Capítulo 4

Testes de Hipóteses

Neste capítulo discutimos um método inferencial designado por *teste de hipóteses*, o qual constitui uma outra análise do problema inferencial atacado no capítulo anterior, na estimação paramétrica.

Definição 1 *Uma hipótese é uma afirmação sobre a distribuição e/ou os parâmetros de uma população.*

Quando a forma da distribuição ou, equivalentemente, a forma da função densidade de probabilidade, é conhecida e as afirmações dizem respeito ao parâmetro, fala-se em *hipóteses paramétricas.* Caso contrário diz-se *hipóteses não paramétricas.* O objectivo de um teste de hipóteses é decidir, com base na informação contida numa amostra retirada da população em causa, qual de duas hipóteses complementares é verdadeira.

Definição 2 *As duas hipóteses complementares de um problema de teste de hipóteses são designadas por hipótese nula e hipótese alternativa, denotadas por H_0 e H_1, respectivamente.*

No caso de uma hipótese paramétrica, a forma geral das hipóteses é $H_0 : \theta \in \Theta_0$ vs $H_1 : \theta \in \Theta_0^1$, onde $\Theta_0 \subseteq \Theta$. H_0 diz-se *hipótese simples* quando especifica *completamente* a f.d.p conjunta de $(X_1, X_2, \ldots, X_n)$; diz-se *hipótese composta* no caso contrário. Existem várias hipóteses compostas. Hipóteses que afirmem que um parâmetro univariado é grande (e.g., $H_0 : \theta \geq \theta_0$) ou pequeno (e.g., $H_0 : \theta \leq \theta_0$) dizem-se *hipóteses unilaterais*; se $H_0 : \theta \neq \theta_0$ então a hipótese diz-se bilateral.

Exemplo 3 *Se $X_i \sim Poisson(\theta)$, para $i = 1, \ldots, n$, então $H_0 : \theta = 143$ é hipótese simples, $H_0 : \theta \leq 143$ é hipótese composta. Se $X_i, i = 1, 2, \ldots, n$ são também i.i.d. então $H_0 : X_i \sim Poi(\theta), \theta \in I\!R^+$ é também uma hipótese composta.*

Note-se que inclusivamente a hipótese de independência pode ser testada, sendo que esta hipótese é composta (e mais vaga que a maior parte das hipóteses testadas).

Definição 4 *Um teste de hipóteses é uma regra que especifica:*

(i) *Para que valores amostrais a decisão deve levar à aceitação de* H_0 *como verdadeira;*

(ii) *Para que valores* H_0 *é rejeitada e* H_1 *é aceite.*

Numa linguagem comumente aceite pela comunidade estatística, dir-se-à *não rejeitar* H_0 em vez de *aceitar* H_0, uma vez que a primeira expressão tem um caracter menos conclusivo e definitivo que a segunda.

As hipóteses estatísticas são avaliadas através de procedimentos que utilizam a informação contida nos dados, qualificando a consistência dos mesmos com a decisão sobre a (não) rejeição da hipótese nula.

De seguida abordamos duas perspectivas de certa forma distintas de analisar um ensaio de hipóteses. Em particular discutiremos os *ensaios de significância* vs *testes de hipóteses.*

4.1 Ensaios de significância e Testes de hipóteses

Num ensaio de significância o problema é colocado da seguinte forma: considere-se a amostra de variáveis aleatórias $(X_1, X_2, \ldots, X_n)$ (quaisquer; não necessariamente independente e identicamente distribuídas) cuja f.d.p. conjunta se admite pertencer à família $\mathcal{F}$, e seja H_0 a hipótese estatística que envolve $\mathcal{F}$. O objectivo de um ensaio de significância é avaliar o suporte que uma observação $(x_1, x_2, \ldots, x_n)$ de $(X_1, X_2, \ldots, X_n)$ dá a H_0.

Nos ensaios de significância puros existem características que os distinguem dos testes de hipóteses, os quais focaremos nesta secção. Em particular num ensaio de significância não há intervenção de hipóteses que não decorram da hipótese H_0; não são tidas em conta quaisquer hipóteses alternativas. A teoria associada aos ensaios de significância assenta na construção de Fisher, e basiea-se na inferência indutiva.

De seguida recordamos alguma terminologia e notação referentes aos ensaios (e igualmente comum aos testes de hipóteses - ou ensaios de hipóteses) usada já no decorrer de cadeiras anteriores. Assim:

(a) A chave de um ensaio de significância é encontrar uma estatística de teste $T = T(X_1, X_2, \ldots, X_N)$ (que na verdade é uma medida de afastamento, e que é uma função das observações), com as seguintes propriedades:

- · A distribuição de T quando H_0 é verdadeira é (aproximadamente) conhecida.
- · Quanto maior for o valor de T, maior será a evidência para rejeitar H_0, i.e., quanto maior for a medida de afastamento das observações em relação à hipótese, mais concludente é a impugnação de H_0. Daí que T deva ser adequadamente escolhida, traduzindo a tal medida de afastamento em relação à hipótese H_0.

(b) Seja $t_{obs} = T(x_1, x_2, \ldots, x_n)$ o valor observado da estatística de teste T para uma dada amostra concreta $(x_1, x_2, \ldots, x_n)$. Define-se a seguinte probabilidade:

$$p_{obs} = P(T \geq t_{obs} | H_0)$$

> O valor p_{obs} (ou, como é hábito designar-se, o valor-p) mede a evidência que os dados fornecem para aceitar H_0 e tem a seguinte interpretação: quanto menor for o valor de p_{obs}, menor será a consistência da hipótese H_0 face aos dados observados.

Note-se que a conclusão de um ensaio de hipóteses não é necessariamente a rejeição ou a não-rejeição de uma hipótese, embora muitas vezes se tome como decisão a rejeição de H_0 quando p_{obs} é menor que determinado limiar (que não é universal, dependendo muitas vezes do tipo de hipótese que está em jogo).

Por vezes a partir do mesmo conjunto de dados efectuam-se vários testes, cada um dos quais correspondendo a uma medida de afastamento (ou a uma estatística de teste). Nesse caso pode-se optar como nível de significância associado à combinação dos ensaios o valor menos favorável à hipótese, i.e.,

$$p_{com} = m \min\{p_{1,obs}, p_{2,obs}, \dots, p_{m,obs}\}$$

onde $p_{i,obs}$ representa o valor-p referente à utilização do i-ésimo ensaio de significância. Por exemplo, se tanto valores elevados da estatística de teste como valores baixos conduzirem a um afastamento em relação à hipótese H_0 então podemos tomar a estatística de teste T como dando igual (ou aproximadamente igual...) importância a valores $+t$ e $-t$, pelo que nesse caso dever-se-á optar como estatística de teste $|T|$, donde em ambos os casos o nível de significância p_{obs} passa a ser

$$p_{obs} = P(|T| \geq |t_{obs}||H_0)$$

No caso de valores elevados de T não terem o mesmo significado que valores baixos, então não se deve dar igual importância aos dois casos, sendo preferível tomar como nível de significância o que resulta da combinação dos dois testes. Assim tomando

$$p_{1,obs} = P(T \geq t_{obs}|H_0), \quad p_{2,obs} = P(T \leq t_{obs}|H_0)$$

então vem que o nível de significância do teste combinado é

$$p_{obs} = 2 \min(p_{1,obs}, p_{1,obs}).$$

No caso dos testes de hipóteses (ou ensaios de hipóteses) existe um caracter essencialmente decisional, dando lugar à decisão entre o *aceitar* e o *rejeitar* da hipótese nula H_0. Assim sendo num teste de hipóteses a hipótese alternativa H_1 desempenha um papel fundamental, ao contrário do que se passa nos ensaios de significância.

