S-value e P-value

Alexandre Galvão Patriota

Departamento de Estatística
Instituto de Matemática e Estatística
Universidade de São Paulo

Roteiro

Teste de Hipóteses

- Testar uma hipótese científica é uma das práticas mais importantes no meio científico.
- Usualmente, definem-se duas hipóteses científicas, a saber: H e sua negação $\neg H$.
- As hipóteses científicas devem ser traduzidas em termos probabilísticos/estatísticos.

Por exemplo: testar se o tempo médio de vida de um tipo de lâmpada é igual a M.

H :*"o tempo médio de vida é igual a M"*
$\neg H$:*"o tempo médio de vida* **não** *é igual a M".*

Qual o experimento a ser feito e qual modelo usar? qual seria a tradução para modelos estatísticos das hipóteses acima para H_0 e H_1?

Procedimentos estatísticos

Na abordagem clássica, há dois procedimentos para testar tais hipóteses, a saber:

- **Procedimento de Neyman-Pearson**[1],

  A hipótese H (H_0) pode ser tanto aceita como rejeitada (testes mais poderosos para um nível de significância fixado), pois o universo de possibilidades é fechado e bem definido.

- **Procedimento Fisheriano**[2].

  A hipótese H (H_0) só pode ser rejeitada, pois o universo de possibilidades é aberto e nem sempre bem definido.

---

[1] Neyman, J., Pearson, E.S. (1933). On the Problem of the Most Efficient Tests of Statistical Hypotheses, *Phil. Trans. R. Soc. Lond. A*, **231**, 289–337.

[2] Fisher, R.A. (1935). The logic of inductive inference, *Journal of the Royal Statistical Society*, **98**, 39–54.

Restritividades das hipóteses de interesses

Em termos das hipóteses, seguem dois exemplos de restritividade:

- se a restritividade da hipótese de interesse é *similar* ao da sua negação hipótese, então faz sentido aceitar ou rejeitar a hipótese de interesse

  H : *"as duas faces da moeda são iguais"*, $\neg H$:*"as duas faces da moeda são diferentes"*.

- se a restritividade da hipótese de interesse é *significativamente maior* que a da sua negação, então a hipótese de interesse nunca deveria ser aceita

  Popper: H : *"todos os cisnes são brancos"*, $\neg H$:*"nem todos os cisnes são brancos"*.

Procedimento de Neyman-Pearson

- Definem-se matematicamente as hipóteses H_0 e H_1 mutuamente **exclusivas** e **exaustivas**.
- Constrói-se uma região de rejeição fixando a probabilidade do Erro Tipo I. A partir de região de rejeição será tomada uma decisão sobre H_0 e H_1
- Se a estatística observada pertencer a região de rejeição, **rejeita-se** a hipótese H_0 para o nível de significância pré-fixado.
- Se a estatística observada não pertencer a região de rejeição, **aceita-se** a hipótese H_0 para o nível de significância pré-fixado. Neste último caso, por influência dos argumentos de Fisher, diz-se atualmente que

  > *"não foi possível encontrar evidências para rejeitar a hipótese nula".*

Procedimento Fisheriano

- Define-se apenas H_0 matematicamente. Fisher considera que a negação de H_0 é **indefinível**.

- Calcula-se o p-valor (probabilidade de algo extremo ocorrer sob as restrições de H_0).

- Esta quantidade é utilizada como termômetro para rejeitar H_0: **quanto menor** for o seu valor mais evidências de que H_0 é falsa.

- Usualmente rejeita-se a hipótese H_0 se o p-valor for menor do que um ponto de corte, α (e.g., $\alpha = 0.05$), caso seja maior diz-se que não há evidências para rejeitar H_0.

A definição para o p-valor será dada adiante.

