

1 Função de Risco e Função de Perda

Esta é uma tradução literal da seção 3.4 (pgs 297 a 299) do Mood & Graybill.

Na seção 3.2 o erro quadrático médio de um estimador foi usado como medida de proximidade de $\tau(\theta)$. Outras medidas são possíveis, por exemplo,

$$E_{\theta}[|T - \tau(\theta)|],$$

chamado de desvio médio absoluto.

A fim de expor e considerar ainda outras medidas de proximidade, vamos pedir emprestado e contar com a linguagem da teoria da decisão.

Baseado em uma amostra aleatória de uma função de densidade ou de probabilidade, o estatístico deve decidir qual estimador para $\tau(\theta)$ ele usará.

Ele pode chamar o valor de qualquer estimador $T = t(X_1, X_2, \dots, X_n)$ de decisão e chamar o próprio estimador de função de decisão desde que diga qual decisão tomar.

Agora a estimativa t de $\tau(\theta)$ pode conter um erro e então, uma medida para avaliar a gravidade do erro é necessária. A palavra perda é usada no lugar de erro e função de perda é usada como uma medida do erro.

Uma definição formal é agora apresentada.

Definição 11 Função de Perda. Considere a estimação de $\tau(\theta)$.

A função de perda, denotada por $l(t; \theta)$ é definida como uma função real satisfazendo:

- (i) $l(t; \theta) \geq 0$, para quaisquer escolhas de t e $\theta \in \Theta$, e
- (ii) $l(t; \theta) = 0$ para $t = \tau(\theta)$.

Assim, $l(t; \theta)$ é a perda cometida quando a estimativa de $\tau(\theta)$, t , é usada e o verdadeiro valor do parâmetro é θ .

Em um problema de estimação teremos de definir uma apropriada função de perda para o problema particular em estudo.

Queremos garantir que a perda seja pequena, ou, dizendo de outra maneira, desejamos assegurar que o erro de estimação seja pequeno ou que a nossa estimativa esteja perto daquilo que queremos estimar.

Exemplo 16: Várias funções de perda são possíveis:

(i) $l_1(t; \theta) = [t - \tau(\theta)]^2$.

(ii) $l_2(t; \theta) = |t - \tau(\theta)|$.

(iii) $l_3(t; \theta) = A > 0$ se $|t - \tau(\theta)| > \epsilon$ e 0 caso contrário.

(iv) $l_4(t; \theta) = \rho(\theta)|t - \tau(\theta)|^r$, para $\rho(\theta) \geq 0$ e $r \geq 0$.

l_1 é chamada de função de perda quadrática e l_2 de função de perda absoluta. Note que l_1 e l_2 crescem quando $t - \tau(\theta)$ crescem em magnitude.

l_3 diz que não há perda alguma se a nossa estimativa estiver ϵ unidades de $\tau(\theta)$ caso contrário perdemos a quantia A .

l_4 é uma função de perda geral que inclui l_1 e l_2 como casos particulares.

Suponha agora que uma apropriada função de perda foi definida para o nosso problema de estimação e pensamos na função de perda como uma medida do erro ou perda.

Nosso objetivo é selecionar um estimador $T = t(X_1, X_2, \dots, X_n)$ que torna o erro ou a perda pequena. (Em geral, a seleção de uma função de perda apropriada não é trivial.)

A função de perda em seu primeiro argumento depende da estimativa t , e t é um valor do estimador T , isto é, $t = t(x_1, x_2, \dots, x_n)$.

Então, a perda depende da nossa amostra X_1, X_2, \dots, X_n . Não podemos fazer a perda pequena para todas as possíveis amostras mas podemos tentar fazer isto, em média

Definição 12: Função de Risco.

Para um dada função de perda l , a função de risco, denotada por $R_l(\theta)$, de um estimador $T = t(X_1, X_2, \dots, X_n)$ é definido por:

$$R_l(\theta) = E_\theta[l(T; \theta)]. \quad (10)$$

A função de risco é a perda esperada. A esperança em (10) pode ser tomada de duas maneiras. Por exemplo, se a densidade $f(x; \theta)$ que foi amostrada é uma função densidade de probabilidade, então

$$\begin{aligned} E_\theta[l(T; \theta)] &= E_\theta[l(T(X_1, X_2, \dots, X_n); \theta)] \\ &= \int \dots \int l(t(x_1, x_2, \dots, x_n); \theta) \prod_{i=1}^n f(x_i; \theta) dx_i. \end{aligned}$$

Ou podemos considerar a variável aleatória T e sua densidade $f_T(t)$. Assim,

$$E_\theta[l(T; \theta)] = \int l(t, \theta) f_T(t) dt.$$

Exemplo 17: Considere as funções de perdas dadas no Exemplo 16. Os correspondentes riscos são dados por:

- (i) $E_\theta[T - \tau(\theta)]^2$, nosso familiar erro médio quadrático.
- (ii) $E_\theta|T - \tau(\theta)|$, que é o erro médio absoluto.
- (iii) $AP_\theta[|t - \tau(\theta)| > \epsilon]$.
- (iv) $\rho(\theta)E_\theta[|T - \tau(\theta)|^r]$

Nosso objetivo agora é selecionar um estimador que torne a perda esperada pequena e idealmente escolher aquele estimador que tenha o menor risco. Para ajudar na nossa prova precisamos usar o conceito de estimadores admissíveis.

