Testes de Hipóteses Bilaterais no contexto da $Cauchy(heta, 1)$

Pedro Henrique Corrêa de Almeida

Janeiro de 2023, Juiz de Fora

Conteúdo

1 Introdução

Um teste de hipótese é um método estatístico com o objetivo de, baseado em evidências, chegar a conclusões acerca de uma hipótese particular. Nesse contexto, os testes de hipóteses são fundamentais em um estudo estatístico, uma vez que estes dão um embasamento e são responsáveis por justificar nossas decisões dentro de um estudo.

Por (BOLFARINE; SANDOVAL, 2001), temos a seguinte definição:

Definição 1.1: Chamamos de hipótese estatística qualquer afirmação acerca da distribuição de probabilidades de uma ou mais variáveis aleatórias.

1.1 Motivação

Ao se trabalhar com testes de hipóteses, muitas vezes, estamos interessados em avaliar hipóteses bilaterais, no entanto, para esses tipos de hipótese não existem testes uniformemente mais poderosos, nesse caso o mais comum é se trabalhar com o método de razão de verossimilhança generalizada, porém existem outros métodos que podem ser utilizados.

Nesse sentido, este trabalho tem o objetivo de estudar esses diferentes métodos, para isso vamos estudar no contexto da distribuição $Cauchy(\theta, 1)$.

1.2 Métodos

Os métodos que serão avaliados nesse estudos são os testes de Wald, Escore e Razão de Verossimilhanças. Um teste terá desempenho quando apresentar baixas probabilidades dos erros do tipo I e II, ou seja, que a hipótese nula tenha baixa probabilidade de rejeição esta for verdadeira, e alta probabilidade caso contrátrio.

Nesse sentido, iremos realizar um estudo de simulação e avaliar empiricamente quais as proporções de vezes que a hipótese nula foi rejeitada em relação ao verdadeiro valor do parâmetro de interesse. Para isso, o método de Monte Carlo será utilizado para estimar o nível de significância, 5000 amostras independentes serão geradas de acordo com o modelo definido, sob a hipótese nula de interesse, e tamanhos amostrais $n = 10,\ 50,\ 100,\ 500,\ 1000,\ 5000\ e\ 10000$.

Nesse estudo de simulação, vamos considerar um nível de significância de 5, aonde as hipóteses serão:

$$H_0 : \theta = 0$$

$$H_1 : \theta \neq 0$$

Além disso, vamos realizar o estudo para diferentes valores do parâmetro de interesse, θ, serão estes:

$$\theta = -5,\ -4,\ -3,\ -2,\ -1,\ 0,\ 1,\ 2,\ 3,\ 4,\ 5$$

1.3 Algoritmo EM

Neste trabalho, vamos trabalhar com a distribuição $Cauchy(\theta, 1)$, logo vamos precisar utilizar métodos iterativos na obtenção do estimador de máxima verossimilhança, uma vez que não conseguimos chegar analiticamente em uma expressão para o estimador de máxima verossimilhança.

Nesse contexto, vamos utilizar o Algoritmo EM (DEMPSTER; LAIRD; RUBIN, 1977). Esse algoritmo tem como base a reformulação do problema utilizando uma formulação de dados aumentados a fim de simplificar a obtenção das estimativas.

Dessa forma, vamos considerar $\mathbf{x}_{obs}$ os dados observados, $\mathbf{x}_{mis}$ os dados não-observáveis, e $\mathbf{x}_{x_c} = (\mathbf{x}_{obs}, \mathbf{x}_{mis})$.

Consideramos $l_c(\theta|\mathbf{x}_{x_c}) = log(f(\mathbf{x}_{obs}))$ a função $log\text{-}verossimilhana$ dos dados completos, e, $Q(\theta|\hat{\theta}) = E(l_c(\theta|\mathbf{x}_{x_c}))$.

A fim de encontrar os estimadores de máxima verossimilhança, cada iteração do algoritmo consiste em dois passos, são estes:

Passo E: Para $\theta = \hat{\theta}^{(t)}$, calcular:

$$Q(\theta|\hat{\theta}) = E[l_c(\theta|\mathbf{x}_{x_c}, \hat{\theta}^{(t)})]$$

Passo M: Obter $\hat{\theta}^{(t+1)}$ que mazimiza $Q(\theta|\hat{\theta})$, tal que

$$Q(\hat{\theta}^{(t+1)}|\hat{\theta}^{(t)}) > Q(\theta|\hat{\theta}^{(t)})$$

Com isso nós alternamos os passos E e M até atingir a convergência, que pode ser medida a partir de um critério de parada. Neste estudo o critério de parada utilizado foi:

$$\frac{l(\widehat{\theta}^{(k+1)}) - l(\widehat{\theta}^{(k)})}{l(\widehat{\theta}^{(k)})} < \epsilon$$