Um ensaio de hipóteses é descrito da seguinte forma: seja $(X_1, X_2, \dots, X_n)$ uma amostra (não necessariamente i.i.d.) da população em causa, com espaço amostral $\Omega_{\underline{X}}$, com f.d.p. conjunta pertencente à família $\mathcal{F}$. Seja ainda $W \subseteq \Omega_{\underline{X}}$ tal que

$$\begin{aligned}(X_1, X_2, \dots, X_n) \in W &\Rightarrow \text{ rejeição de } H_0\\ (X_1, X_2, \dots, X_n) \notin W &\Rightarrow \text{ não rejeição de } H_0\end{aligned}$$

O subconjunto W designa-se por *região crítica* ou *região de rejeição.* $\Omega_{\underline{X}} - W = W^c$ é a *região de aceitação.*

Uma vez que o ensaio de hipóteses leva a uma decisão sobre a rejeição ou não, existe a possibilidade de se tomar uma decisão errada. Ao proceder ao ensaio de H_0 podem cometer-se dois tipos de erros: rejeitar H_0, sendo porém H_0 verdadeira ou, pelo contrário, não rejeitar H_1, sendo H_0 falsa. No primeiro caso diz-se que se cometeu um erro de primeira espécie (*erro tipo I*) enquanto que no segundo caso se cometeu um erro de segunda espécie (*erro tipo II*). O aspecto fundamental reside na escolha de um procedimento que minimize a probabilidade de ocorrência de cada um destes tipos de erros.

Um dado procedimento ou ensaio é designado por δ, sendo W_δ a região crítica associada. A probabilidade de cometer um erro tipo I é dada por

$$\alpha = P\{(X_1, X_2, \ldots, X_n) \in W_\delta | H_0\} = \int \ldots \int_{W_\delta} f_{\underline{X}}(\underline{x}|\theta_0) d\underline{x}$$

enquanto que a probabilidade de ocorrer um erro tipo II é dada por

$$1 - \beta = P\{(X_1, X_2, \ldots, X_n) \notin W_\delta | H_1\} = 1 - \int \ldots \int_{W_\delta} f_{\underline{X}}(\underline{x}|\theta_1) d\underline{x}$$

Ao valor α (probabilidade de cometer um erro de primeira espécie) chama-se *dimensão do teste* (ou tamanho do teste) enquanto que β (probabilidade de não cometer um erro de segunda espécie) chama-se *potência do teste.*

Quando as hipóteses H_0 e H_1 são hipóteses compostas do tipo $H_0 : \theta \in \Theta_0$ vs $H_1 : \theta \in \Theta_1$, então o cálculo de α e de β depende do particular espaço paramétrico Θ_0 ou Θ_1, respectivamente. Se H_0 for uma hipótese composta, por exemplo, então a dimensão do teste é α com

$$\alpha = \sup_{\theta \in \Theta_0} P\{(X_1, X_2, \ldots, X_n) \in W_\delta | \theta\}$$

Em vez de potência, passa a falar-se *função potência*, com

$$\beta_\delta(\theta) = P\{(X_1, X_2, \ldots, X_n) \in W_\delta | \theta\}, \quad \theta \in \Theta$$

Note-se que a função potência está definida em todo o espaço paramétrico, embora seja particularmente interessante quando $\theta \in \Theta_1$.

Recomenda-se reler e estudar a matéria referente a testes de hipóteses dada em PE1 e PE2.

4.2 Propriedades dos testes

Como já vimos na secção anterior, quando se toma uma decisão respeitante à validade de uma hipótese podemos cometer erros. Nesta secção discutimos como esses erros podem ser controlados.

4.2.1 Função potência

Idealmente a função potência $\beta(\theta)$, definida para todo o $\theta \in \Theta$, tal como foi introduzida na secção anterior, deveria ser zero para todo o $\theta \in \Theta_0$. Porém só em casos triviais é que se regista esta situação. Qualitativamente, um bom teste deve ter função potência próxima de zero para quase todo o $\theta \in \Theta_0$, e próxima de um para quase todo o $\theta \in \Theta_1$. Nessa situação é facilmente discernível qual a decisão a adoptar face à veracidade das hipóteses.

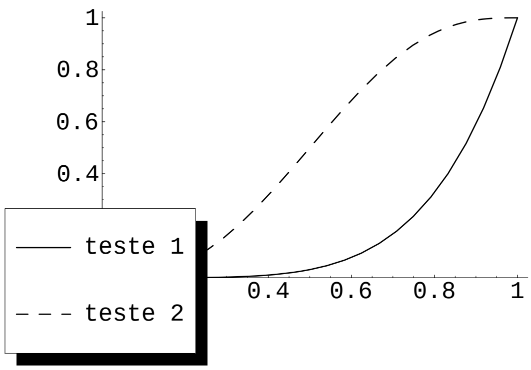
Exemplo 5 *Seja $X \sim Bin(5,\theta)$, e considere-se o teste $H_0 : \theta \leq \frac{1}{2}$ vs $H_1 : \theta > \frac{1}{2}$. A regra de decisão face à validade das hipóteses é: rejeitar H_0 se e só se só se observarem 5 sucessos nas 5 observações. Então a função potência do teste é*

$$\beta_1(\theta) = P(X = 5|\theta) = \theta^5$$

Suponhamos agora que damos outra regra de decisão: rejeitar H_0 se $X \geq 3$. Então

$$\beta_2(\theta) = P(X \geq 3|\theta) = \binom{5}{3}\theta^3(1-\theta)^2 + \binom{5}{4}\theta^4(1-\theta) + \binom{5}{5}\theta^5(1-\theta)^0$$

Representando β_1 e β_2 em função de θ vem: Logo dado que $\beta_2(\theta) > \beta_1(\theta)$ para $\theta \leq \frac{1}{2}$,



a probabilidade de erro tipo I é maior para o segundo teste, enquanto que a probabilidade de erro tipo II é maior para o primeiro teste. Para decidir qual dos dois testes é melhor, terão de ser feitas considerações sobre a estrutura de erro pretendida, i.e., terá de ser decidido se se priveligia os erros tipo I ou os erros tipo II.

Tipicamente a função potência depende da dimensão amostral n, uma vez que a distribuição da estatística de teste depende igualmente da dimensão amostral. Para uma dimensão amostral fixa, é virtualmente impossível fazer ambas as probabilidades de erro arbitrariamente pequenas. Geralmente dá-se primazia aos erros tipo I, pelo que neste sentido os testes não são simétricos, i.e, tende-se a minorar a probabilidade de erro tipo I. Assim sendo a estratégia mais usual é fixar um nível para a probabilidade de erro tipo I e, de entre todos os testes que obedecem a esta restrição, escolher aquele que minimiza a probabilidade de erro tipo II.

Doravante, nesta secção respeitante às propriedades dos testes, assumiremos que o parâmetro desconhecido θ sobre o qual é conduzido o teste de hipóteses é uniparamétrico.

Eventualmente poderá haver outro(s) parâmetro(s) desconhecido(s) que nesta situação são designados por *parâmetros perturbadores.* O caso de hipóteses multiparamétricas será eventualmente tratado em trabalhos.

Definição 6 *Para $0 \leq \alpha \leq 1$, um teste com função potência $\beta(\theta)$ diz-se ser um teste de tamanho α se $\sup_{\theta \in \Theta_0} \beta(\theta) = \alpha$. Um teste com nível α é um teste para o qual $\sup_{\theta \in \Theta_0} \beta(\theta) \leq \alpha$.*

Assim sendo o conjunto de testes de nível α contém o conjunto de testes de tamanho α. A distinção entre estes dois conceitos nem sempre é clara. Mas é especialmente importante quando não é possível construir testes com tamanho α, pelo que nessas situações há um compromisso que leva a aceitar testes de nível α.

Em aplicações é corrente considerar-se $\alpha = 0.01, 0.05, 0.10$. Quando se fixa o valor de α, não estamos a controlar necessariamente a probabilidade de erro tipo II!

Quando se fixa o nível de significância α, uma decisão tem de ser tomada, i.e., tem de ser concluída ou a rejeição ou a não rejeição de H_0. Claramente que quanto menor for α, a decisão de rejeitar H_0 é mais convincente (com óbvias implicações na probabilidade de erro tipo II); mas se α é elevado, a decisão de rejeitar H_0 não é muito convincente pois o teste tem elevada probabilidade de estar a tomar uma decisão incorrecta. Uma outra forma de concluir sobre a aceitabilidade da hipótese nula é através do valor-p, já discutido a propósito dos ensaios de significância. O valor-p para uma amostra particular $\underline{x}$ é o menor valor de α para o qual os resultados amostrais levam à rejeição de H_0. Note-se que o valor-p depende intrinsecamente da amostra, pelo que nova análise tem de ser realizada na sua integralidade se nova amostra for recolhida. Além disso um valor-p **não** é um nível de significância; em particular, e dado que depende das observações, não tem a interpretação como taxa de erro que o valor α tem. Quando se analisa a veracidade de um par de hipóteses através do valor-p, a conclusão sobre aceitação de H_1, quando tal ocorre, é sempre mais convincente que a conclusão sobre a não rejeição de H_0, quando tal ocorre: se p for pequeno, maior será a evidência da veracidade de H_1.

Para além do nível de significância e do valor-p, existem outras características que definem um teste. Por exemplo, seria desejável que um teste rejeitasse H_0 com maior evidência quando o verdadeiro valor $\theta \in \Theta_0^c$ do que quando $\theta \in \Theta_0$.

Definição 7 *Um teste com função potência $\beta(\theta)$ é centrado se $\beta(\theta') \geq \beta(\theta''), \forall \theta' \in \Theta_0^c, \theta'' \in \Theta_0$.*

Exemplo 8 *Seja $X \sim N(\theta, \sigma^2)$, com σ conhecido. Ao efectuar o teste $H_0 : \theta \leq \theta_0$ vs $H_1 : \theta > \theta_0$, considera-se a estatística de teste*

$$\frac{\bar{X} - \theta_0}{\sigma/\sqrt{n}} \sim_{H_0} N(0, 1)$$

e a região crítica

$$W_\delta = \{\frac{\bar{X} - \theta_0}{\sigma/\sqrt{n}} > c\}$$

onde c é uma constante positiva. Para este teste a função potência é dada por

$$\beta(\theta) = P(Z > c + \frac{\theta - \theta_0}{\sigma/\sqrt{n}} | Z \sim N(0,1))$$

Como $\beta(\theta)$ é, para θ_0 fixo, uma função crescente de θ, então segue-se que

$$\beta(\theta) > \beta(\theta_0) = \sup_{t \leq \theta_0} \beta(t), \forall \theta > \theta_0$$

pelo que o teste é centrado.

4.2.2 Testes UMP

Já vimos que na elaboração de testes de hipóteses considera-se frequentemente mais importante minorar a probabilidade de erro tipo I que a de tipo II. Porém um bom teste será certamente um para o qual não só a probabilidade de erro tipo I é baixa mas também a de tipo II, i.e., deverá ter uma função potência elevada para $\theta \in \Theta_0^c$. Um teste com tamanho α para o qual a função potência tenha esta característica é certamente um bom candidato para melhor teste.

Definição 9 *Seja $\mathcal{C}$ a classe de testes para o conjunto de hipóteses $H_0 : \theta \in \Theta_0$ vs $H_1 : \theta \in \Theta_0^c$. Um teste pertencente a $\mathcal{C}$, com função potência $\beta(\theta)$ é um teste* **uniformemente mais potente** *(UMP, de uniformly most powerful) se $\beta(\theta) \geq \beta'(\theta)$ para todo o $\theta \in \Theta_0^c$ e para todo o $\beta'(.)$ (função potência de algum teste pertencente à classe $\mathcal{C}$).*

Nesta definição $\mathcal{C}$ é a classe de todos os testes com nível α.

O problema dos testes UMP é que raramente existem em problemas reais; mas quando existem, então são os melhores testes na classe em questão. O problema reside então em determinar em que condições existem testes UMP e como identificá-los.

Antes de duduzir o resultado, tentaremos motivá-lo. Considere-se o teste $H_0 : \theta = \theta_0$ vs $H_1 : \theta = \theta_1$, e suponhamos que $\Theta = \{\theta_0, \theta_1\}$. A melhor região crítica (i.e., a região crítica correspondente a um teste UMP) de tamanho α deve certamente satisfazer as seguintes condições:

$$P(X \in W_\delta | \theta_0) = \sum_{x \in W_\delta} f_X(x|\theta_0) = \alpha$$
$$P(X \in W_\delta | \theta_1) = \sum_{x \in W_\delta} f_X(x|\theta_1) = \text{máxima}$$

a primeira das quais estabelece a probabilidade de cometer erro tipo I e a segunda maximiza a potência (i.e., a probabilidade de não cometer erro tipo II).

É pois necessário identificar os pontos $x \in \Omega$ que W_δ deve conter. Para isso cada ponto do espaço é avaliado segundo dois critérios: contabiliza-se $f_X(X|\theta_0)$ (i.e. H_0 é verdadeira) e $f_X(X|\theta_1)$ (i.e. H_0 é falsa). No seu conjunto os pontos incluídos em W_δ não devem ter pontuação superior a α quando avaliados pelo primeiro critério e devem ter pontuação máxima quando avaliados pelo segundo critério. Assim a atribuição de um ponto x a W_δ deve ser feita de acordo com a razão:

$$\lambda(x|\theta_0,\theta_1)=\frac{f_X(x|\theta_0)}{f_X(x|\theta_1)}$$

atribuindo então a W_δ tantos pontos x (por ordem decrescente da razão) tantos quantos o permita a condição $P(X\in W_\delta|\theta_0)=\alpha$ ou, na impossibilidade de atingir a igualdade, $P(X\in W_\delta|\theta_0)\leq\alpha$. Assim estaremos a procurar a melhor região crítica de tamanho não superior a α.

