

MAE5911/IME: Fundamentos de Estatística e Machine Learning. Prof.: Alexandre Galvão Patriota

Questão 01: Seja (Z_1, \dots, Z_n) uma amostra aleatória de $Z \sim \text{Ber}(\theta)$, $\theta \in (0, 1)$.

- (a) Encontre o EMV para $g(\theta) = P_\theta(Z = 0)$.
- (b) Encontre o EMV para $g(\theta) = \text{Var}_\theta(Z)$.
- (c) Considere que os dados foram observados $(0, 0, 1, 0, 0, 1)$. Encontre as estimativas de MV nos itens acima.
- (d) Construa o valor-p para a hipótese $H: \theta = 0.1$ usando os dados do item anterior.

Resposta:

O estimador via função de máxima verossimilhança foi proposto por Fisher e comprovado ser superior a outros métodos estimadores, como o método de momentos. Este estimador é utilizado sobretudo quando a informação da distribuição de probabilidade do modelo é conhecida. Ela tem propriedades como eficiência assintótica, consistência, distribuição normal e invariância [Notas de aula pg. 60]. O estimador de máxima verossimilhança representa o valor do parâmetro θ que torna os dados observados o mais plausível possível.

Devido a propriedade de invariância do EMV, basta encontrarmos EMV para a letra (a) e depois calcular a letra (b) utilizando este estimador como parâmetro.

Seja (Z_1, \dots, Z_n) a amostra aleatória de

$$Z_i \sim \text{Bernoulli}(\theta), \quad \theta \in (0, 1).$$

A função de verossimilhança, que é a função de probabilidade do modelo, reinter é dada por

$$L(\theta; z_1, \dots, z_n) = \prod_{i=1}^n \theta^{z_i} (1 - \theta)^{1-z_i}.$$

Para encontrar o estimador de máxima verossimilhança, precisamos maximizar $L(\theta)$ em relação a θ . Mas como esta função se trata de um produto, é mais fácil maximizar a função log-verossimilhança, que por ser monotonicamente crescente, preserva o valor de θ do ponto máximo $L(\theta)$. A log-verossimilhança correspondente é

$$\ell(\theta) = \sum_{i=1}^n [z_i \log \theta + (1 - z_i) \log(1 - \theta)].$$

Derivando em relação a θ e igualando a zero:

$$\frac{\partial \ell(\theta)}{\partial \theta} = \frac{\sum z_i}{\theta} - \frac{n - \sum z_i}{1 - \theta} = 0,$$

obtemos o estimador de máxima verossimilhança para a probabilidade de sucesso θ :

$$\hat{\theta}_{MV} = \bar{Z} = \frac{1}{n} \sum_{i=1}^n Z_i.$$

O teste de hipótese, mede o quão improvável seria observar $\hat{\theta}_{MV}$ se a hipótese nula fosse verdadeira. O valor-p é esta medida de improbabilidade.

Em condições de regularidade, o valor- p segue uma distribuição uniforme em $(0, 1)$. Isso significa que, se o seu valor é muito pequeno, por exemplo $p \leq 0,01$, então o resultado observado é um evento raro sob a hipótese nula H_0 , ou, alternativamente, que a hipótese nula está incorreta. Nessa situação, existem evidências para rejeitar H_0 .

(a) $g(\theta) = P_\theta(Z = 0)$

Pela **propriedade de invariância do EMV**, temos:

$$\hat{g}_{MV} = g(\hat{\theta}_{MV}) = 1 - \bar{Z}.$$

(b) $g(\theta) = \text{Var}_\theta(Z)$

Sabemos que $\text{Var}_\theta(Z) = \theta(1 - \theta)$. Logo, aplicando novamente a invariância:

$$\hat{g}_{MV} = g(\hat{\theta}_{MV}) = \bar{Z}(1 - \bar{Z}).$$

(c) Sejam os dados observados: $(0, 0, 1, 0, 0, 1)$

Temos $n = 6$ e $\sum z_i = 2$, portanto

$$\hat{\theta}_{MV} = \bar{Z} = \frac{2}{6} = \frac{1}{3} \approx 0.3333.$$

Assim:

$$P(\widehat{Z} = 0) = 1 - \hat{\theta}_{MV} = \frac{2}{3} \approx 0.6667,$$

$$\widehat{\text{Var}}(Z) = \hat{\theta}_{MV}(1 - \hat{\theta}_{MV}) = \frac{1}{3} \cdot \frac{2}{3} = \frac{2}{9} \approx 0.2222.$$

(d) Teste de hipótese $H_0 : \theta = 0.1$

Queremos verificar se temos evidência a partir dos dados observados para rejeitar a hipótese nula de que a probabilidade de sucesso é $\theta = 0.1$.

