada por

$$L(\lambda, \phi) = (\phi + (1-\phi)e^{-\lambda})^{n_0} \prod_{\substack{i=1 \\ x_i \neq 0}}^{n} (1-\phi)e^{-\lambda} \frac{\lambda^{x_i}}{x_i!},$$

onde n_0 é o número de observações de zero na amostra. Assim, a função de log-verosimilhança é dada por

$$l(\lambda, \phi) = n_0 \log(\phi + (1-\phi)e^{-\lambda}) + (n - n_0)(\log(1-\phi) - \lambda) + \log(\lambda) \sum_{i=1}^{n} x_i - \sum_{\substack{i=1 \\ x_i \neq 0}}^{n} \log(x_i!).$$

Vamos então obter

$$\frac{\partial l}{\partial \lambda} = -n_0 \frac{(1-\phi)e^{-\lambda}}{\phi + (1-\phi)e^{-\lambda}} - (n - n_0) + \frac{1}{\lambda}\sum_{i=1}^{n} x_i$$
$$\frac{\partial l}{\partial \phi} = n_0 \frac{1 - e^{-\lambda}}{\phi + (1-\phi)e^{-\lambda}} - \frac{n - n_0}{1-\phi}$$

pelo que vem

$$\frac{\partial^2 l}{\partial \lambda^2} = \frac{n_0(1-\phi)\phi e^{-\lambda}}{(\phi + (1-\phi)e^{-\lambda})^2} - \frac{1}{\lambda^2}\sum_{i=1}^{n} x_i = \frac{n_0(1-\phi)\phi e^{-\lambda}}{(\phi + (1-\phi)e^{-\lambda})^2} - \frac{n\overline{X_1}}{\lambda^2}$$
$$\frac{\partial^2 l}{\partial \phi^2} = -\frac{n_0(1-e^{-\lambda})^2}{(\phi + (1-\phi)e^{-\lambda})^2} - \frac{n - n_0}{(1-\phi)^2}$$
$$\frac{\partial^2 l}{\partial \lambda \partial \phi} = \frac{n_0 e^{-\lambda}}{(\phi + (1-\phi)e^{-\lambda})^2}$$

Finalmente, chegamos a

$$I(\lambda, \phi) = \begin{bmatrix} \frac{n\overline{X_1}}{\lambda^2} - \frac{n_0(1-\phi)\phi e^{-\lambda}}{(\phi+(1-\phi)e^{-\lambda})^2} & -\frac{n_0 e^{-\lambda}}{(\phi+(1-\phi)e^{-\lambda})^2} \\ -\frac{n_0 e^{-\lambda}}{(\phi+(1-\phi)e^{-\lambda})^2} & \frac{n_0(1-e^{-\lambda})^2}{(\phi+(1-\phi)e^{-\lambda})^2} + \frac{n-n_0}{(1-\phi)^2} \end{bmatrix},$$

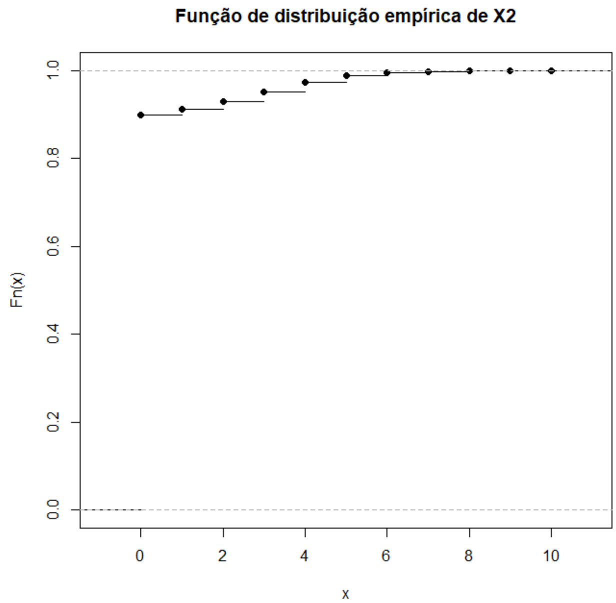
pelo que o intervalo de confiança é dado por

$$\begin{bmatrix} \widehat{\lambda} - \lambda & \widehat{\phi} - \phi \end{bmatrix} \begin{bmatrix} \frac{n\overline{X_1}}{\lambda^2} - \frac{n_0(1-\phi)\phi e^{-\lambda}}{(\phi+(1-\phi)e^{-\lambda})^2} & -\frac{n_0 e^{-\lambda}}{(\phi+(1-\phi)e^{-\lambda})^2} \\ -\frac{n_0 e^{-\lambda}}{(\phi+(1-\phi)e^{-\lambda})^2} & \frac{n_0(1-e^{-\lambda})^2}{(\phi+(1-\phi)e^{-\lambda})^2} + \frac{n-n_0}{(1-\phi)^2} \end{bmatrix} \begin{bmatrix} \widehat{\lambda} - \lambda \\ \widehat{\phi} - \phi \end{bmatrix} < z^2_{\alpha/2}$$

Observação: Não consegui desenvolver a expressão. Obtendo a expressão final, íamos obter um intervalo de confiança para λ e ϕ. Assim íamos obter os limites para estes parâmetros. No fim, para obter o número máximo expectável de pedidos de indemnização, isto é, de $\lambda(1-\phi)$, multiplicávamos os limites do lado direito.

Distribuição de X_2

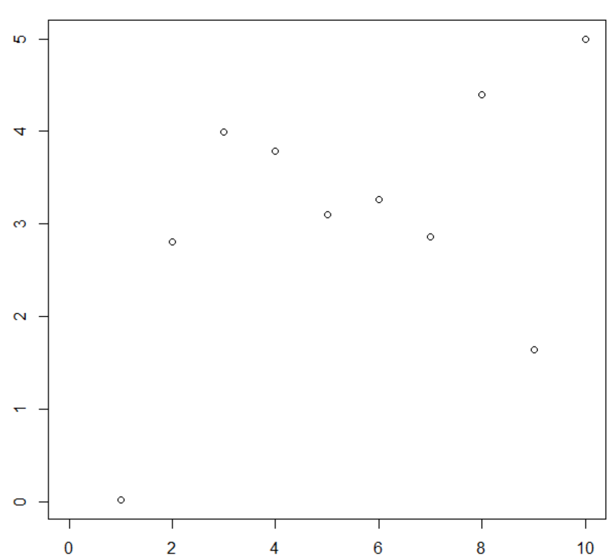
Comecemos por analisar a função de distribuição empírica de X_2.