Exemplo 10 *Uma urna contem bolas pretas e bolas brancas, na proporção de* $3:1$, *desconhecendo-se porém qual a cor dominante. Seja* θ *a probabilidade de tirar uma bola preta da urna. A experiência aleatória consiste em retirar 3 bolas da urna, ao acaso e com reposição, registando o número de bolas pretas observadas. Seja então* X *a v.a. que indica o número de bolas pretas observadas na experiência aleatória. Segue-se que* $X\sim Binomial(3,\theta)$. *Pretendemos testar* $H_0:\theta=\frac{1}{4}$ *vs* $H_1:\theta=\frac{3}{4}$. *Então*

x	0	1	2	3
$f_X(x\|\frac{1}{4})$	0.421875	0.421875	0.140625	0.015625
$\lambda(x\|\frac{1}{4},\frac{3}{4})=\frac{\frac{1}{4}^x\frac{3}{4}^{3-x}}{\frac{3}{4}^x\frac{1}{4}^{3-x}}$	0.0037	0.333	3	27

Suponhamos que $\alpha=0.2$. *Então* $f_X(2|\theta=\frac{1}{4})+f_X(3|\theta=\frac{1}{4})=0.15625$, *sendo a única combinação cuja soma de probabilidades dá menor ou igual a* 0.2, *segue-se que* $W_\delta=\{2,3\}=\{x:\lambda(x|\frac{1}{4},\frac{3}{4})<c\}$, *onde* c *é um valor qualquer entre* 0.333 *e* 3.

O teorema que se segue descreve os testes UMP para um tamanho α desde que tanto H_0 como H_1 sejam hipóteses simples.

Teorema 11 *(***Lema de Neyman-Pearson***) Seja* $\underline{X}$ *uma amostra de uma população com f.d.p.* $f_X(.|\theta)$, *com* $\theta\in\{\theta_0,\theta_1\}$, c *um número real positivo e* W_δ *um conjunto do espaço amostral, tais que*

$$\frac{\prod_{i=1}^n f(x_i|\theta_1)}{\prod_{i=1}^n f(x_i|\theta_0)}\geq c,\quad\forall(x_1,\ldots,x_n)\in W_\delta$$
$$\frac{\prod_{i=1}^n f(x_i|\theta_1)}{\prod_{i=1}^n f(x_i|\theta_0)}\leq c,\quad\forall(x_1,\ldots,x_n)\in W_\delta^c$$
$$P\{(X_1,\ldots,X_n)\in W_\delta|\theta=\theta_0)=\alpha$$

Então W_δ *é a melhor região crítica de tamanho* α *para ensaiar* $H_0:\theta=\theta_0$ *vs* $H_1:\theta=\theta_1$.

Demonstração: (Considera-se apenas o caso contínuo, sendo porém óbvia a extensão ao caso discreto).

Seja δ e δ' dois testes, ambos de nível α, e com função potência $\beta(\theta)$ e $\beta(\theta')$, respectivamente para δ e para δ'. Suponhamos ainda que o teste δ tem região crítica W_δ para a qual

$$\frac{\prod_{i=1}^{n} f(x_i|\theta_1)}{\prod_{i=1}^{n} f(x_i|\theta_0)} \geq c, \quad \forall (x_1, \ldots, x_n) \in W_\delta$$
$$\frac{\prod_{i=1}^{n} f(x_i|\theta_1)}{\prod_{i=1}^{n} f(x_i|\theta_0)} \leq c, \quad \forall (x_1, \ldots, x_n) \in W_\delta^c$$

tal como enunciado no teorema.

Seja agora

$$\phi(\underline{x}) = \begin{cases} 1 & \underline{x} \in W_\delta \\ 0 & c.c. \end{cases} \qquad \phi'(\underline{x}) = \begin{cases} 1 & \underline{x} \in W_{\delta'} \\ 0 & c.c. \end{cases}$$

duas funções de teste (i.e., são as indicatrizes da rejeição correspondende aos testes δ e δ'). Dado que $\phi' \in [0, 1]$, então

$$\big(\phi(\underline{x}) - \phi'(\underline{x})\big)\left(f(\underline{x}|\theta_1) - cf(\underline{x}|\theta_0)\right) \geq 0, \quad \forall \underline{x}$$

uma vez que $\phi = 1$ se $f(\underline{x}|\theta_1) > cf(\underline{x}|\theta_0)$ e $\phi = 0$ caso contrário. Logo o seu integral também deve ser não negativo, i.e.,

$$\begin{aligned}
0 &\leq \int \left[\big(\phi(\underline{x}) - \phi'(\underline{x})\big)\left(f(\underline{x}|\theta_1) - cf(\underline{x}|\theta_0)\right)\right] d\underline{x} \\
&= \int_{W_\delta \cap W_{\delta'}} \left[\big(\phi(\underline{x}) - \phi'(\underline{x})\big)\left(f(\underline{x}|\theta_1) - cf(\underline{x}|\theta_0)\right)\right] d\underline{x} \\
&+ \int_{\Omega - (W_\delta \cup W_{\delta'})} \left[\big(\phi(\underline{x}) - \phi'(\underline{x})\big)\left(f(\underline{x}|\theta_1) - cf(\underline{x}|\theta_0)\right)\right] d\underline{x} \\
&+ \int_{W_\delta \cap W_{\delta'}^c} \left[\big(\phi(\underline{x}) - \phi'(\underline{x})\big)\left(f(\underline{x}|\theta_1) - cf(\underline{x}|\theta_0)\right)\right] d\underline{x} \\
&+ \int_{W_\delta^c \cap W_{\delta'}} \left[\big(\phi(\underline{x}) - \phi'(\underline{x})\big)\left(f(\underline{x}|\theta_1) - cf(\underline{x}|\theta_0)\right)\right] d\underline{x} \\
&= \int_{W_\delta \cap W_{\delta'}^c} \left[\left(f(\underline{x}|\theta_1) - cf(\underline{x}|\theta_0)\right)\right] d\underline{x} - \int_{W_\delta^c \cap W_{\delta'}} \left[\left(f(\underline{x}|\theta_1) - cf(\underline{x}|\theta_0)\right)\right] d\underline{x} \\
&= \beta(\theta_1) - \beta'(\theta_1) - c\big(\beta(\theta_0) - \beta'(\theta_0)\big)
\end{aligned}$$

Logo

$$0 \leq \beta(\theta_1) - \beta'(\theta_1) - c\big(\beta(\theta_0) - \beta'(\theta_0)\big) \leq \beta(\theta_1) - \beta'(\theta_1)$$

pelo que $\beta(\theta_1) \geq \beta'(\theta_1)$ pelo que ϕ tem mais potência que ϕ'. Como ϕ' é um teste arbitrário de nível α, e θ_1 é o único ponto de Θ_0^c, então δ é um teste UMP de nível α. ■

A constante c referida no teorema de Neyman-Pearson deve escolher-se de forma a que a região crítica tenha tamanho α; porém em muitos casos não é necessário determinar esta constante (o que aconteceu no exemplo anterior referente ao teste da proporção de uma Bernoulli), sendo por isso possível determinar a região crítica sem necessitar de c. No caso da população ter distribuição discreta pode não ser possível encontrar uma região crítica

W_δ para a qual $P\{(X_1,\ldots,X_n)\in W_\delta|\theta=\theta_0)=\alpha$; nesse caso esta condição é substituída por

$$P\{(X_1,\ldots,X_n)\in W_\delta|\theta=\theta_0)\leq\alpha$$

Teorema 12 *Se $T=T(X_1,\ldots,X_n)$ é uma estatística suficiente para θ, a região crítica W_δ UMP dada pelo lema de Neyman-Pearson pode exprimir-se em função de T.*

