Distribuição Amostral da média

A importância da média amostral $\bar{X}$ surge de seu uso para tirar conclusões sobre a média da população μ. Alguns dos procedimentos de inferência, usados com mais freqüência, baseiam-se nas propriedades da distribuição de $\bar{X}$. A apresentação inicial dessas propriedades foi feita anteriormente através de cálculos e experimentos de simulação.

Proposição Sejam $X_1, X_2, \ldots, X_n$ uma amostra aleatória de uma população X, com $E(X) = \mu$ e $Var(X) = \sigma^2$. Então,

1- $E(\bar{X}) = \mu = \mu_{\bar{X}}$

2- $Var(\bar{X}) = \sigma^2/n = \sigma^2_{\bar{X}}$ e $\sigma_{\bar{X}} = \sigma/\sqrt{n}$.

Além disso, com $T_o = X_1 + \cdots + X_n$ (o total da amostra), $E(T_o) = n\mu$, $Var(T_o) = n\sigma^2$ e $\sigma_{T_o} = \sqrt{n}\,\sigma$.

Comentário: De acordo com o resultado 1, a distribuição amostral de $\bar{X}$ é centrada precisamente na média da população da qual a amostra foi selecionada. O resultado 2 mostra que a distribuição de $\bar{X}$ se torna mais concentrada em torno de μ à medida que o tamanho da amostra n aumenta.

Com diferença marcante, a distribuição de T_o se dispersa mais ~~[illegible]~~ à medida que n aumenta.

Caso de Distribuição da População Normal

Ao reexaminar o 2º experimento de simulação, observamos que quando a distribuição da população é normal, cada histograma de valores de $\bar{x}$ é bem aproximado pela curva normal. Segue o resultado exato.

Proposição: Sejam $X_1, X_2, \ldots, X_n$ uma amostra aleatória de uma população X que segue distribuição normal com média μ e desvio padrão σ. Então, para qualquer n, $\bar{X}$ é normalmente distribuído com média μ e desvio padrão $\sigma/\sqrt{n}$.

Além disso, T_0 é normalmente distribuído com média $n\mu$ e desvio padrão $\sqrt{n}\,\sigma$.

(ir para as páginas 8, 9 e 10)

Fgm

Exemplo: O tempo que um rato de determinada subespécie, selecionado aleatoriamente leva para encontrar o caminho em um labirinto é uma variável aleatória distribuída normalmente com $\mu = 1,5$ min e $\sigma = 0,35$ min. Suponha que cinco ratos sejam selecionados. Sejam $X_1, \dots, X_5$ seu tempo no labirinto. Assumindo que X_i, $i = 1, 2, \dots, 5$ é uma amostra aleatória dessa distribuição normal, qual é a probabilidade do tempo total dos cinco estar 6 e 8 min?

$T_0 = X_1 + \dots + X_5$

Resolução: Pela proposição, $T_0 \sim N(\underbrace{5 \times 1,5}_{7,5}, 5 \times 0,35^2)$.

$\boxed{\text{PADRONIZAÇÃO de } T_0}$

$$P(6 \leq T_0 \leq 8) = P\left(\frac{6 - 7,5}{\sqrt{5}\, 0,35} \leq \frac{T_0 - \mu}{\sqrt{\sigma_{T_0}}} \leq \frac{8 - 7,5}{\sqrt{5}\, 0,35}\right)$$

$$= P\left(\frac{6 - 7,5}{0,783} \leq Z \leq \frac{8 - 7,5}{0,783}\right), \text{ onde } Z \sim N(0,1).$$

$$= P(-1,92 \leq Z \leq 0,64) = \Phi(0,64) - \Phi(-1,92) =$$

(39)

$0,7389 - 0,0274 = 0,7115.$

EXERCÍCIO!!

Qual a probabilidade do tempo médio da amostra $\bar{X}$ ser no máximo 2 min?

Lembretes

Combinação Linear

DEFINIÇÃO: Dado um conjunto de n variáveis aleatórias $X_1, \ldots, X_n$ e n constantes numéricas $a_1, \ldots, a_n$, a variável aleatória

$$Y = a_1 X_1 + \cdots + a_n X_n = \sum_{i=1}^{n} a_i X_i$$

é denominada combinação linear dos X_i's.

