Tópico 7: Testes de hipóteses

Ben Dêivide

6 de outubro de 2021

Descrição sobre o ponto - Testes de hipóteses: Conceitos básicos; Hipóteses simples: Lema de Neyman-Pearson; Hipóteses compostas: testes uniformemente mais poderosos; Teste da razão de verossimilhanças; Teste t de Student; Teste de qui-quadrado; Teste F de Fisher-Snedecor.

1 Conceitos iniciais

Consideramos uma amostra aleatória $X_1, X_2, \ldots, X_n$ de tamanho $n \geq 2$ definida pelo seu modelo probabilístico $f_X(x)$ que é dependente de um parâmetro θ (pode ser um vetor), em que o conjunto dos valores que θ pode assumir é o espaço paramétrico, denotado por Θ. Denote $\mathcal{X}$ o espaço amostral das observações, isto é, $\mathcal{X} = \{(x_1, x_2, \ldots, x_n) : (x_1, x_2, \ldots, x_n)$ são as possíveis realizações de $(X_1, X_2, \ldots, X_n)\}$.

Uma das etapas da inferência estatística é testar a veracidade de algumas afirmações impostas ao parâmetro θ. Essa etapa é a teoria de decisão conhecida também como teste de hipóteses.

Definição 1 (Hipótese estatística)**.** *Uma hipótese estatística, denotada por $\mathcal{H}$, é uma afirmação ou conjectura feita ao parâmetro θ. Se a hipótese especifica completamente o valor de θ, diz-se uma hipótese simples, caso contrário, é chamada de hipótese composta.* □

Quando a forma da distribuição de probabilidade da população é conhecida e as afirmações feitas são relacionadas aos parâmetros dessa distribuição, dizemos que estamos diante de hipóteses paramétricas. Caso contrário, temos as hipóteses não paramétricas.

Assim, com base numa amostra da população e num teste de hipóteses se decide qual das hipóteses complementares é verdadeira.

Definição 2 (Hipóteses complementares)**.** *As hipóteses complementares de uma hipótese estatística são a hipótese nula, denotada por $\mathcal{H}_0$, e a hipótese alternativa, denotada por $\mathcal{H}_1$. O formato geral das hipóteses complementares é*

$$\begin{cases} \mathcal{H}_0 : \theta \in \Theta_0 \\ \mathcal{H}_1 : \theta \in \Theta_0^c, \end{cases} \tag{1}$$

em que $\Theta_0 \subset \Theta$, e Θ_0^c é seu complemento. □

Definição 3 (Hipótese unilateral e hipótese bilateral). *Uma hipótese $\mathcal{H}$ que faz tais afirmações para um parâmetro univariado θ,*

$$\mathcal{H} : \theta \geq \theta^* \quad ou \qquad \mathcal{H} : \theta < \theta^*, \tag{2}$$

é chamada de hipótese unilateral. Uma hipótese $\mathcal{H}$ que faz a seguinte afirmação de um parâmetro univariado θ,

$$\mathcal{H} : \theta \neq \theta^*, \tag{3}$$

é chamada de hipótese bilateral.

Exemplo 1. *Seja uma amostra aleatória $X_1, X_2, \ldots, X_n$, tal que $X \sim N(\theta, 25)$. Se denotarmos $\mathcal{H} : \theta \leq 17$, $\mathcal{H}$ é chamada de hipótese composta. Caso $\mathcal{H} : \theta = 17$, $\mathcal{H}$ é chamada de hipótese simples.* □

A Hipótese avaliada pelo pesquisador é a hipótese $\mathcal{H}_0$, e ele deve tomar a seguinte decisão: "Rejeitar"ou "Não rejeitar"a hipótese nula. Essa decisão será baseada por um teste.

Definição 4 (Teste de uma hipótese). *Um teste de hipótese, denotado por Υ, é uma regra ou procedimento para rejeitar a hipótese $\mathcal{H}_0$.* □

A decisão de um teste Υ está baseada nas observações $x_1, x_2, \ldots, x_n$ da amostra $X_1, X_2, \ldots, X_n$. E a região para qual o teste rejeita a hipótese $\mathcal{H}_0$ é definida a seguir.

Definição 5 (Região crítica). *Seja Υ um teste que avalia a hipótese estatística $\mathcal{H}_0$. Então se a decisão for: rejeitar $\mathcal{H}_0$ se $(x_1, x_2, \ldots, x_n) \in C_\Upsilon$, em que C_Υ é um subconjunto de $\mathcal{X}$, dizemos que C_Υ é a região crítica do teste Υ.* □

O teste Υ será baseado em uma estatística $W(X_1, X_2, \ldots, X_n)$ que é função da amostra $X_1, X_2, \ldots, X_n$ com $f_X(x; \theta)$ parametrizada por $\theta \in \Theta$. A decisão por meio da estatística do teste $W(X_1, X_2, \ldots, X_n)$ será: se $(X_1, X_2, \ldots, X_n) \in C_\Upsilon$ a hipótese $\mathcal{H}_0$ é rejeitada, caso contrário, não haverá evidências para a sua rejeição. Por exemplo, um teste Υ pode especificar que H_0 será rejeitada se a média amostral $(\bar{X})$ for maior que 3. Nesse caso a estatística do teste é $W(X_1, X_2, \ldots, X_n) = \bar{X}$ e a região crítica é $C_\Upsilon = \{(x_1, x_2, \ldots, x_n) : \bar{X} > 3\}$.

Exemplo 2 (Teste Binomial). *Um professor aplica um teste do tipo certo-errado com 10 questões. Queremos testar a hipótese de que o aluno está advinhando. A hipótese nula é que o aluno acerta ao acaso as questões. A hipótese alternativa é que o aluno tem algum conhecimento. Sendo p a probabilidade (desconhecida) do aluno acertar cada questão, então a hipótese estatística de interesse pode ser dada como $H_0 : p = 1/2$. Como hipótese alternativa $H_1 : p > 1/2$, isto é, o aluno tem algum conhecimento para resolver as questões. Vamos assumir que as questões são independentes e que a variável aleatória X denota o número de acertos entre as 10 questões. Portanto, X tem distribuição binomial com parâmetros $n = 10$ e p desconhecido. Supondo que adotamos a seguinte regra de decisão: o aluno não está advinhando se acertar 8 ou mais questões. Isto equivale a*

$$Rejeitar\ H_0\ se\ X \geq 8.$$

Dessa forma, a região crítica do teste é $C_Y = \{X : X > 8\}$. É possível também que o aluno acerte 8 ou mais questões e esteja advinhando, isto é, podemos rejeitar H_0 dado que ela é verdadeira. A probabilidade de que isso ocorra é:

$$P(X \geq 8|p = 1/2) = \sum_{k=8}^{10} (1/2)^k (1 - 1/2)^{10-k} \approx 0,055.$$

Essa probabilidade é chamada de nível de significância ou erro tipo I, denotada por α. Fica claro que o valor de α depende da regra de decisão, por exemplo se a região crítica fosse $X \geq 7$ teríamos $\alpha \approx 0,171$.

Dessa forma, percebemos que a tomada de decisão do pesquisador sobre θ não está livre de erros.

Definição 6 (Erro tipo I)**.** *Seja um teste* Y *que avalia uma hipótese verdadeira $\mathcal{H}_0$. A decisão de rejeitar a hipótese $\mathcal{H}_0$, dado que ela é verdadeira, é chamada de erro tipo I. O tamanho do erro tipo I é a probabilidade com que o erro é cometido, denotado por α.* □

Veremos como usar α para construir uma regra de decisão.