Demonstração: O resultado decorre do lema de Neyman-Pearson, usando o critério da factorização para estatísticas suficientes. ■

Exemplo 13 *Seja $X\sim N(\theta,\sigma^2)$, com σ conhecido. Pretende-se testar*

$$H_0:\theta=\theta_0 \quad versus \quad H_1:\theta=\theta_1$$

com $\theta_1>\theta_0$, a partir de uma amostra $\underline{x}$ de dimensão n. A f.d.p. conjunta de $\underline{x}$ é dada por

$$f_{\underline{X}}(\underline{x}|\theta_j)=\left(\frac{1}{\sqrt{2\pi}}\right)^n e^{-\frac{\sum_{i=1}^n(x_i-\theta_j)^2}{2\sigma^2}},\quad j=0,1$$

pelo que a razão das f.d.p. é dada por

$$\lambda(\underline{x}|\theta_0,\theta_1)=e^{-\frac{\sum_{i=1}^n(x_i-\theta_1)^2-\sum_{i=1}^n(x_i-\theta_0)^2}{2\sigma^2}}\geq c$$

se e só se

$$-\frac{\sum_{i=1}^n(x_i-\theta_1)^2-\sum_{i=1}^n(x_i-\theta_0)^2}{2\sigma^2}\geq\ln c$$

$$\sum_{i=1}^n(x_i-\theta_1)^2-\sum_{i=1}^n(x_i-\theta_0)^2\geq 2\sigma^2\ln c$$

$$\sum_{i=1}^n x_i^2+n\theta_1^2-2\theta_1\sum_{i=1}^n x_i-\sum_{i=1}^n x_i^2-n\theta_0^2-2\theta_0\sum_{i=1}^n x_i\geq 2\sigma^2\ln c$$

$$\sum_{i=1}^n x_i\geq\frac{2\sigma^2\ln c+(n/2)(\theta_1^2-\theta_0^2)}{\theta_1-\theta_0}=c'$$

pelo que a região crítica vem em função de $\sum_{i=1}^n X_i$ ou, equivalementemente, em função de $\bar{X}$. Seja $c''=\frac{c'}{n}$. Então c'' pode ser calculada a partir de

$$P(\bar{X}\geq c''|\theta=\theta_0)=\alpha$$

donde

$$c''=\theta_0+\frac{\sigma}{\sqrt{n}}\Phi^{-1}(1-\alpha)$$

Note-se que no exemplo anterior não foi preciso determinar c, uma vez que se consegue determinar a região crítica à custa da estatística suficiente $\bar{X}$. Note-se ainda que a região crítica W_δ é independente de θ_1, desde que este seja maior que θ_0. Logo o lema de Neyman-Pearson serve não só para determinar o teste UMP para o par de hipóteses $H_0:\theta=\theta_0$ vs

$H_1 : \theta = \theta_1$ mas também para determinar o teste UMP para $H_0 : \theta = \theta_0$ vs $H_1 : \theta > \theta_0$, passando assim a ser um teste de hipótese simples contra hipótese composta unilateral.

Mas a conclusão anterior pode ainda ser estendida. Note-se que

$$P(\bar{X} \geq c''|\theta = \theta') < P(\bar{X} \geq c''|\theta = \theta_0) = \alpha, \quad \forall \theta' < \theta_0$$

em virtude das propriedades da f.d.p. Normal. Então concluímos que é possível determinar um teste UMP para o conjunto de hipóteses

$$H_0 : \theta \leq \theta_0 \quad \text{vs} \quad H_1 : \theta > \theta_0$$

sendo a região crítica igual à do teste $H_0 : \theta = \theta_0$ vs $H_1 : \theta = \theta_1 (> \theta_0)$.

O mesmo raciocínio se aplica ao teste $H_0 : \theta = \theta_0$ vs $H_1 : \theta = \theta_1$ (com $\theta_0 > \theta_1$), mas sendo que a região de rejeição é agora da forma

$$W_\delta = \{\bar{X} < a\}$$

com

$$a = \theta_1 + \frac{\sigma}{\sqrt{n}} \Phi^{-1}(\alpha)$$

resultado este extensível aos testes $H_0 : \theta = \theta_0$ vs $H_1 : \theta < \theta_0$ e $H_0 : \theta \geq \theta_0$ vs $H_1 : \theta < \theta_0$.

Porém se se pretender ensaiar $H_0 : \theta = \theta_0$ vs $H_1 : \theta \neq \theta_0$ então não existe teste UMP, uma vez que se $\theta < \theta_0$ a região de rejeição é a aba esquerda da v.a. $\bar{X}$ enquanto que se $\theta > \theta_0$ é a aba direita, como acabámos de ver.

Uma grande classe de problemas que admitem testes UMP de tamanho α envolvem testes de hipóteses unilaterais e f.d.p. com função de verosimilhança monótona.

Definição 14 *Uma família de f.d.p. $\mathcal{G} = \{g(.|\theta), \theta \in \Theta\}$ para uma v.a. T unidimensional com parâmetro real θ diz-se ter razão de verosimilhança monótona (MLR, de monotone likelihood ratio) de para qualquer $\theta_2 > \theta_1$, a razão $\frac{g(t|\theta_2)}{g(t|\theta_1)}$ é uma função monótona em t para todo o t pertencente ao suporte da f.d.p. g.*

As f.d.p. que pertencem à família exponencial são MLR, por exemplo, pelo que esta propriedade verifica-se nas distribuições mais vulgares.

Já vimos pois que embora o teorema de Neyman-Pearson seja aplicável, tal como é enunciado, a testes de hipóteses simples vs simples, com alguns argumentos podemos provar que exustem regiões UMP para testes não necessariamente nestas condições, como o exemplo anterior demonstra. Na verdade

Corolário 15 *Para testar $H_0 : \theta \in \Theta_0$ vs $H_1 : \theta \in \Theta_0^c$ admita que existe um teste δ de nível α baseado numa estatística suficiente T, com região de rejeição W_δ, para o qual existe um $\theta_0 \in \Theta_0$ tal que*

$$P(T \in W_\delta|\theta = \theta_0) = \alpha$$

Seja ainda $g(.|\theta)$ a f.d.p. de T. Se existir $c \geq 0$ tal que

$$g(t|\theta') > kg(t|\theta_0), \quad t \in W_\delta$$
$$g(t|\theta') < kg(t|\theta_0), \quad t \in W_\delta^c$$

para qualquer $\theta' \in \Theta_0^c$, então δ é um teste UMP de tamanho α.

Note-se que um teste pode ter tamanho α sem que exista pelo menos um $\theta_0 \in \Theta_0$ para o qual $P(T \in W_\delta|\theta = \theta_0) = \alpha$, pelo que no teorema anterior as condições exigidas são consideravelmente fortes.

Se a família de f.d.p. for MLR, então o resultado anterior pode ser enunciado da seguinte forma:

Corolário 16 *(***teorema de Karlin-Rubin***) Considere-se o par de hipóteses $H_0 : \theta \leq \theta_0$ vs $H_1 : \theta > \theta_0$. Seja T uma estatística suficiente para a família em causa, e suponhamos que a família de f.d.p. de T, $\mathcal{G} = \{g_T(.|\theta), \theta \in \Theta\}$ é MLR. Então para qualquer t_0 o teste δ que rejeita H_0 se e só se $T > t_0$ é um teste UMP de tamanho α, onde $\alpha = P(T > t_0|\theta_0)$.*

Claramente que o corolário anterior também pode ser enunciado se em vez do par de hipóteses apontado tivermos $H_0 : \theta \geq \theta_0$ vs $H_1 : \theta < \theta_0$. Nesse caso o teste que rejeita H_0 sempre que $T < t_0$ (para qualquer t_0) é um teste UMP de tamanho α, com $\alpha = P(T < t_0|\theta_0)$.