Procedimento alternativo

Nesta palestra estudaremos uma terceira alternativa em que é possível tomar três tipos de decisões quando o universo é fechado ($H_1 \equiv \neg H_0$):

1. aceitar H_0 (rejeitar H_1),
2. rejeitar H_0 (aceitar H_1),
3. não rejeitar nem aceitar qualquer uma das hipóteses. A decisão fica suspensa.

- Se o universo for aberto, pode-se simplesmente: rejeitar H_0 ou não rejeitar H_0, pois neste caso a negação de H_0 será sempre muito "grande".

Observe que quando os dados trazem informação ambígua sobre as hipóteses não deveríamos nem aceitar, nem rejeitar H_0, neste caso o mais adequado seria coletar mais dados ou outras informações.

Cálculo do P-valor

Há várias metodologias para o cálculo do p-valor, mas essencialmente tem-se o seguinte:

- fixa-se o experimento que fornecerá os dados experimentais,
- propõe-se uma estatística T (positiva), função dos dados experimentais, tal que:
- quanto maior o valor observado de T, maior deve ser a discrepância de H_0 para explicar os dados observados.

Exemplo: testar se a média populacional é igual a zero, então $T = \bar{X}^2$ é uma estatística que satisfaz as condições acima. Em geral $-2\log(RV)$ satisfaz essas condições, em que RV é a estatística da razão de verossimilhanças.

Definição informal do P-valor

Supondo que H_0 é de fato compatível com os dados observados, o p-valor é a probabilidade de ser observado em outro experimento uma estatística T, no mínimo, tão extrema quanto a que foi observada no experimento atual:

$$p \equiv p(T, \Theta_0) = P(T \geq t; \text{ sob } H_0),$$

em que t é o valor observado da estatística T e "sob H_0" significa que a medida de probabilidade utilizada é aquela que foi especificada na hipótese nula.

Vale ressaltar que essa é uma definição informal que mantém certa simplicidade, porém, como todas as definições informais, sua interpretação pode gerar muitas controvérsias.

Definição formal do P-valor

Seja o modelo estatístico induzido pela variável aleatória X:

$$(\mathcal{X}, \mathcal{F}, \mathcal{P})$$

em que:
$\mathcal{X} \subseteq \mathbb{R}^n$ é o espaço amostral,
$\mathcal{F}$ é uma σ-álgebra de borel de $\mathcal{X}$,
$\mathcal{P}$ é uma família de medidas de probabilidade, ou seja, para cada $P \in \mathcal{P}$, temos que:

> $P(\varnothing) = 0$, $P(\mathcal{X}) = 1$ *e se* $\{A_n\}_{n\geq 1} \subset \mathcal{F}$ *são disjuntos, então* $P(\cup A_n) = \sum_n P(A_n)$.

Um estatística T é uma função que leva os valores de $\mathcal{X}$ em outro espaço, portanto temos também um modelo estatístico induzido por T.

Definição formal do P-valor

O problema de testar hipóteses deve ser escrito em termos de subconjuntos da família $\mathcal{P}$ e da estatística T. Ou seja, deve-se encontrar a família de medidas de probabilidades tal que $H_0 \equiv$ "$P \in \mathcal{P}_{H_0}$"

Por exemplo: suponha que nosso modelo estatístico contém as medidas $\mathcal{P} = \{P_0, P_1, P_2, P_3\}$ e estamos interessados em testar se a medida que gera os dados é P_0, então $\mathcal{P}_{H_0} = \{P_0\}$.

Neste caso, o p-valor é então definido por

$$p = P_0(T \geq t),$$

ou seja, sob H_0 tem-se que a medida que gera os dados é P_0 e portanto deve-se utilizá-la no cálculo do p-valor.

Definição formal do P-valor

Suponha que o interesse está em verificar se a medida que gera os dados é P_0 ou P_1, então escreve-se $H_0 : P \in \mathcal{P}_{H_0}$, sendo $\mathcal{P}_{H_0} = \{P_0, P_1\}$, neste caso há duas medidas para definir o p-valor, deve-se escolher então a que gera a maior probabilidade:

$$p = \max\{P_0(T \geq t), P_1(T \geq t)\}.$$

Ou seja, se em pelo menos uma medida o evento $\{T \geq t\}$ tem probabilidade alta de ocorrer não deveríamos rejeitar a hipótese nula.