Exemplo 3 (Teste Binomial)**.** *Um professor aplica um teste do tipo certo-errado com 10 questões. Queremos testar a hipótese de que o aluno está advinhando. Nossa hipótese nula é que o aluno acerta cada questão ao acaso e a hipótese alternativa é que o aluno tem algum conhecimento na matéria. Denotando por p a probabilidade (desconhecida) do aluno acertar cada questão, a hipótese estatística de interesse pode ser formulada como $H_0 : p = 1/2$. Nesse caso a hipótese alternativa mais adequada é $H_1 : p > 1/2$ indicando que ele tem algum conhecimento da matéria. Temos então 10 repetições do experimento com p constantes e vamos assumir também que as questões são resolvidas de forma independente. Portanto, a variável aleatória X = "número de acertos" tem distribuição binomial com parâmetros $n = 10$ e p desconhecido. Suponha que adotamos a seguinte regra de decisão: o aluno não está advinhando se acertar 8 ou mais questões. Isto equivale a rejeitar H_0 se $X \geq 8$ (região de rejeição ou região crítica) e não rejeitar H_0 se $X < 8$ (região de não rejeição). No entanto, é possível que um aluno acerte 8 ou mais questões e esteja advinhando, ou seja podemos rejeitar H_0 quando ela é verdadeira. A probabilidade de que isso ocorra é*

$$P(X \geq 8|p = 1/2) = \sum_{k=8}^{10} 0,5^k (1 - 0,5)^{10-k} \approx 0,055.$$

Esta probabilidade é o nível de significância α. Fica claro que o valor de α depende da regra de decisão, por exemplo, se a região crítica fosse $X \geq 7$ teríamos $\alpha \approx 0,171$

Definição 7 (Erro tipo II)**.** *Seja um teste* Y *que avalia uma hipótese falsa $\mathcal{H}_0$. A decisão de aceitar a hipótese $\mathcal{H}_0$, dado que ela é falsa, é chamada de erro tipo II. O tamanho do erro tipo II é a probabilidade com que o erro é cometido, denotado por β.* □

Em situações reais esses dois erros podem ocorrer, porém o que se deseja é minimizá-los, uma vez que o valor de α é inversamente proporcional ao valor de β. Assim, fixado um tamanho da amostra (n), a baixa probabilidade de se incorrer no erro tipo I implica numa alta probabilidade de se incorrer no erro tipo II, e o único modo de se causar redução simultânea de ambos é aumentar o tamanho da amostra (n). No desenvolvimento de um teste de hipóteses o que se faz é fixar o valor de α e daí tentar minimizar o valor do erro tipo II. Do ponto de vista teórico, o erro tipo II pode ser minimizado com uma série de ações:

- escolha apropriada do teste e da avaliação do atendimento das pressuposições da utilização do teste. Se o teste é adequado e as pressuposições para sua aplicação estão satisfeitas, existe uma garantia de um maior poder;
- determinação de um tamanho de amostra para que o teste tenha um maior poder possível e que não aumente em demasia o custo da pesquisa a ser realizada; (exemplo: Casella port p. 334)
- a fixação de α entre 0,10 e 0,01, é sempre que possível, uma boa medida por causa da relação inversa entre as taxas do erro tipo I e do erro tipo II.

Se por um lado a probabilidade de se cometer o erro tipo I é conhecida e fixada pelo investigador, por outro a probabilidade do erro tipo II não é conhecida e nem pode ser especificada, uma vez que a hipótese alternativa não tem o seu parâmetro especificado, a menos que estejamos diante de hipóteses simples, algo muito raro nas situações práticas.

A forma de mensurar o tamanho desses erros é avaliado pela função poder.

Definição 8 (Função Poder). *Seja um teste* Y *com região de rejeição* C_Y*, então a função poder, denotada por* $\pi_Y(\theta)$*, é a probabilidade* $\pi_Y(\theta) = P_\theta[(X_1, X_2, \ldots, X_n) \in C_Y]$ *parametrizado por* θ*.* □

Em outras palavras, a função poder é a probabilidade de rejeitar a hipótese $\mathcal{H}_0$ quando a distribuição de que a amostra foi obtida é parametrizada por θ. Dessa forma, o teste Y com região C_Y para as hipóteses

$$\begin{cases} \mathcal{H}_0 : \theta \in \Theta_0 \\ \mathcal{H}_1 : \theta \in \Theta_0^c, \end{cases} \tag{4}$$

dado que a hipótese $\mathcal{H}_0$ é verdadeira, tem probabilidade de erro tipo I dada por $\alpha = \pi_Y(\theta \in \Theta_0) = P_{\theta\in\Theta_0}((X_1, X_2, \ldots, X_n) \in C_Y)$. Considerando a probabilidade do erro tipo II dada por β, a probabilidade de rejeitar a hipótese $\mathcal{H}_0$, dado que $\theta \in \Theta_0^c$, é dado por $\pi_Y(\theta \in \Theta_0^c) = P_{\theta\in\Theta_0^c}((X_1, X_2, \ldots, X_n) \in C_Y)$, do qual chamamos de poder do teste que é na realidade o complemento da probabilidade do erro tipo II, isto é, $1 - \beta$.

Definição 9 (Tamanho do teste). *Seja um teste* Y *que avalia a hipótese* $\mathcal{H}_0 : \theta \in \Theta_0$, $\Theta_0 \subset \Theta$*, sendo* Θ *o espaço paramétrico. Então para* $0 \leq \alpha \leq 1$ *o tamanho do teste é definido por*

$$\alpha = \sup_{\theta\in\Theta_0} \pi_Y(\theta). \tag{5}$$

□

Definição 10 (Nível do teste). *Seja um teste* Y *que avalia a hipótese* $\mathcal{H}_0 : \theta \in \Theta_0$, $\Theta_0 \in \Theta$*, sendo* Θ *o espaço paramétrico. Então para* $0 \leq \alpha \leq 1$ *o nível do teste é definido por*

$$\alpha \geq \sup_{\theta\in\Theta_0} \pi_Y(\theta). \tag{6}$$

□

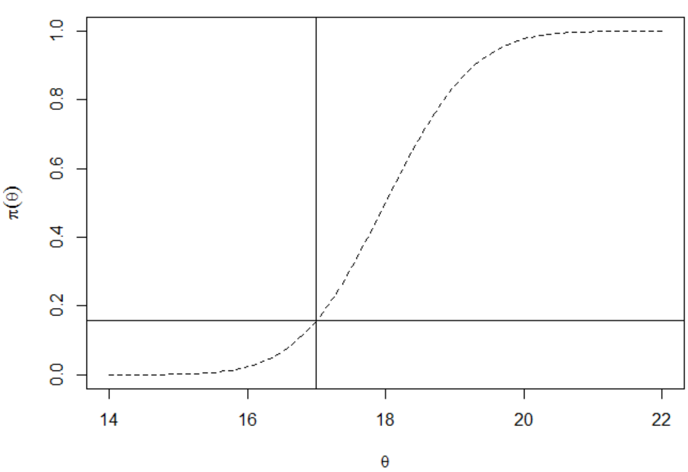
Exemplo 4. *Seja $X_1, X_2, \ldots, X_n$ uma amostra aleatória da distribuição $N(\theta, \sigma^2)$ com $\sigma^2 = 25$ e suponha que queremos testar $H_0 : \theta \leq 17$. Suponha que a regra de decisão consiste em rejeitar H_0 se e somente se $\bar{X} > 17 + \sigma/\sqrt{n}$. Neste caso a função poder é dada por*

$$\pi(\theta) = P(\text{rejeitar } H_0|\theta) = P(\bar{X} > 17 + \sigma/\sqrt{n}) = P\left(Z > \frac{17 + \sigma/\sqrt{n} - \theta}{\sigma/\sqrt{n}}\right)$$

onde $Z \sim N(0,1)$. Para $n = 25$ segue que,

$$\pi(\theta) = P(Z > 18 - \theta)$$

e calculando esta probabilidade para vários valores de θ podemos construir o gráfico para a função poder (Ricardo Ehlers, p. 93-94).