4.2.3 Testes UMPU

Um teste UMP é na prática muito difícil de obter. Assim fixado um nível de significância α, a classe de testes de nível α é tão extensa que muitas vezes não há nenhum teste que domine, em termos de função potência, todos os restantes testes da mesma classe. Nessa situação uma estratégia possível é restringir a classe de testes, considerando uma subclasse e procurando aí um teste UMP. Vejamos de seguida um exemplo em que não existe teste UMP.

Exemplo 17 *Sejam $X_1, \ldots, X_n$ n v.a. i.i.d., com distribuição $N(\theta, \sigma^2)$, sendo σ conhecido. Considere-se o conjunto de hipóteses $H_0 : \theta = \theta_0$ vs $H_1 : \theta \neq \theta_0$. Para um dado valor α, um teste de nível α é tal que satisfaz*

$$P(\text{ rejeitar } H_0|\theta = \theta_0) \leq \alpha$$

Note-se que a hipótese alternativa é bilateral. Suponhamos que consideramos primeiro um ponto $\theta_1 < \theta_0$. Então é possível deduzir, com a ajuda do lema de Neyman-Pearson, que o teste a que corresponde a região $W_1 = \{\bar{X} < -\sigma\Phi^{-1}(1-\alpha/2)/\sqrt{n} + \theta_0\}$ tem a potência máxima em θ_1. Logo trata-se de um teste UMP e que, a menos de um região de probabilidade nula, define a única região UMP.

Suponhamos agora que consideramos um outro teste para o qual a região de rejeição é $\{\bar{X} > \sigma\Phi^{-1}(1-\alpha/2)/\sqrt{n} + \theta_0$*, o qual é igualmente um teste com nível* α*. Seja* $\beta_i(.)$ *a função potência para o i-ésimo (*$i = 1, 2$*) teste. Para qualquer* $\theta_2 > \theta_0$*:*

$$\begin{aligned}
\beta_2(\theta_2) &= P(\bar{X} > \sigma\Phi^{-1}(1-\alpha/2)/\sqrt{n} + \theta_0|\theta = \theta_2) \\
&= P(\frac{\bar{X}-\theta_2}{\sigma/\sqrt{n}} > \Phi^{-1}(1-\alpha/2) + \frac{\theta_0 - \theta_2}{\sigma/\sqrt{n}}|\theta = \theta_2) \\
&> P(Z > \Phi^{-1}(1-\alpha/2)|Z \sim N(0,1)) \quad \textit{pois } \theta_0 - \theta_2 < 0 \\
&= P(Z \leq -\Phi^{-1}(1-\alpha/2)) \\
&> P(\frac{\bar{X}-\theta_2}{\sigma/\sqrt{n}} < -\Phi^{-1}(1-\alpha/2) + \frac{\theta_0 - \theta_2}{\sigma/\sqrt{n}}|\theta = \theta_2) \\
&= P(\bar{X} < -\sigma\Phi^{-1}(1-\alpha/2)/\sqrt{n} + \theta_0|\theta = \theta_2) \\
&= \beta_1(\theta_2)
\end{aligned}$$

i.e., o segundo teste tem potência superior à do teste 1 pelo menos para um dos pontos $\theta \in \Theta_0^c$*; logo o teste 1 não é UMP, pelo que não existe nenhum teste UMP de nível* α *para este problema.*

Na verdade quando se testam hipóteses compostas bilaterais do tipo $H_0 : \theta = \theta_0$ vs $H_1 : \theta \neq \theta_0$ ou $H_0 : \theta \in (\theta_1, \theta_2)$ vs $H_1 : \theta < \theta_1$ ou $\theta > \theta_2$ (com $\theta_2 > \theta_1$), impôr restrições à família de distribuições não permite estabelecer a existência, em termos gerais, de testes UMP, sendo particularmente importante neste caso restringir a classe de testes de determinado nível.

Já vimos anteriormente a noção de teste centrado (ou não enviesado). Recordemos que num teste centrado $\beta(\theta) \leq \beta(\theta'), \forall\theta \in \Theta_0, \theta' \in \Theta_0^c$. A interpretação de um teste centrado é a seguinte: se T é um teste de tamanho α não enviesado, a probabilidade de rejeitar H_0 quando falsa nunca é inferior à probabilidade de rejeitar H_0 quando verdadeira. Um teste que seja uniformemente mais potente de entre os testes centrados diz-se ser um teste *uniformemente mais potente centrado* (UMPU, de *uniformly most powerful unbiased*).

O teorema que se segue é uma generalização do lema de Neyman-Pearson, podendo ser aplicado a funções arbitrárias (e não somente funções densidade de probabilidade).

Teorema 18 *Sejam* $c_1, c_2, \ldots, c_m$ *constantes e* $f_1(\underline{x}), f_2(\underline{x}), \ldots, f_{m+1}(\underline{x})$ *funções reais. Seja ainda* $\mathcal{C}$ *a classe de funções* $\phi(.)$ *para as quais* $0 \leq \phi(\underline{x}) \leq 1$ *para todo o* $\underline{x}$ *e*

$$\int \phi(\underline{x}) f_i(\underline{x}) d\underline{x} = c_i, \quad i = 1, 2, \ldots, m$$

Se $\phi^\star(.)$ *é uma função de* $\mathcal{C}$ *para a qual*

$$\phi(\underline{x}) = \begin{cases} 1 & f_{m+1}(\underline{x}) > \sum_{i=1}^m k_i f_i(\underline{x}) \\ 0 & f_{m+1}(\underline{x}) < \sum_{i=1}^m k_i f_i(\underline{x}) \end{cases}$$

para algumas constantes $k_1, \ldots, k_m$ *então* $\phi^\star$ *maximiza* $\int \phi(\underline{x}) f_{m+1}(\underline{x}) d\underline{x}$ *de entre todas as funções de* $\mathcal{C}$*.*

Note-se que se $m = 1$, $f_1(.) = f(.|\theta = \theta_0)$ e $f_2(.) = f(.|\theta = \theta_1)$ então o teorema anterior corresponde ao lema de Neyman-Pearson.

As funções ψ deste teorema no caso das funções f_i serem f.d.p. correspondem a *funções de teste aleatórias*, com a seguinte interpretação: suponhamos que temos um teste δ que leva à rejeição ou à não-rejeição de uma hipótese H_0. Se a decisão do teste for *não rejeitar* H_0, então é efectuada uma experiência de Bernoulli, independente da decisão do teste, sendo p a probabilidade de se observar um sucesso. Se o sucesso for observado, então a decisão final é rejeitar H_0; caso contrário não se rejeita H_0. Assim sendo este teste depende não só do teste original δ como também do resultado de uma experiência de Bernoulli independente. Reportando-nos ao teorema anterior, $\phi(\underline{x})$ designa a probabilidade de rejeitar H_0 quando a amostra observada é $\underline{x}$.