Por outro lado, se para ambas medidas o evento $\{T \geq t\}$ tem probabilidade baixa de ocorrer, então a hipótese nula pode ser considerada implausível.

Definição formal do P-valor

Usualmente, quando a família de medidas é paramétrica, i.e., existe um espaço finito dimensional $\Theta \subseteq \mathbb{R}^k$ tal que a família de probabilidade seja definda por $\mathcal{P} = \{P_\theta;\ \theta \in \Theta\}$.

Pode-se representar a hipótese nula utilizando a seguinte notação paramétrica $H_0 : \theta \in \Theta_0$, sendo $\mathcal{P}_{H_0} = \{P_\theta;\ \theta \in \Theta_0\}$.

Neste caso o p-valor é então definido seguindo a mesma regra anterior

$$p = \sup_{\theta \in \Theta_0} P_\theta(T \geq t),$$

ou seja, se p for pequeno então todas as medidas de probabilidades que estão definidas em H_0 geram probabilidades menores.

"Inconsistências" do P-valor

O p-valor nos dá uma medida de evidência contra uma hipótese específica: quanto menor o p-valor, mais evidências de que a hipótese nula não se adéqua aos dados observados.

Observe que a estatística $T = T_{H_0}$ (depende da hipótese nula).

Portanto, um cuidado especial deve-se ter ao comparar p-valores de hipóteses diferentes, pois duas hipóteses diferentes induzem duas medidas diferentes no cálculo do p-valor.

Esta característica gera conflitos de interpretação de diferentes graus de gravidade como veremos a seguir.

"Inconsistências" do P-valor

Para ilustrar dois dos conflitos, considere duas populações representadas pelas variáveis X e Y, com esperanças $E(X) = \mu_1$ e $E(Y) = \mu_2$.

Considere as seguintes hipóteses nulas

$$H_{01} : \mu_1 = \mu_2, \quad H_{02} : \mu_1 = \mu_2 = b, \quad H_{03} : \mu_1 = b, \quad H_{04} : \mu_2 = b$$

sendo b uma constante real qualquer. Considere as seguintes conclusões inconsistentes:

1. Há evidências para rejeitar H_{02}, mas não há evidências para rejeitar as hipóteses H_{03} e H_{04} separadamente.
2. Há evidências para rejeitar H_{01}, mas não há evidências para rejeitar H_{02}.

"Inconsistências" do P-valor: 1º caso

"Há evidências para rejeitar $H_{02} : \mu_1 = \mu_2 = b$, mas não há evidências para rejeitar as hipóteses $H_{03} : \mu_1 = b$ e $H_{04} : \mu_2 = b$ separadamente."

Note que teoricamente $H_{02} \equiv H_{03} \wedge H_{04}$, ou seja, se H_{02} é rejeitada espera-se que pelo menos uma das hipóteses H_{03} ou H_{04} seja rejeitada.

Esta inconsistência é particularmente simples de explicar:

- note que na hipótese H_{02} deve-se considerar os dois conjuntos de dados na estatística do teste.
- enquanto que para cada hipótese isolada consideramos apenas um dos conjuntos de dados e excluímos o outro.

Dessa forma, os testes marginais (H_{03} e H_{04}) não consideram uma possível dependência entre X e Y, quanto maior for a dependência maior será a diferenças de conclusões.

Exemplo para o 1º caso

Considere

$$\begin{pmatrix} X \\ Y \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} 1 & 6 \\ 6 & 37 \end{pmatrix} \right).$$

Assuma que uma amostra de tamanho $n = 100$ foi retirada e as médias amostrais são $\bar{x} = 0{,}06$ e $\bar{y} = 0{,}62$.