Note que o valor máximo da função quando H_0 é verdadeira é obtida para $\theta = 17$ e portanto o tamanho do teste é dado por

$$\sup_{\theta \leq 17} \left[P\left(Z > \frac{17 + \sigma/\sqrt{n} - \theta}{\sigma/\sqrt{n}}\right)\right] = \pi(17) = P(Z > 1) \approx 0,159 = \alpha.$$

A função poder é útil em dizer o quão bom é um teste específico. Nesse exemplo, se θ é maior do que 20, o teste é quase certo rejeitar H_0, como deveria. E se θ for menor que 16, o teste é quase certo não rejeitar H_0, como deveria. Por outro lado, se $17 < \theta < 18$, o teste terá menos que a metade da chance de rejeitar H_0.

A escolha do nível de significância para um teste é completamente arbitrária. Além disso, quando a distribuição da estatística do teste é discreta, o nível escolhido pode nem mesmo ser atingido. Vejamos o seguinte exemplo,

Exemplo 5. *Um fornecedor garante que 90% de sua produção não apresenta defeito. Para testar essa afirmação selecionamos ao acaso 10 itens de um lote e contamos o número de defeituosos. Com base nesta amostra tomaremos uma decisão: comprar ou não comprar o lote. É bem intuitivo que devemos decidir comprar o lote se o número de itens sem defeitos for pequeno. O problema é definir o quão pequeno. Seja X a variável aleatória que denota o número de itens sem defeito na amostra de 10 itens. Temos então um distribuição binomial com parâmetros $n = 10$ e p*

desconhecido, e queremos testar $H_0: \ p = 0,9$. Aqui p é a proporção de itens sem defeito no lote e portanto a hipótese alternativa deve ser $H_1: \ p < 0,9$. Suponha que decidimos manter $\alpha \geq 0,025$ e a partir deste valor vamos estabelecer a nossa regra de decisão, ou seja obter o valor da constante c tal que H_0 é rejeitada se $X \geq c$. Para isso vamos calcular α para diferentes regiões críticas,

$$P(X \geq 5|p = 0,9) \sum_{k=0}^{5} 0,9^k(1-0,9)^{10-k} = 0,002$$
$$P(X \geq 6|p = 0,9) \sum_{k=0}^{6} 0,9^k(1-0,9)^{10-k} = 0,013$$
$$P(X \geq 7|p = 0,9) \sum_{k=0}^{7} 0,9^k(1-0,9)^{10-k} = 0,002$$

Portanto, devemos usar a região crítica $X \geq 6$. Isto é, vamos rejeitar o lote se o número de itens defeituosos na amostra for maior ou igual a 4.

Na maioria das aplicações práticas o valor escolhido é 0,05 ou 0,01, mas não há nada que justifique o uso destes valores em particular. Um enfoque consiste em calcular uma quantidade chamada nível crítico, probabilidade de significância ou valor-p.

Definição 11 (*Valor-p*)**.** *Seja uma amostra aleatória X_1, X_2, ..., X_n com $f_X(x;\theta)$, $\theta \in \Theta$, sendo Θ o espaço paramétrico. Considere um teste de hipótese* Υ *que avalia $H_0 : \theta \in \Theta_0$* versus *$H_1 : \theta \in \Theta_0^c$, em que $\Theta_0 \subset \Theta$, com região de crítica C_Υ e tamanho α. Então o valor-p é uma estatística, denotada por $p(\boldsymbol{X})$, que satisfaz $0 \leq p(\boldsymbol{X}) \leq 1$ tal que*

$$p(\boldsymbol{X}) = \inf\{\alpha : \boldsymbol{X} \in C_\Upsilon\}, \tag{7}$$

em que $\boldsymbol{X} = [X_1, X_2, \ldots, X_n]'$.

Em geral, o *valor-p* é o menor nível de significância α para que rejeite a hipótese H_0 com base em um teste de tamanho α.

O cálculo do *valor-p* pode ser dado como segue. Considere que o teste Υ é da forma: "Rejeita-se H_0 se $W(\boldsymbol{X}) > c$", sendo W uma estatística do teste, então sob H_0, o *valor-p* é a probabilidade $P(W > w)$, sendo w o valor observado de W.

Exemplo 6. *Do Exemplo* **??** *suponha que o número observado de questões certas foi $X = 9$. Então o valor-p será*

$$P(X \geq 9|p = 1/2) = \binom{10}{9} 0,5^9(1-0,5)^{10-9} + \binom{10}{10} 0,5^{10}(1-0,5)^{10-10} = 0,0107,$$

e rejeitamos H_0 para todo nível de significância maior do que este valor.

Portanto, o valor-p é a probabilidade de observar resultados tão extremos quanto aqueles que foram obtidos se a hipótese nula for verdadeira. A ideia é que se o valor-p for grande ele fornece evidência de que H_0 é verdadeira, enquanto que um valor-p pequeno indica que existe evidência nos dados contra H_0.

Da forma como a metodologia clássica de testes de hipóteses foi desenvolvida, podemos ter a impressão de que estamos calculando probabilidades a respeito de uma hipótese. De fato, algumas vezes é incorretamente afirmado que rejeitar H_0 ao nível α

indica que a probabilidade de H_0 ser verdadeira é menor do que α. Essa interpretação não é válida e o valor-p calculado em um teste não fornece nenhuma indicação sobre qualquer probabilidade a respeito de H_0.

Por exemplo, um valor-p próximo de zero nos fornece (do ponto de vista clássico) muita evidência contra H_0, porém não significa de maneira alguma $P(H_0$ ser verdadeira) seja próximo de zero. Esta última afirmação probabilística sequer faz sentido na inferência clássica, embora seja exatamente isto que gostaríamos de calcular.

Para que esta interpretação fosse válida, teríamos que usar a abordagem Bayesiana. Basicamente, teríamos que atribuir uma probabildiade *a priori*, isto é, antes de observar os dados, para a hipótese H_0. Após a observação dos dados amostrais esta probabilidade seria atualizada, segundo regras da inferência Bayesiana, e teríamos uma probabilidade *a posteriori* para a hipótese H_0. (Ehlers, pag. 100)

2 Método para obtenção de testes de hipótese

Existem diversos métodos na literatura para obtenção de testes de hipóteses, tais como testes bayesianos, testes união-interseção e interseção-união, testes baseados em distribuições assintóticas, etc. Entretanto, daremos ênfase a testes baseados na função de verossimilhança. Posteriormente, discutiremos sobre o teste assintótico baseado na razão de verossimilhanças.

Para uma amostra aleatória $X_1, X_2, \ldots, X_n$ independentes e identicamente distribuídas de uma função de probabilidade (fp) ou função densidade de probabilidade (fdp) $f_X(x;\theta)$, $\theta \in \Theta$, desejamos um teste sob as hipóteses simples, isto é, $\mathcal{H}_0 : \theta = \theta_0$ versus $\mathcal{H}_1 : \theta = \theta_1$.