Vejamos como aplicar este resultado na determinação de testes UMPU. Seja X uma população com f.d.p. $f_X(.|\theta)$, onde θ é um parâmetro desconhecido. Considere-se o conjunto de hipóteses $H_0 : \theta = \theta_0$ vs $H_1 : \theta \neq \theta_0$. Por argumentos similares ao teorema de Karlin-Rubin, o teste UMPU deverá pertencer à classe de testes cuja região crítica depende de uma estatística suficiente T, com f.d.p. $f_T(.)$. Seja $\theta_1 \neq \theta_0$, $\underline{x}$ uma amostra de X e $t = T(\underline{x})$ o valor observado da estatística de teste. Tome-se:

$$f_3(\underline{x}) = f_T(t|\theta = \theta_1), \quad f_2(\underline{x}) = \left.\frac{\delta}{\delta\theta} f_T(t|\theta)\right|_{\theta=\theta_0}, \quad f_1(\underline{x}) = f_T(t|\theta = \theta_0)$$

Então o teste $\phi^\star$, com

$$\phi^\star(\underline{x}) = \begin{cases} 1 & f_T(t|\theta = \theta_1) > k_1 f_T(t|\theta = \theta_0) + k_2 \left.\frac{\delta}{\delta\theta} f_T(t|\theta)\right|_{\theta=\theta_0} \\ 0 & c.c. \end{cases}$$

para algumas constantes k_1, k_2 é um teste UMPU.

Na verdade a região crítica $W_\delta = \{t \in \Omega_T : t \leq a \text{ ou } t \geq b\}$, onde a e b são tais que

$$P(T \leq a \vee T \geq b|\theta = \theta_0 = \alpha$$
$$E\left[T.I_{W_\delta}(T)|\theta = \theta_0\right] = \alpha E\left[T|\theta = \theta_0\right]$$

onde $I_{W_\delta}(T)$ designa a função indicatriz da região crítica no domínio de T é UMPU.

Exemplo 19 *Seja $\underline{X}$ uma amostra proveniente de uma população $N(0, \sigma^2)$, para a qual se pretende testar $H_0 : \sigma = \sigma_0$ vs $H_1 : \sigma \neq \sigma_0$, usando para tal uma região crítica da forma $W = \{t : t \leq a \vee t \geq b\}$.*

Para esta família uma estatística suficiente é $T(\underline{X}) = \sum_{i=1}^n X_i^2$, sendo que a sua distribuição quando H_0 é verdadeira dada por $\frac{T}{\sigma_0^2} \sim \chi^2_n$. Então os valores a e b são *calculados através de*

$$\begin{aligned}
P(\sum_{i=1}^n X_i^2 \leq a \vee \sum_{i=1}^n X_i^2 \geq b|\sigma=\sigma_0) &= P(\frac{\sum_{i=1}^n X_i^2}{\sigma_0^2} \leq \frac{a}{\sigma_0^2} \vee \frac{\sum_{i=1}^n X_i^2}{\sigma_0^2} \geq \frac{b}{\sigma_0^2}|\sigma=\sigma_0) \\
&= 1 - \int_{a/\sigma_0^2}^{b/\sigma_0^2} f_{\chi_n^2}(s)ds \\
&= \alpha \\
E\left[T.I_{W_\delta}(T)|\sigma=\sigma_0\right] &= \sigma_0^2 E\left[\frac{\sum_{i=1}^n X_i^2}{\sigma_0^2}.I_{W_\delta}(T)|\sigma=\sigma_0\right] \\
&= \sigma_0^2 E\left[\frac{\sum_{i=1}^n X_i^2}{\sigma_0^2}|\sigma=\sigma_0\right] - \sigma_2^2 \int_{a/\sigma_0^2}^{b/\sigma_0^2} t f_{\chi_n^2}(s)ds \\
&= n\sigma_0^2 - \sigma_0^2 \int_{a/\sigma_0^2}^{b/\sigma_0^2} t f_{\chi_n^2}(s)ds
\end{aligned}$$

Decorre de propriedades da distribuição do qui-quadrado que

$$s f_{\chi_n^2}(s) = n f_{\chi_{n+2}^2}(s)$$

pelo que

$$\int_{a/\sigma_0^2}^{b/\sigma_0^2} t f_{\chi_n^2}(s)ds = \int_{a/\sigma_0^2}^{b/\sigma_0^2} f_{\chi_{n+2}^2}(s)ds = 1-\alpha$$

Logo os valores a e b são obtidos à custa da solução das equações

$$\begin{aligned}
\int_{a/\sigma_0^2}^{b/\sigma_0^2} f_{\chi_n^2}(s)ds &= 1-\alpha \\
E\left[T.I_{W_\delta}(T)|\sigma=\sigma_0\right] &= (n-1-\alpha)\sigma_0^2
\end{aligned}$$

ou, equivalentemente,

$$\begin{aligned}
\int_{a/\sigma_0^2}^{b/\sigma_0^2} f_{\chi_n^2}(s)ds &= 1-\alpha \\
\int_{a/\sigma_0^2}^{b/\sigma_0^2} f_{\chi_{n+2}^2}(s)ds &= n(1-\alpha)
\end{aligned}$$

Exemplo 20 *Consideramos a mesma situação que no exemplo anterior, sendo porém que agora as hipóteses são $H_0 : \sigma_1^2 \leq \sigma^2 \leq \sigma_2^2$ vs $H_1 : \sigma < \sigma_1^2$ ou $\sigma > \sigma_2^2$, para o qual se considera uma região crítica da forma $\{W : t \leq a \vee t \geq b\}$. Então as constantes a e b correspondentes aos testes UMPU são determinadas a partir das equaçõea:*

$$\begin{aligned}
\int_{a/\sigma_1^2}^{b/\sigma_1^2} f_{\chi_n^2}(s)ds &= 1-\alpha \\
\int_{a/\sigma_2^2}^{b/\sigma_2^2} f_{\chi_n^2}(s)ds &= 1-\alpha
\end{aligned}$$

4.3 Métodos de construção de testes

Nesta secção discutimos métodos de construção de testes, os quais permitem lidar com diferentes situações.

4.3.1 Teste da razão de verosimilhanças

Trata-se do método mais geral, sendo quase sempre aplicável. Está relacionado com os estimadores de máxima verosimilhança, tema explorado no capítulo anterior. Tem ainda a vantagem de resultar em testes a que correspondem regiões críticas intuitivas, embora não conduza necessariamente a testes UMP ou UMPU. Podem ainda estar definidos no caso multi-paramétrico.

Seja X uma população com f.d.p. $f_X(.|\theta)$, onde θ é um parâmetro desconhecido, e $\underline{x} = (x_1, \ldots, x_n)$ uma amostra proveniente de X. Recordemos que a função de verosimilhança, definida para todo o espaço paramétrico Θ, é dada por

$$L(\theta|\underline{x}) = f(\underline{x}|\theta) = \prod_{i=1}^{n} f_X(x_i|\theta)$$

Definição 21 *A razão de verosimilhanças para testar $H_0 : \theta \in \Theta_0$ vs $H_1 : \theta \in \Theta_0^c$ baseada numa amostra $\underline{x}$ é dada por*

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

Um teste da razão de verosimilhanças é um teste cuja região de rejeição W é dada por $\{\underline{x} : \lambda(\underline{x}) \leq c\}$, onde $c \in [0, 1]$.

