

Introdução à Teoria da Decisão

Prova didática apresentada como parte do concurso para obtenção do título de
Livre-Docente

Victor Fossaluza

Departamento de Estatística - IME-USP

6 de outubro de 2025

- Introduzir os conceitos básicos de teoria da decisão;
- Discutir como obter soluções ótimas para problemas de decisão, mostrando que essas soluções são naturalmente bayesianas (no sentido de serem apenas resumos da distribuição posterior);
- Sob essa abordagem, apresentar os temas usualmente tratados na inferência estatística: estimação pontual, estimação por região e teste de hipóteses;
- Ilustrar os problemas descritos anteriormente através de exemplos tão simples quanto possível.

- A teoria da decisão é uma das possíveis formas de embasar a inferência bayesiana.
- Considere uma **medida de probabilidade** que representa (quantifica) sua *incerteza*, isto é, sob algumas suposições, se você acredita em B pelo menos tanto em quanto em A , então existe uma única medida de probabilidade P tal que $P(A) \leq P(B)$.
- De modo similar, considere também uma **função de perda** (ou **função de utilidade**) que quantifica numericamente as *consequências* (ou *preferências*) de sua decisão para um dado valor do parâmetro. Essa quantificação de preferências novamente é subjetiva e, dado um conjunto de suposições, existe uma função de perda que representa numericamente o prejuízo causado por uma decisão para cada possível valor do parâmetro.
- Essas construções não serão apresentadas aqui mas podem ser encontradas no livro *Optimal Statistical Decisions* (DeGroot, M.H.).

- d : **decisão** - uma particular afirmação, por exemplo, sobre θ . No contexto inferencial, uma decisão pode ser uma estimativa (pontual ou intervalar) para θ ou a escolha de uma hipótese específica em um teste de hipóteses.
- $\mathcal{D}$: **espaço de decisões** - conjunto de todas as possíveis decisões (afirmações sobre θ).
- θ : **estado da natureza** - quantidade desconhecida ou parâmetro (no contexto de inferência estatística).
- Θ : **espaço dos estados da natureza** - espaço paramétrico ($\mathcal{A}$: σ -álgebra de Θ).
- $P : \mathcal{A} \rightarrow [0, 1]$: **medida de probabilidade** - quantifica numericamente sua incerteza sobre θ .
- $L : \mathcal{D} \times \Theta \rightarrow \mathbb{R}$: **função de perda** - $L(d, \theta)$ que representa o prejuízo de uma decisão d quando o estado da natureza é θ .

Exemplo 1

- Suponha que você está saindo de casa pela manhã e precisa tomar uma importante decisão: levar ou não seu guarda-chuva.
 - $\mathcal{D} = \{G, \bar{G}\}$, em que G : levar guarda-chuva.
 - $\Theta = \{C, \bar{C}\}$, em que C : chove durante o dia.
- Suponha que carregar o guarda-chuva é algo que não lhe agrada. Por outro lado, você odeia ficar molhado e acredita que a pior situação seria não levá-lo e tomar chuva. Mesmo se chover, você ficará incomodado se levar o guarda-chuva pois, além de tê-lo carregado, voltou para casa com o tênis molhado. Note que, nessas circunstâncias, o cenário preferido por você seria não levar o guarda-chuva e não chover.

Exemplo 1

- Para quantificar suas preferências, considere uma função de perda $L : \mathcal{D} \times \Theta \rightarrow \mathbb{R}$, de modo que, quanto mais algum cenário lhe gera incômodo, maior sua perda. Um exemplo é apresentado a seguir.

Decisão	Estados da Natureza	
	C	$\bar{C}$
G	2 (ruim)	1 (bom)
$\bar{G}$	3 (pior)	0 (melhor)
$P(\theta)$	p	$1 - p$

- Uma possível maneira de tomar uma decisão é escolher a decisão “menos prejudicial”. Se levar o guarda-chuva, no pior caso, sua perda é $\max_{\theta} L(G, \theta) = 2$ e, se não levá-lo, a maior perda possível é $\max_{\theta} L(\bar{G}, \theta) = 3$. A decisão que tem a menor dentre as maiores perdas é levar o guarda-chuva. Esse procedimento para tomada de decisões é chamado *min-max* e consiste em escolher a decisão d' tal que $d' = \arg \min_d \max_{\theta} L(d, \theta)$.

Exemplo 1

- Sendo um pouco mais otimista, você pode escolher a decisão que tenha a maior dentre as menores perdas. Esse procedimento é chamado *max-min* e consiste em escolher a decisão $d' = \arg \max_d \min_{\theta} L(d, \theta)$. No nosso exemplo, esse procedimento também sugere que você sempre carregue o guarda-chuva.
- Note que as decisões escolhidas pelos dois procedimentos descritos anteriormente sugere que você sempre deve carregar o guarda-chuva. Contudo, isso pode não ser razoável. Imagine que você estava lendo notícias antes de sair de casa e viu que a probabilidade de chuva era 0.01. Nesse caso, não parece fazer sentido você levar o guarda-chuva, já que isso vai te trazer um desconforto e a chance de chover é muito baixa. Portanto, a probabilidade de chover deveria ser levada em consideração em sua tomada de decisão.