Sabemos que, sob H_0 , a estatística

$$X = \sum_{i=1}^n Z_i$$

segue uma distribuição Binomial(n, θ).

Com $n = 6$ e $x_{\text{obs}} = 2$ sucessos observados, temos:

$$X \sim \text{Binomial}(6, 0.1).$$

O valor-p é a probabilidade de observar um valor da estatística $T(Z_n)$ *tão ou mais extremo quanto o observado*, assumindo que H_0 é verdadeira:

$$p = P_{H_0}(T(Z_n) \geq T(z_n)).$$

No contexto binomial, a estatística de teste natural é o número de sucessos X , e portanto:

$$p = P_{H_0}(X \geq 2).$$

Calculando o valor-p:

$$P(X = 0) = (0.9)^6 = 0.531441, \quad P(X = 1) = \binom{6}{1}(0.1)(0.9)^5 = 0.354294.$$

Logo:

$$p = 1 - [P(X = 0) + P(X = 1)] = 1 - (0.531441 + 0.354294) = 0.114265.$$

$$p = 0.114265$$

Este resultado **não é improvável** sob H_0 (<0.05). Portanto, não há evidência para rejeitar H_0 , isto é, não evidências de que $\theta = 0.1$ seja incompatível com os dados.

Conclusão: Não há evidências suficientes para rejeitar $H_0 : \theta = 0.1$.

Questão 02: Seja (Z_1, \dots, Z_n) uma amostra aleatória de $Z \sim \text{Exp}(\theta)$, $\theta \in (0, \infty)$.

- (a) Encontre o EMV para $g(\theta) = P_\theta(Z > 1)$.
- (b) Encontre o EMV para $g(\theta) = P_\theta(0.1 < Z < 1)$.
- (c) Encontre o EMV para $g(\theta) = \text{Var}_\theta(Z)$.
- (d) Considere que os dados foram observados (0.2, 0.6, 0.3, 0.2, 0.8, 0.12). Encontre um IC aproximado de 95% de confiança para $g(\theta)$ nos itens acima.
- (e) Faça uma simulação de Monte Carlo para verificar se os IC's aproximados obtidos no passo anterior têm cobertura próxima do nível de confiança estabelecido. Caso não tenham, proponha um tamanho amostral que produza IC's mais confiáveis para cada caso.

Resposta:

Questão 03: Seja (Z_1, \dots, Z_n) uma amostra aleatória de $Z \sim N(\mu, \sigma^2)$, $\theta = (\mu, \sigma^2) \in (-\infty, \infty) \times (0, \infty)$.

- (a) Encontre o EMV para $g(\theta) = E_\theta(Z)$.
- (b) Encontre o EMV para $g(\theta) = P_\theta(Z < 2)$.
- (c) Encontre o EMV para $g(\theta) = P_\theta(2.6 < Z < 4)$.
- (d) Encontre o EMV para $g(\theta) = \text{Var}_\theta(Z)$.
- (e) Considere que os dados foram observados (2.4, 2.7, 2.3, 2, 2.5, 2.6). Encontre as estimativas de MV nos itens acima.

Resposta:

Questão 04: Seja (Z_1, \dots, Z_n) uma amostra aleatória de $Z \sim f_\theta$, $\theta \in (0, \infty)$, tal que a função densidade de probabilidade é dada por

$$f_\theta(x) = \theta x^{\theta-1}, \quad x \in (0, 1),$$

e $f_\theta(x) = 0$, caso contrário.

- (a) Encontre o EMV para $g(\theta) = E_\theta(Z)$.
- (b) Encontre o EMV para $g(\theta) = P_\theta(Z > 0.3)$.
- (c) Encontre o EMV para $g(\theta) = P_\theta(0 < Z < 0.1)$.
- (d) Encontre o EMV para $g(\theta) = \text{Var}_\theta(Z)$.
- (e) Considere que os dados foram observados (0.12, 0.50, 0.20, 0.23, 0.30, 0.11). Encontre as estimativas de MV para os itens acima.

Resposta:

