IND 1115
Inferência Estatística
Aula 11
Novembro 2004

Mônica Barros

Conteúdo

- Método dos Momentos
- Medidas da Qualidade dos Estimadores
- Informação de Fisher
- Limite Inferior de Cramér e Rao

Método dos Momentos

- É uma alternativa ao método de máxima verossimilhança para encontrar estimadores de maneira simples.
- Idéia Geral
- Igualar os momentos da distribuição aos momentos da amostra. Isso é razoável pois a distribuição empírica converge estocasticamente para a função de distribuição F(X).

Método dos Momentos

- Considere uma amostra aleatória da distribuição com densidade $f(x,\theta_1,\theta_2,...,\theta_p)$ onde $(\theta_1,\theta_2,...,\theta_p)\in\Omega$.
- Seja $E(X^k)$ o k-ésimo momento da distribuição (k =1,2,...).
- Seja $M_k=\frac{1}{n}\sum_{i=1}^{n}X_i^k$ o k-ésimo momento amostral
- Faça $E(X^k) = M_k$ para k = 1,2,....
- Faça isto para quantos k's forem necessários até obter soluções únicas para os parâmetros desconhecidos

Método dos Momentos

- Nota
- Uma formulação equivalente do método dos momentos usa os momentos centrais, igualando os momentos centrais da distribuição e da amostra até obter soluções únicas para todos os parâmetros desconhecidos.
- Ou seja, a outra formulação do método dos momentos é igualar:

$$E(X-\mu)^k = m_k = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^k \text{ para } k = 1,2.....$$

- Até encontrar uma solução única. Muitas vezes a resposta pelas duas formulações será igual.

Método dos Momentos

- Na maneira dos casos práticos, os estimadores pelo método dos momentos são consistentes.

- O grande problema com este método é que ele não fornece estimadores únicos para um certo parâmetro.

- Por exemplo, se procuramos estimar a média e variância de uma distribuição é possível encontrar estimadores diferentes pelo método de momentos dependendo de quais momentos amostrais são igualados aos momentos da distribuição.

Medidas da Qualidade dos Estimadores

- Já estudamos dois métodos para encontrar estimadores de um parâmetro desconhecido, o método da máxima verossimilhança e o método dos momentos.

- Estes são os métodos mais comuns para encontrar estimadores.

- Nosso objetivo agora é desenvolver alguns critérios para saber se um dado estimador é "bom" ou "ruim" segundo estes critérios.

Medidas da Qualidade dos Estimadores

- Na seção anterior definimos estimadores não tendenciosos e consistentes, e comentamos que muitas vezes o MLE é um estimador consistente.

- Também definimos o erro quadrático médio, e veremos que esta quantidade pode ser usada como critério para dizer se um estimador qualquer é "bom" ou "ruim".

- Seja T = $T = T(X_1, X_2, \ldots, X_n)$ um estimador de $\tau(\theta)$ onde $\tau(\theta)$ é uma função qualquer de um parâmetro de interesse θ.

Medidas da Qualidade dos Estimadores

- Como comparar um número potencialmente infinito de estimadores tendenciosos e não tendenciosos?
- A comparação entre estimadores será geralmente feita dentro de uma mesma "classe", por exemplo, tentaremos comparar estimadores não tendenciosos, já que o fato de um estimador ser não tendencioso é uma propriedade desejável.

Medidas da Qualidade dos Estimadores

- A seguir definimos o nosso estimador "ideal". A definição, no entanto, não garante a existência deste estimador.
- Este estimador será chamado de UMVUE (Uniformly Minimum Variance Unbiased Estimator, ou Estimador não tendencioso de variância mínima ou melhor estimador não tendencioso).

Medidas da Qualidade dos Estimadores

- Um estimador T* é o UMVUE , ou *melhor estimador não tendencioso* para $\tau(\theta)$ se:
- i) T* é não tendencioso para $\tau(\theta)$.
- ii) $VAR(T^*) \leq VAR(T)$ para qualquer outro estimador não tendencioso T de $\tau(\theta)$.
- Ou seja, o melhor estimador não tendencioso é o "melhor" no sentido de possuir a menor variância dentre os estimadores não tendenciosos.

Medidas da Qualidade dos Estimadores

- Note que o critério apresentado é equivalente à minimização do erro quadrático médio para estimadores não tendenciosos, pois o erro quadrático médio de um estimador não tendencioso é a variância deste estimador.

Medidas da Qualidade dos Estimadores

- **Exemplo 1**
- **Seja T a duração de um componente eletrônico até sua primeira falha. Suponha que T tem distribuição exponencial com média 1/β, ou seja, sua densidade é:**

$f(t) = \beta . e^{-\beta . t}$ para t > 0 e $\beta > 0$.

- **Encontre um estimador não tendencioso da média de T a partir de uma amostra aleatória de tamanho n.**

Exemplo 1

- **Mostre que os estimadores** $\overline{T} = \frac{1}{n}\sum_{i=1}^{n} T_i$ e $T^* = nZ$ **onde** $Z = \min(T_1, T_2 ..., T_n)$ **são não tendenciosos para a média de T, 1/β, e compare as variâncias destes estimadores.**
- **Solução**
- $\overline{T}$ **é claramente não tendencioso, pois:**

$$E(\overline{T}) = \frac{1}{n}\sum_{i=1}^{n} E(T_i) = \frac{1}{n}\sum_{i=1}^{n}(1/\beta) = 1/\beta$$

- **Precisamos descobrir qual a distribuição de Z para podermos calcular sua média e variância.**

Exemplo 1

$$\Pr[Z \le z] = 1 - \Pr[Z > z] = 1 - \Pr[\min(T_1, T_2 ..., T_n) \ge z] = 1 - \{\Pr[T_1 \ge z]\}^n$$

- **pois os T_i 's são independentes e identicamente distribuídos com densidade Exponencial de média 1/β. Mas,**

$$\Pr[T > z] = \int_z^\infty \beta_e^{-\beta t} dt = \beta \frac{e^{-\beta t}}{-\beta}\bigg|_z^\infty = e^{-\beta . z}$$

- **Assim:**

$$\Pr[Z \le z] = 1 - \{\Pr[T_1 \ge z]\}^n = 1 - e^{-\beta . z . n}$$

- **Mas, esta é a função de distribuição de uma variável aleatória Exponencial com média 1/nβ. Ou seja, a densidade de** $Z = \min(T_1, T_2 ..., T_n)$ **é:**

Exemplo 1

$$g(z) = \frac{1}{\beta n} e^{-n\beta z}, \beta > 0, \; z > 0$$

- **E portanto a média e variância de Z são, respectivamente, 1/nβ e 1/(nβ)².**
- **Daí:** $E(T^*) = E(nZ) = nE(Z) = \frac{1}{\beta} = E(T)$

e T^* também é não tendencioso para E(T).

- **A comparação das variâncias dos estimadores nos dá:**

Exemplo 1

- $VAR(\overline{T}) = VAR(T_i)/n = 1/n\beta^2$ e

$VAR(T^*) = n^2 VAR(Z) = 1/\beta^2 > VAR(\overline{T})$

- para todo n>1. Assim, é preferível usar a média amostral como estimador, pois sua variância é menor que a de T^* sempre que a amostra tem tamanho maior que 1.

Medidas da Qualidade dos Estimadores

- Nota
- O fato de T ser não tendencioso para $\tau(\theta)$ ***não implica*** em g(T) não tendencioso para $g(\tau(\theta))$ quando g(.) é uma função ***não linear***.

- Às vezes quando procuramos um estimador para $\tau(\theta)$ encontramos T tal que $E(T) = c.\tau(\theta)$ onde c é uma constante.
- Obviamente então: $E(T/c) = \tau(\theta)$ e T/c é um estimador não tendencioso para $\tau(\theta)$.
- No entanto, se T é "quase" não tendencioso para $\tau(\theta)$, dividir por c pode "deteriorar" o erro quadrático médio.

Medidas da Qualidade dos Estimadores

- Exemplo 2
- Sejam $X_1, X_2, \ldots, X_n$ uma amostra da densidade Normal(μ, σ^2).Considere o MLE para a variância:

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2 = S^2 = \text{variância amostral}$$

- Já mostramos antes que $E(S^2) = ((n-1)/n)\,\sigma^2$ e assim S^2 é tendencioso.