Exemplo 1

- Uma maneira de considerar a chance de chuva é utilizar a **perda esperada**. Note que θ é uma quantidade desconhecida e, pelo que já foi discutido anteriormente, você deve descrever sua incerteza em relação a essa quantidade em termos de probabilidade. Suponha que no exemplo $P(C) = p$, $0 \leq p \leq 1$.

Perda Esperada ou Risco da decisão d contra a priori P

Para cada decisão $d \in \mathcal{D}$, é possível calcular o valor esperado da função de perda como

$$\rho(d, P) = E[L(d, \theta)] = \int_{\Theta} L(\theta) dP(\theta) .$$

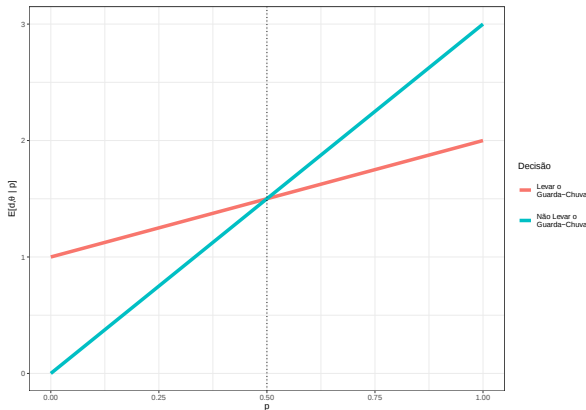
- No exemplo, temos
 - $E[L(G, \theta)] = L(G, C)P(C) + L(G, \bar{C})P(\bar{C}) = 2p + 1(1 - p) = p + 1$;
 - $E[L(\bar{G}, \theta)] = L(\bar{G}, C)P(C) + L(\bar{G}, \bar{C})P(\bar{C}) = 3p + 0(1 - p) = 3p$.

Exemplo 1

- Deste modo, as perdas esperadas associadas a cada decisão dependem da probabilidade de chuva p e, para cada possível valor de p , deve-se tomar a decisão que tem menor perda esperada.
- Por exemplo, se a probabilidade de chuva é $p = 0.1$, temos que as perdas esperadas para as decisões de levar ou não o guarda-chuva são, respectivamente, $E[L(G, \theta)] = 1.1$ e $E[L(\bar{G}, \theta)] = 0.3$. Sob essa abordagem, sua decisão seria de não levar o guarda-chuva nesse caso.
- Por outro lado, se a probabilidade de chuva for $p = 0.9$, suas perdas esperadas seriam respectivamente $E[L(G, \theta)] = 1.9$ e $E[L(\bar{G}, \theta)] = 2.7$, de modo que a decisão ótima seria levar o guarda-chuva.

Exemplo 1

- O gráfico a seguir apresenta as perdas para cada decisão d e para cada valor de p . É possível notar que a decisão ótima é levar o guarda-chuva quando $p > 0.5$ e não levá-lo caso contrário.



- Um **problema de decisão** será denotado por $(\Theta, \mathcal{D}, L, P)$, onde Θ é o espaço paramétrico, $\mathcal{D}$ é o espaço de decisões, $L : \mathcal{D} \times \Theta \rightarrow \mathbb{R}$ é uma função de perda e P é a distribuição de probabilidade que representa sua crença sobre a quantidade desconhecida θ . Equivalentemente, a função de perda L poderia ser substituída por uma *função de utilidade* U (por exemplo, tome $U = -L$).
- A solução de um problema de decisão $(\Theta, \mathcal{D}, L, P)$ é chamada **decisão de Bayes** (ou *decisão ótima*), $d^* \in \mathcal{D}$, tal que $\rho(d^*, P) = \inf_{d \in \mathcal{D}} \rho(d, \theta)$.
- O **risco de bayes** é a perda esperada da decisão de Bayes,
 $\rho(d^*, P) = \inf_{d \in \mathcal{D}} \rho(d, P)$.
- Para uma argumentação mais formal sobre a escolha pela decisão que minimiza a perda esperada, ver *Optimal Statistical Decisions* (DeGroot, M.H.).

- Suponha que antes de escolher uma decisão $d \in \mathcal{D}$, é possível observar um v.a. X que (supostamente) está relacionada com θ (isto é, X traz alguma informação sobre θ).
- Desde modo, considere a família $\mathcal{P} = \{f(\cdot|\theta) : \theta \in \Theta\}$ de funções de distribuição condicionais para X , isto é, para cada $\theta \in \Theta$ é possível determinar a distribuição condicional de $X|\theta$. Essa distribuição, juntamente com a distribuição a priori $f(\theta)$, determina totalmente uma distribuição conjunta $f(x, \theta) = f(x|\theta)f(\theta)$.
- O objetivo agora é encontrar uma regra de decisão que, para cada possível observação $x \in \mathfrak{X}$, será associada uma decisão em $\mathcal{D}$.

- Defina uma **função de decisão** $\delta : \mathfrak{X} \longrightarrow \mathcal{D}$ que associa a cada resultado experimental $x \in \mathfrak{X}$ uma decisão $d \in \mathcal{D}$. Denote o conjunto de todas as possíveis funções de decisão por Δ .
- O risco $\rho(\delta, P)$ da função de decisão $\delta \in \Delta$ é dado por

$$\rho(\delta, P) = E[L(\delta, \theta)] = \int_{\Theta} \int_{\mathfrak{X}} L(\delta(x), \theta) dP(x, \theta) .$$

- A **função de decisão de Bayes**, $\delta^* \in \Delta$, é tal que $\rho(\delta^*, P) = \inf_{\delta \in \Delta} \rho(\delta, P)$.

