Testes de hipóteses assintóticos no contexto da $Cauchy(\theta, 1)$

Pedro Henrique Corrêa de Almeida

Janeiro, 2023

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.

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 $Cauchy(\theta, 1)$.

Metodologia

Os métodos que serão avaliados nesse estudo são os testes de:

- Wald
- Escore
- Razão de Verossimilhanças

Iremos realizar um estudo de simulação gerando 5000 amostras independentes de uma variável aleatória $Cauchy(\theta, 1)$, sob a hipótese nula de interesse, e tamanhos amostrais $n = 10,\ 50,\ 100,\ 500,\ 1000,\ 5000\ e\ 10000.$

Hipótese

Nossa hipótese será testar se:

$$H_0 : \theta = 0$$

$$H_1 : \theta \neq 0$$

Dessa forma, para avaliar os testes em relação ao tamanho do teste e ao poder, vamos simular dados variando o valor de θ entre:

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

Distribuição Cauchy

Uma variável aleatória X tem distribuição $Cauchy$ com parâmetro de locação $\theta \in \mathbb{R}$, 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 $U \sim Gama(\frac{1}{2}, \frac{1}{2})$

Função Escore

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

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

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

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

Informação de Fisher

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

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

Pela definição do valor esperado, temos:

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

Informação de Fisher

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:

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

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

Estimadores de máxima verossimilhança

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 $\mathbf{x}_{x_c} = (\mathbf{x}_{obs}, \mathbf{x}_{mis})$.

Algoritmo EM

Consideramos $l_c(\theta|\mathbf{x}_{x_c}) = log(f(\mathbf{x}_{obs}))$ a função *log-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.

Algoritmo EM

Seja $X_1, ..., X_n$ uma amostra aleatória simples de X, temos que estas serão independentes e podem ser representadas por:

$$X_i = \theta + U_i^{-\frac{1}{2}} Z_i \qquad \forall\ i = 1, ..., n$$

onde $Z_i \sim N(0, 1)$ e $U_i \sim Gama(\frac{1}{2}, \frac{1}{2})$, logo, como Z_i pertence à família escala:

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

Etapa E

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

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

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

Etapa E

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:

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

Etapa E

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

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:

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

Etapa M

Após a Etapa E, temos a maximização da verossimilhança da função $Q(\theta) = E[l_c(\theta)]$, dessa forma, temos:

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

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

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

Etapa M

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:

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

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

Etapa M

A fim de confirmar se $\hat{\theta}$ 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, $\hat{\theta} = \frac{\sum_{i=1}^{n} \hat{u}_i x_i}{\sum_{i=1}^{n} \hat{u}_i}$ é ponto de máximo.

EM Cauchy

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

$k = 1$

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 $\left|\frac{l(\widehat{\theta}^{(k+1)}) - l(\widehat{\theta}^{(k)})}{l(\widehat{\theta}^{(k)})}\right|$, $\epsilon = 10^{-10}$ e $c = Mediana(x)$, o chute inical.

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:

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

Wald

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:

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

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:

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

Escore

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:

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

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:

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

Razão de Verossimilhança Generalizada

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:

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

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.

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.