2 Distribuição Cauchy

2.1 Introdução

Uma variável aleatória X tem distribuição $Cauchy$ com parâmetro de locação $\theta \; \epsilon \; \mathbb{R}$, se a função de densidade de probabilidades é dada por:

$$f_X(x) = \frac{1}{\pi[1 + (x-\theta)^2]} I_{(-\infty,\infty)}(x)$$

Além disso, uma variável aleatória X tem distribuição $Cauchy(\theta, 1)$ tem representação estocástica:

$$X = \theta + U^{-\frac{1}{2}} Z$$

onde $Z \sim N(0,1)$ e $U \sim Gama(\frac{1}{2}, \frac{1}{2})$

Dessa forma, temos a função de verossimilhança como:

$$L(\theta) = \pi^{-n} \prod_{i=1}^{n} (1 + (x-\theta)^2)^{-1}$$

Consequentemente, a log-$verossmilhana$:

$$l(\theta) = log(\pi^{-n} \prod_{i=1}^{n} (1 + (x_i - \theta)^2)^{-1}) l(\theta) = -nlog(\pi) + \sum_{i=1}^{n} -log(1 + (x_i - \theta)^2)$$

Derivando $l(\theta)$ em função de θ, temos a função $Escore$ como:

$$S(\theta) = \sum_{i=1}^{n} \frac{-2(x_i - \theta)}{(1 + (x_i - \theta)^2)}$$

2.2 Informação de Fisher Esperada

$$I_1(\theta) = -E[(\sum_{i=1}^{n} \frac{-2(X_i - \theta)}{(1 + (X_i - \theta)^2)})^2]$$

$$I_1(\theta) = -\sum_{i=1}^{n} E[(\frac{4(X_i - \theta)^2}{[(1 + (X_i - \theta)^2)]^2})]$$

Pela definição do valor esperado, temos:

$$E[\frac{4(X - \theta)^2}{[(1 + (X - \theta)^2)]^2}] = \int_{-\infty}^{\infty} f(x) \frac{4(x - \theta)^2}{[(1 + (x - \theta)^2)]^2} dx$$

Seja $u = x - \theta$, temos $du = -dx$, logo:

$$E[\frac{4(X - \theta)^2}{[(1 + (X - \theta)^2)]^2}] = \frac{-4}{\pi} \int_{-\infty}^{\infty} \frac{u^2}{[(1 + u^2)]^3} du$$

$$E[\frac{4(X - \theta)^2}{[(1 + (X - \theta)^2)]^2}] = \frac{-4}{\pi} \frac{\pi}{8} = \frac{-1}{2}$$

Dessa forma, concluímos que:

$$I_1(\theta) = -E[\frac{4(X - \theta)^2}{[(1 + (X - \theta)^2)]^2}] = \frac{1}{2}$$

$$I(\theta) = \frac{n}{2}$$

Com isso, pelo teorema da consistência do estimador de máxima verossimilhança, sabemos que:

$$\hat{\theta} \sim^a N(\theta, \frac{1}{I_1(\theta) n})$$

$$\hat{\theta} \sim^a N(\theta, \frac{2}{n})$$

2.3 Algoritmo EM

A fim de utilizar o Algoritmo EM no contexto da $Cauchy$, vamos partir de sua representação estocástica:

$$X = \theta + U^{-\frac{1}{2}}Z$$

onde $Z \sim N(0,1)$ e $U \sim Gama(\frac{1}{2}, \frac{1}{2})$

Seja $X_1, ..., X_n$ uma amostra aleatória simples de X, temos que estas serão independentes, logo como a distribuição $Cauchy$ pertence à família escala, temos:

$$X_i|U_i = u_i \sim N(\theta, \frac{1}{u_i})$$

2.3.1 Etapa E

A fim de encontrar a distribuição de $U_i|X_i = x_i$ vamos utilizar o teorema de geral da probabilidade, onde:

$$f(u_i|x_i) = \frac{f(u_i, x_i)}{f(x_i)} = \frac{f(x_i|u_i)f(u_i)}{f(x_i)}$$

$$f(u_i|x_i) = \frac{(2\pi\ \frac{1}{u_i})^{\frac{-1}{2}} e^{-\frac{u_i}{2}(x_i-\theta)^2} \sqrt{\frac{1}{2}} \frac{1}{\Gamma(\frac{1}{2})\sqrt{u_i}} e^{\frac{u_i}{2}}}{\frac{1}{\pi[1+(x_i-\theta)^2]}}$$

Como $\Gamma(\frac{1}{2}) = \sqrt{\pi}$

$$f(u_i|x_i) = \frac{\sqrt{\frac{1}{2}}\sqrt{u_i}\ e^{-\frac{u_i}{2}(x_i-\theta)^2 + \frac{u_i}{2}}}{\frac{\sqrt{2\pi}}{\pi[1+(x_i-\theta)^2]}\sqrt{\pi}\sqrt{u_i}}$$

$$f(u_i|x_i) = \frac{e^{-\frac{u_i}{2}[1+(x_i-\theta)^2]}}{\frac{2}{[1+(x_i-\theta)^2]}}$$

Seja $d_i = (x_i - \theta)$, temos:

$$f(u_i|x_i) = \frac{e^{-\frac{u_i}{2}[1+d_i^2]}}{\frac{1}{[1+d_i^2]}}$$

$$f(u_i|x_i) = \frac{[1+d_i^2]}{2} e^{-\frac{u_i}{2}[1+d_i^2]}$$

Ou seja, temos, pela função de densidade, que $U_i|X_i = x_i \sim Gama(1, \frac{1+d_i^2}{2})$