Definição 12 (Função de verossimilhança). *A função de verossimilhança, denotada por $L(\theta; x_1, x_2, \ldots, x_n)$, é definida como*

$$L(\theta; x_1, x_2, \ldots, x_n) = L(\theta; \boldsymbol{x}) = f_{\boldsymbol{X}}(\boldsymbol{x};\theta) = \prod_{i=1}^{n} f_X(x_i;\theta), \tag{8}$$

considerando que as realizações $x_1, x_2, \ldots, x_n$ vieram de uma amostra aleatória $X_1, X_2, \ldots, X_n$ independente e identicamente distribuída com fdp ou fp conjunta $f_{\boldsymbol{X}}(\boldsymbol{x};\theta)$. □

Definição 13 (Teste da razão de verossimilhança simples - TRVS). *Seja $L(\theta_i; \boldsymbol{x})$ a função de verossimilhança para uma amostra aleatória $X_1, X_2, \ldots, X_n$ com fp conjunta ou fdp conjunta $f_{\boldsymbol{X}}(\boldsymbol{x};\theta_i)$, para $i = 0, 1$, sendo $\theta_i \in \Theta$. Então o teste da razão de verossimilhança simples que avalia as hipóteses $\mathcal{H}_0 : \theta = \theta_0$ versus $\mathcal{H}_1 : \theta = \theta_1$, denotado por λ, é definida por*

- *Rejeita H_0 se $\lambda \leq k$,*
- *Não rejeita H_0 se $\lambda > k$,*

em que

$$\lambda = \lambda(x_1, x_2, \ldots, x_n) = \lambda(\boldsymbol{x}) = \frac{L_0(\theta_0; \boldsymbol{x})}{L_1(\theta_1; \boldsymbol{x})},$$

e k é uma constante não-negativa. □

Para cada k diferente teremos um teste diferente. Fixado um k, o teste rejeita a hipótese $\mathcal{H}_0$ se λ for um número pequeno, isto é, a verossimilhança L_1 é maior que L_0, assim é mais provável que a amostra com $f_X(x;\theta)$ tenha sido parametrizada por $\theta = \theta_1$ do que por $\theta = \theta_0$. Algo bem intuitivo.

Um teste Υ ideal seria aquele em que a função poder $\alpha = \pi_\Upsilon(\theta \in \Theta_0) = 0$ e $1 - \beta = \pi_\Upsilon(\theta \in \Theta_0^c) = 1$. Entretanto, em situações práticas isso é praticamente impossível. O que se faz é controlar a probabilidade do erro tipo I, e dentro dessa classe encontrar os melhores testes, isto é, um teste mais poderoso.

Definição 14 (Teste Mais Poderoso)**.** *Um teste* Υ *que avalia as hipóteses* $\mathcal{H}_0 : \theta = \theta_0$ *versus* $\mathcal{H}_1 : \theta = \theta_1$ *é definido mais poderoso de tamanho* α*, para* $0 < \alpha < 1$ *se*

i) $\pi_\Upsilon(\theta_0) = \alpha$;

ii) $\pi_\Upsilon(\theta_1) \geq \pi_{\Upsilon^*}(\theta_1)$ *para algum outro teste* Υ^* *de nível* α*, isto é,* $\pi_{\Upsilon^*}(\theta_0) \leq \alpha$.

Assim, apresentamos o seguinte teorema,

Teorema 1 (Lema de Neyman-Pearson)**.** *Seja uma amostra aleatória* $X_1, X_2, \ldots, X_n$ *iid com modelo probabilístico dado por* $f_{\boldsymbol{X}}(\boldsymbol{x};\theta)$*, em que* $\theta \in \Theta = \{\theta_0, \theta_1\}$ *e* α *fixado, para* $0 < \alpha < 1$*. Considere ainda k uma constante não negativa e* C_Υ *um subconjunto de* $\mathcal{X}$ *que satisfaz:*

i) $P_{\theta_0}[(X_1, X_2, \ldots, X_n) \in C_\Upsilon] = \alpha$;

ii) $\Upsilon = \dfrac{L(\theta_0; x_1, x_2, \ldots, x_n)}{L(\theta_1; x_1, x_2, \ldots, x_n)} = \dfrac{L_0}{L_1} \leq k$ *se* $(x_1, x_2, \ldots, x_n) \in C_\Upsilon$;

iii) $\Upsilon > k$ *se* $(x_1, x_2, \ldots, x_n) \in \bar{C}_\Upsilon = \mathcal{X} - C_\Upsilon$.

Então o teste Υ *correspondente a região crítica* C_Υ *é mais poderoso de tamanho* α *que avalia as hipóteses* $\mathcal{H}_0 : \theta = \theta_0$ *versus* $\mathcal{H}_1 : \theta = \theta_1$. □

Demonstração. A prova será mostrada para o caso contínuo. Sob H_0, considere $C_\Upsilon \in \mathcal{X}$ a região crítica, então

$$\begin{aligned}\pi_\Upsilon(\theta_1) =& P_{\theta_1}(\{X_1, X_2, \ldots, X_n\} \in C_\Upsilon) \\ =& \int_{C_\Upsilon} \ldots \int_{C_\Upsilon} \prod_{i=1}^{n} f_X(x_i;\theta) dx_i, \qquad \text{caso contínuo e iid}\end{aligned}$$

é o poder do teste.

Denotamos Υ^* um outro teste de nível α, e C_{Υ^*} sua respectiva região crítica, tal que

$$\pi_{\Upsilon^*}(\theta_0) = P_{\theta_0}[(\{X_1, X_2, \ldots, X_n\} \in C_{\Upsilon^*})] \leq \alpha.$$

Devemos mostrar que $\pi_\Upsilon(\theta_1) \geq \pi_{\Upsilon^*}(\theta_1)$ para completar a prova, que é equivalente a

$$\int_{C_\Upsilon} L_1 = \int_{C_\Upsilon} \ldots \int_{C_\Upsilon} \prod_{i=1}^{n} f_X(x_i;\theta_1) dx_i \geq \int_{C_{\Upsilon^*}} L_1 = \int_{C_{\Upsilon^*}} \ldots \int_{C_{\Upsilon^*}} \prod_{i=1}^{n} f_X(x_i;\theta_1) dx_i.$$

Se conseguirmos mostrar que

$$\int_{C_\Upsilon} L_1 - \int_{C_{\Upsilon^*}} L_1 \geq 0,$$

então o Teorema está provado. Observe que se $C_\Upsilon, C_{\Upsilon^*} \in \mathcal{X}$, podem existir valores dessas regiões que são comuns entre si. Vejamos o gráfico demonstrativo.

