Teoria da Decisão: NEYMAN-PEARSON e BAYES

1- Introdução

Teoria da Decisão é uma extensão das teorias de testes de hipóteses e estimação, e permite modelar situações nas quais as teorias clássicas são inadequadas.

Os elementos da Teoria da Decisão são muito semelhantes aos da teoria matemática dos jogos desenvolvida por Von Neumann e Morgenstern (1944) mas, para finalidades estatísticas, um dos "jogadores" é a própria natureza (ao invés de um "jogador" propriamente dito) e que "age" através de uma experiência aleatória.

A análise Bayesiana e a Teoria da Decisão fornecem visões unificadas da Estatística, e torna-se bastante natural pensar nas duas em conjunto. No entanto, grande parte das respectivas teorias desenvolveu-se em separado. No lado Bayesiano, houve a expansão da teoria de inferência estatística (subjetiva e objetiva) e o reconhecimento de que a análise estatística deve ser encarada condicionalmente (ou seja, os dados observados devem ser tratados como **conhecidos**). Este desenvolvimento ocorreu mesmo sem a incorporação formal da função de perda à análise.

Por outro lado, existem contribuições importantes à teoria da decisão feitas por estatísticos freqüentistas (por exemplo, Wald). Na Teoria da Decisão freqüentista evita-se a utilização formal de distribuições a priori.

Aqui adotaremos na maior parte do tempo o paradigma Bayesiano para introduzir as idéias de Teoria da Decisão.

Quais os elementos fundamentais num problema de teoria de decisão?

1) Um conjunto não vazio Θ de possíveis *estados da natureza* θ *(o espaço paramétrico), a partir dos quais decisões devem ser tomadas*;
2) Um conjunto não vazio A de TODAS as possíveis decisões que podem ser tomadas;
3) Uma função perda L que determina a perda $L(\theta, a)$ sofrida pelo indivíduo quando ele toma a ação $\underline{a}$ (*que é uma função do resultado observado da experiência aleatória*) e do real estado da natureza $\underline{\theta}$. *Esta perda é expressa como um número real (cuja interpretação física é uma perda monetária).*

Um **problema de decisão estatística** é um jogo (Θ, A, L) acoplado a uma experiência aleatória cujo resultado x tem densidade $p(x|\theta)$ que depende de um dos "estados da natureza" θ no espaço paramétrico Θ.

Monica Barros
e-mails: barrosm@alumni.utexas.net
monica@mbarros.com

Em termos práticos, a ação(ou decisão) a geralmente indica o valor da estimativa usada para o parâmetro θ.

Suponha que, com base no resultado x da experiência aleatória, o estatístico escolha uma ação a, a = T(x), que resulta numa perda L(θ, a).

Definição 1.1.
Pode-se calcular o **valor esperado da função perda** sobre todos os resultados possíveis da experiência, o que nos dá:

$$R(\theta,T) = E\{L(\theta,T(x))\} = \int L(\theta,T(x))p(x|\theta)dx \qquad (1)$$

A integral (1) depende do verdadeiro estado da natureza θ e da forma da função T(x), que serve para determinar que decisão será tomada uma vez observado o resultado da experiência aleatória.

R(θ,T) definido pela equação (1) é chamado de **função risco.**

Um problema potencial...
- A integral (1) pode não existir, mas iremos ignorar este caso!
- Definiremos uma **regra de decisão** ou **função de decisão** como qualquer função T para a qual R(θ, T) existe e é finita para todo $\theta \in \Theta$.

A escolha da função perda e, conseqüentemente, o cálculo do risco a ela associado é, até certo ponto arbitrária. Existem algumas escolhas "naturais" (ou, pelo menos, mais comuns), como as funções perda quadrática e erro absoluto. Entretanto, estas não são as únicas funções possíveis.