A tabela seguinte indica o número de apólices em função dos pedidos de indemnização que a apresentaram:

No. pedidos (k)	0	1	2	3	4	5	6	7	8	9	10	Total
No. apólices (n_k)	5905	82	115	153	145	90	49	20	11	2	1	6573
$k\frac{n_k}{n_{k-1}}$	–	0.01	2.80	3.99	3.79	3.10	3.27	2.86	4.40	1.64	5.00	

A representação dos valores registados na última linha é

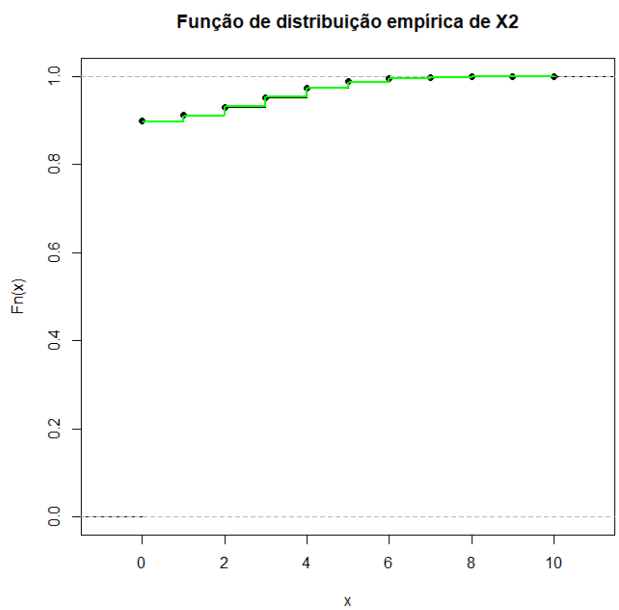


Esta representação é aproximadamente linear pelo que é aceitável que a distribuição subjacente seja de classe $(a, b, 0)$. A disposição dos pontos indica que o declive da reta subjacente, isto é, a aproximação para a, possa ser 0. Trata-se assim de uma distribuição que aparenta ser da família Poisson. No entanto, como podemos observar na tabela apresentada acima, n_0 é muito superior aos restantes n_k. Isto sugere que consideremos as distribuições "zero-inflanted Poisson". Para estimar os parâmetros da distribuição podemos utilizar o método dos momentos.

Analogamente ao estudo feito anteriormente vamos obter os estimadores

$$\widehat{\lambda} = \frac{\mathrm{Var}(X_2)}{\overline{X_2}} + \overline{X_2} - 1, \quad \widehat{\phi} = -\frac{\widehat{\lambda} - \overline{X_2}}{\widehat{\lambda}}.$$

Aplicando aos dados obtemos $\widehat{\lambda} = 3.33876$, $\widehat{\phi} = 0.8928718$. Uma comparação entre a função de distribuição empírica e a função de distribuição zero-inflated Poisson correspondente aos parâmetros obtidos é descrita na figura abaixo:



Como podemos observar, a distribuição zero-inflated Poisson parece aproximar bem a distribuição de X_2. No entanto, o teste de Anderson-Darling leva à rejeição desta distribuição (pois o p-valor obtido é inferior aos níveis de significância usuais).

Observação: Fiz várias tentativas para aproximar a distribuição de X_2 à distribuição zero-inflated Binomial mas não consegui determinar os estimadores pelo que optei por aceitar a distribuição zero-inflated Poisson para continuar o exercício.

Intervalo de confiança para os parâmetros de X_2

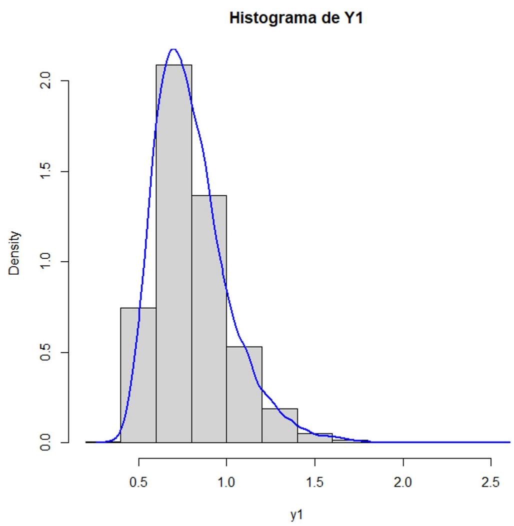
Análogo à análise com o mesmo comentário final.

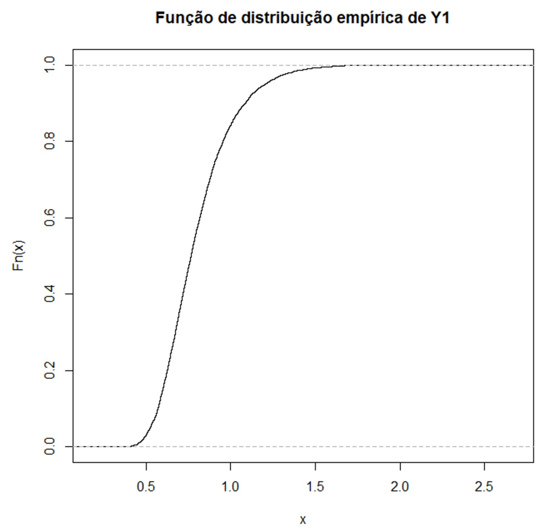
Exercício 2.

Queremos uma aproximação para as distribuições de Y_1 e Y_2.

Distribuição de Y_1

Comecemos por analisar o histograma e a função de distribuição empírica de Y_1 .





A análise da figura sugere que comparemos o comportamento da variável Y_1 com o das distribuições LogNormal, Gamma e Weibull. Para estimar os parâmetros de cada distribuição vamos utilizar o método da máxima verosimilhança.

– **Distribuição Lognormal** (μ, σ^2)

densidade $f(x) = \dfrac{1}{\sigma x\sqrt{2\pi}} \exp\left(-\dfrac{(\log x - \mu)^2}{2\sigma^2}\right)$

A função de verosimilhança é dada por

$$L(\mu, \sigma) = \frac{1}{\sigma^n(\sqrt{2\pi})^n \prod_{i=1}^n x_i} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^n (\log x_i - \mu)^2\right)$$

pelo que a função de log-verosimilhança é

$$l(\mu, \sigma) = -\sum_{i=1}^n \log x_i - n\log(\sigma) - n\log(\sqrt{2\pi}) - \frac{1}{2\sigma^2}\sum_{i=1}^n (\log x_i - \mu)^2.$$