Para testar $H_{02} : \mu_1 = \mu_2 = 0$ tem-se que o logaritmo da estatística da razão de verossimilhanças é

$$T_2 = n \begin{pmatrix} \bar{X} \\ \bar{Y} \end{pmatrix}^\top \Sigma^{-1} \begin{pmatrix} \bar{X} \\ \bar{Y} \end{pmatrix} \sim \chi_2^2, \quad \text{sob } H_{02}.$$

Utilizando os valores observados ($t_2 = 7{,}12$), o p-valor é dado por $p_2 = P(\chi_2^2 > t_2) = 0{,}03$ (evidência para rejeitar H_{02}).

Exemplo para o 1º caso

Considere

$$\begin{pmatrix} X \\ Y \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} 1 & 6 \\ 6 & 37 \end{pmatrix}\right).$$

Assuma que uma amostra de tamanho $n = 100$ foi retirada e as médias amostrais são $\bar{x} = 0{,}06$ e $\bar{y} = 0{,}62$.

Para testar $H_{03} : \mu_1 = 0$ e $H_{04} : \mu_2 = 0$, tem-se que os logaritmos das estatísticas da razão de verossimilhanças são

$$T_3 = n\bar{X}^2 \overset{H_{03}}{\sim} \chi_1^2, \quad T_4 = n\frac{\bar{Y}^2}{37} \overset{H_{04}}{\sim} \chi_1^2,$$

Utilizando os valores observados ($t_3 = 0{,}36$ e $t_4 = 1{,}04$), os respectivos p-valores são $p_3 = P(\chi_1^2 > t_2) = 0{,}55$ e $p_4 = P(\chi_1^2 > t_4) = 0{,}31$ (não há evidências para rejeitar H_{03} e H_{04} separadamente).

"Inconsistências" do P-valor: 2º caso

"Há evidências para rejeitar $H_{01} : \mu_1 = \mu_2$, mas não há evidências para rejeitar $H_{02} : \mu_1 = \mu_2 = b$."

Note que teoricamente $H_{02} \rightarrow H_{01}$, ou seja, se H_{02} for verdadeira então H_{01} também deveria ser verdadeira.

Teoricamente rejeitar H_{01} implica em rejeitar H_{02}, pois se rejeitamos que $\mu_1 = \mu_2$ em particular devemos rejeitar que $\mu_1 = \mu_2 = b$.

Note que o problema agora é diferente, ambas hipóteses são conjuntas e as estatísticas são construídas levando em conta toda a informação dos dados.

Essas conclusões inconsistentes podem ocorrer na prática?

Exemplo para o 2º caso

Considere $X \sim N(\mu_1, 1)$ e $Y \sim N(\mu_2, 1)$ independentes para simplificar o problema. Assuma que uma amostra de tamanho $n = 100$ foi retirada e as médias amostrais são $\bar{x} = 2{,}14$ e $\bar{y} = 1{,}84$,

Para testar

$$H_{01} : \mu_1 = \mu_2,$$

tem-se que o logaritmo da estatística da razão de verossimilhanças é

$$T_1 = \frac{n}{2}(\bar{X} - \bar{Y})^2 \sim \chi^2_1, \quad \text{sob } H_{01},$$

Utilizando os valores observados ($t_1 = 4{,}50$), o p-valor é dado por $p_1 = P(\chi^2_1 > t_1) = 0{,}03$.

Há evidências para rejeitar H_{01}, a 5% de significância estatística.