Podemos representar $C_Y = (C_Y \cap \bar{C}_{Y^*}) \cup (C_Y \cap C_{Y^*})$ e $C_{Y^*} = (C_{Y^*} \cap \bar{C}_Y) \cup (C_{Y^*} \cap C_Y)$. Logo,

$$\begin{aligned}
\int_{C_Y} L_1 - \int_{C_{Y^*}} L_1 &= \left(\int_{C_Y \cap \bar{C}_{Y^*}} L_1 + \int_{C_Y \cap C_{Y^*}} L_1\right) - \left(\int_{C_{Y^*} \cap \bar{C}_Y} L_1 + \int_{C_{Y^*} \cap C_Y} L_1\right)\\
&= \int_{C_Y \cap \bar{C}_{Y^*}} L_1 - \int_{C_{Y^*} \cap \bar{C}_Y} L_1.
\end{aligned}$$

Agora,

$$\int_{C_Y \cap \bar{C}_{Y^*}} L_1 - \int_{C_{Y^*} \cap \bar{C}_Y} L_1 \geq \frac{1}{k} \int_{C_Y \cap \bar{C}_{Y^*}} L_0 - \frac{1}{k} \int_{C_{Y^*} \cap \bar{C}_Y} L_0,$$

uma vez que $L_1 \geq L_0/k$ em C_Y, logo $L_1 \geq L_0/k$ em $C_Y \cap \bar{C}_{Y^*}$, já que $\{C_Y \cap \bar{C}_{Y^*}\} \subset C_Y$. Como também $L_1 \leq L_0/k$ em $\bar{C}_Y$ que implica em $-L_1 \geq -L_0/k$, logo $-L_1 \geq -L_0/k$ em $C_{Y^*} \cap \bar{C}_Y$, uma vez que $\{C_{Y^*} \cap \bar{C}_Y\} \subset C_{Y^*}$. Entretanto, sabemos que

$$\begin{aligned}
\frac{1}{k} \int_{C_Y \cap \bar{C}_{Y^*}} L_0 - \frac{1}{k} \int_{C_{Y^*} \cap C_Y} L_0 &= \frac{1}{k} \left(\int_{C_Y \cap \bar{C}_{Y^*}} L_0 - \int_{C_{Y^*} \cap C_Y} L_0\right)\\
&= \frac{1}{k} \left(\underbrace{\int_{C_Y \cap \bar{C}_{Y^*}} L_0 + \int_{C_Y \cap C_{Y^*}} L_0}_{\int_{C_Y} L_0} - \underbrace{\int_{C_{Y^*} \cap C_Y} L_0 + \int_{C_{Y^*} \cap C_Y} L_0}_{\int_{C_{Y^*}} L_0} \right)
\end{aligned}$$

Logo,

$$\begin{aligned}
\int_{C_Y} L_1 - \int_{C_{Y^*}} L_1 &\geq (1/k) \left(\int_{C_Y} L_0 - \int_{C_{Y^*}} L_0\right)\\
&\geq 1/k[\alpha - \text{ nível do teste } Y^*]\\
&\geq 0, \qquad (\text{nível do teste} Y^* \leq \alpha),
\end{aligned}$$

concluindo que o poder do teste Y, $\pi_Y(\theta_1)$, é no mínimo igual a $\pi_{Y^*}(\theta_1)$, sendo o teste mais poderoso de tamanho α para as hipóteses $\mathcal{H}_0 : \theta = \theta_0$ versus $\mathcal{H}_1 : \theta = \theta_1$. $\square$

Implicações do Lema de Neyman-Peason:

1. Pelas hipóteses lançadas, hoje esse lema não é um muito útil nas aplicações estatísticas (Justificar);
2. Nem sempre é possível determinar a região crítica C_Y e nem mesmo a constante k (Exemplo 8, Mood(1974), pag. 413);
3. Em algumas aplicações esse lema pode ser usado em hipóteses compostas;

Exemplo 7. *Seja $X_1, X_2, \ldots, X_n$ uma amostra aleatória de $f_X(x;\theta) = \theta e^{-\theta x} I_{(0,\infty)}(x)$, em que $\Theta = \{\theta : \theta > 0\}$. Teste $\mathcal{H}_0 : \theta = \theta_0$ versus $\mathcal{H}_1 : \theta = \theta_1$, considerando que θ_0 e θ_1 são números fixos tal que $\theta_0 < \theta_1$. Pelo lema de Neyman-Pearson, reitamos $\mathcal{H}_0$ se*

$$Y = \frac{L_0}{L_1} = \frac{\theta_0^n \exp\{-\theta_0 \sum_{i=1}^n x_i\}}{\theta_1^n \exp\{-\theta_1 \sum_{i=1}^n x_i\}} = \frac{\theta_0^n}{\theta_1^n} \exp\left\{(\theta_1 - \theta_0) \sum_{i=1}^n x_i\right\} \leq k,$$

que é equivalente a

$$\sum_{i=1}^n x_i \leq \frac{1}{(\theta_1 - \theta_0)} \exp\{n \log \theta_1 - n \log \theta_0\} k = k',$$

em que k' é alguma constante. Sabemos que $\sum_{i=1}^n x_i$ tem distribuição gama com parâmetros n e θ, então

$$\begin{aligned} P_{\theta_0}[(X_1, X_2, \ldots, X_n) \in C_Y] = P_{\theta_0}\left[Y = \sum_{i=1}^n x_i \leq k'\right] &= \alpha \\ = \int_0^{k'} \frac{1}{\Gamma(n)} \theta_0 y^{n-1} e^{-x\theta_0} dy &= \alpha, \end{aligned}$$

uma equação em que k' pode ser determinado. Assim, o teste mais poderoso de tamanho α de $\mathcal{H}_0 : \theta = \theta_0$ versus $\mathcal{H}_1 : \theta = \theta_1$, tal que $\theta_0 < \theta_1$ é: rejeitar $\mathcal{H}_0$ se $\sum_{i=1}^n x_i \leq k'$, em que k' é o α-ésimo quantil superior da distribuição gama com parâmetros n e θ_0.

Desejamos agora um teste sob hipóteses compostas $\mathcal{H}_0 : \theta \in \Theta_0$ versus $\mathcal{H}_1 : \theta \in \Theta_0^c$.

Definição 15 (Teste da razão de verossimilhança generalizada - TRVG)**.** *Seja $L(\theta; x_1, x_2, \ldots, x_n)$ a função de verossimilhança para uma amostra aleatória $X_1, X_2, \ldots, X_n$ com fp conjunta ou fdp conjunta $f_{X_1,X_2,\ldots,X_n}(x_1, x_2, \ldots, x_n; \theta)$, para $\theta \in \Theta$. Então o teste da razão de verossimilhança generalizada denotada por λ, é definida por*

$$\lambda = \lambda(X_1, X_2, \ldots, X_n) = \frac{\sup_{\theta \in \Theta_0} L(\theta; X_1, X_2, \ldots, X_n)}{\sup_{\theta \in \Theta} L(\theta; X_1, X_2, \ldots, X_n)}.$$

O teste tem região de rejeição $C_\lambda = \{(x_1, x_2, \ldots, x_n) : \lambda(x_1, x_2, \ldots, x_n) \leq \lambda_0\}$, em que $0 \leq \lambda_0 \leq 1$. □

Algumas considerações sobre o TRVG:

- Embora pensemos que o TRVG seja uma generalização do TRVS, o TRGV não reduz-se ao TRVS para $\Theta = \{\theta_0, \theta_1\}$;

- λ necessariamente satisfaz $0 \leq \lambda \leq 1$;
- $\lambda \geq 0$ uma vez que temos uma razão não negativa de quantidade, e $\lambda \leq 1$ uma vez que o supremo é tomado sobre um conjunto maior de parâmetros no denominador do que no numerador, então o numerador não poderá ser maior que o denominador;
- o parâmetro θ poderá ser um vetor;
- o denominador do TRVG é a função de verossimilhança avaliado no estimador de máxima verossimilhança.