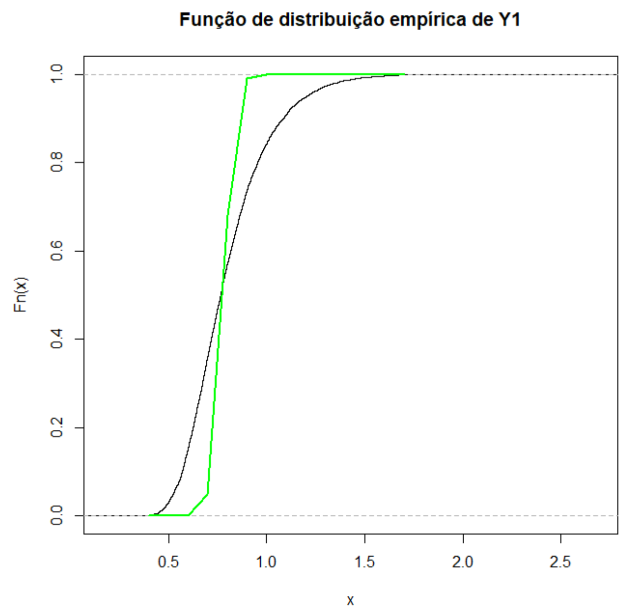
Derivando para obter as condições de primeira ordem encontramos o sistema de equações:

$$\begin{cases} \dfrac{1}{\sigma^2}\displaystyle\sum_{i=1}^n (\log x_i - \mu) = 0 \\ -\dfrac{n}{\sigma} + \dfrac{1}{\sigma^3}\displaystyle\sum_{i=1}^n (\log x_i - \mu)^2 = 0 \end{cases}$$

pelo que obtemos os estimadores

$$\widehat{\mu} = \frac{1}{n}\sum_{i=1}^{n} \log x_i, \quad \widehat{\sigma^2} = \frac{1}{n}\sum_{i=1}^{n}(\log x_i - \widehat{\mu})^2.$$

Aplicando aos dados obtemos $\widehat{\mu} \approx -0.2526797$ e $\widehat{\sigma^2} \approx 0.06306341$. Uma comparação entre a função de distribuição empírica e a função de distribuição Lognormal correspondente aos parâmetros obtidos é descrita na figura abaixo:



– **Distribuição Gamma** $\Gamma(\alpha, \theta)$

densidade $f(x) = \dfrac{x^{\alpha-1}}{\theta^\alpha \Gamma(\alpha)} e^{-\frac{x}{\theta}} \mathbb{I}_{[0,+\infty)}(x)$

A função de verosimilhança é dada por

$$L(\alpha, \theta) = \frac{\prod_{i=1}^{n} x_i^{\alpha-1} \exp(-\frac{1}{\theta}\sum_{i=1}^{n} x_i)}{\theta^{n\alpha}\Gamma(\alpha)^n}$$

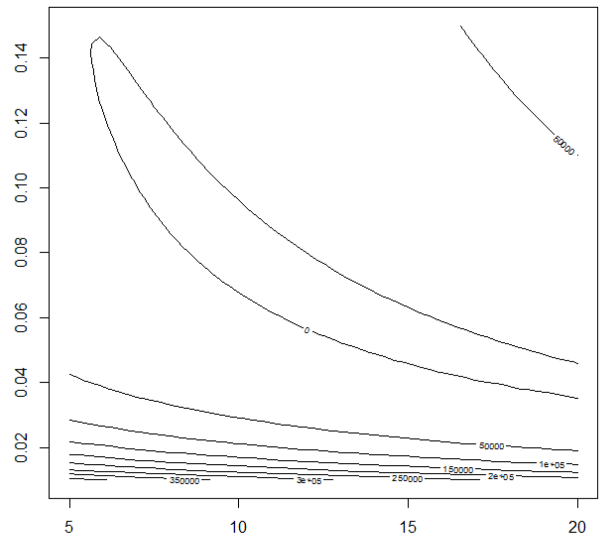
pelo que a função de log-verosimilhança é

$$l(\alpha, \theta) = (\alpha - 1)\sum_{i=1}^{n} \log x_i - n\alpha \log(\theta) - n \log(\Gamma(\alpha)) - \frac{1}{\theta}\sum_{i=1}^{n} x_i$$

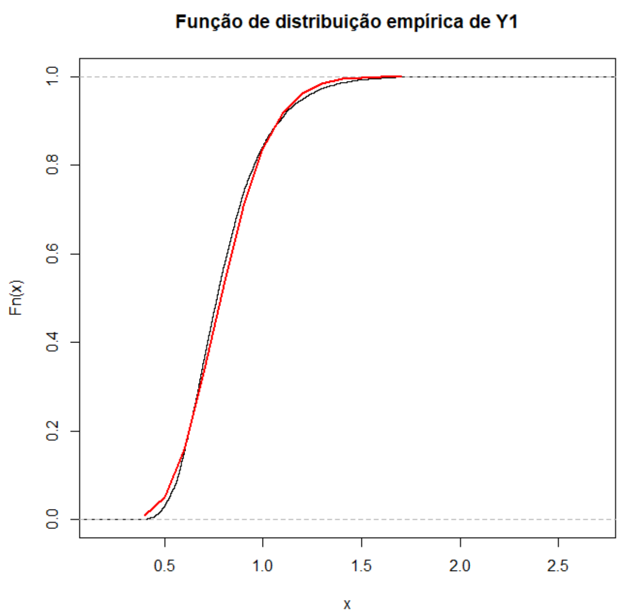
Derivando para obter as condições de primeira ordem encontramos o sistema de equações:

$$\begin{cases} \sum_{i=1}^{n} \log x_i - n\log(\theta) - n\dfrac{\Gamma'(\alpha)}{\Gamma(\alpha)} = 0 \\ -\dfrac{n\alpha}{\theta} + \dfrac{1}{\theta^2}\sum_{i=1}^{n} x_i = 0 \end{cases}$$

Este sistema é de difícil resolução analítica pelo que abordaremos a sua solução utilizando métodos numéricos. Por forma a determinar um ponto inicial para o método iterativo construímos o gráfico de curvas de nível para função de verosimilhança que se mostra ao lado. Esta representação indica que o máximo da função de verosimilhança ocorre quando α é um pouco maior do que 15 e θ perto de 0.05. Assim, adotamos como ponto de partida $(\alpha, \theta) = (10,\ 0.01)$.



Obtém-se como resultado a aproximação $(\widehat{\alpha}, \widehat{\theta}) \approx (15.51480321, 0.05172029)$.
Uma comparação entre a função de distribuição empírica e a função de distribuição Gamma correspondente aos parâmetros obtidos é descrita na figura abaixo:



– **Distribuição de Weibull** (τ, θ)

densidade $f(x) = \dfrac{\tau x^{\tau-1} \exp(-(\frac{x}{\theta})^{\tau})}{\theta^{\tau}} \mathbb{I}_{[0,+\infty)}(x)$

A função de verosimilhança é dada por

$$L(\theta, \tau) = \frac{\tau^n \prod_{i=1}^{n} x_i^{\tau-1} \exp(-\frac{1}{\theta^{\tau}} \sum_{i=1}^{n} x_i^{\tau})}{\theta^{n\tau}}$$