Definição 13 Estimador Admissível.

Para dois estimadores $T_1 = t_1(X_1, X_2, \dots, X_n)$ e $T_2 = t_2(X_1, X_2, \dots, X_n)$, estimador T_1 é definido como um melhor estimador do que T_2 se e só se

$$R_{T_1}(\theta) \leq R_{T_2}(\theta), \forall \theta \in \Theta,$$

e

$$R_{T_1}(\theta) < R_{T_2}(\theta), \quad ,$$

para pelo menos um $\theta \in \Theta$.

Um estimador $T = t_1(X_1, X_2, \dots, X_n)$ é dito admissível se e somente se não há estimador melhor.

Em geral, dados dois estimadores T_1 e T_2 não há um estimador melhor do que outro, isto é, suas funções de risco se cruzam com funções de θ .

Este mesmo fenômeno foi observado quando estudamos erro quadrático médio. Aqui, como lá, não existe em geral, um estimador com risco uniformemente mínimo. O problema é a dependência da função de risco de θ

Outra maneira de remover a dependência da função de risco de θ é trocar a função de risco por seu valor máximo e comparar os estimadores olhando os máximos de suas funções de risco e naturalmente preferindo aquele estimador com o menor risco máximo. Tal estimador é dito ser um estimador minimax.

Definição 14: Estimador Minimax.

Um estimador T^* é definido como um estimador minimax se e somente se

$$\sup_{\theta} R_{T^*}(\theta) \leq \sup_{\theta} R_T(\theta),$$

para qualquer estimador T .

Estimadores de Bayes serão discutidos na seção 7.

2 7: Estimadores de Bayes

Agora vamos traduzir a seção 7 do Mood(pgs-339 a 351)

Em nossas considerações sobre o problema nas seções anteriores deste capítulo, assumimos que a nossa amostra vinha de alguma densidade $f(\cdot; \theta)$, onde a função f era suposta conhecida. Além disso, supomos também que o parâmetro θ era fixado, embora desconhecido.

Em algumas situações do mundo real que a densidade $f(\cdot; \theta)$ representa, existe frequentemente informação adicional sobre θ (não somente a informação de que θ assume valores em Θ).

Por exemplo, o pesquisador pode ter evidências que o próprio θ seja uma variável aleatória e ele esteja apto a formular uma função densidade para θ bem realística. Por exemplo, uma máquina que faz peças para um carro é examinada para medir a fração θ de defeituosos que produz.

Certo dia, 10 peças da produção são examinadas e que são representadas por X_1, X_2, \dots, X_{10} , onde $X_i = 1$ se a i -ésima peça é defeituosa e $X_i = 0$ se a peça não é defeituosa. Isto pode ser visto como uma amostra aleatória de tamanho 10 da Bernoulli

$$f(x; \theta) = \theta^x (1 - \theta)^{1-x} I_{\{0,1\}}, \text{ para } 0 \leq \theta \leq 1,$$

que indica que a probabilidade de uma dada peça ser defeituosa é igual ao número desconhecido θ .

A distribuição conjunta da amostra, isto é, das 10 variáveis aleatórias X_1, X_2, \dots, X_{10} é dada por:

$$f(\cdot; \theta) = \theta^{\sum x_i} (1 - \theta)^{n - \sum x_i} \prod_{i=1}^{10} I_{\{0,1\}}(x_i), \text{ para } 0 \leq \theta \leq 1.$$

Olhando a distribuição conjunta da amostra como uma função de θ , $L(\theta)$ obtemos a função de verossimilhança da amostra que quando maximizada nos fornece o estimador de máxima verossimilhança de θ que é dado por:

$$\hat{\Theta}_{MV} = \bar{X},$$

que é a média amostral.

O método dos momentos nos fornece o mesmo estimador, como veremos a seguir:

$$E_{\theta}(X) = \bar{X},$$

como a esperança da Bernoulli é θ

$$\hat{\Theta}_{MM} = \bar{X}.$$

Suponha, agora que o pesquisador tenha alguma informação adicional sobre θ . Ele observou vários dias que o valor de θ varia e ele supõe que essa variação pode ser representada por uma variável aleatória com densidade

$$g(\theta) = 6\theta(1 - \theta) I_{(0,1)}(\theta).$$

Uma questão importante aparece: O que fazer com a informação adicional a respeito de θ na estimação de θ_0 , onde θ_0 é o valor que Θ daquele dia em que a amostra foi retirada?

