Probabilidade II

Departamento de Estatística

Universidade Federal da Paraíba

Lei dos Grandes Números

Lei dos Grandes Números

Certos procedimentos estatísticos são válidos quando o tamanho da amostra é grande. Deste modo, é importante o estudo das distribuições de variáveis aleatórias definidas para amostras grandes.

Introduziremos noção de convergência para sequências de variáveis aleatórias e apresentaremos a lei dos grandes números e o teorema do limite central.

Definição 20.1: A sequência de variáveis aleatórias X_n converge em probabilidade para a variável aleatória X se para todo $\epsilon > 0$ tem-se

$$lim_{n \to \infty} P[|X_n - X| \geq \epsilon] = 0.$$

Indicaremos que X_n converge em probbilidade para X com a notação $X_n \to^p X$.

Assim, $X_n \to^p X$ significa que se n for suficientemente grande, então a probabilidade que X_n difira de X por mais que ϵ é muito pequena.

Lei dos Grandes Números

Lei dos Grandes Números

Definição 20.2: (Lei Fraca dos Grandes Números) Sejam $X_1, X_2, \ldots$ uma sequência de variáveis aleatórias independentes e identicamente distribuídas com média μ e variância σ^2. Seja $S_n = \sum_{i=1}^{n} X_i$. Tem-se

$$\frac{S_n}{n} \rightarrow^{p} \mu.$$

DEMONSTRAÇÃO:

Lei dos Grandes Números

Lei dos Grandes Números

Um caso particular importante é aquele em que $X_I = 1$ (sucesso) com probabilidade p e $X_i = 0$ (fracasso) com probabilidade $1 - p$, para $i = 1, 2, \ldots$, que corresponde a uma sequência de ensaios de bernoulli.

$\frac{S_n}{n}$ representa a frequência relativa de sucessos em n ensaios. Temos então que $\mu = p$ e $\sigma^2 = p(1-p)$. Assim

$$P\left[|\frac{S_n}{n} - p| \geq \epsilon \right] \leq \frac{p(1-p}{n\epsilon^2}.$$

A expressão do lado direito tende a zero para $n \longrightarrow \infty$.

A lei fraca diz que $\frac{S_n}{n}$, a frequência relativa de sucessos, converge em probabilidade para p.

Teorema Central do Limite

TEOREMA CENTRAL DO LIMITE

Se uma variável aleatória X puder ser representada pela soma de quaisquer n variáveis aleatórias independentes, que satisfaçam certas condições gerais, então esta soma, para n suficientemente grande, terá distribuição aproximadamente normal.

Teorema: (Teorema Central do Limite - variáveis aleatórias i.i.d.) Seja $X_1, X_2, \ldots, X_n$ uma sequência de n variáveis aleatórias independentes e identicamente distribuídas (i.i.d.) com $E(X_i) = \mu$ e $Var(X_i) = \sigma^2$. Então, para $S_n = \sum_{i=1}^{n} X_i$, tem-se

$$Z_n = \frac{S_n - E(S_n)}{\sqrt{Var(S_n)}} = \frac{S_n - n\mu}{\sigma\sqrt{n}}$$

tem uma distribuição aproximada **N(0,1)** na medida em que n se aproxima do infinito. Se F_n é a função de distribuição de Z_n, então

$$\lim_{n\to\infty} \frac{F_n(z)}{\Phi(z)} = 1, \qquad \text{para todo z.}$$

Distribuição Normal

TEOREMA CENTRAL DO LIMITE

O fato de S_n ser aproximadamente normalmente distribuída quando os termos X_i podem ter qualquer distribuição é a razão básica para a importância da distribuição normal.

Em numerosas aplicações, a variável aleatória considerada pod ser representada como a soma de n variáveis aleatórias independentes, algumas das quais podem se dever a erros de medidas, algumas se devem a considerações físicas, entre outras, de modo que a distribuição normal fornece uma boa aproximação.

Distribuição Normal

EXEMPLO 1: Peças são embaladas em engradados com capacidade para 250 peças. Os pesos das peças são variáveis aleatórias independentes, com uma média de 0.5 Kg e um desvio padrão de 0.10 Kg. Vinte engradados são carregados para uma bandeja. Qual a probabilidade de que as peças na bandeja excederão 2510 Kg? (Despreze os pesos da bandeja e do engradado).

$$Y = X_1 + X_2 + \ldots + X_{5000}$$

$$E(Y) = E(X_1 + X_2 + \ldots + X_{5000}) = 5000 \times 0.5 = 2500$$
$$Var(Y) = Var(X_1 + X_2 + \ldots + X_{5000}) = 5000 \times 0.01 = 50$$

$$\begin{aligned} P(Y > 2510) &= P\left(\frac{Y - E(Y)}{\sqrt{Var(Y)}} > \frac{2510 - E(Y)}{\sqrt{Var(Y)}}\right) \\ &= P\left(Z > \frac{2510 - 2500}{\sqrt{50}}\right) = 1 - \Phi(1.41) = 0.08. \end{aligned}$$

Distribuição Normal

EXEMPLO 2: Suponha que temos n voltagens (V_i) de ruídos independentes que são recebidas por um somador. Suponha que cada V.A. V_i seja uniformemente distribuída sobre o intervalo $[0,10]$. Para $n=20$, qual a probabilidade de que a voltagem total exceda 105 volts?

Distribuição Normal

Aproximação Normal para a Distribuição Binomial:

- Uma consequência do Teorema Central do Limite é a aproximição de cálculos de probabilidade da Binomial pela distribuição Normal.
- Sendo $X \sim B(n, p)$, desejamos calcular $P(a \leq X \leq b)$, com a e b inteiros e $0 \leq a$, $b \leq n$.
- Para n suficientemente grande, a aproximação será feita através da variável $Y \sim N(\mu = np, \sigma^2 = np(1-p))$.
- Em geral, essa aproximação é aceitável sempre que $np \geq 5$ e $np(1-p) \geq 5$ (alguns autores consideram o limite igual a 10).

EXEMPLO 3: Suponha que 25% de todos os motoristas habilitados da PB não possuam seguro. Em uma amostra de 50 motoristas, qual a probabilidade de no máximo 5 terem seguro? (o valor exato dessa probabilidade é 0.007046)