Exemplo para o 2º caso

Considere $X \sim N(\mu_1, 1)$ e $Y \sim N(\mu_2, 1)$ independentes para simplificar o problema. Assuma que uma amostra de tamanho $n = 100$ foi retirada e as médias amostrais são $\bar{x} = 2{,}14$ e $\bar{y} = 1{,}84$,

Para testar

$$H_{02} : \mu_1 = \mu_2 = 2,$$

tem-se que o logaritmo da estatística da razão de verossimilhanças é

$$T_2 = n[(\bar{X} - 2)^2 + (\bar{Y} - 2)^2] \sim \chi^2_2, \quad \text{sob } H_{02},$$

Utilizando os valores observados ($t_2 = 4{,}52$), o p-valor é dado por $p_2 = P(\chi^2_2 > t_2) = 0{,}10.$

Não há evidências para rejeitar H_{02}, a 5% de significância estatística.

"Inconsistências" do P-valor

Em resumo: deve-se ter cuidado ao utilizar p-valores em hipóteses aninhadas.

Se houver evidências para rejeitar uma hipótese H_0, não se pode afirmar que há evidências para rejeitar nenhuma outra hipótese H_0' que implica H_0.

Mais detalhes podem ser estudados em:

- Schervish, M.J.(1996). P Values: What they are and what they are not, Am. Stat. 50, 203–206.
- Izbicki, R., et al. (2012). Testing allele homogeneity: the problem of nested hypotheses, BMC Genet. 13, 103.
- Patriota, A.G. (2013). A classical measure of evidence for general null hypotheses, FSS, 233, 74–88.
- `http://eranraviv.com/blog/on-p-value/`
- `http://analisereal.com/2013/11/07/voce-e-obeso-mas-nao-e-gordo-2-ou-mais-sobre-p-valores/`

Alerta

Note que:

Os resultados anteriores não descrevem uma contradição do método clássico em si.

Mostram apenas algumas limitações do uso do p-valor como medida de evidência.

O p-valor não deve ser interpretado como uma medida de suporte sobre o espaço paramétrico, ele é uma medida de evidência para cada Hipótese fixada.

O p-valor não é uma probabilidade condicional!

Uma medida de evidência alternativa

Para evitar as "Inconsistências" relatadas, precisamos definir uma medida que utilize a mesma métrica para qualquer subconjunto de Θ. Faremos isso usando regiões de confiança:

A região de confiança baseada na razão de verossimilhanças com nível de significância α é definida por

$$\Lambda_\alpha = \{\theta \in \Theta : T_\theta(X) \leq F_{\theta,\alpha}\},$$

em que:
$T_\theta(X) = 2(\ell_{\widehat{\theta}}(X) - \ell_\theta(X))$, $\widehat{\theta} = \arg\max_\theta \ell_\theta(X)$,
$\ell_\theta(X) = \log(L_\theta(X))$
e $F_{\theta,\alpha}$ é tal que $F_\theta(F_{\theta,\alpha}) = 1 - \alpha$ com F_θ sendo (um)a função (aproximada) de distribuição de $T_\theta(X)$ (podendo ser a função de distribuição acumulada da qui-quadrado com $k =$ dim(Θ) graus de liberdade).

Uma medida de evidência alternativa

A medida de evidência para $H_0 : \theta \in \Theta_0$ é o valor

$$s = s(\Theta_0) \equiv s(X, \Theta_0) = \max\{0, \sup\{\alpha \in (0,1) : \Lambda_\alpha \cap \Theta_0 \neq \varnothing\}\}.$$

Interpretação: s é o maior nível de significância possível para que a região de confiança e o conjunto $\overline{\Theta_0}$ tenham ao menos um ponto em comum.

Valores altos de s indicam que existe pelo menos um ponto em Θ_0 que está perto do EMV.

Valores baixos de s indicam que todos os pontos em Θ_0 estão longe do EMV.

