NOÇÕES DE INFERÊNCIA ESTATÍSTICA

2014

Problemas de inferência

Inferir significa fazer afirmações sobre algo desconhecido.

A inferência estatística tem como objetivo fazer afirmações sobre uma característica de uma população a partir do conhecimento de dados de uma parte desta população (isto é, uma amostra de n observações).

A população é representada por uma distribuição de probabilidade com parâmetro(s) cujo(s) valor(es) é (são) desconhecido(s).

Fazemos inferências sobre o(s) parâmetro(s).

Problemas de inferência

Se θ é um parâmetro da distribuição de uma v. a. X e $X_1,...,X_n$ é uma amostra desta distribuição, encontramos três problemas típicos:

1. Estimação pontual

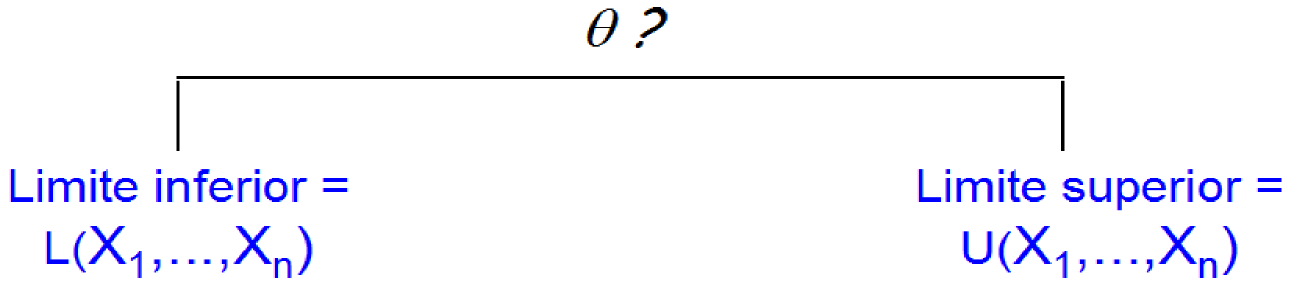
Apresentar um valor para θ, que é uma função da amostra $X_1,...,X_n$ ("cálculo" de θ), chamada de estimador de θ.

Espera-se que o estimador tenha boas propriedades: (i) em média esteja próximo de θ, (ii) o estimador se aproxima de θ quando n aumenta, ...b

Problemas de inferência

2. Estimação intervalar

Apresentar um intervalo de possíveis valores para θ, chamado de intervalo de confiança. Os limites do intervalo são funções da amostra $X_1,...,X_n$ (são aleatórios).



A probabilidade de que o intervalo contenha θ deve ser alta.vv

A amplitude do intervalo deve ser tão pequena quanto possível (intervalo mais preciso).

Problemas de inferência

3. Teste de hipóteses

Uma hipótese estatística (H) é uma afirmação sobre o valor de θ. Pode ser verdadeira ou falsa.

Se θ é a probabilidade de sucesso no modelo binomial, H: $\theta = ½$, H: $\theta \neq ½$ e H: $\theta > ¾$ são exemplos de hipóteses.

Com base na amostra $X_1,...,X_n$, formulamos uma regra de decisão que permita concluir pela rejeição ou não rejeição (aceitação) de H. A decisão pode ser correta ou errada.

Estimação pontual – método de substituição

(a). Distribuição binomial. X ~ B(n, p). Vimos que E(X) = np.

$$\text{Um estimador para } p: \; \overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i = \text{proporção amostral de sucessos.}$$

(b). Distribuição de Poisson. X ~ Po(μ). Vimos que E(X) = μ.

$$\text{Um estimador para } \mu: \; \overline{X}.$$

(c). Distribuição exponencial. X ~ Ex(λ). Vimos que E(X) = 1 / λ.

$$\text{Um estimador para } \lambda: \; = \frac{1}{\overline{X}}.$$

(d). Distribuição normal. X ~ N(μ, σ^2). Vimos que E(X) = μ e Var(X) = σ^2.

$$\text{Um estimador para } \mu: \; \overline{X}. \qquad \text{Um estimador para } \sigma^2: \; s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2.$$

Obs. Existem outros métodos de estimação.

Teorema central do limite

Se X_1, X_2, ..., X_n é uma amostra aleatória de tamanho n de uma distribuição com média μ e desvio padrão σ ($0 < \sigma < \infty$), então a distribuição aproximada de

$$Z = \frac{\sqrt{n}(\overline{X} - \mu)}{\sigma}$$ é normal padrão N(0,1),

sendo que $\overline{X} = (X_1 + X_2 + \cdots + X_n)/n$ é a média amostral.

Observações.