Exemplo 2

- Seja $\Theta = \{\theta_1, \theta_2\}$, $\mathcal{D} = \{a_1, a_2\}$, $\mathfrak{X} = \{0, 1\}$.
 $X|\theta_1 \sim \text{Ber}(3/4)$, $X|\theta_2 \sim \text{Ber}(1/3)$ e, a priori, $P(\theta = \theta_1) = P(\theta = \theta_2) = 1/2$.
Considere a função de perda a seguir.

L	θ_1	θ_2
a_1	0	5
a_2	10	0

- Temos que $|\Delta| = 2^2 = 4$, de modo que as possíveis funções de decisão são

$$\delta_1(x) = \begin{cases} a_1, & x = 1 \\ a_2, & x = 0 \end{cases}$$

$$\delta_2(x) = \begin{cases} a_1, & x = 0 \\ a_2, & x = 1 \end{cases}$$

$$\delta_3(x) = a_1$$

$$\delta_4(x) = a_2$$

Exemplo 2

Para a função δ_1 , temos

x	θ	$\delta(x)$	$L(\delta_1(x), \theta)$	$P(x \theta)$	$P(\theta)$	$P(x, \theta)$
0	θ_1	a_2	10	1/4	1/2	1/8 = 3/24
0	θ_2	a_2	0	2/3	1/2	2/6 = 8/24
1	θ_1	a_1	0	3/4	1/2	3/8 = 9/24
1	θ_2	a_1	5	1/3	1/2	1/6 = 4/24

$$\rho(\delta_1, P) = \sum_{x=0}^1 \sum_{i=1}^2 L(\delta_1(x), \theta_i) \underbrace{P(X=x|\theta_i)P(\theta_i)}_{P(x,\theta)} = 10 \frac{3}{24} + 5 \frac{4}{24} = \frac{50}{24}$$

De forma análoga, $\rho(\delta_2, P) = 130/24$, $\rho(\delta_3, P) = 60/24$, $\rho(\delta_4, P) = 120/24$.

$$\delta^*(x) = \delta_1(x) = \begin{cases} a_1, & x = 1 \\ a_2, & x = 0 \end{cases}$$

Risco de Bayes:

$$\rho(\delta^*, P) = 50/24.$$

- Em problemas mais complicados, pode ser muito trabalhoso (ou impossível) obter a função de decisão dessa forma, chamada de **forma normal**. Sob essa abordagem, é necessário encontrar a função de decisão de bayes δ^* dentre todas as possíveis funções de decisão em Δ .
- Nesses casos, pode ser mais fácil resolver o problema usando a **forma extensiva** em que, para cada $x \in \mathfrak{X}$, obtém-se a decisão de Bayes d_x^* que minimiza o *risco posterior*, definido por $r_x(d) = \int_{\Theta} L(d, \theta) dP(\theta|x)$.
- Assim, é possível obter uma decisão de Bayes d_x^* para um específico ponto x observado ou, ainda, construir uma função de decisão de Bayes, fazendo $\delta^*(x) = d_x^*$ para cada $x \in \mathfrak{X}$. A seguir, é mostrado que essa duas formas produzem resultados que minimizam o risco.

$$\rho(\delta, P) = E[L(\delta, \theta)] = \int_{\Theta} \int_{\mathfrak{X}} L(\delta(x), \theta) dP(x, \theta) = \int_{\mathfrak{X}} \left[\underbrace{\int_{\Theta} L(\delta(x), \theta) dP(\theta|x)}_{r_x(\delta(x))} \right] dP(x).$$

Note que a integral interna (em θ) pode ser resolvida para cada x fixado. Para cada x , considere a decisão d_x^* tal que $r_x(d_x^*) = \inf_{d \in \mathcal{D}} r_x(d)$. Assim,

$$\begin{aligned} \rho(\delta, P) &= \int_{\mathfrak{X}} \left[\underbrace{\int_{\Theta} L(\delta(x), \theta) dP(\theta|x)}_{r_x(\delta(x))} \right] dP(x) = \int_{\mathfrak{X}} r_x(\delta(x)) dP(x) \\ &\geq \int_{\mathfrak{X}} r_x(d_x^*) dP(x). \end{aligned}$$

De modo que, a função $\delta^* : \mathfrak{X} \longrightarrow \mathcal{D}$ tal que $\delta^*(x) = d_x^*$ é uma *função de decisão de Bayes*.