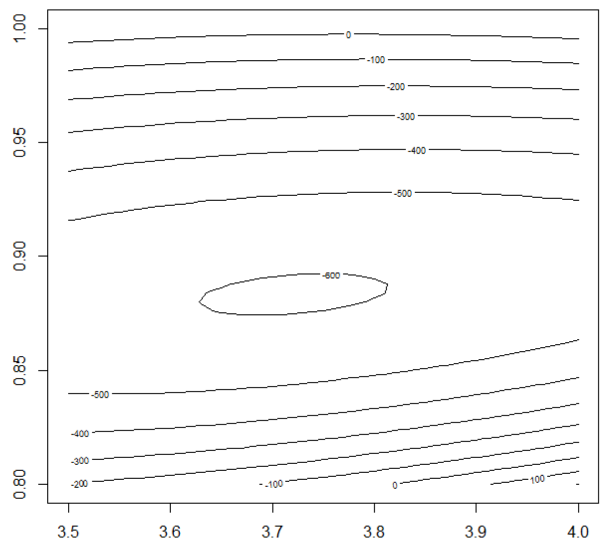
pelo que a função de log-verosimilhança é

$$l(\theta,\tau) = n\log(\tau) - n\tau\log(\theta) + (\tau-1)\sum_{i=1}^{n}\log x_i - \frac{1}{\theta^\tau}\sum_{i=1}^{n} x_i^\tau$$

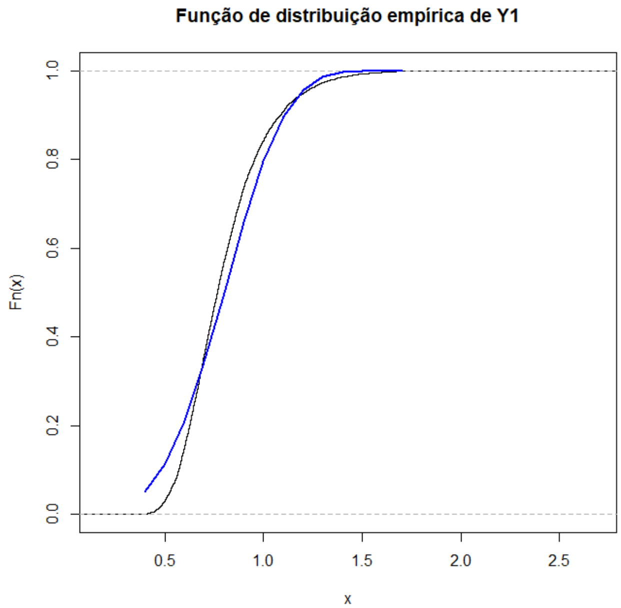
Derivando para obter as condições de primeira ordem encontramos o sistema de equações:

$$\begin{cases} \dfrac{n}{\tau} - n\log(\theta) + \displaystyle\sum_{i=1}^{n}\log x_i - \sum_{i=1}^{n}\log\left(\frac{x_i}{\theta}\right)\left(\frac{x_i}{\theta}\right)^\tau = 0 \\ -\dfrac{n\tau}{\theta} - \dfrac{\tau}{\theta}\displaystyle\sum_{i=1}^{n}\left(\frac{x_i}{\theta}\right)^\tau = 0 \end{cases}$$

Este sistema é de difícil resolução analítica pelo que abordaremos a sua solução utilizando métodos numéricos. Por forma a determinar um ponto inicial para o método iterativo construímos o gráfico de curvas de nível para função de verosimilhança que se mostra ao lado. Esta representação indica que o máximo da função de verosimilhança ocorre quando τ está perto de 3.7 e θ perto de 0.88. Assim, adotamos como ponto de partida $(\tau,\theta) =$ (3.5, 0.8).



Obtém-se como resultado a aproximação $(\widehat{\tau},\widehat{\theta}) \approx (3.7208823, 0.8832828)$.
Uma comparação entre a função de distribuição empírica e a função de distribuição de Weibull correspondente aos parâmetros obtidos é descrita na figura abaixo:



Como podemos observar, as distribuições que melhor se aproximam são a de Weibull e a Gamma. No entanto, os testes de Anderson-Darling e de Kolmogorov-Smirnov levam à rejeição de ambas as distribuições (pois os p-valores obtidos são inferiores aos níveis de significância usuais). Assim, fomos analisar as distribuições de Weibull inversa e Gamma inversa.

– **Distribuição Gamma inversa** $\Gamma(\alpha, \theta)$

densidade $f(x) = \dfrac{\theta^\alpha e^{-\frac{\theta}{x}}}{\Gamma(\alpha) x^{\alpha+1}} \mathbb{I}_{(0,+\infty)}(x)$

A função de verosimilhança é dada por

$$L(\alpha, \theta) = \frac{\theta^{n\alpha}}{\Gamma(\alpha)^n} \prod_{i=1}^{n} \frac{1}{x_i^{\alpha+1}} \exp\left(-\theta \sum_{i=1}^{n} \frac{1}{x_i}\right)$$

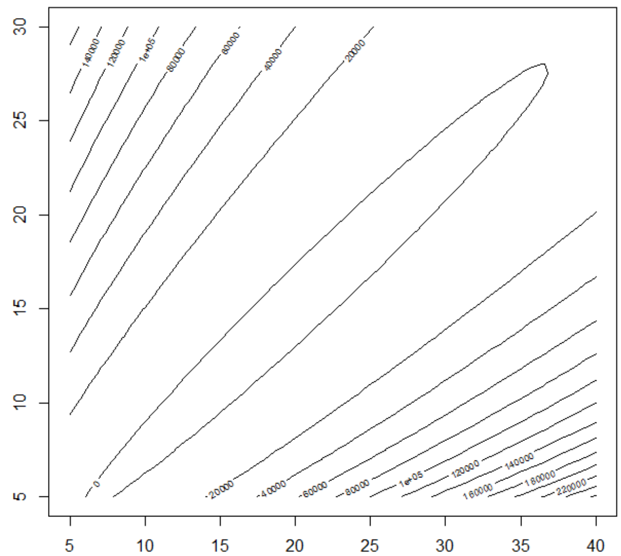
pelo que a função de log-verosimilhança é

$$l(\alpha, \theta) = n\alpha \log(\theta) - (\alpha + 1) \sum_{i=1}^{n} \log x_i - n \log(\Gamma(\alpha)) - \sum_{i=1}^{n} \frac{\theta}{x_i}$$