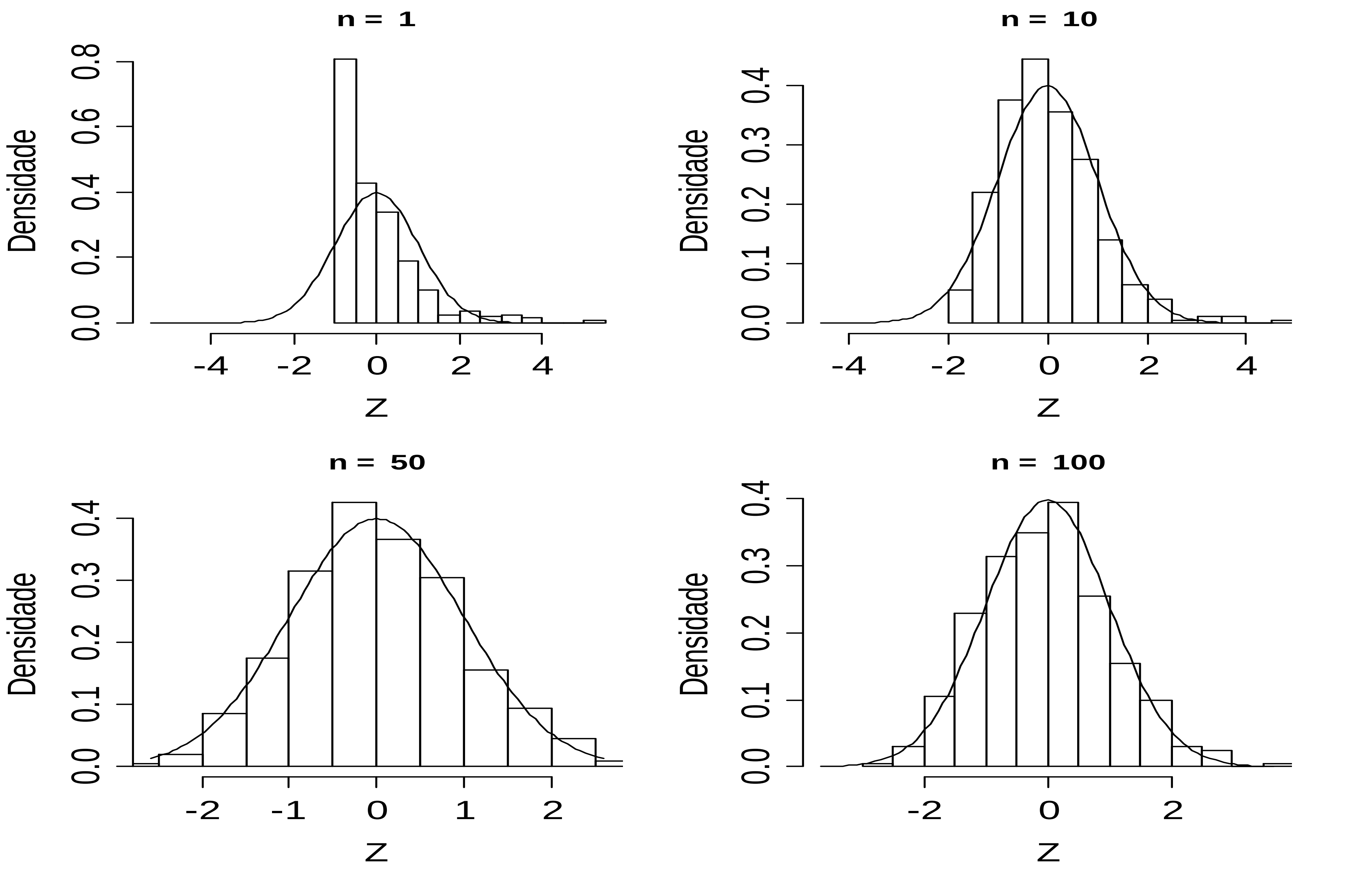
(1) Quanto maior n, melhor a aproximação.

(2) A distribuição das variáveis X pode ser discreta ou contínua.