Algumas escolhas usuais para a função perda são:

Função perda quadrática	L(θ, a) = $(\theta - a)^2$
Função perda erro absoluto	L(θ, a) = $\lvert\theta - a\rvert$
Função perda erro quadrático ponderado	L(θ, a) = $w(\theta).(\theta - a)^2$
Função perda 0-1	L(θ, a) = $\begin{cases} 0 & \text{se } \theta = a \\ 1 & \text{se } \theta \neq a \end{cases}$

Agora suponha que temos crenças prévias sobre o estado da natureza, expressas na forma de uma distribuição a priori $\pi(\theta)$. O **risco de Bayes** r(T) da regra de decisão T é definido como o valor esperado de R(θ,T) para todos os possíveis estados da natureza θ.

Definição 1.2. (Risco de Bayes)

$$r(T) = E\{R(\theta,T)\} = \int R(\theta,T)\pi(\theta)d\theta$$

r(T) é um número real, e portanto pode-se comparar dois estimadores com base nos seus respectivos riscos de Bayes, preferindo aquele com o menor risco, isto é, estabelecer uma regra de decisão.

Uma idéia....
Que tal minimizar as perdas através da minimização do risco de Bayes?

$$r(T) = \int\int L(\theta,T(x))p(x|\theta)\pi(\theta)dxd\theta = \int\int L(\theta,T(x))p(x,\theta)dxd\theta = \\ = \int\left\{\int L(\theta,T(x))p(\theta|x)d\theta\right\}p(x)dx \qquad (2)$$

A **perda esperada a posteriori** de uma ação a é definida como:

$$\rho(a,x) = \int L(\theta,a)p(\theta|x)d\theta \qquad (3)$$

Então, o risco de Bayes é minimizado se a regra de decisão T é escolhida de tal forma que ρ(T(x), x) seja um mínimo para todo x.

2- Estimação Pontual

Agora aplicamos a teoria bayesiana de decisão para encontrar estimadores pontuais de parâmetros de interesse.

Uma regra de decisão bayesiana neste contexto é chamada de **estimador de Bayes**. Deve ser óbvio da expressão (3) que os estimadores de Bayes dependem de qual função perda é usada !!!

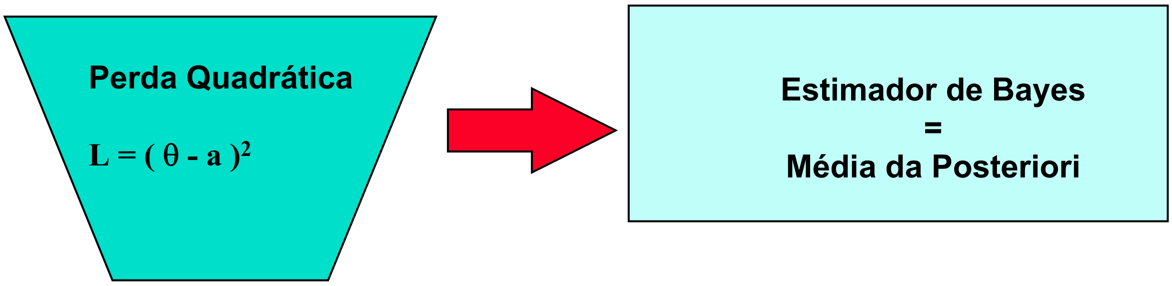
Seja $\underset{\sim}{X} = (X_1, X_2, ..., X_n)$ uma amostra aleatória da densidade f(x, θ). Seja T(X) um estimador para θ e x o valor observado de $\underset{\sim}{X}$.

O estimador de Bayes de θ com respeito à função perda L(θ, T) é aquele com o menor risco de Bayes.

Monica Barros
e-mails: barrosm@alumni.utexas.net
monica@mbarros.com

1) Função de perda quadrática

Se a função perda empregada é a quadrática, isto é, $L(\theta, a) = (\theta - a)^2$, então: ρ(a, x) reduz-se ao erro quadrático médio, e o estimador de Bayes é apenas a média da distribuição a posteriori de θ.