Voltando ao Exemplo 2

$\mathcal{D} = \{a_1, a_2\}$, $X|\theta_1 \sim \text{Ber}(3/4)$, $X|\theta_2 \sim \text{Ber}(1/3)$ e $P(\theta_1) = P(\theta_2) = 1/2$.

$$P(\theta_1|X=0) = \frac{P(X=0|\theta_1)P(\theta_1)}{P(X=0|\theta_1)P(\theta_1) + P(X=0|\theta_2)P(\theta_2)} = \frac{\frac{1}{4} \frac{1}{2}}{\frac{1}{4} \frac{1}{2} + \frac{2}{3} \frac{1}{2}} = \frac{3}{11}$$

$$P(\theta_2|X=0) = \frac{8}{11}$$

$$r_x(a_1, P) = \sum_{i=1}^2 L(a_1, \theta_i) P(\theta_i|X=0) = 0 P(\theta_1|X=0) + 10 P(\theta_2|X=0) = \frac{80}{11}$$

$$r_x(a_2, P) = 5 P(\theta_1|X=0) + 0 P(\theta_2|X=0) = \frac{15}{11}$$

Logo, para $x=0$, $d_0^* = a_2$. De forma análoga, para $x=1$, $d_1^* = a_1$ e, portanto,

$$\delta^*(x) = \begin{cases} a_2, & x=0 \\ a_1, & x=1 \end{cases}.$$

- Todos os problemas de inferência estatística podem ser vistos como um caso particular de Teoria da Decisão.
- Por exemplo, um problema de estimação pontual consiste em encontrar um “chute” para o valor do parâmetro θ , de modo que o espaço de decisões é $\mathcal{D} = \Theta$.
- Nesse tipo de problema é usual considerar funções de perda da forma $L(d, \theta) = s(\theta)\Delta(d, \theta)$, onde Δ é uma distância (ou uma medida de discrepância) relacionada ao erro por tomar a decisão d quando o valor do parâmetro é θ e s é uma função não-negativa relacionada à gravidade do erro para cada θ (pode ser constante).

Exemplo 3 - Estimação Pontual para Perda Quadrática

- Considere um problema de estimação pontual, isto é, $\mathcal{D} = \Theta$, onde a função de perda é dada por $L(d, \theta) = (d - \theta)^2$, conhecida como *perda quadrática*.

$$\begin{aligned} r_x(d) &= \int_{\Theta} (d - \theta)^2 dP(\theta|x) = \int_{\Theta} (d^2 - 2d\theta + \theta^2) dP(\theta|x) \\ &= d^2 \int_{\Theta} dP(\theta|x) - 2d \int_{\Theta} \theta dP(\theta|x) + \int_{\Theta} \theta^2 dP(\theta|x) \\ &= d^2 - 2d E[\theta|x] + E[\theta^2|x] \end{aligned}$$

- Como $r_x(d)$ é um polinômio de grau 2 em d , o mínimo ocorre em $d_x^* = E[\theta|x]$.
- No caso em que $E[\theta|x] \notin \Theta$, d_x^* será o ponto do espaço paramétrico mais próximo de $E[\theta|x]$, de modo que $r_x(d_x^*) = \inf_{d \in \Theta} r_x(d)$.
- Um estimador para θ contra a perda quadrática é $\delta^*(X) = E[\theta|X]$.

- Estimador de Bayes para θ contra diferentes funções de perda:

① Perda Quadrática: $L_2(d, \theta) = (d - \theta)^2 \implies \delta_2^*(X) = E[\theta|X];$

② Perda Absoluta: $L_1(d, \theta) = |d - \theta| \implies \delta_1^*(X) = Med(\theta|X);$

③ Perda 0-1: $L_0(d, \theta) = c \mathbb{I}(d \neq \theta) \implies \delta_0^*(X) = Moda(\theta|X).$

Exemplo 3 - Voltando à Perda Quadrática

- O risco a posteriori para o estimador $\delta_2^*(X) = E[\theta|X]$ é

$$r_x(E[\theta|x]) = \int_{\Theta} L(\delta^*(x), \theta) dP(\theta|x) = \int_{\Theta} (\theta - E[\theta|x])^2 dP(\theta|x) = \text{Var}(\theta|x) ,$$

de modo que o risco de Bayes é dado por

$$\rho(\delta^*(X), P) = \int_{\mathfrak{X}} \underbrace{\int_{\Theta} (\theta - E[\theta|x])^2 dP(\theta|x)}_{\text{Var}[\theta|x]} dP(x) = E[\text{Var}(\theta|X)] .$$

Exemplo 3 - Voltando à Perda Quadrática

- A variância da posteriori $Var(\theta|x)$ pode ser vista como uma medida de informação, no sentido que quanto menor essa variância, mais concentrada é a distribuição e há “*menos incerteza*” sobre θ . Nesse sentido, espera-se que ao observar $X = x$, a variância $Var(\theta|x)$ diminua em relação a variância da priori $Var(\theta)$.

$$\underbrace{Var(\theta)}_{\text{constante}} = \underbrace{E[Var(\theta|X)]}_{\Downarrow} + \underbrace{Var[E(\theta|X)]}_{\Uparrow}$$

- Aparentemente, como espera-se que a variância da posteriori diminua, a variância do estimador deveria aumentar. Algumas vezes isso é colocado como se o objetivo fosse encontrar o estimador de maior variância, em contradição com a abordagem frequentista. Contudo, há outra interpretação desse resultado: deseja-se obter um estimador que varie bastante de acordo com o valor observado de X , isto é, a informação trazida pela amostra muda sua opinião sobre θ .