Para a implementação da TRVG, os seguintes passos devem ser seguidos:

1. Obter o estimador de máxima verossimilhança (EMV) $\hat{\theta}$ de θ, quando $\theta \in \Theta$;
2. Obter o EMV $\hat{\theta}_0$ de θ, quando $\theta \in \Theta_0$;
3. Calcular $\lambda(X_1, X_2, \ldots, X_n) = \frac{L(\hat{\theta}_0; X_1, X_2, \ldots, X_n)}{L(\hat{\theta}; X_1, X_2, \ldots, X_n)}$.
4. Caso não seja simples encontrar a distribuição de $\lambda(X_1, X_2, \ldots, X_n)$, pode-se usar uma função h estritamente crescente no domínio de λ, geralmente a função log, tal que $h(\lambda(X_1, X_2, \ldots, X_n))$ tenha uma forma simples e uma distribuição conhecida e tabelada sob a hipótese $\mathcal{H}_0$.
5. Obter λ_0, resolvendo a equação $\alpha = P_{\theta_0}(h(\lambda(X_1, X_2, \ldots, X_n)) \leq \lambda_0)$.

Inserir o exemplo do mood p. 420 (Exemplo 11), inclusive com detalhes dessa resolução do caderno de probabilidade feito por Lucas (UFLA).

Em busca de propriedades ótimas para um teste de hipóteses compostas, definimos

Definição 16 (Teste Uniformemente Mais Poderoso - TUMP). *Um teste* Y *que avalia as hipóteses* $\mathcal{H}_0 : \theta \in \Theta_0$ *versus* $\mathcal{H}_1 : \theta \in \Theta_0^c$*, sendo* $\Theta_0 \subset \Theta$ *e* Θ *o espaço paramétrico, é uniformemente mais poderoso de tamanho* α*, para* $0 < \alpha < 1$ *se*

i) $\sup_{\theta \in \Theta_0} \pi_{\mathrm{Y}}(\theta) = \alpha$;

ii) $\pi_{\mathrm{Y}}(\theta) \geq \pi_{\mathrm{Y}^*}(\theta), \forall \theta \in \Theta_0^c$, *sendo* Y^* *algum outro teste de nível* α*, isto é,* $\sup_{\theta \in \Theta_0} \pi_{\mathrm{Y}^*}(\theta) \leq \alpha$. □

Um teste Y é uniformemente mais poderoso de tamanho α se para qualquer outro teste de nível α o seu poder é maior para todos os outros valores de θ sob H_1. O advérbio "uniformemente" refere-se a todos os valores de θ sob H_1. O TUMP não existe para todos os problemas sobre testes de hipóteses, mas em algumas situações especiais poderemos encontrá-lo.

Inicialmente, se quisermos um teste para as seguintes hipóteses $H_0 : \theta = \theta_0$ vs $H_1 : \theta > \theta_0$, poderemos obtê-lo usando o lema de Neyman-Pearson. Vejamos o seguinte exemplo,

Exemplo 8. *Seja $X_1, X_2, \ldots, X_n$ uma amostra aleatória de $f_X(x;\theta) = \theta e^{-\theta x} I_{(0,\infty)}(x)$, em que $\Theta = \{\theta : \theta > 0\}$. Teste $\mathcal{H}_0 : \theta = \theta_0$ versus $\mathcal{H}_1 : \theta > \theta_0$. No Exemplo* **??**, *pelo lema de Neyman-Pearson, o teste $\mathcal{H}_0 : \theta = \theta_0$ versus $\mathcal{H}_1 : \theta = \theta_1$, $\theta_1 > \theta_0$, rejeita $\mathcal{H}_0$ se $\sum_{i=1}^n x_i \leq k'$, em que k' é dado pela equação*

$$\int_0^{k'} \frac{1}{\Gamma(n)} \theta_0 y^{n-1} e^{-x\theta_0} dy = \alpha.$$

Observe que em nenhuma passagem dependeu de θ_1, a menos que $\theta_1 > \theta_0$. Portanto, nós poderemos conseguir o mesmo teste para algum $\theta_1 > \theta_0$, tal que o teste agora é uniformemente mais poderoso!

Testes uniformemente mais poderosos existem para hipóteses unilaterais se a fdp ou fp tem uma razão de verossimilhança monótona em alguma estatística.

Definição 17 (Razão de verossimilhança monótona)**.** *Uma família de densidades $\{f_X(x;\theta) : \theta \in \Theta\}$ é dita ter razão de verossimilhança monótona se existe uma estatística $T = t(X_1, X_2, \ldots, X_n)$, tal que a razão $L(\theta'; x_1, x_2, \ldots, x_n)/L(\theta''; x_1, x_2, \ldots, x_n)$ é uma função monótona não crescente de $t(x_1, x_2, \ldots, x_n)$ para todo $\theta' < \theta''$ ou não-decrescente de $t(x_1, x_2, \ldots, x_n)$ para todo $\theta' < \theta''$.* □

Note que o termo "razão de verossimilhança monótona" a razão de verossimilhança não é um razão de verossimilhança generalizada, e sim uma razão de duas funções de verossimilhança.

Teorema 2. *Seja uma amostra aleatória $X_1, X_2, \ldots, X_n$ com fp ou fdp $f_X(x;\theta)$, $\theta \in \Theta$, em que Θ é algum intervalo. Assuma que a família tal que $\{f_X(x;\theta) : \theta \in \Theta\}$ tem uma razão de verossimilhança monótona na estatística $t(\boldsymbol{x})$:*

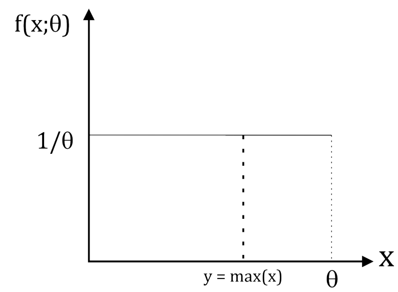
i) Se a razão de verossimilhança em $t(\boldsymbol{x})$ é monótona não-decrescente e se k tal que $P_{\theta_0}(t(\boldsymbol{x}) < k) = \alpha$, então o teste Υ com uma região crítica $C_\Upsilon = \{(x_1, x_2, \ldots, x_n) : t(\boldsymbol{x}) < k\}$ é um teste uniformemente mais poderoso de tamanho α de $\mathcal{H}_0 : \theta \leq \theta_0$ versus $\mathcal{H}_1 : \theta > \theta_0$.

ii) Se a razão de verossimilhança em $t(\boldsymbol{x})$ é monótona não-crescente e se k tal que $P_{\theta_0}(t(\boldsymbol{x}) > k) = \alpha$, então o teste Υ com uma região crítica $C_\Upsilon = \{(x_1, x_2, \ldots, x_n) : t(\boldsymbol{x}) > k\}$ é um teste uniformemente mais poderoso de tamanho α de $\mathcal{H}_0 : \theta \leq \theta_0$ versus $\mathcal{H}_1 : \theta > \theta_0$.

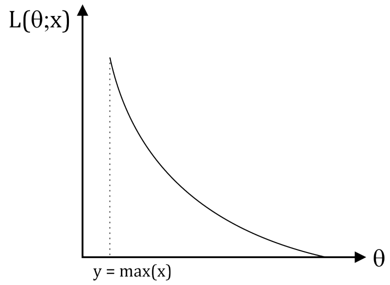
Exemplo 9. *Seja $X_1, X_2, \ldots, X_n$ uma amostra aleatória com $f_X(x;\theta) = 1/\theta I_{(0,\theta)}(x)$, em que $\theta > 0$. Teste $H_0 : \theta \leq \theta_0$ versus $H_1 : \theta > \theta_0$. Observe inicialmente que*

$$L(\theta; x_1, x_2, \ldots, x_n) = \prod_{i=1}^n (1/\theta) = (1/\theta)^n I_{(0,\theta)}(x).$$

Verificando o gráfico da função densidade abaixo,



observe que o valor de θ deve ser no mínimo tal que $\theta > y = \max(x)$. Logo o estimador de máxima verossimilhança para θ é $\hat{\theta} = y = \max(x)$, pois graficamente vemos que $L(\theta; \boldsymbol{x})$ atinge o máximo quando θ assumir o menor valor possível que é $y = \max(x)$.