Note-se que por construção $\lambda(\underline{x}) \in (0, 1]$, uma vez que $0 < \sup_{\theta \in \Theta_0} L(\theta|\underline{x}) \leq \sup_{\theta \in \Theta} L(\theta|\underline{x})$. Quanto maior for $\lambda(\underline{x})$ mais plausível é a hipótese H_0; pelo contrário se está próximo de zero é depositada pouca confiança na veracidade da hipótese nula.

Se $T(\underline{X})$ for uma estatística suficiente para θ com f.d.p. $g_T(.|\theta)$, então pode ser aconselhável construir o teste da razão de verosimilhanças com base em T e na sua função de verosimilhança $L^\star(\theta|t) = g_T(t|\theta)$. A questão pertinente é saber se os testes construidos com base na razão $\lambda(\underline{x})$ e com base em $\lambda^\star = \frac{\sup_{\theta \in \Theta_0} L^\star(\theta|T(\underline{x})=t)}{\sup_{\theta \in \Theta} L^\star(\theta|T(\underline{x})=t)}$ se são equivalentes ou não. De facto

Teorema 22 *Se $T(\underline{X})$ é uma estatística suficiente para θ e se $\lambda^\star(T(\underline{x}) = t)$ e $\lambda(\underline{x})$ são as razões de verosimilhanças baseadas em T e em $\underline{X}$, respectivamente, então $\lambda^\star(T(\underline{x}) = \lambda(\underline{x})$ para todo o $\underline{x} \in \Omega_{\underline{X}}$, pelo que o teste baseado na razão de verosimilhanças de T é idêntico ao teste baseado na razão de verosimilhanças de $\underline{X}$.*

Demonstração: Decorre do teorema de factorização que $f(\underline{x}|\theta) = g_T(T(\underline{x})|\theta)h(\underline{x})$. Logo

$$\begin{aligned}
\lambda(\underline{x}) &= \frac{\sup_{\theta\in\Theta_0} L(\theta|\underline{x})}{\sup_{\theta\in\Theta} L(\theta|\underline{x})} \\
&= \frac{\sup_{\theta\in\Theta_0} f(\underline{x}|\theta)}{\sup_{\theta\in\Theta} f(\underline{x}|\theta)} \\
&= \frac{\sup_{\theta\in\Theta_0} g_T(T(\underline{x})|\theta)h(\underline{x})}{\sup_{\theta\in\Theta} g_T(T(\underline{x})|\theta)h(\underline{x})} \\
&= \frac{\sup_{\theta\in\Theta_0} g_T(T(\underline{x})|\theta)}{\sup_{\theta\in\Theta} g_T(T(\underline{x})|\theta)} \\
&= \frac{\sup_{\theta\in\Theta_0} L^\star(\theta|T(\underline{x}))}{\sup_{\theta\in\Theta} L^\star(\theta|T(\underline{x}))} \\
&= \lambda^\star(T(\underline{x})) \quad \blacksquare
\end{aligned}$$

Exemplo 23 *Seja $X \sim N(\theta, \sigma^2)$, e $X_1, \ldots, X_n$ uma amostra aleatória proveniente de X. Suponhamos que $H_0 : \theta \leq \theta_0$ vs $H_1 : \theta > \theta_0$, sendo pois σ (desconhecido) uma parâmetro perturbador. A função de verosimilhança é dada por*

$$L(\theta, \sigma|\underline{x}) = \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n e^{-\frac{\sum_{i=1}^n (x_i-\theta)^2}{2\sigma^2}}$$

Então

$$\begin{aligned}
\lambda(\underline{x}) &= \frac{\sup_{\{\theta,\sigma^2:\theta<\theta_0,\sigma^2>0\}} L(\theta, \sigma|\underline{x})}{\sup_{\{\theta,\sigma^2:\theta\in I\!R,\sigma^2>0\}} L(\theta, \sigma|\underline{x})} \\
&= \frac{\sup_{\{\theta,\sigma^2:\theta<\theta_0,\sigma^2>0\}} L(\theta, \sigma|\underline{x})}{L(\hat{\theta}, \hat{\sigma}|\underline{x})}
\end{aligned}$$

onde $\hat{\theta}$ e $\hat{\sigma}$ são os estimadores de máxima verosimilhança de θ e σ, respectivamente. Se $\hat{\theta} \leq \theta_0$ então o máximo restrito é igual ao máximo irrestrito; caso contrário o máximo restrito é igual a $L(\theta_0, \tilde{\sigma}|\underline{x})$, com $\tilde{\sigma}^2 = \frac{\sum_{i=1}^n (x_i-\theta_0)^2}{n}$. Note-se que

$$\begin{aligned}
L(\theta_0, \tilde{\sigma}|\underline{x}) &= \left(\frac{1}{\sqrt{2\pi\tilde{\sigma}^2}}\right)^n e^{-\frac{\sum_{i=1}^n (x_i-\theta_0)^2}{2\tilde{\sigma}^2}} \\
&= \left(\frac{n}{2\pi\sum_{i=1}^n (x_i-\theta_0)^2}\right)^{\frac{n}{2}} e^{-\frac{n}{2}} \\
L(\hat{\theta}, \hat{\sigma}|\underline{x}) &= \left(\frac{1}{\sqrt{2\pi\hat{\sigma}^2}}\right)^n e^{-\frac{\sum_{i=1}^n (x_i-\hat{\theta})^2}{2\hat{\sigma}^2}} \\
&= \left(\frac{n}{2\pi\sum_{i=1}^n (x_i-\bar{x})^2}\right)^{\frac{n}{2}} e^{-\frac{n}{2}}
\end{aligned}$$

Logo

$$\lambda(\underline{x}) = \begin{cases} 1 & \hat{\theta} \leq \theta_0 \\ \frac{L(\theta_0,\tilde{\sigma}|\underline{x})}{L(\hat{\theta},\hat{\sigma}|\underline{x})} & c.c. \end{cases} = \begin{cases} 1 & \hat{\theta} \leq \theta_0 \\ \left(\frac{\sum_{i=1}^n (x_i-\bar{x})^2}{\sum_{i=1}^n (x_i-\theta_0)^2}\right)^{n/2} & c.c. \end{cases}$$

Dado que

$$\sum_{i=1}^{n}(x_i-\theta_0)^2=\sum_{i=1}^{n}(x_i-\bar{x})^2+n(\bar{x}-\theta_0)^2$$

vem que no caso de $\hat{\theta}>\theta_0$,

$$\begin{aligned}\lambda(\underline{X}) &= \left(\frac{1}{1+n\frac{(\bar{X}-\theta_0)^2}{\sum_{i=1}^{n}(X_i-\bar{X})^2}}\right)^{n/2}\\ &= \left(\frac{1}{1+\frac{t_{n-1}^2}{n-1}}\right)^{n/2}\end{aligned}$$

onde t_{n-1} *designa uma v.a. com distribuição* T*-student com* $n-1$ *graus de liberdade. Concluímos assim que a razão de verosimilhanças é, no caso de* $\hat{\theta}>\theta_0$*, uma função monótona de* t_{n-1} *pelo que a região crítica vem determinada em função de quantis de probabilidade de uma* T*-Student, como estamos habituados a fazer.*

Existem outros métodos de construção de testes, que serão oportunamente estudados, no âmbito de trabalhos fornecidos aos alunos.