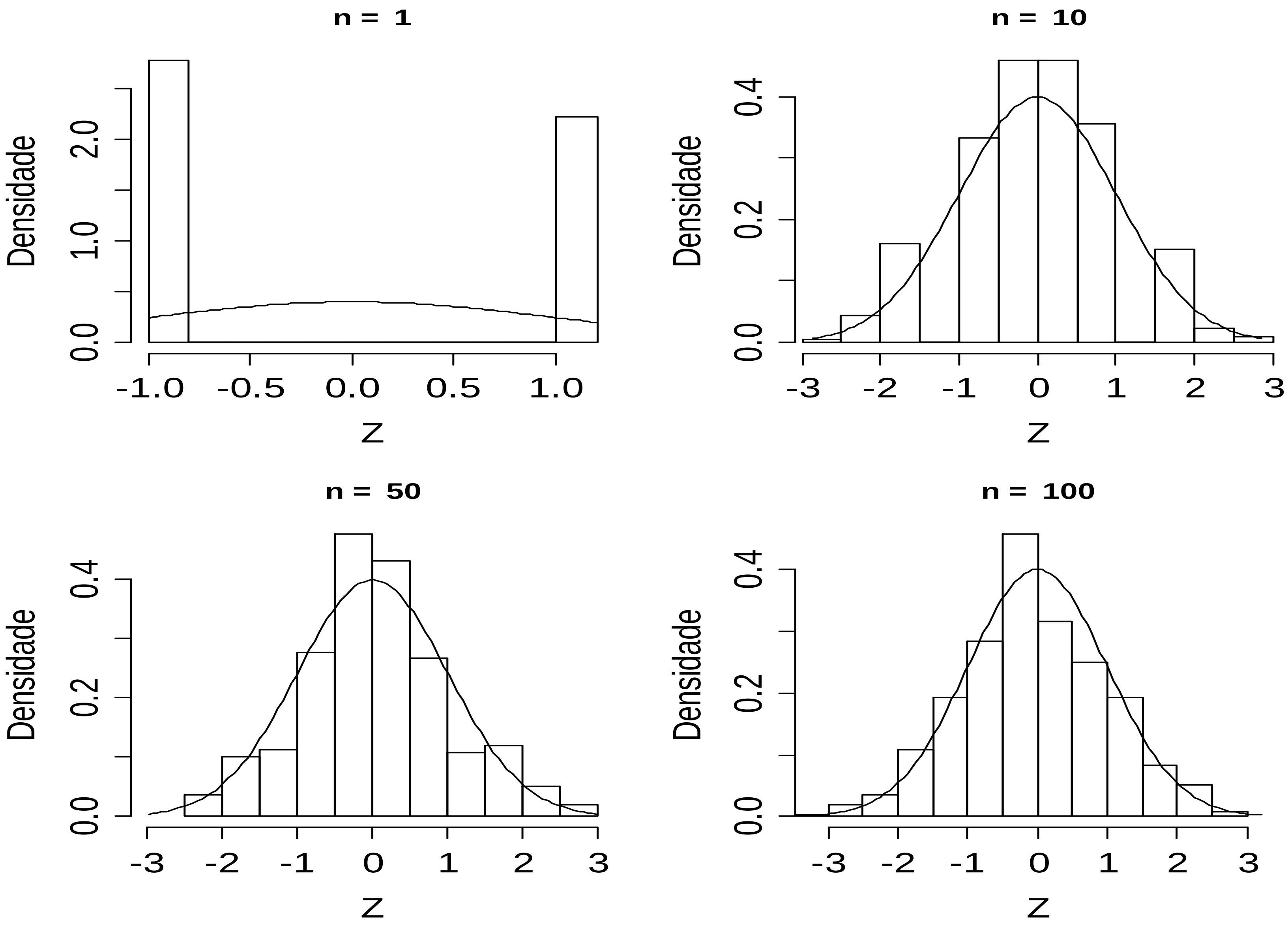
(3) A distribuição aproximada de

$$\sum_{i=1}^{n} X_i \text{ é } \mathrm{N}(n\mu, n\sigma^2).$$

Teorema central do limite – Distribuição exponencial



Teorema central do limite – Distribuição Bernoulli (p = 0,45)



Exemplo

Após arredondamento para o inteiro mais próximo, 48 números são somados. Os erros de arredondamento individuais são uniformemente distribuídos no intervalo (-0,5; 0,5). Qual a probabilidade de que a soma dos números arredondados seja diferente da verdadeira soma por mais de 3 unidades (em ambos os sentidos) ?

Solução. Utilizando o teorema central do limite obtemos uma solução aproximada.

X_i, i = 1,...,48 são os erros de arredondamento tais que $X_i \sim U(-0,5; 0,5)$,

$E(X_i) = (-0,5 + 0,5) / 2 = 0$ e $Var(X_i) = [0,5 – (-0,5)]^2 / 12 = 1 / 12$ (veja lâmina 2).

O erro de arredondamento E é dado por $E = X_1 + X_2 + ... + X_{48}$, sendo que a distribuição aproximada é $E \sim N(48 \times 0, 48 \times 1/12) = N(0,4)$.

Devemos calcular $P((E < -3) \cup (E > 3))$, que é igual a $P(E < -3) + P(E > 3)$.

Usando a distribuição aproximada, $P(E < -3) + P(E > 3) = 2 P(E < -3)$

$$= 2P\left(\frac{E - 0}{2} < \frac{-3-0}{2}\right) = 2P(Z < -1{,}50) = 2 \times 0{,}0668 = 0{,}1336.$$