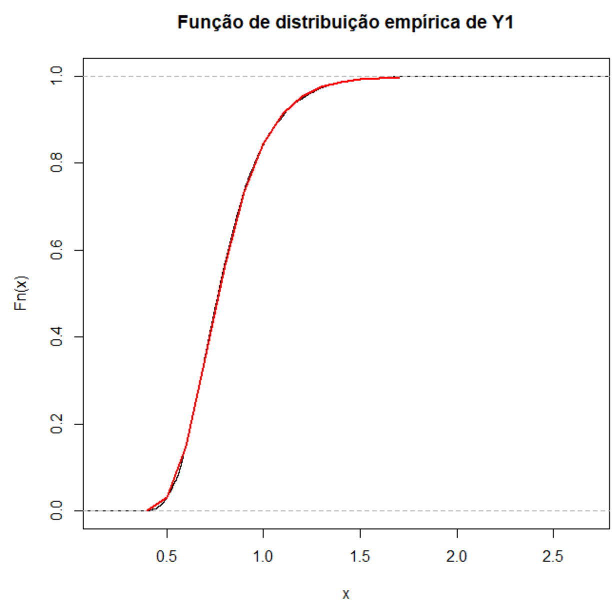
Derivando para obter as condições de primeira ordem encontramos o sistema de equações:

$$\begin{cases} n \log(\theta) + \displaystyle\sum_{i=1}^{n} \log x_i - n \frac{\Gamma'(\alpha)}{\Gamma(\alpha)} = 0 \\ \dfrac{n\alpha}{\theta} - \displaystyle\sum_{i=1}^{n} \frac{1}{x_i} = 0 \end{cases}$$

Este sistema é de difícil resolução analítica pelo que abordaremos a sua solução utilizando métodos numéricos. Por forma a determinar um ponto inicial para o método iterativo construímos o gráfico de curvas de nível para função de verosimilhança que se mostra ao lado. Esta representação indica que o máximo da função de verosimilhança ocorre quando α é um pouco maior do que 15 e θ perto de 10. Assim, adotamos como ponto de partida $(\alpha, \theta) = (5,\, 5)$.



Obtém-se como resultado a aproximação $(\widehat{\alpha}, \widehat{\theta}) \approx (16.41887, 12.36647)$.
Uma comparação entre a função de distribuição empírica e a função de distribuição Gamma inversa correspondente aos parâmetros obtidos é descrita na figura abaixo:



– Distribuição de Weibull inversa (τ, θ)

densidade $f(x) = \dfrac{\tau\theta^\tau \exp(-\left(\frac{\theta}{x}\right)^\tau)}{x^{\tau+1}} \mathbb{I}_{(0,+\infty)}(x)$

A função de verosimilhança é dada por

$$L(\theta, \tau) = \frac{\tau^n \theta^{n\tau} \exp(-\theta^\tau \sum_{i=1}^n \frac{1}{x_i^\tau})}{\prod_{i=1}^n x_i^{\tau+1}}$$

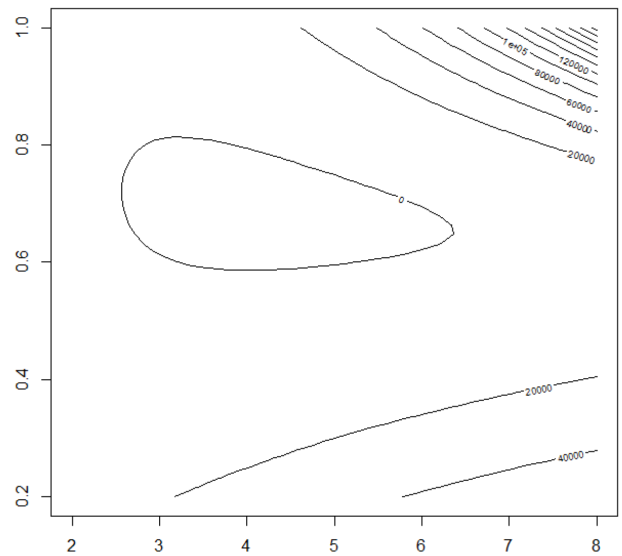
pelo que a função de log-verosimilhança é

$$l(\theta,\tau) = n\log(\tau) + n\tau\log(\theta) - (\tau+1)\sum_{i=1}^{n}\log x_i - \sum_{i=1}^{n}\left(\frac{\theta}{x_i}\right)^{\tau}$$

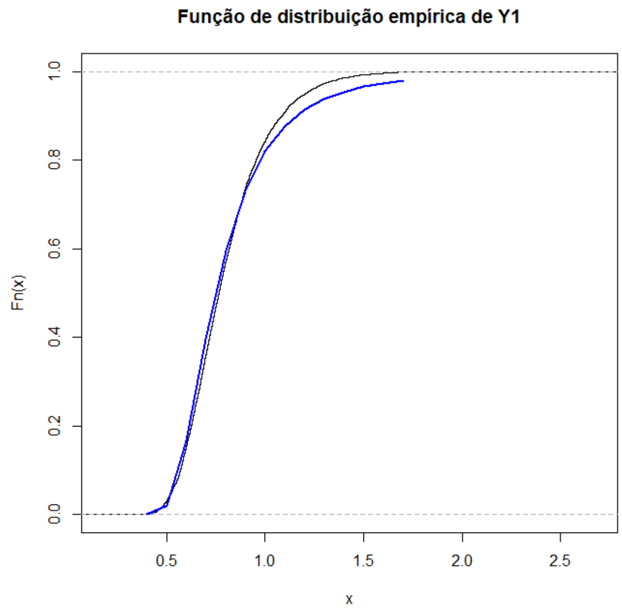
Derivando para obter as condições de primeira ordem encontramos o sistema de equações:

$$\begin{cases} \dfrac{n}{\tau} + n\log(\theta) - \displaystyle\sum_{i=1}^{n}\log x_i - \sum_{i=1}^{n}\log\left(\frac{\theta}{x_i}\right)\left(\frac{\theta}{x_i}\right)^{\tau} = 0 \\ \dfrac{n\tau}{\theta} - \dfrac{\tau}{\theta}\displaystyle\sum_{i=1}^{n}\left(\frac{\theta}{x_i}\right)^{\tau} = 0 \end{cases}$$

Este sistema é de difícil resolução analítica pelo que abordaremos a sua solução utilizando métodos numéricos. Por forma a determinar um ponto inicial para o método iterativo construímos o gráfico de curvas de nível para função de verosimilhança que se mostra ao lado. Esta representação indica que o máximo da função de verosimilhança ocorre quando τ está perto de 4 e θ perto de 0.7. Assim, adotamos como ponto de partida $(\tau,\theta) = (2,\, 0.2)$.