Exemplo 2.1.

(Estimador de Bayes para o parâmetro de uma distribuição de Bernoulli)

Sejam $X_1, X_2, \ldots, X_n \overset{iid}{\sim}$ Bernoulli(θ). Considere a priori $\pi(\theta)$ = Beta(α,β) onde α e β são conhecidos. Suponha que usamos uma função de perda quadrática. Encontre o estimador de Bayes para θ.

Solução

A distribuição a posteriori para θ é uma Beta($\alpha+\Sigma X_i$, β + n-ΣX_i). A média desta posteriori é:

$$\frac{\alpha + \sum X_i}{\alpha + \sum X_i + \beta + n - \sum X_i} = \frac{\alpha + \sum X_i}{\alpha + \beta + n}$$

O estimador de Bayes usando uma função de perda quadrática é:

$$\Gamma^*(\underset{\sim}{X}) = \frac{\alpha + \sum X_i}{\alpha + \beta + n} = \frac{\alpha + n\overline{X}}{\alpha + \beta + n} = \frac{n}{\alpha + \beta + n}\overline{X} + \frac{(\alpha + \beta)}{\alpha + \beta + n}\left(\frac{\alpha}{\alpha + \beta}\right)$$

Note que $\overline{X}$ é o estimador de máxima verossimilhança para θ, e $\alpha/(\alpha+\beta)$ é a média da priori. Assim, o estimador de Bayes é uma soma ponderada da média a priori e do estimador de máxima verossimilhança de θ, e os pesos desta soma dependem dos parâmetros α e β da priori e do número de observações na amostra (n).

Monica Barros
e-mails: barrosm@alumni.utexas.net
monica@mbarros.com

Ao escolher uma priori para θ devemos nos preocupar não apenas com α e β que nos dêem a média a priori desejada, mas também com o valor dos pesos a posteriori. A soma $\alpha + \beta$ nos dá o peso equivalente ao de uma amostra de tamanho n.

Assim, se a nossa opinião a priori tem tanto peso quanto uma amostra de tamanho 20, e se a nossa média a priori é 3/4, devemos escolher α e β tais que:

$$\frac{\alpha}{\alpha+\beta} = \frac{3}{4} \;\; e \;\; \alpha+\beta = 20 \;\Rightarrow\; \alpha = 15 \text{ e } \beta = 5$$

Exemplo 2.2. (Estimador de Bayes para a média de uma Normal com perda quadrática)

Sejam $X_1, X_2, \ldots, X_n \overset{iid}{\sim} N(\theta,\sigma^2)$, onde σ^2 é conhecido. Considere a seguinte priori para θ:

$\pi(\theta) = N(\mu_0, \sigma_0^2)$, onde μ_0, σ_0^2 são conhecidos, e suponha que usamos uma função de perda quadrática. Encontre o estimador de Bayes para θ.

Solução

O estimador de Bayes para θ é μ_p, a média a posteriori , isto é:

$$T^*(\underset{\sim}{X}) = \mu_p = \frac{\sigma^2\mu_0 + n\sigma_0^2\overline{X}}{\sigma^2 + n\sigma_0^2} = \frac{\pi_0}{\pi_p}\mu_0 + \frac{\pi_L}{\pi_p}\overline{X}$$

onde π_0, π_L, π_p são as precisões (recíprocos das variâncias).