Para estudar este problema, vamos supor, além de que nossa amostra aleatória venha da densidade $f(\cdot; \theta)$, que θ é desconhecido e é um valor de alguma variável aleatória Θ . Estamos ainda interessados na estimação de alguma função de θ , $\tau(\theta)$.

Sejam $G(\theta)$, a função de distribuição acumulada de θ e $g(\theta)$ sua densidade e assumiremos que elas não contenham parâmetros desconhecidos.

Vamos assumir que a distribuição de Θ é conhecida, isto é, temos informações adicionais. Portanto, uma questão importante é: como essas informações adicionais podem ser usadas na estimação?

2.1 7.1- Distribuição a Posteriori

Daqui por diante usaremos a notação $f(x; \theta)$ para indicar a densidade da variável aleatória X para cada $\theta \in \Theta$. Agora queremos indicar que o parâmetro θ é um valor da variável aleatória Θ , assim vamos escrever a densidade de X como $f(x|\theta)$ ao invés de $f(x; \theta)$.

Agora podemos usar $f(x|\theta)$ como uma densidade condicional, isto é, a densidade de $X|\Theta = \theta$. Uma notação mais completa poderia ser:

$$f_{X|\Theta=\theta}(x|\theta).$$

Seja X_1, X_2, \dots, X_n uma amostra aleatória de tamanho n da densidade $f(x|\theta)$ onde θ é um valor da variável aleatória Θ , suponha que a densidade de Θ , $g(\theta)$ é conhecida e sem parâmetros desconhecidos e suponha que desejamos estimar $\tau(\theta)$. Como incorporar a informação adicional do conhecimento de $g(\theta)$ em nossos procedimentos de estimação.

No passado, pensamos que a função de verossimilhança era uma simples expressão que continha toda a nossa informação. A função de verossimilhança incluía a amostra observada x_1, x_2, \dots, x_n bem como a forma da densidade $f(x; \theta)$ da qual foi amostrada.

Agora precisamos de uma expressão que carregue toda a informação que a verossimilhança contenha mais a informação adicionada da densidade conhecida $g(\theta)$. $g(\theta)$ é conhecida como distribuição a priori de Θ .

Ela sumariza o que sabemos de θ antes da amostra aleatória ser retirada.

Agora vamos procurar uma expressão que sumariza o conhecimento a respeito de θ após a amostra aleatória ser retirada.

Vamos procurar a distribuição a posteriori de $\Theta|X_1 = x_1, \dots, X_n = x_n$.

Definição 29 Distribuições: a Priori e a Posteriori

A densidade $g(\theta)$ é conhecida como distribuição a priori de Θ .

A densidade condicional de $\Theta|X_1 = x_1, \dots, X_n = x_n$, denotada por $f_{\Theta=X_1=x_1, \dots, X_n=x_n}(\theta|x_1, \dots, x_n)$ é chamada de distribuição a posteriori de Θ .

Nota:

$$\begin{aligned} f_{\Theta=X_1=x_1, \dots, X_n=x_n}(\theta|x_1, \dots, x_n) &= \frac{f_{X_1, \dots, X_n|\Theta=\theta}(x_1, \dots, x_n|\theta)g_{\Theta}(\theta)}{f_{X_1, \dots, X_n}(x_1, \dots, x_n)} \\ &= \frac{\prod_{i=1}^n [f(x_i|\theta)] g_{\Theta}(\theta)}{\int \prod_{i=1}^n [f(x_i|\theta)] g_{\Theta}(\theta) d\theta}. \end{aligned}$$

A distribuição a posteriori toma o lugar da função de verossimilhança como uma expressão que incorpora a informação toda.

Se queremos estimar θ e paralelamente o desenvolvimento do estimador de máxima verossimilhança pode ser feito maximizando a distribuição a posteriori, isto é, estimar θ pela moda da posteriori.

Entretanto, ao contrário, da função de verossimilhança (como função de θ), a distribuição a posteriori é uma distribuição de probabilidade e podemos estimar θ pela mediana ou a média da posteriori.

Podemos usar a média da distribuição a posteriori como nossa estimativa de θ e, em geral, podemos estimar $\tau(\theta)$ como a média de $\tau(\Theta)$ dado $X_1 = x_1, \dots, X_n = x_n$, isto é, tomar

$$E[\tau(\Theta)|X_1 = x_1, \dots, X_n = x_n],$$

como nossa estimativa de $\tau(\theta)$.