Dessa forma, temos que o valor de esperado dado X será:

$$\hat{u_i} = E(U_i|X_i) = \frac{2}{1+d_i^2}$$

2.3.2 Etapa M

Após a Etapa E, temos a maximização da verossimilhança da função $Q(\theta) = E[l_c(\theta)]$, dessa forma, uma vez que $X_i|U_i = u_i \sim N(\theta, \frac{1}{u_i})$

$$Q(\theta) = E[c - \frac{1}{2}\sum_{i=1}^{n} U_i(x_i - \theta)^2]$$

$$Q(\theta) = -\frac{1}{2}\sum_{i=1}^{n}(\hat{u_i}x_i^2 - 2\hat{u_i}x_i\theta + \hat{u_i}\theta^2)$$

Derivando a função Q a fim de encontrar o ponto de máximo:

$$\frac{dQ(\theta)}{d\theta} = -\frac{1}{2}\sum_{i=1}^{n}(-2\hat{u_i}x_i + 2\hat{u_i}\theta)$$

Igualando a expressão a 0, temos:

$$-\frac{1}{2}\sum_{i=1}^{n}(-2\hat{u_i}x_i + 2\hat{u_i}\hat{\theta}) = 0$$

$$\sum_{i=1}^{n}\hat{u_i}x_i = \sum_{i=1}^{n}\hat{u_i}\hat{\theta}$$

$$\widehat{\theta} = \frac{\sum_{i=1}^{n} \hat{u}_i x_i}{\sum_{i=1}^{n} \hat{u}_i}$$

A fim de confirmar se $\widehat{\theta}$ maximiza $Q(\theta)$, vamos avaliar a segunda derivada:

$$\frac{d^2Q(\theta)}{d\theta^2} = -\sum_{i=1}^{n} \hat{u}_i$$

Uma vez que $\hat{u_i} = \frac{2}{1+d_i^2}$, este só assume valores positivos, logo $\frac{d^2Q(\theta)}{d\theta^2}$ será negativa para todo θ, ou seja, $\widehat{\theta} = \frac{\sum_{i=1}^{n} \hat{u}_i x_i}{\sum_{i=1}^{n} \hat{u_i}}$ é ponto de máximo.

Ou seja, a partir do chute inicial de $\widehat{\theta}$, vamos estimar $\hat{u_i}$, com isso vamos estimar $\widehat{\theta}$ e repetir esse processo iterativamente até atingir a convergência.

2.3.3 EM Cauchy

Dessa forma, vamos utilizar o Algoritmo EM da seguinte forma:

$\widehat{\theta}^{(1)} = c$

$k = 1$

Enquanto critério > ϵ:

$\hat{u_i}^{(k)} = \frac{2}{1+(x_i-\widehat{\theta}^{(k)})^2}$

$\widehat{\theta}^{(k+1)} = \frac{\sum_{i=1}^{n} \hat{u_i}^{(k)} x_i}{\sum_{i=1}^{n} \hat{u_i}^{(k)}}$

$k = k+1$

Neste estudo vamos utilizar o critério de parada como $\frac{l(\widehat{\theta}^{(k+1)})-l(\widehat{\theta}^{(k)})}{l(\widehat{\theta}^{(k)})}$, $\epsilon = 10^{-10}$ e $c = Mediana(x)$, o chute inical.

2.4 Estatística Wald

A estatística Wald pode ser definida como:

$$Q_W = \frac{n(\hat{\theta} - \theta_0)}{I^{-1}(\hat{\theta})}$$

Essa estatística pode ser utilizada na construções de testes de hipóteses assintóticos, uma vez que temos o resultado que, sob H_0:

$$Q_W \sim^{n \to \infty} \chi^2_1$$

No caso da $Cauchy(\theta, 1)$, temos que pode ser calculada como:

$$Q_W = \frac{n(\hat{\theta} - \theta_0)}{2}$$

Com isso vamos rejeitar H_0 se:

$$R_c = \{x,\ Q_W \leq \chi^2_{1;1-\frac{\alpha}{2}}\ \ ou\ \ Q_W \geq \chi^2_{1;\frac{\alpha}{2}}\}$$

2.5 Estatística Escore

Da mesma forma que a estatística Wald, a estatística Escore pode ser utilizada na obtenção de testes assintóticos. Esta é definida da seguinte maneira:

$$Q_S = \frac{S(\theta_0)^2}{I(\theta_0)}$$

E da mesma forma, temos o resultado assintótico sob H_0:

$$Q_W \sim^{n \to \infty} \chi^2_1$$

No caso da $Cauchy(\theta, 1)$, temos:

$$Q_S = \frac{S(\theta_0)^2}{I(\theta_0)} = \frac{2}{n}[\sum_{i=1}^{n} \frac{-2(x_i - \theta_0)}{(1 + (x_i - \theta_0)^2)}]^2$$