Dessa forma percebemos, que $L(\theta; \boldsymbol{x})$ é não-crescente em θ, e por consequência, não-crescente em $y = \max((x))$, uma vez que $\theta > y$. Assim,

$$\frac{L(\theta'; (x))}{L(\theta''; (x))} = \frac{(1/\theta')^n I_{(0,\theta')}(x)}{(1/\theta'')^n I_{(0,\theta'')}(x)} = \frac{(1/\theta')^n I_{(0,\theta')}(y)}{(1/\theta'')^n I_{(0,\theta'')}(y)}.$$

Logo,

$$\frac{L(\theta'; (x))}{L(\theta''; (x))} = \begin{cases} \left(\frac{\theta''}{\theta'}\right)^n & \textit{para } 0 < y < \theta' \\ 0 & \textit{para } \theta' \leq y < \theta'' \end{cases},$$

sendo que essa razão de verossimilhança é uma função monótona não crescente de $y = \max(\boldsymbol{x})$.

*De acordo com o Teorema **??**, um TUMP de tamanho α tem a seguinte decisão: Rejeita H_0 se $y > k$, em que k é dado pela solução*

$$\alpha = P_{\theta_0}[Y > k] = \int_k^{\theta_0} f_Y(y) dy.$$

Como a densidade do máximo é dada por $f_Y(y) = n[F_X(y)]^{n-1} f_X(y)$, logo

$$\begin{aligned} \alpha = P_{\theta_0}[Y > k] &= \int_k^{\theta_0} n[y/\theta_0]^{n-1} \frac{dy}{\theta_0} = \frac{1}{\theta_0^n} \int_k^{\theta_0} n y^{n-1} dy \\ &= \frac{1}{\theta_0^n}[\theta_0^n - k^n] = 1 - \left(\frac{k}{\theta_0}\right)^n, \end{aligned}$$

que implica que $k = \theta_0 \sqrt[n]{1-\alpha}$.

Exemplo 10. *Uma densidade da família exponencial uniparamétrica de distribuição com fp ou fdp*

$$f_X(x; \theta) = a(\theta) h(x) \exp\{c(\theta) d(x)\},$$

considerando $\theta' < \theta''$, tem razão de verossimilhança

$$\begin{aligned} \frac{L(\theta'; \boldsymbol{x})}{L(\theta''; \boldsymbol{x})} &= \frac{\prod_{i=1}^n a(\theta') h(x) \exp\{c(\theta') d(x_i)\}}{\prod_{i=1}^n a(\theta'') h(x) \exp\{c(\theta'') d(x_i)\}} \\ &= \frac{a^n(\theta') \exp\{c(\theta') \sum_{i=1}^n d(x_i)\}}{a^n(\theta'') \exp\{c(\theta'') \sum_{i=1}^n d(x_i)\}} \\ &= \frac{a^n(\theta')}{a^n(\theta'')} \exp\left\{[c(\theta') - c(\theta'')] t(\boldsymbol{x})\right\}, \end{aligned}$$

sendo $t(\boldsymbol{x}) = \sum_{i=1}^{n} d(x_i)$. Observe que fixado os valores de θ' e θ'' a razão de verossimilhança é monótona não-crescente em $t(\boldsymbol{x})$. Fixado a amostra $(x_1, x_2, \ldots, x_n)$, a razão de verossimilhança é monótona não-crescente em $c(\theta)$. □

Por esse exemplo, podemos apresentar o seguinte teorema,

Teorema 3. *Seja uma amostra aleatória $X_1, X_2, \ldots, X_n$ com fp ou fdp $f_X(x;\theta)$, $\theta \in \Theta$, em que Θ é algum intervalo. Assuma que $f_X(x;\theta) = a(\theta)h(x)\exp\{c(\theta)d(x)\}$ e $t(\boldsymbol{x}) = \sum_{i=1}^{n} d(x_i)$.*

i) Se $c(\theta)$ é uma função monótona crescente em θ e se existe um k tal que $P_{\theta_0}(t(\boldsymbol{x}) > k) = \alpha$, então o teste Y com uma região crítica $C_Y = \{(x_1, x_2, \ldots, x_n) : t(\boldsymbol{x}) > k\}$ é um teste uniformemente mais poderoso de tamanho α de $\mathcal{H}_0 : \theta \leq \theta_0$ versus $\mathcal{H}_1 : \theta > \theta_0$ ou $\mathcal{H}_0 : \theta = \theta_0$ versus $\mathcal{H}_1 : \theta > \theta_0$.

ii) Se $c(\theta)$ é uma função monótona decrescente em θ e se existe um k tal que $P_{\theta_0}(t(\boldsymbol{x}) < k) = \alpha$, então o teste Y com uma região crítica $C_Y = \{(x_1, x_2, \ldots, x_n) : t(\boldsymbol{x}) < k\}$ é um teste uniformemente mais poderoso de tamanho α de $\mathcal{H}_0 : \theta \leq \theta_0$ versus $\mathcal{H}_1 : \theta > \theta_0$ ou $\mathcal{H}_0 : \theta = \theta_0$ versus $\mathcal{H}_1 : \theta > \theta_0$.

Demonstração. Veja Exemplo **??**. □

Observe que o Teorema **??** é consequência do Teorema **??**. Vejamos o seguinte exemplo.

Exemplo 11. *Seja $X_1, X_2, \ldots, X_n$ uma amostra aleatória de $f_X(x;\theta) = \theta e^{-\theta x} I_{(0,\infty)}(x)$, em que $\Theta = \{\theta : \theta > 0\}$. Teste $\mathcal{H}_0 : \theta \leq \theta_0$ versus $\mathcal{H}_1 : \theta > \theta_0$. Observe que*

$$\begin{aligned} f_X(x;\theta) &= \theta I_{(0,\infty)}(x)\exp\{-\theta x\} \\ &= a(\theta)h(x)\exp\{c(\theta)d(x)\}, \end{aligned}$$

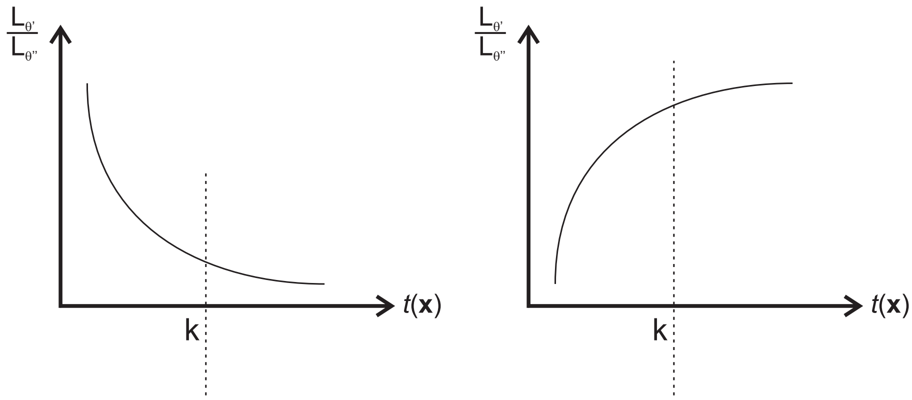
sendo $a(\theta) = \theta$, $h(x) = I_{(0,\infty)}(x)$, $c(\theta) = -\theta$ e $d(x) = x$. Então $f_X(x;\theta)$ é membro da família exponencial. A função de verossimilhança pode ser dado por

$$L(\theta; x_1, x_2, \ldots, x_n) = \theta^n (I_{(0,\infty)}(x))^n \exp\left\{-\theta \sum_{i=1}^{n} x_i\right\}.$$