Resultado EMV

Tamanho amostral	θ	Viés($\hat{\theta}$)	s^2	$\frac{1}{nI_1(\hat{\theta})}$	Tamanho amostral	θ	Viés($\hat{\theta}$)	s^2	$\frac{1}{nI_1(\hat{\theta})}$	Tamanho amostral	θ	Viés($\hat{\theta}$)	s^2	$\frac{1}{nI_1(\hat{\theta})}$
10	-5	0.0111	0.2923	0.20	100	-1	0.0037	0.0205	0.020	1000	3	-9e-04	0.002	0.002
10	-4	-0.0102	0.3073	0.20	100	0	0.0049	0.0210	0.020	1000	4	0.001	0.002	0.002
10	-3	0.0021	0.2940	0.20	100	1	0.0030	0.0213	0.020	1000	5	-4e-04	0.002	0.002
10	-2	0.0009	0.2943	0.20	100	2	-0.0035	0.0208	0.020	5000	-5	-6e-04	4e-04	4e-04
10	-1	0.0043	0.2867	0.20	100	3	0.0029	0.0198	0.020	5000	-4	2e-04	4e-04	4e-04
10	0	-0.0045	0.3017	0.20	100	4	-0.0021	0.0202	0.020	5000	-3	3e-04	4e-04	4e-04
10	1	-0.0038	0.2864	0.20	100	5	-0.0022	0.0203	0.020	5000	-2	-2e-04	4e-04	4e-04
10	2	0.0097	0.2885	0.20	500	-5	-0.0014	0.0041	0.004	5000	-1	-1e-04	4e-04	4e-04
10	3	-0.0147	0.2976	0.20	500	-4	-0.0003	0.0042	0.004	5000	0	-2e-04	4e-04	4e-04
10	4	0.0098	0.2752	0.20	500	-3	-0.0005	0.0039	0.004	5000	1	1e-04	4e-04	4e-04
10	5	0.0150	0.3067	0.20	500	-2	0.0012	0.0039	0.004	5000	2	-1e-04	4e-04	4e-04
50	-5	-0.0013	0.0421	0.04	500	-1	-0.0013	0.0040	0.004	5000	3	2e-04	4e-04	4e-04
50	-4	-0.0003	0.0435	0.04	500	0	0.0011	0.0040	0.004	5000	4	-1e-04	4e-04	4e-04
50	-3	-0.0010	0.0428	0.04	500	1	0.0004	0.0039	0.004	5000	5	0	4e-04	4e-04
50	-2	0.0041	0.0433	0.04	500	2	0.0004	0.0041	0.004	10000	-5	-1e-04	2e-04	2e-04
50	-1	-0.0029	0.0433	0.04	500	3	0.0004	0.0039	0.004	10000	-4	-1e-04	2e-04	2e-04
50	0	-0.0014	0.0422	0.04	500	4	0.0003	0.0040	0.004	10000	-3	0	2e-04	2e-04
50	1	-0.0035	0.0439	0.04	500	5	-0.0016	0.0040	0.004	10000	-2	1e-04	2e-04	2e-04
50	2	0.0057	0.0433	0.04	1000	-5	-0.0004	0.0020	0.002	10000	-1	2e-04	2e-04	2e-04
50	3	0.0023	0.0422	0.04	1000	-4	0.0005	0.0021	0.002	10000	0	-1e-04	2e-04	2e-04
50	4	0.0002	0.0422	0.04	1000	-3	-0.0002	0.0020	0.002	10000	1	2e-04	2e-04	2e-04
50	5	-0.0026	0.0416	0.04	1000	-2	0.0003	0.0020	0.002	10000	2	1e-04	2e-04	2e-04
100	-5	-0.0007	0.0200	0.02	1000	-1	0.0007	0.0020	0.002	10000	3	0	2e-04	2e-04
100	-4	-0.0015	0.0212	0.02	1000	0	0.0002	0.0020	0.002	10000	4	-1e-04	2e-04	2e-04
100	-3	-0.0018	0.0212	0.02	1000	1	-0.0005	0.0020	0.002	10000	5	-2e-04	2e-04	2e-04
100	-2	0.0022	0.0200	0.02	1000	2	0.0002	0.0020	0.002					

Estudos de simulação

Agora, vamos utilizar esses resultados e calcular as estatísticas de Wald, Escore e Razão de Verossimilhança para cada amostra gerada.

Dessa forma vamos comparar os resultados da rejeição de H_0 em cada caso.

Resultados

	n = 10			n = 50			n = 100			n = 500		
	W	E	Q	W	E	Q	W	E	Q	W	E	Q
θ = -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
θ = -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
θ = -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
θ = -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
θ = -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
θ = 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
θ = 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
θ = 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
θ = 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
θ = 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
θ = 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

