Estatísticas Suficientes

Como resumir a amostra sem perder nenhuma informação sobre o parâmetro?

ESTAT0078 – Inferência I
Prof. Dr. Sadraque E. F. Lucena
http://sadraquelucena.github.io/inferencia1

UNIVERSIDADE FEDERAL DE SERGIPE

Definição 9.1: Estatística Suficiente

Dizemos que a estatística $T = T(X_1, \ldots, X_n)$ é suficiente para θ, quando a distribuição condicional de $X_1, \ldots, X_n$ dado T for independente de θ.

- **Intuição:** Se a amostra não traz nada novo sobre o parâmetro θ depois que sabemos T, é porque T já absorveu toda a informação sobre θ contida na amostra.
- **Exemplo intuitivo:** Desempenho do Aluno no Semestre ao longo do período.
  - Suficiente $\longrightarrow$ A soma (ou a média) das 3 notas
  - Não Suficiente $\longrightarrow$ Guardar apenas a maior nota (joga fora a informação sobre a consistência e o desempenho global do aluno)

Verificação via Distribuição Condicional

Para verificarmos se T é uma estatística suficiente para θ no caso discreto:

1. Calculamos a probabilidade condicional:

$$P(X_1 = x_1, \ldots, X_n = x_n \mid T = t) = \frac{P(X_1 = x_1, \ldots, X_n = x_n)}{P(T = t)}$$

2. Se a expressão resultante **não envolver** θ, conclui-se que T é suficiente para θ.

Exemplo 9.1

Seja $X_1, \ldots, X_n$ uma amostra aleatória da distribuição de Bernoulli (θ). Verifique se $T = \sum_{i=1}^{n} X_i$ é suficiente para θ.

Lembretes:

1. $X \sim \text{Bernoulli}(p)$: $P(X = x) = p^x (1-p)^{1-x}, x = 0, 1.$
2. A soma de n v.a. de Bernoulli(p) tem distribuição Binomial(n, p).
3. $Y \sim \text{Binomial}(n, p)$: *[Math Processing Error]*, $y = 0, 1, 2, \ldots, n.$

Exemplo 9.2

Considere o exemplo anterior, com $n = 3$ e $T = X_1 + 2X_2 + X_3$. Verifique que T não é suficiente para θ.

- **Dica:** Considere $X_1 = 1, X_2 = 0$ e $X_3 = 1$.

Exemplo 9.3

Seja $X_1, \ldots, X_n$ uma amostra aleatória da distribuição de Poisson com parâmetro θ. Verifique se $T = \sum_{i=1}^{n} X_i$ é suficiente para θ.

Lembrete:

- $X \sim$ Poisson(θ): $P(X = x) = \frac{e^{-\theta}\theta^x}{x!}, x = 0, 1, 2, \ldots$

Próximo Passo: Teorema da Fatoração

E se a distribuição for **contínua** (ex: Normal, Exponencial)?

- O cálculo da densidade condicional $f_{\mathbf{X}|T}(\mathbf{x}|t)$ pode ser extremamente trabalhoso ou algebricamente complexo.
- **Na próxima aula:** Veremos o **Teorema da Fatoração de Neyman-Fisher**, que nos permite identificar estatísticas suficientes por simples inspeção da função de verossimilhança!

Fim