Com isso vamos rejeitar H_0 se:

$$R_c = \{x,\ Q_S \leq \chi^2_{1;1-\frac{\alpha}{2}}\ \ ou\ \ Q_S \geq \chi^2_{1;\frac{\alpha}{2}}\}$$

2.6 Razão de Verossimilhança Generalizada

A razão de verossimilhança generalizada é o teste mais utilizada e tem como base o teste da razão de verossimilhança usual. No entanto, como estamos trabalhando com testes bilaterais e temos uma gama de valores possíveis na hipótese alternativa, essa razão será definida como:

$$\lambda(x) = \frac{sup_{\ \theta\ \in\ \Theta_0}\ L(\theta)}{sup_{\ \theta\ \in\ \Theta}\ L(\theta)}$$

Onde $\Theta_0 = \{\theta,\ \theta = \theta_0\} = \{\theta,\ \theta = 0\}$
e $\Theta = \theta,\ \theta \neq 0$

Ou seja, pela definição dos estimadores de verossimilhança, temos que:

$$\lambda(x) = \frac{L(\theta_0)}{L(\hat{\theta})}$$

Aplicando a função $log(.)$, temos que:

$$\log[\lambda(x)] = l(\theta_0) - l(\hat{\theta})$$

Dessa forma, nós vamos utilizar o seguinte resultado:

$$2\ log[\lambda] = 2[l(\theta_0) - l(\hat{\theta})]\ \sim\ \chi^2_1$$

Com isso vamos rejeitar H_0 se:

$$R_c = \{x,\ Q_R \leq \chi^2_{1;1-\frac{\alpha}{2}}\ \ ou\ \ Q_R \geq \chi^2_{1;\frac{\alpha}{2}}\}$$

3 Estudos de Simulação

Como foi estabelecido anteriormente, a fim de estudar o desempenho dos testes assintóticos vamos realizar um estudo de simulação para diferentes tamanhos amostrais e diferentes valores de θ.

3.1 Estimadores de máxima verossimilhança

Primeiramente, após gerar 5000 vezes as amostras para cada tamanho amostral e cada valor de θ vamos avaliar o comportamento do Algoritmo EM na obtenção dos estimadores de máxima verossimilhança.

Tamanho amostral	θ	Viés($\hat{\theta}$)	s^2	$\frac{1}{nI_1(\hat{\theta})}$	Tamanho amostral	θ	Viés($\hat{\theta}$)	s^2	$\frac{1}{nI_1(\hat{\theta})}$
10	-5	0.0111	0.2923	0.200	500	1	4e-04	0.0039	0.004
10	-4	-0.0102	0.3073	0.200	500	2	4e-04	0.0041	0.004
10	-3	0.0021	0.2940	0.200	500	3	4e-04	0.0039	0.004
10	-2	0.0009	0.2943	0.200	500	4	3e-04	0.004	0.004
10	-1	0.0043	0.2867	0.200	500	5	-0.0016	0.004	0.004
10	0	-0.0045	0.3017	0.200	1000	-5	-4e-04	0.002	0.002
10	1	-0.0038	0.2864	0.200	1000	-4	5e-04	0.0021	0.002
10	2	0.0097	0.2885	0.200	1000	-3	-2e-04	0.002	0.002
10	3	-0.0147	0.2976	0.200	1000	-2	3e-04	0.002	0.002
10	4	0.0098	0.2752	0.200	1000	-1	7e-04	0.002	0.002
10	5	0.0150	0.3067	0.200	1000	0	2e-04	0.002	0.002
50	-5	-0.0013	0.0421	0.040	1000	1	-5e-04	0.002	0.002
50	-4	-0.0003	0.0435	0.040	1000	2	2e-04	0.002	0.002
50	-3	-0.0010	0.0428	0.040	1000	3	-9e-04	0.002	0.002
50	-2	0.0041	0.0433	0.040	1000	4	0.001	0.002	0.002
50	-1	-0.0029	0.0433	0.040	1000	5	-4e-04	0.002	0.002
50	0	-0.0014	0.0422	0.040	5000	-5	-6e-04	4e-04	4e-04
50	1	-0.0035	0.0439	0.040	5000	-4	2e-04	4e-04	4e-04
50	2	0.0057	0.0433	0.040	5000	-3	3e-04	4e-04	4e-04
50	3	0.0023	0.0422	0.040	5000	-2	-2e-04	4e-04	4e-04
50	4	0.0002	0.0422	0.040	5000	-1	-1e-04	4e-04	4e-04
50	5	-0.0026	0.0416	0.040	5000	0	-2e-04	4e-04	4e-04
100	-5	-0.0007	0.0200	0.020	5000	1	1e-04	4e-04	4e-04
100	-4	-0.0015	0.0212	0.020	5000	2	-1e-04	4e-04	4e-04
100	-3	-0.0018	0.0212	0.020	5000	3	2e-04	4e-04	4e-04
100	-2	0.0022	0.0200	0.020	5000	4	-1e-04	4e-04	4e-04
100	-1	0.0037	0.0205	0.020	5000	5	0	4e-04	4e-04
100	0	0.0049	0.0210	0.020	10000	-5	-1e-04	2e-04	2e-04
100	1	0.0030	0.0213	0.020	10000	-4	-1e-04	2e-04	2e-04
100	2	-0.0035	0.0208	0.020	10000	-3	0	2e-04	2e-04
100	3	0.0029	0.0198	0.020	10000	-2	1e-04	2e-04	2e-04
100	4	-0.0021	0.0202	0.020	10000	-1	2e-04	2e-04	2e-04
100	5	-0.0022	0.0203	0.020	10000	0	-1e-04	2e-04	2e-04
500	-5	-0.0014	0.0041	0.004	10000	1	2e-04	2e-04	2e-04
500	-4	-0.0003	0.0042	0.004	10000	2	1e-04	2e-04	2e-04
500	-3	-0.0005	0.0039	0.004	10000	3	0	2e-04	2e-04
500	-2	0.0012	0.0039	0.004	10000	4	-1e-04	2e-04	2e-04
500	-1	-0.0013	0.0040	0.004	10000	5	-2e-04	2e-04	2e-04
500	0	0.0011	0.0040	0.004					