Exemplos: $a_1 = a_2 = \cdots = a_n = 1 \Rightarrow Y = X_1 + \cdots + X_n$

$= T_0$.

$a_1 = a_2 = \cdots = a_n = \frac{1}{n} \Rightarrow$

$$Y = \frac{1}{n} X_1 + \cdots + \frac{1}{n} X_n = \bar{X}.$$

{ Observe que não é requerido que os X_i's sejam independentes ou identicamente distribuídos }.

Proposição: Suponha que $X_1, \ldots, X_n$ tenham valores médios $\mu_1, \ldots, \mu_n$, respectivamente, e variâncias $\sigma_1^2, \ldots, \sigma_n^2$, respectivamente.

1 - Se os X_i's são ou não ~~são~~ independentes,

$$E\left(\sum_{i=1}^{n} a_i X_i\right) = \sum_{i=1}^{n} a_i E(X_i) = \sum_{i=1}^{n} a_i \mu_i .$$

2 - Se $X_1, \ldots, X_n$ são independentes,

$$Var\left(\sum_{i=1}^{n} a_i X_i\right) = \sum_{i=1}^{n} a_i^2 Var(X_i) =$$

$$\sum_{i=1}^{n} a_i^2 \sigma_i^2 .$$

3 - Para qualquer $X_1, \ldots, X_n$,

$$Var(a_1 X_1 + \cdots + a_n X_n) = \sum_{i=1}^{n} \sum_{j=1}^{n} a_i a_j Cov(X_i, X_j).$$

Diferença Entre Duas Variáveis Aleatórias

Um caso especial importante de combinação linear resulta de $n = 2$, $a_1 = 1$ e $a_2 = -1$

é $y = a_1 X_1 + a X_2 = X_1 - X_2$.

Corolário: $E(X_1 - X_2) = E(X_1) - E(X_2)$.

Se X_1 e X_2 são independentes,

$Var(X_1 - X_2) = Var(X_1) + Var(X_2)$.

O valor esperado de uma diferença é a diferença entre os dois valores esperados, mas a variância da diferença entre duas variáveis independentes é a soma.

Há tanta variabilidade em $X_1 - X_2$ como em $X_1 + X_2$.

Caso de Variáveis Aleatórias Normais

Proposição: Se $X_1, \ldots, X_n$ são v.a.'s independentes, distribuídas normalmente (com médias e/ou variâncias possivelmente diferentes), então qualquer combinação linear das X_i's também tem distribuição normal.

Em particular, a própria diferença $X_1 - X_2$ entre duas variáveis independentes e normalmente distribuídas é normalmente distribuída.

Proposição:

Se $X_1, \dots, X_n$ são v.a.'s independentes distribuídas normalmente (com médias e/ou variâncias possivelmente diferentes), então qualquer combinação linear dos X_i's também tem distribuição normal.

Isto é,

$$Y = \sum_{i=1}^{n} a_i X_i \sim N\left(\sum_{i=1}^{n} a_i \mu_i, \sum_{i=1}^{n} a_i^2 \sigma_i^2\right),$$

onde $X_1, \dots, X_n$ são v.a.'s independentes $N(\mu_i, \sigma_i^2)$, $i = 1, \dots, n$.

Caso $X_1, \ldots, X_n$ fossem i.i.d $N(\mu, \sigma^2)$,

então

$$Y \sim N\left(\mu \sum_{i=1}^{n} a_i, \quad \sigma^2 \sum_{i=1}^{n} a_i^2\right) \, q.$$

Considere $Y = \bar{X} = \dfrac{X_1 + \cdots + X_n}{n}$, onde

$a_i = 1/n$, $i = 1, \ldots, n$. Temos que

$\bar{X} = Y \sim N\left(\mu, \sigma^2/n\right)$, como já provamos

em sala de aula.

Além disso, considere $Y = T_0 = X_1 + \cdots + X_n$,

onde $a_i = 1$, $i = 1, \ldots, n$. Temos que

$T_0 = Y \sim N(n\mu, n\sigma^2)$, como já

provamos em sala de aula.

caminho $X_1 - X_2$?

q: prova via fgm

RELEMBRANDO!