Exemplo 4 - Perda Quadrática

- $X_1, \dots, X_n$ c.i.i.d. tais que $X_i|\theta \sim \text{Poisson}(\theta)$ e, a priori, $\theta \sim \text{Gama}(a, b)$.
- A função de verossimilhança é $f(x|\theta) \propto \prod \theta^{x_i} e^{-\theta}$ e a priori é $f(\theta) \propto \theta^{a-1} e^{-b\theta}$.
- A posteriori é $f(\theta|x) \propto \theta^{\sum x_i} e^{-n\theta} \cdot \theta^{a-1} e^{-b\theta} \propto \theta^{a+\sum x_i-1} e^{-(b+n)\theta}$, de modo que $\theta|X = x \sim \text{Gama}(a + \sum x_i, b + n)$.
- Como visto anteriormente, o estimador de Bayes contra a perda quadrática é o valor esperado de θ dado X e, neste caso,

$$\delta^*(X) = E[\theta|X] = \frac{a + \sum_{i=1}^n X_i}{b + n}.$$

Exemplo 4 - Perda Quadrática

- Note que $E[X_i|\theta] = \theta$, de modo que $E[X_i] = E[E(X_i|\theta)] = E[\theta] = a/b$.
- Além disso, $Var(\theta) = \frac{a}{b^2}$ e $Var(\theta|X) = \frac{a + \sum X_i}{(b + n)^2}$. Então,

$$\rho(\delta^*, P) = E[Var(\theta|X)] = E\left[\frac{a + \sum_i X_i}{(b + n)^2}\right] = \frac{(a + \sum E[X_i])}{(b + n)^2} = \frac{\frac{a}{b}(b + n)}{(b + n)^2} = \frac{a}{b(b + n)}.$$

- Por fim, note que a decisão de Bayes pode ser escrita como uma combinação linear da média da distribuição a priori e do estimador de máxima verossimilhança (EMV)

$$E[\theta|X] = \frac{a + \sum_i X_i}{b + n} = \frac{b}{b + n} \left(\frac{a}{b}\right) + \frac{n}{b + n} \bar{X}.$$

Resultado

Seja $L(d, \theta) = \mathbb{I}(|\theta - d| > \varepsilon) = 1 - \mathbb{I}(d - \varepsilon \leq \theta \leq d + \varepsilon)$, $\varepsilon > 0$. Então, $\delta^*(X)$ é centro do intervalo modal, isto é, o intervalo de tamanho 2ε de maior probabilidade a posteriori. Em particular, quando $\varepsilon \downarrow 0$, temos que $\delta^*(X) = \text{Moda}(\theta|X)$.

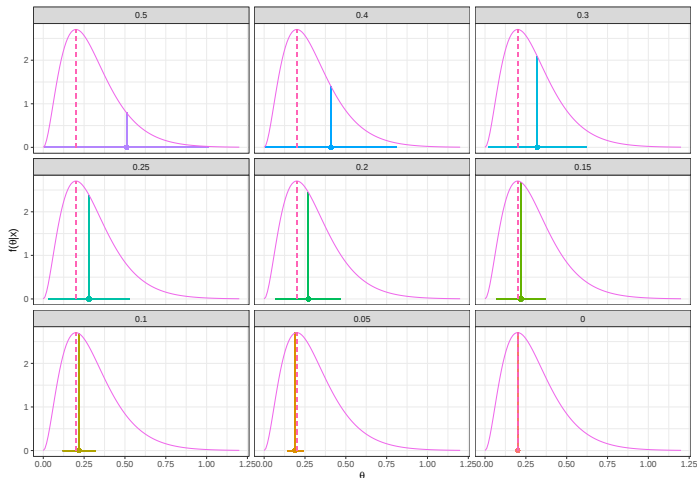
Demo: O risco posterior de uma decisão d é

$$\begin{aligned} r_x(d) &= E[L(d, \theta)|x] = E[\mathbb{I}(|\theta - d| > \varepsilon)] = E[1 - \mathbb{I}(d - \varepsilon \leq \theta \leq d + \varepsilon)|x] \\ &= 1 - P(d - \varepsilon \leq \theta \leq d + \varepsilon|x) \end{aligned}$$

O risco $r_x(d)$ é mínimo quando a probabilidade $P(d - \varepsilon \leq \theta \leq d + \varepsilon|x)$ é máxima. Assim, basta tomar o intervalo $[d_x^* - \varepsilon ; d_x^* + \varepsilon]$ com maior probabilidade a posteriori e a decisão de Bayes nesse caso será o valor central desse intervalo, d_x^* .

Exemplo 5 - Função de Perda Não Usual

Considere o Exemplo 4 em que $\theta|x \sim \text{Gama}(a + \sum x_i, b + n)$ e a função de perda do resultado anterior, $L(d, \theta) = \mathbb{I}(|\theta - d| > \varepsilon)$. Para diferentes valores de ε , tem-se:



Exemplo 5 - Função de Perda Não Usual

- Quando $\varepsilon \downarrow 0$, temos que essa função de perda é equivalente a Perda 0-1, $L_0(d, \theta) = \mathbb{I}(d \neq \theta)$ e

$$\delta^*(x) = \text{Moda}(\theta|x) = \sup_{\theta} f(\theta|x) = \frac{a + \sum x_i - 1}{b + n}.$$

- Similarmente ao exemplo da perda quadrática, é possível escrever o estimador de Bayes desse exemplo como combinação linear da moda da priori e do EMV.