2) Função de Perda Erro Absoluto

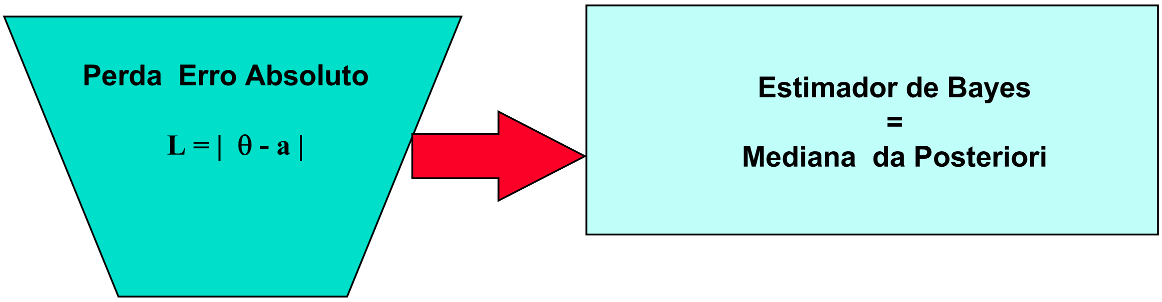
Também é usada com freqüência na prática.

$$L(\theta, a) = |\theta - a|$$

Para qualquer valor observado $\underset{\sim}{X}$, o estimador de Bayes deve minimizar: $E[|\theta - a| \,|\, \underset{\sim}{X}]$ Pode-se mostrar que para qualquer distribuição para θ, min $E[|\theta - a| \,|\, \underset{\sim}{X}]$ ocorre quando a é a ***mediana*** da distribuição de θ.

Logo, o ***estimador de Bayes*** $T^*(\underset{\sim}{X})$ para θ usando a ***perda erro absoluto*** é a ***mediana*** da distribuição a posteriori de θ.

Monica Barros
e-mails: barrosm@alumni.utexas.net
monica@mbarros.com



Exemplo 2.3. (Bernoulli)
Considere a mesma situação do exemplo 2.1., mas suponha que usamos a função perda erro absoluto. O estimador de Bayes para θ é a mediana da posteriori, ou seja, a mediana de uma distribuição Beta($\alpha+\Sigma X_i$, $\beta+n-\Sigma X_i$). Não existe uma expressão simples para esta mediana, ela tem que ser obtida por métodos numéricos.

Em alguns casos simples é possível chegar a uma solução analítica do problema. Por exemplo, suponha que a priori para θ é Beta(α = 2, β = 1), e que observamos n = 5 repetições de Bernoulli que resultam em ΣX_i = 3 sucessos.

A densidade a posteriori é Beta($\alpha+\Sigma X_i$, $\beta+n-\Sigma X_i$) = Beta(5,3), e devemos descobrir a mediana <u>m</u> desta densidade. Note que m satisfaz:

$$\int_0^m \frac{\Gamma(8)}{\Gamma(3)\Gamma(5)}\theta^{5-1}(1-\theta)^{3-1}d\theta = 50\% = \frac{1}{2}$$

e m é um número no intervalo (0,1).

Ou seja:

$$\frac{7!}{2!4!}\int_0^m \theta^4(1-\theta)^2 d\theta = \frac{1}{2}$$

$$\frac{5040}{2(24)}\int_0^m \theta^4(1-2\theta+\theta^2)d\theta = \frac{1}{2}$$

$$210\int_0^m (\theta^4 - 2\theta^5 + \theta^6)d\theta = 1$$

Monica Barros
e-mails: barrosm@alumni.utexas.net
monica@mbarros.com

$$210\left[\frac{m^5}{5}-\frac{2m^6}{6}+\frac{m^7}{7}\right]=1$$

$$210\left[\frac{42m^5-70m^6+30m^7}{210}\right]=1$$

$$42m^5-70m^6+30m^7=1$$

$30m^7 - 70m^6 + 42m^5 - 1 = 0$ é a equação a ser resolvida.

A solução desta equação é: m = 0.6359 e é este o valor do estimador de Bayes usando a perda erro absoluto.