Obtém-se como resultado a aproximação $(\widehat{\tau},\widehat{\theta}) \approx (4.2958066, 0.6873333)$.
Uma comparação entre a função de distribuição empírica e a função de distribuição de Weibull inversa correspondente aos parâmetros obtidos é descrita na figura abaixo:

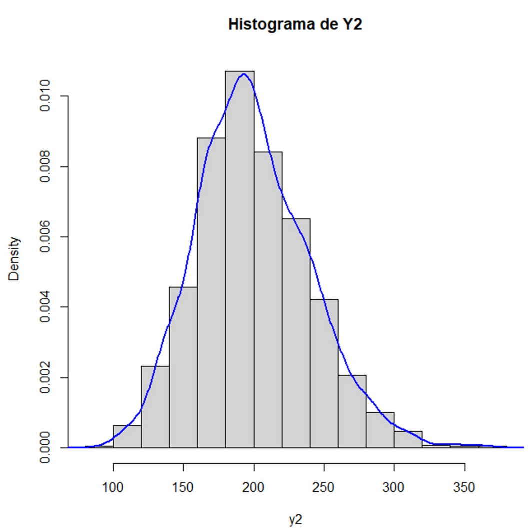


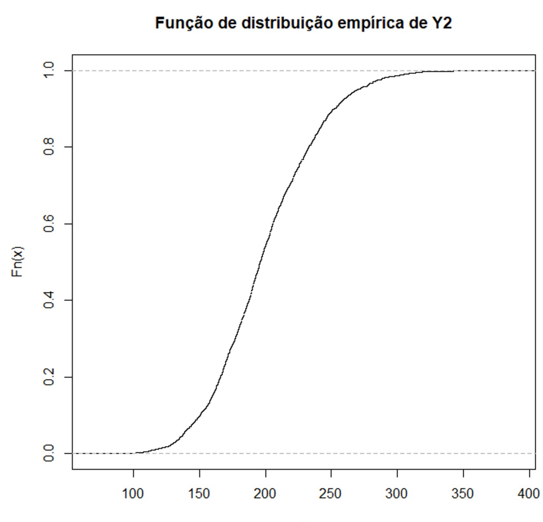
Como podemos observar, a distribuição Gamma inversa parece ser uma boa aproximação para a distribuição de Y_1. Realizando os testes de Anderson-Darling e de Kolmogorov-Smirnov somos agora levados a aceitar esta distribuição (pois os p-valores obtidos são superiores aos níveis de significância usuais).

Assim, concluímos que a distribuição de Y_1 é bem aproximada por uma distribuição Gamma inversa de parâmetros $(\alpha, \theta) \approx (16.41887, 12.36647)$.

Distribuição de Y_2

Comecemos por analisar o histograma e a função de distribuição empírica de Y_2.





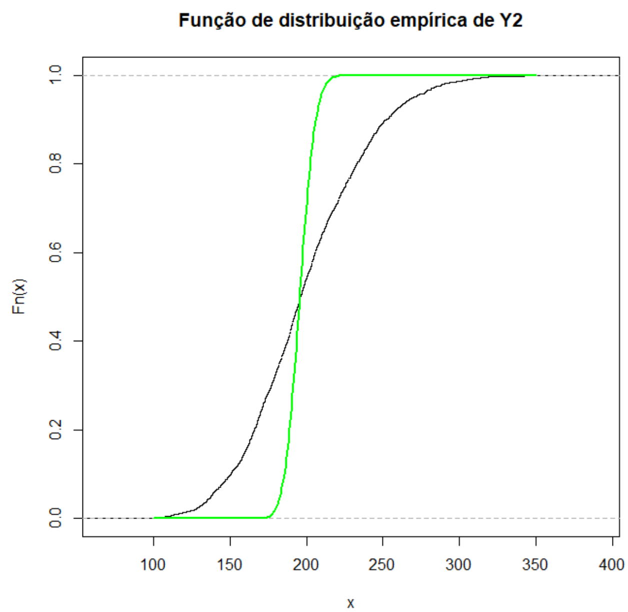
A análise da figura sugere que comparemos o comportamento da variável Y_2 com o das distribuições LogNormal, Gamma e Weibull. Para estimar os parâmetros de cada distribuição vamos utilizar o método da máxima verosimilhança.

– **Distribuição Lognormal** (μ, σ^2)

Analogamente à análise feita para a distribuição de Y_1, obtemos os estimadores

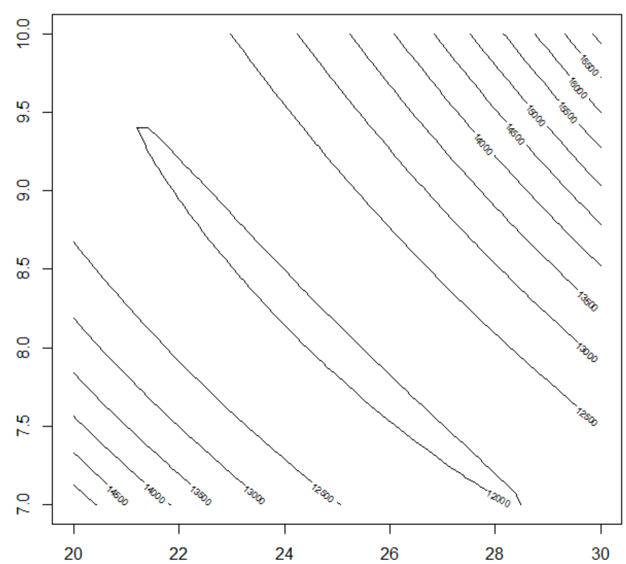
$$\widehat{\mu} = \frac{1}{n}\sum_{i=1}^{n} \log x_i, \quad \widehat{\sigma^2} = \frac{1}{n}\sum_{i=1}^{n}(\log x_i - \widehat{\mu})^2.$$

Aplicando aos dados obtemos $\widehat{\mu} \approx 5.276025$ e $\widehat{\sigma^2} \approx 0.04129443$. Uma comparação entre a função de distribuição empírica e a função de distribuição Lognormal correspondente aos parâmetros obtidos é descrita na figura abaixo:



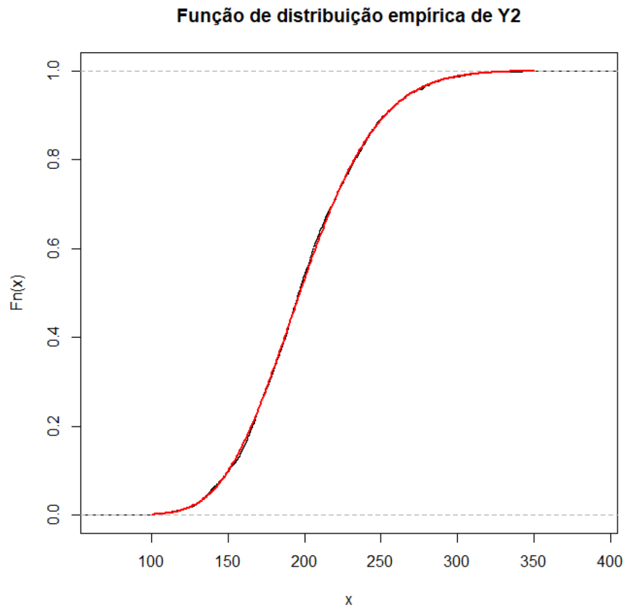
– **Distribuição Gamma** $\Gamma(\alpha, \theta)$

Analogamente à análise feita para a distribuição de Y_1, obtemos um sistema de difícil resolução analítica pelo que abordaremos a sua solução utilizando métodos numéricos. Por forma a determinar um ponto inicial para o método iterativo construímos o gráfico de curvas de nível para função de verosimilhança que se mostra ao lado. Esta representação indica que o máximo da função de verosimilhança ocorre quando α é um pouco maior do que 24 e θ perto de 8. Assim, adotamos como ponto de partida $(\alpha, \theta) = (20,\ 7)$.