- A tendência ("BIAS") de S^2 é:

$$E(S^2 - \sigma^2) = -\frac{1}{n}\sigma^2$$

Exemplo 2

- O erro quadrático médio de S^2 é:

$$MSE(S^2) = VAR(S^2) + [BIAS\,(S^2)]^2 =$$

$$= \frac{2(n-1)\sigma^2}{n^2} + \left(-\frac{\sigma^2}{n}\right)^2 = \frac{(2n-1)\sigma^4}{n^2}$$

- Agora considere o estimador:

$$S^{*2} = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2 = \frac{nS^2}{n-1}$$

- Note que S^{*2} é um estimador não tendencioso para a variância da Normal.

Exemplo 2

- Mas:

$$MSE(S^{*2}) = \mathrm{VAR}(S^{*2}) = \frac{n^2}{(n-1)^2}\,\mathrm{VAR}(S^{*2}) = \frac{2}{n-1}\sigma^4$$

- que é maior que o MSE de S^2 (estimador tendencioso) para qualquer inteiro positivo n.
- Assim, embora S^{*2} seja não tendencioso, se a população é Normal, S^{*2} é pior que S^2 no sentido de ter maior MSE.

Informação de Fisher

- Sejam $X_1, X_2, \ldots, X_n$ variáveis aleatórias com verossimilhança $L(\theta)$ e log-verossimilhança $l(\theta)$.
- Definição (Função Score)

$$S(X,\theta) = \frac{\partial}{\partial\theta}\left[\log L(\theta)\right] = \frac{\partial \ell(\theta)}{\partial\theta}$$

- Por esta definição, vemos que o estimador de máxima verossimilhança é geralmente obtido fazendo-se o score igual a zero.

Informação de Fisher

- Definição (Informação de Fisher)

A informação de Fisher é definida como:

$I(\theta) = \mathrm{VAR}[S(\boldsymbol{X},\theta)]$

- isto é, a Informação é a variância da função score.
- Veremos depois que a idéia de informação de Fisher corresponde intuitivamente à nossa idéia de "informação" contida numa amostra.

Exemplo 3

- Sejam $X_1, X_2, \ldots, X_n$ uma amostra aleatória da distribuição Poisson(θ).
- Calcule a função score e a informação.
- Solução
- A verossimilhança é:

$$L(\theta) = \prod_{i=1}^{n} \frac{\theta^{x_i}\, e^{-\theta}}{x_i!} = \frac{\theta^{\sum x_i}\, e^{-n\theta}}{\prod_{i=1}^{n} x_i!}$$

- A log verossimilhança é:

$$\ell(\theta) = \log L(\theta) = \left(\sum_{i=1}^{n} x_i\right)\log\theta - n\theta - \log\left[\prod_{i=1}^{n} x_i!\right]$$

Exemplo 3

- A função score é:

$$S(\underset{\sim}{X},\theta) = \frac{d\ell(\theta)}{d\theta} = \frac{n\overline{X}}{\theta} - n$$

- Note que VAR(X_i) = θ (i = 1,2,...,n), e assim VAR($\overline{X}$) = θ/n. Logo, a informação de Fisher é:

$$I(\theta) = VAR(S(\underset{\sim}{X},\theta)) = VAR\left[\frac{n\overline{x}}{\theta} - n\right] =$$

$$= VAR\left(\frac{n\overline{x}}{\theta}\right) = \frac{n^2}{\theta^2} VAR(\overline{X}) = \frac{n^2}{\theta^2}\left(\frac{\theta}{n}\right) = \frac{n}{\theta}$$

- Note que a informação é o inverso da variância de $\overline{X}$

Informação de Fisher

- De alguma maneira, a informação de Fisher mede a "precisão" da amostra. Quanto maior a variância de $\overline{X}$, menor a "precisão" (e menor a informação), e vice-versa.

- A informação de Fisher é um dos elementos fundamentais da Estatística Clássica e Bayesiana.

Informação de Fisher

- A idéia de que a informação mede a "precisão" de uma amostra é empregada em problemas bem mais complicados que os já vistos aqui.

- É importante notar que a informação de Fisher depende apenas da distribuição dos dados, e não de qualquer valor dos X's da amostra. Isto pode ser comprovado algebricamente se notarmos que a informação é a variância do Score, que, por sua vez, é função dos X's e de θ. Ao calcularmos esta variância encontramos uma constante que só depende dos parâmetros da distribuição.