Ilustração no Plano

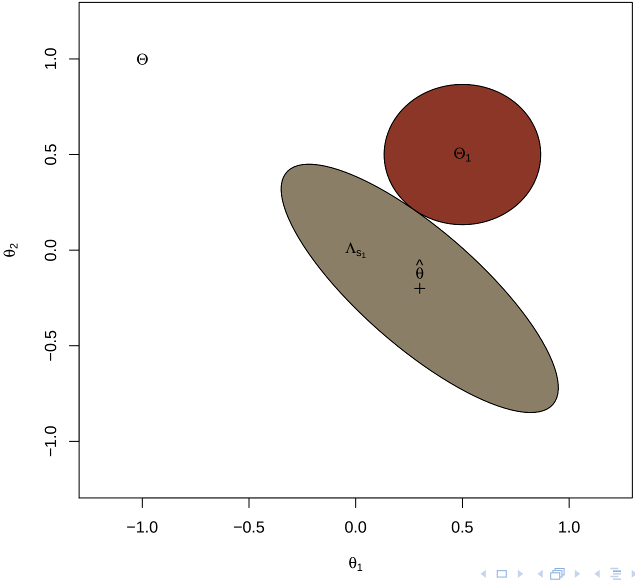
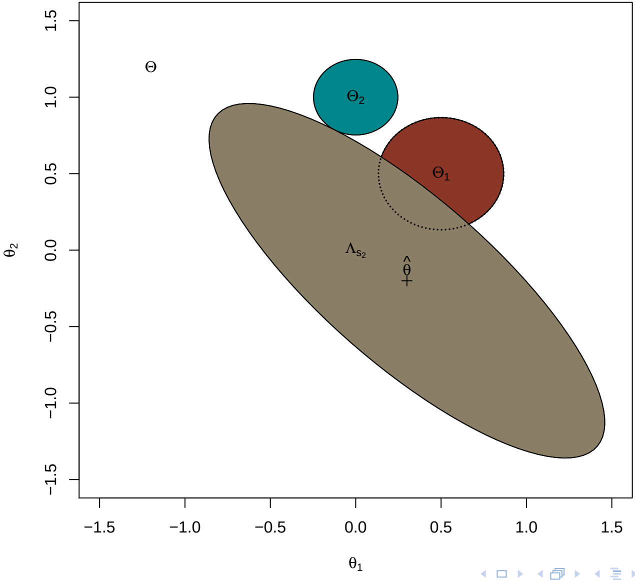


Ilustração no Plano



Propriedades

Algumas propriedades importantes do s-valor são listadas abaixo:

1. $s(\varnothing) = 0$ e $s(\Theta) = 1$
2. Se $\Theta_1 \subseteq \Theta_2$, então $s(\Theta_1) \leq s(\Theta_2)$
3. Para quaisquer $\Theta_1, \Theta_2 \subseteq \Theta$, tem-se $s(\Theta_1 \cup \Theta_2) = \max\{s(\Theta_1), s(\Theta_2)\}$
4. Se $\Theta_1 \subseteq \Theta$, então $s(\Theta_1) = \sup_{\theta \in \Theta_1} s(\{\theta\})$
5. $s(\Theta_1) = 1$ ou $s(\Theta_1^c) = 1$. Se $\widehat{\theta} \in \overline{\Theta_1}$ (fecho de Θ_1), então $s(\Theta_1) = 1$, se $\widehat{\theta} \in \overline{\Theta_1^c}$, então $s(\Theta_1^c) = 1$.

As propriedades acima mostram que s é uma medida possibilística definida nos subconjuntos do espaço paramétrico Θ.

Decisões sobre H_0

Com respeito a decisão sobre aceitação ou rejeição da hipótese nula, considere que o universo de possibilidades é fechado: $H_0 : \theta \in \Theta_0$ contra $H_1 : \theta \in \Theta - \Theta_0$ com $\Theta_0 \subset \Theta$.

Definimos a função Φ, o "grau de crença" em subconjuntos de Θ, da seguinte forma:

$$\Phi(\Theta_0) = \langle s(\Theta_0), s(\Theta_0^c) \rangle$$

em que:

- $\Phi(\Theta_0) = \langle 1, 0 \rangle$ é o valor maximal (aceitamos imediatamente H_0)
- $\Phi(\Theta_0) = \langle 0, 1 \rangle$ é o valor minimal (rejeitamos imediatamente H_0).
- $\Phi(\Theta_0) = \langle 1, 1 \rangle$ é o valor de ignorância total com respeito a aceitação/rejeição de H_0
- $\Phi(\Theta_0) = \langle 1, b \rangle$, com $b \in (0, 1]$, se $\widehat{\theta} \in \Theta_0$. Neste caso, se b for suficientemente pequeno podemos aceitar H_0.
- $\Phi(\Theta_0) = \langle a, 1 \rangle$, com $a \in (0, 1]$, se $\widehat{\theta} \in \Theta_0^c$. Neste caso, se a for suficientemente pequeno podemos rejeitar H_0.

Comentários sobre as decisões

Considere o modelo normal e um universo de possibilidades fechado. Suponha por exemplo que observamos médias amostrais ($n = 1000$) $\bar{x} = 1000$ e $\bar{y} = 7000$ com variâncias $s_x^2 = 0{,}2$ e $s_y^2 = 1{,}3$.

- se queremos testar se μ_1 e μ_2 são ambas positivas, então faz pleno sentido aceitar a hipótese de que as médias populacionais são positivas.
- se queremos testar se μ_1 e μ_2 são ambas negativas, então faz pleno sentido rejeitar a hipótese de que as médias populacionais são negativas.

Por outro lado, se as médias amostrais estiverem perto de pelo menos um dos eixos do plano cartesiano, então não poderemos concluir com tanta certeza sobre aceitação e rejeição de que ambas as médias são positivas/negativas.

Observações sobre as decisões

- **Observação 1:** se a dimensão de Θ_0 for menor do que a dimensão de Θ, então teremos sempre o caso $\Phi(\Theta_0) = \langle a, 1 \rangle$ com $a \in (0, 1]$.
- **Observação 2:** se a dimensão de Θ_0^c for menor do que a dimensão de Θ, então teremos sempre o caso $\Phi(\Theta_0) = \langle 1, b \rangle$ com $b \in (0, 1]$.
- **Observação 3:** se a dimensão de Θ_0 for igual a dimensão de Θ_0^c, então o caso dependerá se $\widehat{\theta} \in \Theta_0$ ou $\widehat{\theta} \in \Theta_0^c$.

Como encontrar valores de corte para decidir?

Problema em aberto!

Pode-se tentar *via* funções de perda.

Pode-se usar um critério frequentista, porém o ponto de corte dependerá da hipótese nula:

Propriedade: Assuma que função de log-verossimilhança estritamente côncava, $F_\theta \equiv F$ e H_0 é uma hipótese simples. Então:

$$s = 1 - F(F_{H_0}^{-1}(1-p)),$$

sendo $p = 1 - F_{H_0}(t)$.

Assim, se α e o ponto de corte para o p-valor, $\alpha_0 = 1 - F(F_{H_0}^{-1}(1-\alpha))$ será o ponto de corte frequentista para o s-valor.

Aplicação da medida de evidência

Considere uma amostra $X = (X_1, \dots, X_n)$ em que $X_1 \sim N_2(\mu, \Sigma)$ com $\mu = (\mu_1, \mu_2)^\top$ e Σ uma matriz de variâncias-covariâncias **conhecida**.

O espaço paramétrico é $\Theta = \{(\mu_1, \mu_2) : \ \mu_1, \mu_2 \in \mathbb{R}\} = \mathbb{R}^2$.

Pode-se mostrar que

$$2(\ell_{\widehat{\theta}}(X) - \ell_\theta(X)) = n(\bar{X} - \mu)^\top \Sigma^{-1} (\bar{X} - \mu) \overset{P_\theta}{\sim} \chi^2_2.$$

Portanto, as regiões de confiança são

$$\Lambda_\alpha = \{(\mu_1, \mu_2) \in \mathbb{R}^2 : \ n(\bar{X} - \mu)^\top \Sigma^{-1} (\bar{X} - \mu) \leq F_\alpha\}$$

em que F_α é o $(1 - \alpha)$-quantil da χ^2_2.