Se tivéssemos usado a função perda quadrática, o estimador de Bayes seria apenas a média da posteriori que é:

$$\frac{\alpha+\sum X_i}{\alpha+\sum X_i+\beta+n-\sum X_i}=\frac{2+3}{5+1+5+3}=\frac{5}{8}=0.625$$

A moda da densidade a posteriori é a moda de uma densidade Beta(5, 3), que é o ponto tal que: $\frac{\Gamma(8)}{\Gamma(3)\Gamma(5)}\theta^{5-1}(1-\theta)^{3-1}$ seja máximo. Se notarmos que o logaritmo desta posteriori tem o mesmo máximo que a densidade, podemos derivar o logaritmo desta densidade com relação a θ e igualar a zero para encontrar : $4.(1/\theta) + 2.\{ -1/(1-\theta)\} = 0$ e assim $\theta = 2/3$.

Logo, neste caso 3 estimadores possíveis são:

média	0.625
mediana	0.6359
moda	2/3 = 0.666

Em todos os exemplos já exibidos aqui, é importante ressaltar uma característica fundamental da análise bayesiana de qualquer problema: aqui, ao contrário do que ocorre quando adotamos uma solução freqüentista, a "solução" do problema é uma distribuição de probabilidade, e não apenas uma estimativa pontual. Na verdade, o paradigma bayesiano fornece uma "resposta" ao problema de estimação pontual muito mais completo que a obtida através da solução clássica, pois nos permite visualizar "por inteiro" a distribuição a posteriori.

Exemplo 2.4. (Normal) – perda erro absoluto
Considere a situação do exemplo 2.2. mas suponha que usamos a função perda erro absoluto. O estimador de Bayes $T^*(\underset{\sim}{X})$ é agora a mediana da posteriori. Mas, a densidade a posteriori é Normal, e neste caso a média e a mediana coincidem, e assim o estimador de Bayes é o mesmo usando-se a perda quadrática ou a perda erro absoluto.

3- Limitações

Muitos estatísticos bayesianos têm sérias restrições contra toda a idéia de estimação pontual. Alguns pontos que foram apresentados são realmente questionáveis, por exemplo:

- Num caso específico, por que uma determinada função perda deve representar as "reais" penalidades econômicas incorridas ao tomar uma decisão incorreta?
- Certamente a mesma função perda não deve ser válida em todos os casos.
- Muitas vezes o erro quadrático médio é infinito, mesmo quando empregamos o estimador mais razoável, o que nos leva a "mágicas" para garantir que as integrais convirjam... ("adhockery" segundo Lindley)
- Se a distribuição a posteriori é bimodal, não faz muito sentido falar em um único estimador pontual....

4- Inferência e Teoria de Decisão

O problema de inferência discutido aqui é básico para qualquer problema de decisão, pois este último só poderá ser resolvido a partir da correta especificação do conhecimento sobre o estado da natureza θ. Mas, o papel da inferência estatística é EXATAMENTE prover o conhecimento sobre θ que permita a tomada de decisões, através do fornecimento da distribuição a posteriori (ou de um sumário apropriado). *Os problemas de decisão e inferência são, a princípio, dissociados, e não precisam ser resolvidos pelo mesmo indivíduo – por exemplo, (Lindley), um cientista não considera as decisões que deverão ser tomadas a partir de suas descobertas; sua tarefa é descrever de maneira precisa os parâmetros da sua experiência.*

5- Relação entre Teoria da Decisão e Testes de Hipóteses Clássicos

É possível reformular os testes de hipótese na linguagem da teoria de decisão.

Suponha que desejamos testar H_0: $\theta \in \Theta_0$ versus H_1: $\theta \in \Theta_1$. Existem duas decisões possíveis, a saber:

Monica Barros
e-mails: barrosm@alumni.utexas.net
monica@mbarros.com

$a_0 \Rightarrow$ aceitar H_0

$a_1 \Rightarrow$ aceitar H_1 (rejeitar H_0)

Sejam π_0 e π_1 as probabilidades a priori para H_o e H_1 respectivamente, e p_0, p_1 as probabilidades a posteriori.