Estimador de Bayes a Posteriori:

Seja X_1, X_2, \dots, X_n uma amostra aleatória da densidade $f(x|\theta)$, onde θ é um valor da variável aleatória Θ com função densidade $g(\theta)$. O estimador de a Posteriori de Bayes de $\tau(\theta)$ com respeito a priori $g(\theta)$ é definida por:

$$E[\tau(\Theta)|X_1 = x_1, \dots, X_n = x_n].$$

Nota:

$$\begin{aligned} E[\tau(\Theta)|X_1 = x_1, \dots, X_n = x_n] &= \int \tau(\theta) f_{X_1, \dots, X_n|\Theta=\theta}(x_1, \dots, x_n|\theta) g_{\Theta}(\theta) d\theta \\ &= \frac{\int \tau(\theta) \prod_{i=1}^n [f(x_i|\theta)] g_{\Theta}(\theta) d\theta}{\int \prod_{i=1}^n [f(x_i|\theta)] g_{\Theta}(\theta) d\theta}. \end{aligned}$$

Podemos notar a similaridade entre o estimador de Bayes a Posteriori de $\tau(\theta) = \theta$ e o estimador de Pitman do parâmetro de localização. (Veja a equação 17).

Exemplo 43: Seja X_1, X_2, \dots, X_n uma amostra aleatória da Bernoulli de parâmetro θ . Suponha que a distribuição a priori para Θ é uniforme em $(0, 1)$.

Queremos estimar θ e $\tau(\theta) = \theta(1 - \theta)$.

A distribuição a posteriori de Θ é dada por:

$$\begin{aligned} f(\theta|x_1, \dots, x_n) &= \frac{\theta^s(1-\theta)^{n-s}}{\int_0^1 \theta^s(1-\theta)^{n-s} d\theta} I_{(0,1)}(\theta) \\ &= \frac{\theta^s(1-\theta)^{n-s}}{Beta(s+1, n-s+1)} I_{(0,1)}(\theta) \end{aligned}$$

, onde $s = \sum_{i=1}^n x_i$.

A posteriori de $\Theta|X \sim \text{Beta}(a = s + 1, b = n - s + 1)$.

O estimador a posteriori de Bayes de θ é dado pela média da posteriori ou seja

$$E[\Theta|X_1, \dots, X_n] = \frac{a}{a+b} = \frac{s+1}{n+2}.$$

Assim o estimador de Bayes a posteriori com respeito a priori uniforme é dado por:

$$T_B = \frac{\sum_{i=1}^n X_i + 1}{n+2}.$$

$$\text{Como } E(T_B) = \frac{\sum_{i=1}^n E(X_i) + 1}{n+2} = \frac{n\theta + 1}{n+2},$$

segue que o estimador de Bayes é um estimador viciado de θ . O estimador de máxima verossimilhança é \bar{X} que também é o UMVUE de θ .

Para obter o estimador de Bayes a posteriori de $\tau(\theta) = \theta(1-\theta)$ vamos proceder assim:

$$\begin{aligned} E[\tau(\Theta)|X_1, \dots, X_n] &= E[\Theta(1-\Theta)|X_1, \dots, X_n] \\ &= \frac{\int_0^1 \theta(1-\theta) \theta^s(1-\theta)^{n-s} d\theta}{\text{Beta}(s+1, n-s+1)} \\ &= \frac{\int_0^1 \theta^{s+1}(1-\theta)^{n-s+1} d\theta}{\text{Beta}(s+1, n-s+1)} \\ &= \frac{\text{Beta}(s+2, n-s+2)}{\text{Beta}(s+1, n-s+1)} \\ &= \frac{\Gamma(s+2)\Gamma(n-s+2)}{\Gamma(n+4)} \frac{\Gamma(n+2)}{\Gamma(s+1)\Gamma(n-s+1)} \\ &= \frac{(s+1)(n-s+1)}{(n+3)(n+2)}, \end{aligned}$$

O estimador de Bayes a posteriori de $\tau(\theta) = \theta(1-\theta)$ com respeito a priori uniforme é

$$T_B = \frac{\left(\sum_{i=1}^n X_i + 1\right) \left(n - \sum_{i=1}^n X_i + 1\right)}{(n+3)(n+2)}.$$

Notamos mais uma vez que o estimador obtido é viciado.

A próxima nota afirma que em geral que o estimador a posteriori de Bayes é viciado.

Nota : Se

2.2 7.2: Abordagem por Função de Perda

Definição 31:Risco de Bayes

Teorema 13:

Corolário:

Exemplo 44:

Exemplo 45:

2.3 7.3: Estimador Minimax:

Teorema 14:

Exemplo 46:

3 Intervalos de Confiança Bayesianos:

Este material está nas páginas 396,397 e 398 do Mood:

Exemplo 9:

4