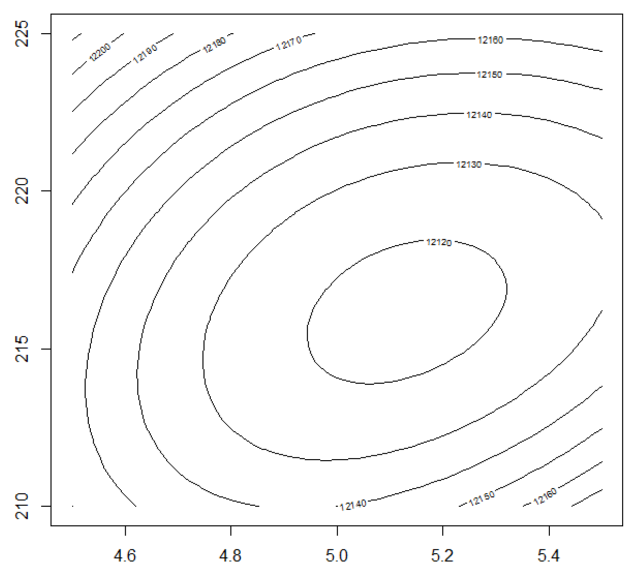
Obtém-se como resultado a aproximação $(\widehat{\alpha}, \widehat{\theta}) \approx (24.686715, 8.086131)$.

Uma comparação entre a função de distribuição empírica e a função de distribuição Gamma correspondente aos parâmetros obtidos é descrita na figura abaixo:

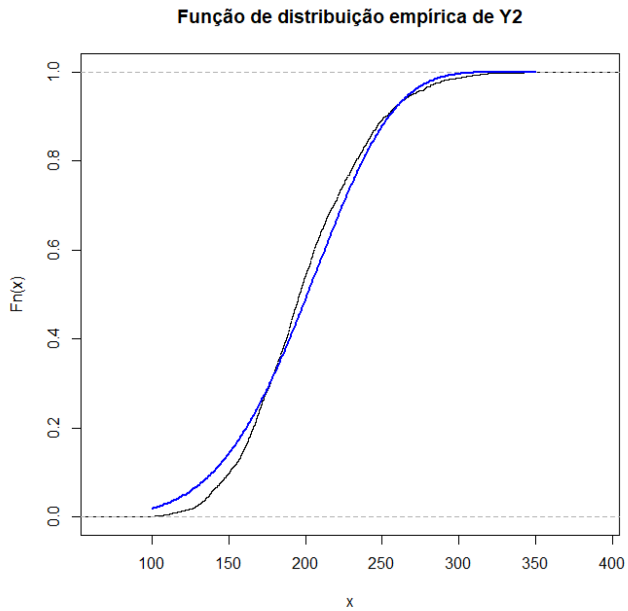


– Distribuição de Weibull (τ, θ)

Analogamente à análise feita para a distribuição de Y_1, obtemos um sistema de difícil resolução analítica pelo que abordaremos a sua solução utilizando métodos numéricos. Por forma a determinar um ponto inicial para o método iterativo construímos o gráfico de curvas de nível para função de verosimilhança que se mostra ao lado. Esta representação indica que o máximo da função de verosimilhança ocorre quando τ está perto de 5 e θ perto de 215. Assim, adotamos como ponto de partida $(\tau, \theta) = (4.5,\ 210)$.



Obtém-se como resultado a aproximação $(\widehat{\tau}, \widehat{\theta}) \approx (5.130443, 216.163986)$.
Uma comparação entre a função de distribuição empírica e a função de distribuição de Weibull correspondente aos parâmetros obtidos é descrita na figura abaixo:



Como podemos observar, as distribuições que melhor se aproximam são a de Weibull e a Gamma, especialmente a Gamma. Realizando os testes de Anderson-Darling e de Kolmogorov-Smirnov somos levados a aceitar a distribuição Gamma (pois os p-valores obtidos são superiores aos níveis de significância usuais para esta distribuição).

Assim, concluímos que a distribuição de Y_2 é bem aproximada por uma distribuição Gamma de parâmetros $(\alpha, \theta) \approx (24.686715, 8.086131)$.

Exercício 3.

Seja $Y = Y_1 + Y_2$. Queremos uma aproximação para $\mathrm{E}(Y)$ e $\mathrm{Var}(Y)$. Considerando as variáveis Y_1 e Y_2 independentes temos

$$\mathrm{E}(Y) = \mathrm{E}(Y_1) + \mathrm{E}(Y_2), \quad \mathrm{Var}(Y) = \mathrm{Var}(Y_1) + \mathrm{Var}(Y_2).$$

Se $X \sim$ Gamma inversa (α, θ) então

$$\begin{aligned}\mathrm{E}(X^k) &= \int_0^{+\infty} x^k \frac{\theta^\alpha e^{-\frac{\theta}{x}}}{\Gamma(\alpha) x^{\alpha+1}}\, dx = \frac{\theta^\alpha}{\Gamma(\alpha)} \int_0^{+\infty} x^{k-\alpha-1} e^{-\frac{\theta}{x}}\, dx \\ &=^* \frac{\theta^\alpha}{\Gamma(\alpha)} \int_0^{+\infty} \left(\frac{\theta}{t}\right)^{k-\alpha-1} e^{-t} \frac{\theta}{t^2}\, dt = \frac{\theta^k \Gamma(\alpha - k)}{\Gamma(\alpha)}, \quad \alpha \geq k\end{aligned}$$

pelo que, para $\alpha \geq k$,

$$\mathrm{E}(X) = \frac{\theta}{\alpha - 1}, \quad \mathrm{E}(X^2) = \frac{\theta^2}{(\alpha-1)(\alpha-2)}$$

$*$ mudança de variável: $t = \theta / x$

Se $X \sim$ Gamma (α, θ) então

$$\mathrm{E}(X^k) = \frac{\theta^k \Gamma(\alpha + k)}{\Gamma(\alpha)}$$

pelo que,

$$\mathrm{E}(X) = \theta\alpha, \quad \mathrm{E}(X^2) = \theta^2(\alpha+1)\alpha$$

Como $Y_1 \sim$ Gamma inversa $(16.41887, 12.36647)$ vem

$$\begin{aligned}
\mathrm{E}(Y_1) &= \frac{12.36647}{16.41887-1} \approx 0.8020347795 \\
\mathrm{Var}(Y_1) &= \mathrm{E}(Y_1^2) - [\mathrm{E}(Y_1)]^2 = \\
&= \frac{12.36647^2}{(16.41887-1)(16.41887-2)} - 0.8020347795^2 \\
&\approx 0.04461235774
\end{aligned}$$

Como $Y_2 \sim$ Gamma $(24.686715, 8.086131)$, temos

$$\begin{aligned}
\mathrm{E}(Y_2) &= 24.686715 \times 8.086131 \approx 199.6200114 \\
\mathrm{Var}(Y_2) &= \mathrm{E}(Y_2^2) - [\mathrm{E}(Y_2)]^2 = \\
&= 8.086131^2 \times (24.686715+1) \times 24.686715 - 199.6200114^2 \\
&\approx 1614.153583
\end{aligned}$$

Assim, vem

$$\begin{aligned}
\mathrm{E}(Y) &= 0.8020347795 + 199.6200114 \approx 200.4220462 \\
\mathrm{Var}(Y) &= 0.04461235774 + 1614.153583 \approx 1614.198195
\end{aligned}$$

Exercício 4.