$$\delta^*(X) = \text{Moda}(\theta|X) = \frac{a + \sum X_i - 1}{b + n} = \frac{b}{b + n} \left(\frac{a - 1}{b} \right) + \frac{n}{b + n} \bar{X}.$$

- Em um problema de estimação por regiões (ou estimação intervalar, no caso univariado), o objetivo é obter um conjunto de valores razoáveis para θ .
- Mais formalmente, temos um problema aonde a decisão consiste em escolher um subconjunto do espaço paramétrico, de modo que $\mathcal{D} = \mathcal{A}$, onde $\mathcal{A}$ é σ -álgebra de subconjuntos de Θ .
- Uma estimativa por região é comumente chamada na literatura Bayesiana de **região de credibilidade** ou **região de probabilidade** $\gamma = 1 - \alpha$.

Função de Perda para Estimação por Regiões

- Uma *função de perda* razoável para um problema de estimação por região deve levar em consideração dois fatores:
 - O tamanho da região $d \in \mathcal{A}$;
 - Pertinência de θ na região d .
- Assim, considere uma função de perda da forma

$$L(d, \theta) = \lambda(d) - k \mathbb{I}(\theta \in d) ,$$

em que $\lambda(d)$ é o “tamanho” da região d e $\mathbb{I}$ é a função indicadora.

- Por exemplo, λ pode ser a medida de Lebesgue, no caso contínuo, ou a medida de contagem, no caso discreto (no caso geral, considere uma medida que domina a distribuição a posteriori, $P(\theta|x) \ll \lambda$).

Estimação por Regiões - Caso Absolutamente Contínuo

- No caso em que a distribuição a posteriori é absolutamente contínua com relação a medida de Lebesgue, o risco a posteriori de uma decisão $d \in \mathcal{A}$ é

$$\begin{aligned} r_x(d) &= \int_{\Theta} [\lambda(d) - k \mathbb{I}(\theta \in d)] dP(\theta|x) = \int_{\Theta} \mathbb{I}(\theta \in d) d\theta - \int_{\Theta} k \mathbb{I}(\theta \in d) f(\theta|x) d\theta \\ &= \int_d (1 - kf(\theta|x)) d\theta . \end{aligned}$$

- Esse risco é mínimo quando

$$d = \{\theta \in \Theta : 1 - kf(\theta|x) \leq 0\} = \{\theta \in \Theta : f(\theta|x) \geq 1/k\} .$$

- Assim, a decisão de Bayes contra essa função de perda consiste em escolher uma região $d \in \mathcal{A}$ que contém os pontos do espaço paramétrico com maior densidade a posteriori.

- A região $R \subseteq \Theta$ é dita uma **região HPD** (Highest Posterior Density) de probabilidade γ se
 - i. $P(\theta \in R|x) = \gamma$;
 - ii. $\forall \theta \in R$ e $\forall \theta' \notin R$, $f(\theta|x) \geq f(\theta'|x)$.
- Na prática, é comum escolher um valor da probabilidade γ da região R . Isso é equivalente a fixar um k_γ na função de perda $L(d, \theta) = \lambda(d) - k_\gamma \mathbb{I}(\theta \in d)$.

Exemplo 6 - Intervalo de Credibilidade para a Normal

- Seja $X_1, \dots, X_n$ c.i.i.d. tais que $X_i|\theta \sim N(\theta, 1/\tau)$, $\tau = 1/\sigma^2$, com τ conhecido.
- Considere que, a priori, $\theta \sim N(m, 1/v)$ então

$$\theta|x \sim N\left(\underbrace{\frac{vm + n\tau\bar{x}}{v + n\tau}}_{M_x}, \underbrace{\frac{1}{v + n\tau}}_{V_x}\right)$$

e, assim,

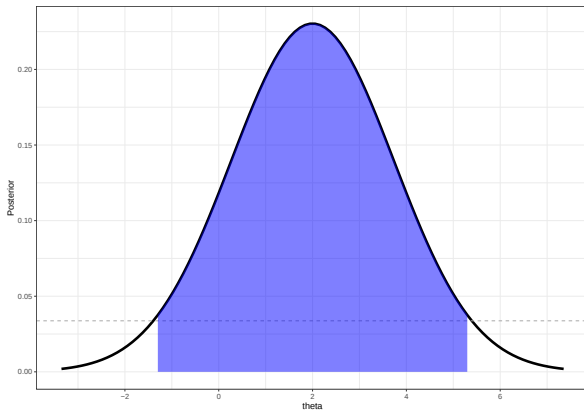
$$Z = \frac{\theta - M_x}{\sqrt{V_x}} \Bigg| x \sim N(0, 1) .$$

- O intervalo HPD de probabilidade $\gamma = 1 - \alpha = 0.95$ é

$$ICr(1-\alpha) = [M_x - z_{\alpha/2}\sqrt{V_x}; M_x + z_{\alpha/2}\sqrt{V_x}] = \left[\frac{vm + n\tau\bar{x}}{v + n\tau} \pm 1.96\sqrt{\frac{1}{v + n\tau}} \right] .$$

Exemplo 6 - Intervalo de Credibilidade para a Normal

- Uma possível forma de representar falta de informação a priori é tomar o limite $v \downarrow 0$ ($1/v \uparrow \infty$). Dessa forma, tem-se $\theta|x \sim N(\bar{x}, 1/(n\tau)) \sim N(\bar{x}, \sigma^2/n)$.
- Neste caso, o intervalo HPD coincide com o I.C. frequentista $ICr(1 - \alpha) = \left[\bar{x} \pm 1.96 \frac{\sigma}{\sqrt{n}} \right]$.