Informação de Fisher

- Propriedades do Score e da Informação de Fisher

$$E[S(\underset{\sim}{X},\theta)] = 0$$

$$I(\theta) = E\left\{[S(\underset{\sim}{X},\theta)]^2\right\} = E\left[-\partial S(\underset{\sim}{X},\theta)/\partial\theta\right] = E\left[-\partial^2\ell(\theta)/\partial\theta^2\right]$$

- Se $X_1, X_2, \ldots, X_n$ é um conjunto de observações independentes (mas não necessariamente identicamente distribuídas) e $I_i(\theta)$ é a informação em X_i então: $I(\theta) = \sum_{i=1}^{n} I_i(\theta).$

Informação de Fisher

- Ou seja, a informação total contida nos dados é a soma das informações trazidas por cada uma das observações.
- Se, além de independentes, os X_i's são identicamente distribuídos, então: $I(\theta) = n.I_1(\theta)$
- Se os X_i's são identicamente distribuídos, todos os I_i 's são iguais e então a informação contida nas n observações é $I(\theta) = n.I_1(\theta)$ onde $I_1(\theta)$ é a contribuição de X_1 na verossimilhança. Isto é, a informação numa amostra aleatória é a soma das informações de cada elemento da amostra.

Informação de Fisher

- Este ítem é de fundamental importância na prática, pois nos fornece duas maneiras de calcular a informação de Fisher.
- A primeira maneira decorre da definição de informação, ou seja, como a variância do Score. A outra maneira envolve o cálculo da segunda derivada da log-verossimilhança, e do valor esperado desta quantidade. É claro que o resultado fornecido deve ser o mesmo, independente da fórmula usada para obter a informação.

O Limite Inferior de Cramér e Rao

- O próximo teorema (Cramér e Rao) fornece um limite inferior para a variância de qualquer estimador não tendencioso de $\tau(\theta)$.
- Este limite é importante, pois nos permite dizer que a variância de qualquer estimador não tendencioso é igual ou superior a este limite.
- Em particular, se T é um estimador não tendencioso de $\tau(\theta)$ cuja variância coincide com este limite, T é o *melhor* estimador não tendencioso possível para $\tau(\theta)$ pois nenhum outro estimador não tendencioso de poderá ter variância menor que a de T.

O Limite Inferior de Cramér e Rao

- O teorema de Cramér e Rao nos fornece o valor mínimo para a variância de um estimador não tendencioso de $\tau(\theta)$.
- Logo, se for possível encontrar um estimador não tendencioso de $\tau(\theta)$ e este tiver a variância indicada no próximo teorema, o estimador encontrado é o UMVUE (melhor estimador não tendencioso) para $\tau(\theta)$.

O Limite Inferior de Cramér e Rao

- Teorema (Desigualdade da informação ou desigualdade de Cramer e Rao)
- Seja T um estimador *não tendencioso* de $\tau(\theta)$. Então, sob certas condições de regularidade:

$$VAR(T) \geq \frac{[\tau'(\theta)]^2}{I(\theta)} = \frac{[\tau'(\theta)]^2}{n.E\left\{\left(\frac{\partial \log f(x,\theta)}{\partial\theta}\right)^2\right\}}$$

- e a igualdade ocorre se, e somente se, a verossimilhança tem a forma:

$$L(\theta) = \exp[A(\theta)T(X) + K(\theta) + u(X)]$$

O Limite Inferior de Cramér e Rao

- Esta desigualdade é conhecida como *Desigualdade de Cramér e Rao*, e o limite inferior para a variância de T é chamado de limite inferior de Cramér e Rao (CRLB = Cramér Rao Lower Bound).
- O termo $(\tau'(\theta))^2 / I(\theta)$ é chamado de *Limite Inferior de Cramer e Rao*. O grande problema com este limite é que, em muitos casos, ele é muito pequeno.
- Exceto em casos "triviais", não é possível encontrar estimadores não tendenciosos cuja variância seja igual ao limite inferior de Cramer e Rao.

O Limite Inferior de Cramér e Rao

- Definição (Estimador Eficiente)
- T é chamado um *estimador eficiente* de $\tau(\theta)$ se:
- I) T é não tendencioso para $\tau(\theta)$.
- II) $VAR[T(\underline{x})] = [\tau'(\theta)]^2 / I(\theta)$

  isto é, o estimador T atinge o limite inferior de Cramér e Rao.