Queremos uma aproximação para $\mathrm{VaR}_{0.95}(Y)$. $\mathrm{VaR}_{0.95}(Y)$ é solução de

$$\mathrm{P}(Y \leq \mathrm{VaR}_{0.95}(Y)) = 0.95$$

Pelo teorema do limite central vem

$$\mathrm{P}\left(\frac{Y - \mathrm{E}(Y)}{\sqrt{\mathrm{Var}(Y)}} \leq \frac{\mathrm{VaR}_{0.95}(Y) - \mathrm{E}(Y)}{\sqrt{\mathrm{Var}(Y)}}\right) = 0.95.$$

Como

$$Z = \frac{Y - \mathrm{E}(Y)}{\sqrt{\mathrm{Var}(Y)}} \sim N(0,1)$$

temos

$$\mathrm{P}\left(Z \leq \frac{\mathrm{VaR}_{0.95}(Y) - \mathrm{E}(Y)}{\sqrt{\mathrm{Var}(Y)}}\right) = 0.95 \Leftrightarrow \frac{\mathrm{VaR}_{0.95}(Y) - \mathrm{E}(Y)}{\sqrt{\mathrm{Var}(Y)}} \approx z_{0.95}$$

$$\mathrm{VaR}_{0.95}(Y) \approx z_{0.95}\sqrt{\mathrm{Var}(Y)} + \mathrm{E}(Y)$$

onde $z_{0.95}$ é o quantil de ordem 0.95 de Z.

Pelo exercício 3 vem

$$\mathrm{VaR}_{0.95}(Y) \approx 1.644854 \times \sqrt{1614.198195} + 200.4220462 \approx 266.5074859$$

Exercício 5.

Queremos calcular os valores de prémio máximos que os clientes estariam dispostos a pagar por cada uma das apólices vendidas, tendo em conta que utilizam a função de utilidade $u(x) = -e^{x/1000}$.

Dado um capital inicial w e um risco de prejuízo descrito por uma variável aleatória X, pelo princípio da utilidade nula, temos a seguinte relação

$$u(w - P^+) = \mathrm{E}(u(w - X))$$

onde P^+ representa o valor máximo que o cliente está disposto a pagar. Assim, obtemos

$$u(w - P^+) = \mathrm{E}(u(w - X)) \Leftrightarrow e^{P^+/1000} = \mathrm{E}(e^{X/1000})$$

$$\Leftrightarrow P^+ = 1000 \ln\Big(\mathrm{E}(e^{X/1000})\Big)$$

Queremos determinar P^+ quando $X = Y_1$ e $X = Y_2$.

Para a variável Y_1,

$$\mathrm{E}(e^{Y_1/1000}) = \int_0^{+\infty} e^{x/1000} \frac{\theta^\alpha e^{-\frac{\theta}{x}}}{\Gamma(\alpha) x^{\alpha+1}}\ dx$$

com $(\alpha, \theta) = (16.41887, 12.36647)$.

Utilizando a função integrate do R obtemos

$$\mathrm{E}(e^{Y_1/1000}) \approx 1.000802$$

Assim, o prémio máximo que os clientes estariam dispostos a pagar pela apólice 1 é

$$P_1^+ = 1000 \ln(1.000802) \approx 0.8016785698$$

Para a variável Y_2,

$$\mathrm{E}(e^{Y_2/1000}) = \int_0^{+\infty} e^{x/1000} \frac{x^{\alpha-1}}{\theta^\alpha \Gamma(\alpha)} e^{-\frac{x}{\theta}}\ dx$$

com $(\alpha, \theta) = (24.686715, 8.086131)$.

Utilizando a função integrate do R obtemos

$$\mathrm{E}(e^{Y_2/1000}) \approx 1.22193$$

Assim, o prémio máximo que os clientes estariam dispostos a pagar pela apólice 2 é

$$P_2^+ = 1000 \ln(1.22193) \approx 200.4316$$

Se o prémio a cobrar disser respeito às duas apólices em conjunto estamos interessados em calcular P^+ quando $X = Y$. Notemos que

$$\mathrm{E}(e^{Y/1000}) = M_Y\left(\frac{1}{1000}\right).$$

Como $Y = Y_1 + Y_2$ e Y_1 e Y_2 independentes então

$$\mathrm{E}(e^{Y/1000}) = M_{Y_1}\left(\frac{1}{1000}\right) \times M_{Y_2}\left(\frac{1}{1000}\right) = \mathrm{E}(e^{Y_1/1000}) \times \mathrm{E}(e^{Y_2/1000})$$

Assim, vem

$$P^+ = 1000 \ln\left(\mathrm{E}(e^{Y_1/1000}) \times \mathrm{E}(e^{Y_2/1000})\right) = 1000\left[\ln\left(\mathrm{E}(e^{Y_1/1000})\right) + \ln\left(\mathrm{E}(e^{Y_2/1000})\right)\right]$$

pelo que o prémio máximo que os clientes estariam dispostos a pagar pelas duas apólices em conjunto é

$$P^+ = P_1^+ + P_2^+ \approx 201.2332545$$

Exercício 6.

Se a seguradora utilizar uma política de reembolsos para as apólices de tipo 1 impondo uma franquia $d = 10$, estamos perante um contrato do tipo *stop-loss*, isto é, os reembolsos são descritos por

$$(Y_1 - 10)_+ = \begin{cases} 0, & \text{se } Y_1 \leq 10 \\ Y_1 - 10, & \text{se } Y_1 > 10 \end{cases}$$

Queremos determinar $\mathrm{E}\left[(Y_1 - 10)_+\right]$. Como Y_1 tem distribuição Gamma inversa vem

$$\mathrm{E}\left[(Y_1 - 10)_+\right] = \int_{10}^{+\infty} (x - 10) \times \frac{\theta^\alpha e^{-\frac{\theta}{x}}}{\Gamma(\alpha) x^{\alpha+1}}\, dx$$

com $(\alpha, \theta) = (16.41887, 12.36647)$

Utilizando a função integrate do R obtemos

$$\mathrm{E}\left[(Y_1 - 10)_+\right] \approx 1.04738 \times 10^{-13}$$