A partir da tabela acima podemos ver que os estimadores de máxima verossimilhança a partir do Algoritmo EM obtiveram um viés muito baixo e foram diminuindo conforme o tamanho amostral aumentou. Além disso, vemos que a variância amostral dos estimadores foi muito próxima de $\frac{n}{I_1(\hat{\theta})}$.

3.2 Resultados

Após gerar as amostras e obter os estimadores de máxima verossimilhança, podemos calcular as estatísticas de Wald, Escore e Razão de Verossimilhança e comparar o resultado das três.

Para isso, vamos considerar um nível de significância de 5 e calcular qual a proporção de vezes que a hipótese nula foi rejeitada.

	n = 10			n = 50			n = 100			n = 500		
	W	E	Q	W	E	Q	W	E	Q	W	E	Q
$\theta =$ -5	0.9998	0.0136	0.9994	1.0000	0.9974	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
$\theta =$ -4	0.9992	0.1498	0.9976	1.0000	0.9994	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
$\theta =$ -3	0.9968	0.4008	0.9904	1.0000	0.9998	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
$\theta =$ -2	0.9722	0.5342	0.9380	1.0000	0.9998	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
$\theta =$ -1	0.5044	0.3370	0.4852	0.9950	0.9722	0.9918	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
$\theta = 0$	0.0870	0.0522	0.0584	0.0570	0.0454	0.0500	0.0508	0.0448	0.0478	0.0506	0.0492	0.0496
$\theta = 1$	0.5034	0.3266	0.4874	0.9942	0.9708	0.9890	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
$\theta = 2$	0.9696	0.5190	0.9340	1.0000	0.9996	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
$\theta = 3$	0.9976	0.4040	0.9884	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
$\theta = 4$	0.9998	0.1444	0.9988	1.0000	0.9990	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
$\theta = 5$	1.0000	0.0146	0.9998	1.0000	0.9978	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

	n = 1000			n = 5000			n = 10000		
	W	E	Q	W	E	Q	W	E	Q
$\theta =$ -5	1.0000	1.0000	1.000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
$\theta =$ -4	1.0000	1.0000	1.000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
$\theta =$ -3	1.0000	1.0000	1.000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
$\theta =$ -2	1.0000	1.0000	1.000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
$\theta =$ -1	1.0000	1.0000	1.000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
$\theta = 0$	0.0466	0.0466	0.046	0.0484	0.0482	0.0478	0.0522	0.0524	0.0526
$\theta = 1$	1.0000	1.0000	1.000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
$\theta = 2$	1.0000	1.0000	1.000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
$\theta = 3$	1.0000	1.0000	1.000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
$\theta = 4$	1.0000	1.0000	1.000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
$\theta = 5$	1.0000	1.0000	1.000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

A partir da tabela, vemos que, para um tamanho amostral baixo, $n = 10$ o teste que melhor obteve um resultado quando do tamanho do teste, ou seja, foi mais próximo do nível de significância, foi o teste de $Escore$, no entanto, este apresentou o poder com um comportamento muito ruim.

Por outro lado, o teste de Wald apresentou um poder alto, no entanto o tamanho do teste foi bem maior que o nível de significâcia. Por fim, temos que para $n = 10$ o teste de Razão de Verossimilhança, de forma geral, apresentou um melhor resultado, com um tamanho do teste mais próximo do nível de significância, e, um alto poder do teste.

Quando aumentamos o tamanho amostral, considerando $n = 50$, temos que o desempenho de todos os testes melhoram, havendo uma grande melhora, principalmente no poder do teste Escore. No entanto, ainda vemos que o melhor teste de forma geral foi o teste de Razão de Verossimilhança, com o poder convergindo para 1, e um tamanho do teste bem próximo de 5.

A partir de $n = 100$, vemos que em relação ao poder todos os testes convergem para 1, para todos os valores de $\theta \neq 0$ utilizados neste estudo. E vemos que o tamanho do teste converge para valores bem próximos do nível de significância.

3.3 Poder do teste

Após avaliar na tabela os resultados, vamos avaliar graficamente o poder dos testes. Dessa forma, vamos visualizar no gráfico a proporção de vezes que H_0 foi rejeitada em cada teste, onde cada gráfico se refere à um tamanho amostral.

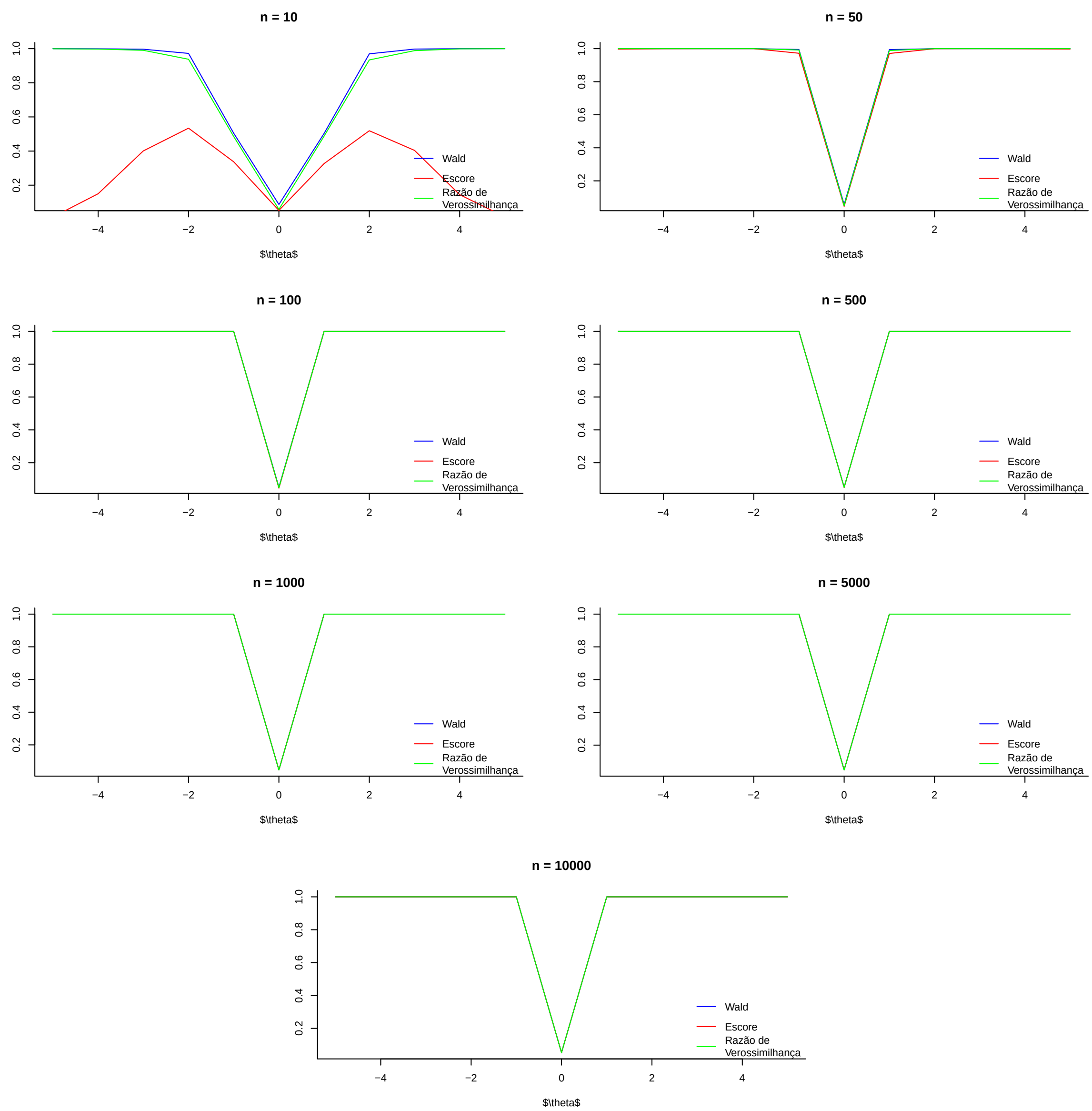

Figure 1: Poder

A partir do gráfico, podemos visualizar o que percebemos a partir da tabela. Onde o teste Escore obteve uma performance muito ruim para um tamanho amostral baixo, porém a partir de $n = 100$ os três testes apresentam um resultado semelhante.

3.4 Convergência

Por fim, vamos avaliar a convergência das estatísticas para a distribuição χ_1^2, uma vez que, utilizamos resultados assintóticos na construção dos testes.

3.4.1 Análise gráfica

Primeiramente, a fim de avaliar a convergência, vamos comparar a distribuição acumulada empírica com a respectiva distribuição teórica. Dessa forma vamos comparar os quantis das estatísticas sob H_0, ou seja, $\theta = 0$ com os quantis teóricos da distribuição χ_1^2.

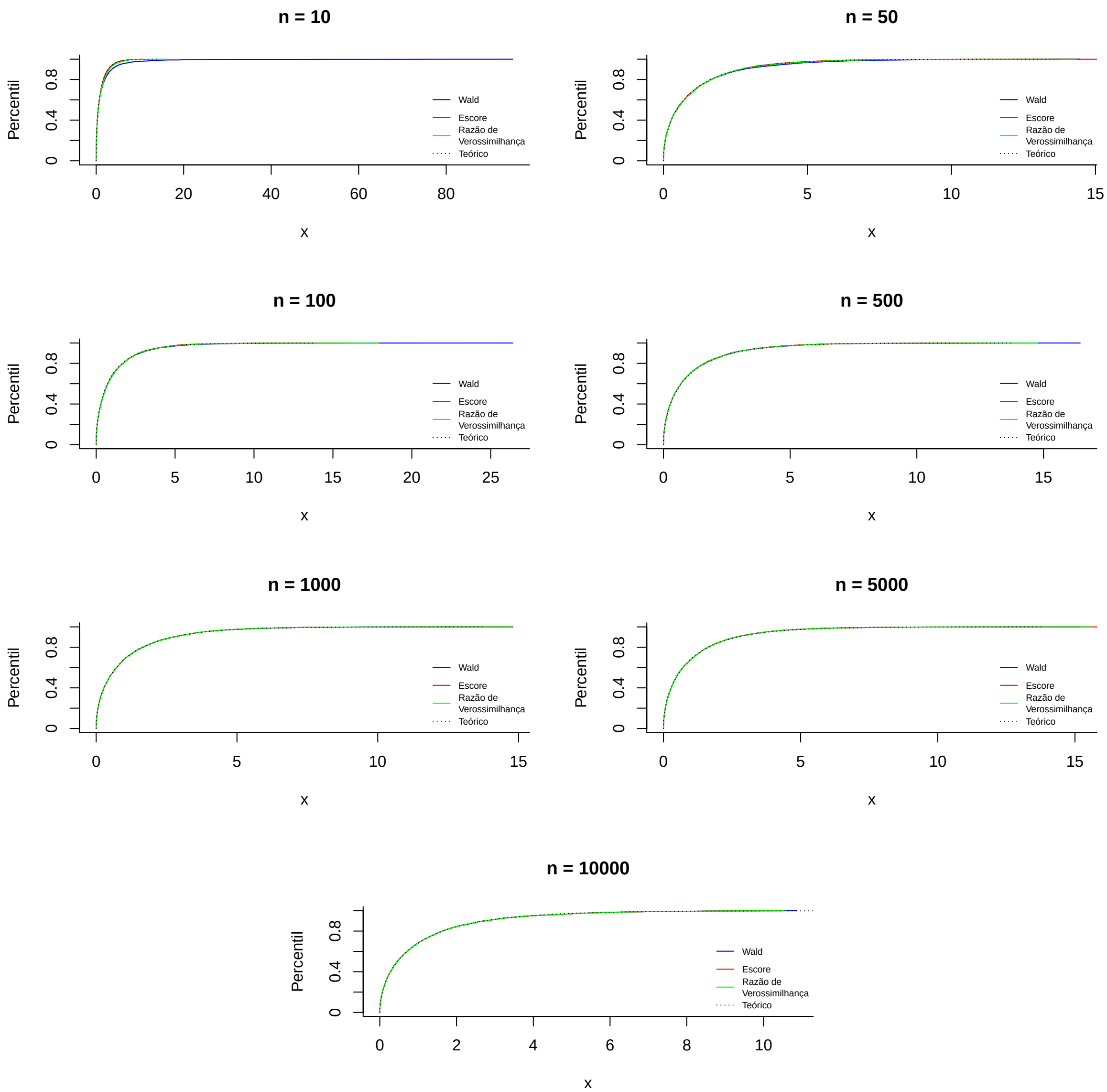

Figure 2: Convergência

Pelo gráfico podemos perceber que, para $n = 10$ a estatística que mais se aproxima da curva teórico é a estatística de Escore, enquanto a estatística Wald é a que mais se distancia, o que concorda com os resultados dos tamanhos dos testes obtidos.

À medida que o tamanho amostral aumenta vemos uma convergência das três estatísticas, em que, a partir de $n = 50$ já vemos um bom resultado de todas.

3.4.2 Kolmogorov-test

A fim de confirmar o que foi visto graficamente, vamos aplicar um teste de hipótese para testar se podemos ou não rejeitar a hipótese dos dados seguirem uma distribuição χ^2.

Nesse contexto vamos utilizar o teste de $Kolmogorov$-$Smirnov$, este teste tem como base a distância da distribuição acumulada empírica em relação a distribuição acumulada de interesse. Com isso aplicamos o teste e calculamos o $Valor$-p, e, utilizando um nível de significância de 5 decidimos sobre nossa hipótese.

Tamanho Amostral	W	S	Q
n = 10	0***	0.3001	3e-04***
n = 50	0.2778	0.4929	0.5336
n = 100	0.0154**	0.0149**	0.0088***
n = 500	0.6637	0.7268	0.804
n = 1000	0.7458	0.8195	0.8312
n = 5000	0.6848	0.7679	0.6951
n = 10000	0.6245	0.5879	0.5709

Podemos ver que, a partir de $n = 500$ para todas as estatística não rejeitamos a hipótese dos dados seguirem distribuição χ_1^2. Porém para os demais tamanhos amostrais vemos que a estatística Escore, mesmo com um tamanho amostral de $n = 10$, obteve um p-valor mais alto.

4 Conclusões

A partir deste trabalho conseguimos concluir que, se considerarmos que a variável de interesse segue distribuição $Cauchy(\theta, 1)$, para amostras pequenas, é fundamental sabermos qual o nosso problema e o que nós iríamos priorizar.

Enquanto a estatística Wald possui um comportamento melhor em relação ao poder do teste, a estatística de Razão de Verossimilhança possui resultados melhores em relação ao tamanho do teste. Por outro lado, o teste Escore não seria recomendado para amostras de tamanho pequeno.

Para tamanhos amostrais maiores, a partir de 100, nenhum teste apresentou uma melhora significativa, e, de forma geral o melhor teste foi o de Razão de Verossimilhança.

5 Referências

BOLFARINE, H.; SANDOVAL, M. C. **Introdução à inferência estatística**. [s.l.] SBM, 2001. v. 2

DEMPSTER, A. P.; LAIRD, N. M.; RUBIN, D. B. Maximum likelihood from incomplete data via the EM algorithm. **Journal of the Royal Statistical Society: Series B (Methodological)**, v. 39, n. 1, p. 1–22, 1977.

XIE, Y.; DERVIEUX, C.; RIEDERER, E. **R markdown cookbook**. [s.l.] Chapman; Hall/CRC, 2020.

ZELLER, C. B. Distribuições misturas de escala skew-normal: Estimação e diagnóstico em modelos lineares. 2009.

6 Algoritmos

6.1 Algoritmo EM

```
em_cauchy = function(x, theta_inicial = NULL, tol = 1E-10){

  if(is.null(theta_inicial)) theta_inicial = median(x)

  crit = 1
  i = 1
  theta = theta_inicial
  l = sum(log(dcauchy(x, location = theta_inicial)))

  while(crit > tol){

    ui = 2/(1+((x-theta)^2))

    i = i+1
    theta = sum(ui*x)/sum(ui)

    l_antes = l
    l = sum(log(dcauchy(x, location = theta)))
    crit = abs((l - l_antes)/l_antes)
  }

  return(list(theta = theta,
```

```
                  iteracoes = i))
}
```

6.2 Simulações

```
teta = seq(-5, 5)
alpha = 0.05

set.seed(8719)
B = 5000
n = c(10, 50, 100, 500, 1000, 5000, 10000)

x = teta |>
  lapply(
    function(teta)
      n |>
        lapply(function(ni) replicate(B, rcauchy(ni, loc = teta)))
    ) |>
  setNames(teta)
```

6.3 Estimadores máxima verossimilhança

```
emv = x |>
  lapply(function(x)
    x |>
      lapply(function(xi) apply(xi, 2, function(v) em_cauchy(v)$theta)))|>
  setNames(teta)
```

6.4 Estatística Wald

```
wald_estatistica = function(teta_chapeu, teta_0, fisher_n){
  fisher_n*(teta_chapeu - teta_0)**2
}
```

```
W = emv |>
  lapply(function(emv_i)
    mapply(wald_estatistica, teta_chapeu = emv_i,
           fisher_n = n/2, MoreArgs = list(teta_0 = 0))) |>
  setNames(teta)

pw = do.call(rbind, W |>
  lapply(function(w) w |>
           apply(2, function(w)
             mean(w >= qchisq(0.975, df = 1) | w <= qchisq(0.025, df = 1)))))
```

6.5 Estatística Escore

```
escore_teste = function(x, teta_0 = 0){
  2*(sum(2*(x-teta_0)/(1+(x-teta_0)**2))**2)/length(x)
}

S = seq_along(teta) |>
  lapply(function(i)
    do.call(cbind, seq_along(n) |>
              lapply(
                function(ni)
                  sapply(1:B,
                         function(j)
                           escore_teste(x[[i]][[ni]][,j]))
              )) |>
  setNames(teta)

ps = do.call(rbind, S |>
  lapply(function(s) s |>
           apply(2, function(s) mean(s >= qchisq(0.975, df = 1) | s <= qchisq(0.025,
```

6.6 Razão de verossimilhança

```
razao_vero = function(x, emv, teta_0 = 0){
  -2*(sum(log(dcauchy(x, loc = teta_0)))-sum(log(dcauchy(x, loc = emv))))
}

Q = seq_along(teta) |>
  lapply(function(i)
    do.call(cbind, seq_along(n) |>
              lapply(
                function(ni)
                  sapply(1:B,
                         function(j)
                           razao_vero(x[[i]][[ni]][,j], emv[[i]][[ni]][j]))
              )) |>
  setNames(teta)

pq = do.call(rbind, Q |>
  lapply(function(q) q |>
           apply(2, function(q) mean(q >= qchisq(0.975, df = 1) | q <= qchisq(0.025,
  as.data.frame()
```