Considerando $\theta' \in \Theta_0 = \{\theta : \theta \leq \theta_0\}$ e $\theta'' \in \Theta_0^c = \{\theta : \theta > \theta_0\}$, portanto $\theta' < \theta''$. Assim,

$$\begin{aligned} \frac{L_{\theta'}}{L_{\theta''}} = \frac{L(\theta'; x_1, x_2, \ldots, x_n)}{L(\theta''; x_1, x_2, \ldots, x_n)} &= \frac{(\theta')^n \exp\{-\theta' \sum_{i=1}^{n} x_i\}}{(\theta'')^n \exp\{-\theta'' \sum_{i=1}^{n} x_i\}} \\ &= \frac{(\theta')^n}{(\theta'')^n} \exp\left\{(\theta'' - \theta') \sum_{i=1}^{n} x_i\right\}. \end{aligned}$$

Sabemos que $c(\theta)$ é monótona decrescente em θ. Isso por consequência, tornou a razão de verossimilhança monótona crescente em $t(x_1, x_2, \ldots, x_n) = t(\boldsymbol{x}) = \sum_{i=1}^{n} x_i$. Como tomada de decisão para um teste, observe graficamente:



Percebemos que fixado um k para a razão de verossimilhança monótona crescente em $t(\boldsymbol{x})$ (lado direito), se $t(\boldsymbol{x}) > k$, o mais provável é que a amostra com $f_X(x;.)$ tenha sido parametrizada por θ' do que por θ''. Dessa forma não haveria indícios para rejeitar a hipótese $\mathcal{H}_0$. Agora se $t(\boldsymbol{x}) \leq k$ o mais provável é que a amostra com $f_X(x;.)$ tenha sido parametrizada por θ'' do que por θ'. Dessa forma teríamos indícios para rejeitar a hipótese $\mathcal{H}_0$. Assim, para um k tal que

$$P_{\theta_0}(t(\boldsymbol{x}) \leq k) = \alpha,$$

o teste correspondendo a região crítica $C = \{(x_1, x_2, \ldots, x_n : t(\boldsymbol{x}) \leq k)\}$ é o teste uniformemente mais poderoso de tamanho α de $\mathcal{H}_0 : \theta \leq \theta_0$ versus $\mathcal{H}_1 : \theta > \theta_0$.
Por outro lado, fixado um k para a razão de verossimilhança monótona decrescente em $t(\boldsymbol{x})$ (lado esquerdp), se $t(\boldsymbol{x}) > k$, o mais provável é que a amostra com $f_X(x;.)$ tenha sido parametrizada por θ'' do que por θ'. Dessa forma haveria indícios para rejeitar a hipótese $\mathcal{H}_0$. Agora se $t(\boldsymbol{x}) \leq k$ o mais provável é que a amostra com $f_X(x;.)$ tenha sido parametrizada por θ' do que por θ''. Dessa forma não teríamos indícios para rejeitar a hipótese $\mathcal{H}_0$. Assim, para um k tal que

$$P_{\theta_0}(t(\boldsymbol{x}) > k) = \alpha,$$

o teste correspondendo a região crítica $C = \{(x_1, x_2, \ldots, x_n : t(\boldsymbol{x}) > k)\}$ é o teste uniformemente mais poderoso de tamanho α de $\mathcal{H}_0 : \theta \leq \theta_0$ versus $\mathcal{H}_1 : \theta > \theta_0$. □

Os testes uniformemente mais poderosos baseados na razão de verossimilhança nem sempre pode ser obtida, pois muitas vezes a distribuição de $\lambda(X_1, X_2, \ldots, X_n)$ é intratável (Exemplo da igualdade de variâncias, Mood(1974), pág. 439). Entretanto, para grandes amostras a distribuição assintótica de $\lambda(X_1, X_2, \ldots, X_n)$ pode ser obtida.

Teorema 4 (Distribuição assintótica do TRV - H_0 simples - Casella, port pág 437)**.** *Seja uma amostra aleatória $X_1, X_2, \ldots, X_n$ com $f_X(x;\theta)$. Para testar $\mathcal{H}_0 : \theta = \theta_0$ versus $\mathcal{H}_1 : \theta \neq \theta_1$, à medida que $n \longrightarrow \infty$,*

$$-2\log\lambda(\boldsymbol{x}) \xrightarrow{\text{d}} \chi^2_1, \tag{9}$$

em que χ^2_1 é uma variável aleatória com distribuição de qui-quadrado com 1 grau de liberdade. □

O teste será: rejeita $\mathcal{H}_0 : \theta = \theta_0$ se $-2\log\lambda(\boldsymbol{x}) > \chi^2_{1,\alpha}$, tal que

$$\lim_{n\to\infty} P_{\theta_0}(\text{rejeitar } \mathcal{H}_0) = \alpha.$$

Teorema 5 (Distribuição assintótica do TRV - Caso geral - Casella, port pág 438). *Seja uma amostra aleatória $X_1, X_2, \ldots, X_n$ com $f_X(x;\theta)$. Para testar $\mathcal{H}_0 : \theta \in \Theta_0$ versus $\mathcal{H}_1 : \theta \in \Theta_0^c$, sendo $\Theta \subset \Theta$, à medida que $n \longrightarrow \infty$,*

$$-2\log\lambda(\boldsymbol{x}) \xrightarrow{\text{d}} \chi^2_\nu, \tag{10}$$

em que χ^2_ν é uma variável aleatória com distribuição de qui-quadrado com ν grau de liberdade. O valor de ν é igual a diferença entre o número de parâmetros livres especificados por $\theta_0 \in \Theta$ e o número de parâmetros livres especificados por $\theta \in \Theta_0$. □

Em geral o cálculo dos graus de liberdade para a estatística do teste é direto. Considere que $\Theta = \{\theta_1, \theta_2, \ldots, \theta_q\}$ e $\Theta_0 = \{\theta_1, \theta_2, \ldots, \theta_p\}$, então $\nu = q - p$, para $p < q$.

O teste será: rejeita $\mathcal{H}_0 : \theta \in \Theta_0$ se $-2\log\lambda(\boldsymbol{x}) > \chi^2_{\nu,\alpha}$, tal que

$$\lim_{n\to\infty} P_\theta(\text{rejeitar } \mathcal{H}_0) = \alpha, \qquad \text{para cada } \theta \in \Theta_0.$$

Depois que um teste de hipótese foi realizado, as conclusões devem ser relatadas de algum modo estatisticamente significativo. Um método para relatar os resultados de um teste de hipótese é expor o tamanho α, do teste utilizado e a decisão de rejeitar ou não rejeitar H_0. O tamanho do teste fornece informações importantes. Se α for pequeno, a decisão de rejeitar H_0 é bastante convincente, mas se α for grande, a decisão de rejeitar H_0 não é muito convincente porque o teste tem uma grande probabilidade de levar, incorretamente, a esta decisão.