Resultados

	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

Poder

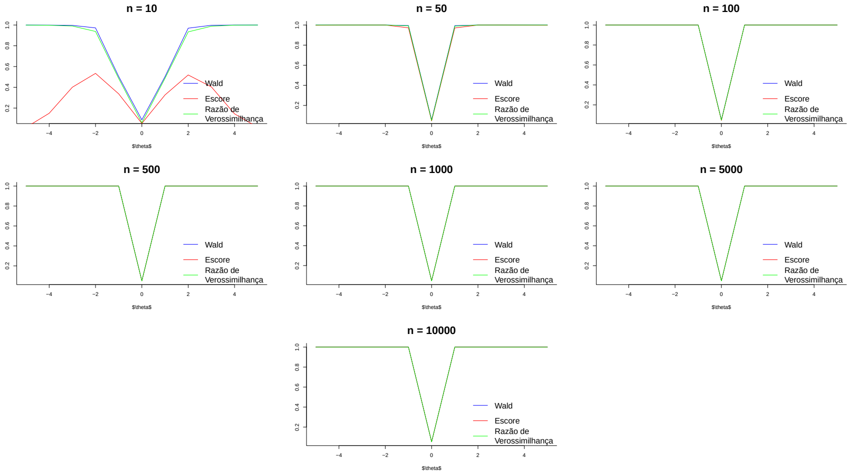
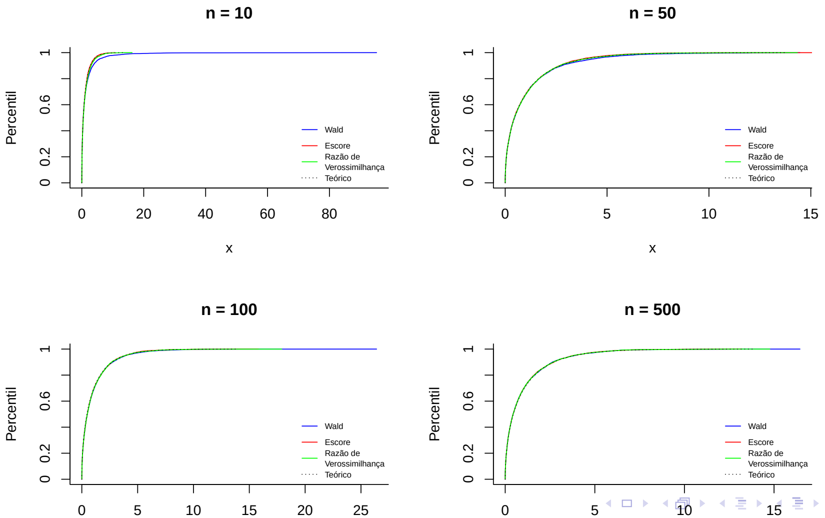
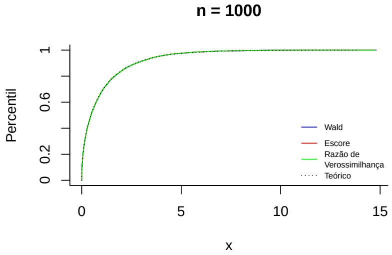


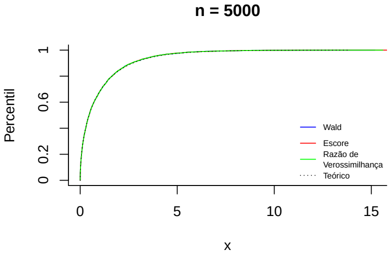
Figure 1: Poder

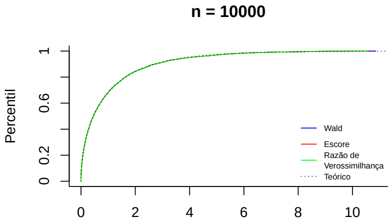
Convergência distribuição acumulada



Convergência distribuição acumulada







Convergência Kolmogorov-Smirnov

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 $Kolmogorov\text{-}Smirnov$. Com isso aplicamos o teste e calculamos o $Valor\text{-}p$.

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

Convergência

A partir do gráfico e do teste de $Kolmogorov$-$Smirnov$, vemos que a estatística que obteve melhor resultado na convergência foi a estatística Escore.

Além disso, percebemos que a partir de $n = 500$ todas as estatísticas convergem para a distribuição χ_1^2. No entanto, mesmo com tamahos amostrais menores, já é possível obter resultados satisfatórios, como foi observado anteriormente.

Conclusões

A partir deste trabalho conseguimos chegar as seguintes conclusões sobre testes de hipóteses bilaterias quando consideramos uma variável com distribuição $Cauchy(\theta, 1)$:

- O teste de Escore não é recomendado para tamanhos amostrais pequenos
- O teste de Wald possui um poder muito alto mesmo para amostras pequenas
- O teste de Razão de Verossimilhança obteve, de forma geral, os melhores resultados e seria o mais indicado
- A partir de uma amostra de tamanho $n = 100$ o resultados de todos os testes se aproximam bastante

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.