- Corolário
- Se é um estimador eficiente de $\tau(\theta)$ então é o melhor estimador não tendencioso de $\tau(\theta)$.

O Limite Inferior de Cramér e Rao

- **Exemplo 4**
- Sejam $X_1, X_2, \ldots, X_n \overset{iid}{\sim} N(0,\theta)$
- Encontre o MLE de θ. Calcule o score e a informação de Fisher e mostre que o MLE é um estimador eficiente de θ.
- Solução
- Note que θ, a variância de distribuição, é o único parâmetro desconhecido aqui.
- A verossimilhança é:

$$L(\theta) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\theta}} \exp\left\{-\frac{1}{2\theta}(x_i^2)\right\} = (2\pi\theta)^{-n/2} \exp\left\{-\frac{1}{2\theta}\sum_{i=1}^{n} x_i^2\right\}$$

Exemplo 4

- A log-verossimilhança é: $\ell(\theta) = -\frac{n}{2}\log(2\pi\theta) - \frac{\sum x_i^2}{2\theta}$
- O Score é:

$$S(\underset{\sim}{x},\theta) = \frac{d\ell(\theta)}{d\theta} = -\frac{n}{2}\frac{(2\pi)}{2\pi\theta} - \frac{\sum x_i^2}{2}\left(-\frac{1}{\theta^2}\right)$$

- O estimador de máxima verossimilhança de θ é encontrado fazendo-se o Score igual a zero. Daí:

$$\frac{d\ell}{d\theta} = 0 \Leftrightarrow \frac{\sum x_i^2}{\theta} = n \;\Leftrightarrow\; \hat{\theta} = \frac{1}{n}\sum_{i=1}^{n} x_i^2 = MLE \text{ para } \theta$$

- A informação de Fisher é:

Exemplo 4

$$I(\theta) = E\left[-\frac{d^2\ell(\theta)}{d\theta^2}\right] = -E\left[\frac{d}{d\theta}\left\{\frac{(\sum x_i^2}{2\theta^2} - \frac{n}{2\theta}\right\}\right]$$

$$= -E\left[\frac{\sum x_i^2}{2}\left(-\frac{2}{\theta^2}\right) - \frac{n}{2}\left(-\frac{1}{\theta^2}\right)\right] = E\left\{\frac{\sum x_i^2}{\theta^2} - \frac{n}{2\theta^2}\right\}$$

$$= \frac{1}{\theta^2}\sum_{i=1}^{n} E(x_i^2) - \frac{n}{2\theta_2} = \frac{1}{\theta^3}\sum_{i=1}^{n}\theta - \frac{n}{2\theta^2} = \frac{n}{\theta^2} - \frac{n}{2\theta^2}$$

$$\frac{2n}{2\theta^2} - \frac{n}{2\theta^2} \Rightarrow I(\theta) = \frac{n}{2\theta^2}$$

Exemplo 4

- O limite inferior de Cramér e Rao para um estimador de θ é: 1/I(θ) = 2θ²/n.
- Logo, se um estimador não tendencioso de θ tem como variância este valor, este estimador é eficiente e é o melhor estimador não tendencioso de θ.
- Considere o estimador de máxima verossimilhança. Para provar que ele é eficiente, precisamos verificar se ele é não tendencioso e que sua variância é 1/I(θ) = 2θ²/n.

Exemplo 4

- Mas, o MLE de θ é: $\hat{\theta} = \sum x_i^2 / n$

$$E(\hat{\theta}) = E\left(\frac{1}{n}\sum x_i^2\right) = \frac{1}{n}\sum_{i=1}^{n} E(x_i^2) = \frac{1}{n}\sum_{i=1}^{n}\theta = \frac{n\theta}{n} = \theta$$

- e então o MLE é não tendencioso para θ.

$$VAR(\hat{\theta}) = VAR\left[\frac{1}{n}\sum x_i^2\right] = \frac{1}{n^2}\sum_{i=1}^{n} VAR(x_i^2) = \frac{1}{n^2}\sum_{i=1}^{n} 2\theta^2$$

$$= \frac{2n\theta^2}{n^2} = \frac{2\theta^2}{n} = \frac{1}{I(\theta)}, \text{ ou seja, } \hat{\theta} \text{ é } \textbf{eficiente.}$$