Hipóteses nulas

Primeira hipótese nula:
$H_0' : \theta \in \Theta_0'$, em que $\Theta_0' = \{(0,0)\}$

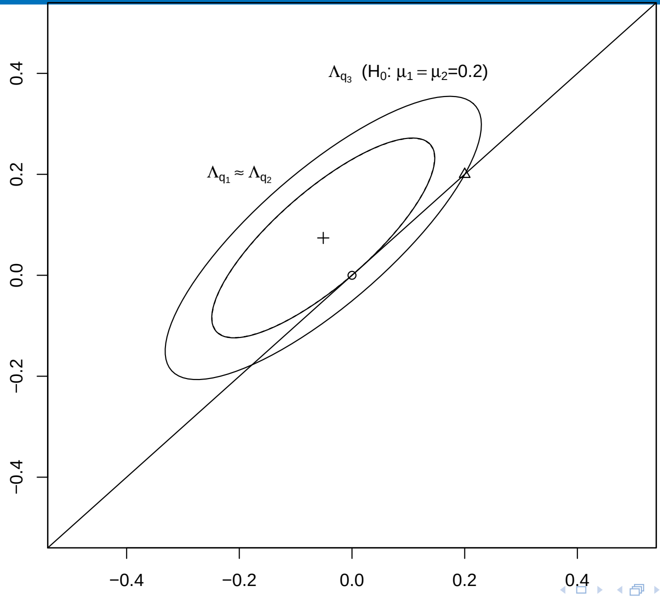
Segunda hipótese nula:
$H_0'' : \theta \in \Theta_0''$, em que $\Theta_0'' \equiv \{(\mu_1, \mu_2) : \ \mu_1 = \mu_2, \ \mu_1, \mu_2 \in \mathbb{R}\}$,

Terceira hipótese nula:
$H_0''' : \theta \in \Theta_0'''$, em que $\Theta_0''' = \{(0.2, 0.2)\}$

Dados Observados: Para $n = 100$, $\bar{x} = (-0.051, 0.074)^\top$ e $\Sigma = \begin{pmatrix} 1 & 0.8 \\ 0.8 & 1 \end{pmatrix}$. Temos que os s-valores são, respectivamente,

$s' = 0.141 \ (p' = 0.141)$,
$s'' = 0.142 \ (p'' = 0.048)$ e
$s''' = 0.0195 \ (p''' = 0.0197)$.

Ilustração gráfica



Comentários Finais

A medida proposta:

- pode ser aplicada imediatamente sempre que o logaritmo da função de verossimilhança for estritamente côncavo $(q = 1 - F(F_{H_0}(1 - p)))$
- é uma medida possibilística e pode ser estudada utilizando o ABC (Darwiche, Ginsberg, 1992).
- pode ser justificada usando *desiderata* (axiomas mais básicos).
- precisa ser estudada com mais detalhes para definir pontos de cortes seguindo algum critério.
- é uma alternativa clássica ao FBST (Pereira, Stern, 1998).

Referências:

Darwiche, A.Y., Ginsberg, M.L. (1992). A symbolic generalization of probability theory, AAAI-92, Tenth National Conference on Artificial Intelligence.

Patriota, AG (2014). Uma medida de evidência alternativa para testar hipóteses gerais, Ciência e Natura, 36, 14–22.

Patriota, AG (2013). A classical measure of evidence for general null hypotheses, Fuzzy Sets and Systems, 233, 74-88.

Pereira, C.A.B., Stern, J.M. (1999). Evidence and credibility: Full Bayesian significance test for precise hypotheses, Entropy, 1, 99–110.