O **fator de Bayes** é definido como:

$$B = \frac{(p_0 / p_1)}{(\pi_0 / \pi_1)}$$

Sejam:

$\rho_0(\theta) = \pi(\theta)/\pi_0$ e
$\rho_1(\theta) = \pi(\theta)/\pi_1$

onde $\pi(\theta)$ é a densidade a priori.

Suponha que existe uma função perda $L(\theta, a)$ definida como:

$a \mid \theta$	$\theta \in \Theta_0$	$\theta \in \Theta_1$
a_0	0	1
a_1	1	0

Então o uso de uma regra de decisão T(x) resulta numa função perda esperada a posteriori:

$\rho(a_0, x) = p_1$
$\rho(a_1, x) = p_0$

Assim, uma decisão T(x) que minimiza a perda esperada a posteriori é apenas uma decisão que ACEITA a hipótese com MAIOR PROBABILIDADE A POSTERIORI. Mas, esta era exatamente a forma de escolher entre hipóteses quando a idéia de testes de hipóteses foi originalmente concebida.

De forma mais geral, se a função perda é uma função "0-K_i", isto é:

$a \mid \theta$	$\theta \in \Theta_0$	$\theta \in \Theta_1$
a_0	0	K_o
a_1	K_1	0

As perdas esperadas a posteriori das duas decisões são:

Monica Barros
e-mails: barrosm@alumni.utexas.net
monica@mbarros.com

$\rho(a_0, x) = p_1.K_0$
$\rho(a_1, x) = p_0.K_1$

Uma regra de decisão de Bayes resulta na **rejeição da hipótese nula** se, e somente se: $\mathbf{p_0.K_1 < p_1.K_0}$, ou seja, se a **perda esperada da decisão a_1 é menor que a perda esperada da decisão a_0**.

Ou seja, em termos do fator de Bayes:

$$B = \frac{(p_0 / p_1)}{(\pi_0 / \pi_1)} < \frac{(k_0 / k_1)}{(\pi_0 / \pi_1)}$$

Na terminologia da estatística clássica, isto corresponde ao uso da região crítica:

$$R = \left\{ x : B < \frac{(k_0 / k_1)}{(\pi_0 / \pi_1)} \right\}$$

No caso de duas hipóteses simples $\Theta_0 = \{\theta_0\}$ e $\Theta_1 = \{\theta_1\}$, o teorema de Bayes implica em:

$B = \dfrac{p(x|\theta_0)}{p(x|\theta_1)}$, a razão das verossimilhanças e a região crítica toma a forma:

$R = \left\{ x : \dfrac{p(x|\theta_0)}{p(x|\theta_1)} < \dfrac{(k_0 / k_1)}{(\pi_0 / \pi_1)} \right\}$ que é o teste de razão de verossimilhança sugerido pela teoria de Neyman e Pearson. A diferença é que, na teoria de Neyman e Pearson, o "valor crítico"da região de rejeição é determinado fixando-se α, ou seja, pela probabilidade de x cair na região de rejeição se a hipótese nula for verdadeira.

Ao contrário, na abordagem da teoria de decisão, o valor crítico é fixado em termos da função perda e das probabilidades a priori.

Monica Barros
e-mails: barrosm@alumni.utexas.net
monica@mbarros.com

Referências

Bickel, P. & Doksum, K. (1977) – Mathematical Statistics: Basic Ideas and Selected Topics, Holden Day, Oakland.

Lee, P. M. (1989) – Bayesian Statistics: An Introduction, Oxford University Press, London.

Lindley, D.V. (1970) – Introduction to Probability and Statistics from a Bayesian Viewpoint – part 2 – Inference, Cambridge University Press, London.

Mood, A.M., Graybill, F.A. & Boes, D.C. (1974) – Introduction to the Theory of Statistics, 3rd Edition, McGraw-Hill, New York.

Monica Barros
e-mails: barrosm@alumni.utexas.net
monica@mbarros.com