Exemplo 7 - Intervalo de Credibilidade para Posteriori Assimétrica

- Considere $X_1, \dots, X_n$ c.i.i.d. tais que $X_i|\theta \sim Unif(0, \theta)$. Se $\theta \sim Pareto(a, b)$, então $\theta|x \sim Pareto(a + n, \max\{x_{(n)}, b\})$

$$f(\theta|x) = \frac{(a + n)[\max\{x_{(n)}, b\}]^{a+n}}{\theta^{a+n+1}} \mathbb{I}(\max\{x_{(n)}, b\} \leq \theta < \infty) .$$

- Note que a função de densidade a posteriori é estritamente decrescente de modo que o extremo inferior da região HPD é $\max\{x_{(n)}, b\}$. A função de distribuição a posteriori é

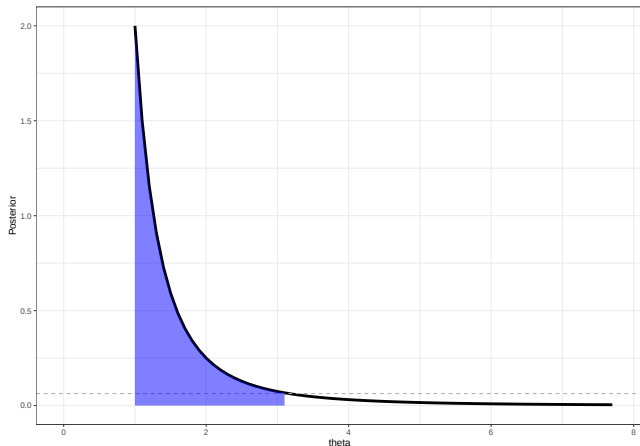
$$F(\theta|x) = 1 - \left(\frac{\max\{x_{(n)}, b\}}{\theta} \right)^{a+n} ,$$

de modo que o extremo superior do intervalo pode ser obtido fazendo

$$1 - \left(\frac{\max\{x_{(n)}, b\}}{\theta} \right)^{a+n} = \gamma \Leftrightarrow \frac{\max\{x_{(n)}, b\}}{\theta^*} = (1 - \gamma)^{1/(a+n)} \Leftrightarrow \theta^* = \frac{\max\{x_{(n)}, b\}}{(1 - \gamma)^{1/(a+n)}} .$$

Exemplo 7 - Intervalo de Credibilidade para Posteriori Assimétrica

$$ICr(1 - \alpha) = \left[\max\{x_{(n)}, b\}, \frac{\max\{x_{(n)}, b\}}{\alpha^{1/(a+n)}} \right]$$



- Uma **hipótese estatística** é uma afirmação sobre o parâmetro θ (ou a família $\mathcal{P}$). No caso usual, tem-se duas hipóteses: $H_0 : \theta \in \Theta_0$, chamada de **hipótese nula**, e $H_1 : \theta \in \Theta_1 = \Theta_0^c$, chamada **hipótese alternativa**.
- Um **teste de hipótese** é uma regra de decisão $\varphi : \mathfrak{X} \longrightarrow \{0, 1\}$, onde $\varphi(x) = 1$ significa rejeitar H_0 (d_1 : decidir por H_1) e $\varphi(x) = 0$, não rejeitar H_0 (d_0 : decidir por H_0).
- Se rejeita-se H_0 (aceita-se H_1) quando H_0 é verdadeira, comete-se um **erro do tipo I**. Por outro lado, se não rejeita-se H_0 (aceita H_0) quando H_0 é falso, ocorre um **erro do tipo II**.
- O conjunto $\varphi^{-1}(1) = \{x \in \mathfrak{X} : \varphi(x) = 1\}$ recebe o nome de **região de rejeição** (ou **região crítica**).
- Uma hipótese é dita **simples** se contém apenas um ponto, $H : \theta = \theta_0$. Caso contrário é chamada de **hipótese composta**. No caso em que $H : \theta \in \Theta_0$ é tal que $\dim(\Theta_0) < \dim(\Theta)$, diz-se que H é uma **hipótese precisa** (“sharp”).

Testes de Hipóteses como um Problema de Decisão

- Considere que $\Theta = \Theta_0 \dot{\cup} \Theta_1$ e deseja-se testar $H_0 : \theta \in \Theta_0$ contra $H_1 : \theta \in \Theta_1$.
- Novamente, temos um problema de decisão em que d_0 : decidir por H_0 e d_1 : decidir por H_1 . Considere então a função de perda a seguir, com $a_0 \leq a_1$ e $b_0 \leq b_1$.

$L(d, \theta)$	Θ_0	Θ_1
d_0	a_0	b_1
d_1	a_1	b_0

- O risco posterior de cada uma das decisões é

$$\begin{aligned} r_x(d_0, \theta) &= a_0 P(\theta \in \Theta_0 | x) + b_1 P(\Theta_1 | x) = a_0 P(\theta \in \Theta_0 | x) + b_1 [1 - P(\Theta_0 | x)] \\ &= a_0 P(\theta \in \Theta_0 | x) + b_1 - b_1 P(\Theta_0 | x) , \end{aligned}$$

$$r_x(d_1, \theta) = a_1 P(\theta \in \Theta_0 | x) + b_0 - b_0 P(\Theta_0 | x) .$$

Testes de Hipóteses como um Problema de Decisão

- De modo que rejeita-se H_0 , $\varphi(x) = 1$, se $r_x(d_1, P) \leq r_x(d_0, P)$

$$\Leftrightarrow (a_1 - b_0)P(\Theta_0|x) + b_0 \leq (a_0 - b_1)P(\Theta_0|x) + b_1$$

$$\Leftrightarrow P(\Theta_0|x) \leq \frac{(b_1 - b_0)}{(a_1 - a_0) + (b_1 - b_0)} .$$

- O teste de bayes nesse caso é

$$\varphi(x) = \begin{cases} 1, & P(\Theta_0|x) \leq \frac{(b_1 - b_0)}{(a_1 - a_0) + (b_1 - b_0)} \\ 0, & c.c. \end{cases} .$$

- É comum adotar-se perda igual a zero quando a decisão tomada é correta, $a_0 = b_0 = 0$, de modo que rejeita-se H_0 se $P(\Theta_0|x) \leq \frac{b_1}{a_1 + b_1}$.

Exemplo 8 - Teste de Hipóteses para a Média da Normal

- $X_1, \dots, X_n$ c.i.i.d. tais que $X_i|\theta \sim N(\theta, \sigma_0^2)$ com σ_0^2 conhecido.
- Considere que o objetivo é testar $H_0 : \theta \leq \theta_0$ contra $H_1 : \theta > \theta_0$.
- Suponha que, a priori, $\theta \sim N(m, v^2)$, de modo que

$$\theta|X \sim N\left(\frac{\sigma_0^2 m + nv^2 \bar{x}}{\sigma_0^2 + nv^2}, \frac{\sigma_0^2 v^2}{\sigma_0^2 + nv^2}\right).$$

- Sendo Φ a função de distribuição da $N(0, 1)$, temos que

$$\begin{aligned} P(\theta \in \Theta_0|x) &= P(\theta \leq \theta_0|x) = P\left(Z \leq \frac{\theta_0 - \frac{\sigma_0^2 m + nv^2 \bar{x}}{\sigma_0^2 + nv^2}}{\sqrt{\frac{\sigma_0^2 v^2}{\sigma_0^2 + nv^2}}} \mid \bar{x}\right) \\ &= \Phi\left(\frac{\theta_0 - \frac{\sigma_0^2 m + nv^2 \bar{x}}{\sigma_0^2 + nv^2}}{\sqrt{\frac{\sigma_0^2 v^2}{\sigma_0^2 + nv^2}}}\right) = \Phi\left(\frac{(\sigma_0^2 + nv^2)\theta_0 - \sigma_0^2 m - nv^2 \bar{x}}{\sigma_0 v \sqrt{\sigma_0^2 + nv^2}}\right). \end{aligned}$$

Exemplo 8 - Teste de Hipóteses para a Média da Normal

- Utilizando o resultado anterior, deve-se rejeitar H_0 se

$$\begin{aligned}P(\theta \in \Theta_0 | x) &\leq \frac{(b_1 - b_0)}{(a_1 - a_0) + (b_1 - b_0)} = K \\ \Leftrightarrow \Phi \left(\frac{(\sigma_0^2 + nv^2)\theta_0 - \sigma_0^2 m - nv^2 \bar{x}}{\sigma_0 v \sqrt{\sigma_0^2 + nv^2}} \right) &\leq K \\ \Leftrightarrow \bar{x} &\geq \frac{(\sigma_0^2 + nv^2)\theta_0 - \sigma_0^2 m}{nv^2} - \Phi^{-1}(K) \frac{\sigma_0 \sqrt{\sigma_0^2 + nv^2}}{nv} \\ \Leftrightarrow \bar{x} &\geq \frac{\sigma_0^2(\theta_0 - m) + nv^2\theta_0}{nv^2} - \Phi^{-1}(K) \frac{\sigma_0 \sqrt{\sigma_0^2 + nv^2}}{nv} .\end{aligned}$$

- Se $a_0 = b_0 = 0$ e $a_1 = b_1$, então $K = 0.5$, $\Phi^{-1}(0.5) = 0$ e rejeita-se H_0 se

$$\bar{x} \geq \frac{\sigma_0^2(\theta_0 - m) + nv^2\theta_0}{nv^2} \xrightarrow[v \uparrow \infty \text{ ou } n \uparrow \infty]{m = \theta_0} \theta_0 .$$

- Foram introduzidos os conceitos básicos de teoria da decisão;
- Discutiu-se como obter soluções ótimas para problemas de decisão com e sem dados, mostrando no segundo caso a equivalência entre a forma normal e extensiva;
- Como casos particulares da teoria da decisão, tem-se os temas usualmente abordados pela inferência estatística: estimação pontual, estimação por região e teste de hipóteses;
- Foram apresentados diversos exemplos ilustrativos para os problemas descritos anteriormente.