$X \sim$ Bernoulli(p)

$$m_X(t) = 1-p+e^t p.$$

$X \sim$ Binomial(n,p)

$$m_X(t) = \left(1-p+e^t p\right)^n.$$

$X \sim$ Poisson(λ)

$$m_X(t) = e^{\lambda(e^t-1)}$$

EXEMPLO:

$X_1, \dots, X_n$ a.a. $X \sim$ Bernoulli(p). Encontre a distribuição de $T = \sum_{i=1}^{n} X_i$.

Pelas propriedades da fgm, temos que

$$m_T(t) = \left[m_X(t)\right]^n = \left(1-p+e^t p\right)^n.$$ Assim,

$$T \sim \text{Binomial}(n,p).$$

Qual a distribuição de $\bar{X}$?

$$m_{\bar{X}}(t) = \left(m_X(t/n)\right)^n = \left(1-p+e^{t/n}p\right)^n.$$

↳ nenhuma distribuição conhecida !!

ver folhas 52 e 53

EXEMPLO: $X_1, \dots, X_n$ a.a. $X \sim \text{Poisson}(\lambda)$.

Encontre a distribuição de $T = \sum_{i=1}^{n} X_i$.

$$m_T(t) = \left(m_X(t)\right)^n = \left[e^{\lambda(e^t-1)}\right]^n = e^{n\lambda(e^t-1)}.$$

$\therefore T \sim \text{Poisson}(n\lambda)$.

ver folhas 54 e 55

Distribuição Exata da média amostral para alguns casos específicos!!

Exemplo 1: Se $X_1, \dots, X_n$ é uma amostra aleatória de uma população X que segue uma distribuição de Bernoulli (p). Desejamos encontrar a distribuição exata de $\bar{X}$.

Resolução: $X \sim$ Bernoulli (p), então

$$p(x) = p^x (1-p)^{1-x} \, I_{(0,1)}(x).$$

$$I(x) = \begin{cases} 1 & \text{se } x = 0 \text{ ou } 1 \\ 0 & \text{c.c} \end{cases}$$

Sabemos que $\sum_{i=1}^{n} X_i \sim$ Binomial (n, p), isto é,

$$P\left(\sum_{i=1}^{n} X_i = k\right) = \binom{n}{k} p^k (1-p)^{n-k} \, I_{(0,1,\dots,n)}(k).$$

$$P(\overline{X} = y) = P\Big(\underbrace{\sum_{i=1}^{n} X_i = ny}_{K}\Big) = \binom{n}{k} (1-p)^{n-k} I_{\{0,1,\ldots,n\}}(k),$$

$$ny = k \Rightarrow y = k/n$$

tal que a distribuição amostral de $\overline{X}$ é obtida da distribuição de $\sum_{i=1}^{n} X_i$.

Portanto, a média amostral proveniente de uma amostra aleatória retirada de uma população com distribuição Bernoulli, assume os valores

$0, \frac{1}{n}, \frac{2}{n}, \ldots, 1$ com as respectivas probabilidades

$$\binom{n}{0} p^0 q^n, \quad \binom{n}{1} p^1 q^{n-1}, \quad \binom{n}{2} p^2 q^{n-2}, \ldots,$$

$$\binom{n}{n} p^n q^0, \text{ onde } q = 1-p.$$

Exemplo 2: Se $X_1, \ldots, X_n$ é uma amostra aleatória de uma população X que segue uma distribuição de Poisson(λ), então sabemos que $\sum_{i=1}^{n} X_i \sim$ Poisson$(n\lambda)$. Assim,

$$P(x) = \frac{e^{-\lambda}\lambda^{x}}{x!} I_{(0,1,2,\ldots)}(x)$$

$$P\left(\sum_{i=1}^{n} X_i = k\right) = \frac{e^{-n\lambda}(n\lambda)^{k}}{k!} I_{(0,1,2,\ldots)}(k)$$

$$P(\bar{X} = y) = P\left(\sum_{i=1}^{n} X_i = \underbrace{ny}_{k}\right) = \frac{e^{-n\lambda}(n\lambda)^{k}}{k!} I_{(0,1,2,\ldots)}(k),$$

onde $y = k/n$. Portanto, a média amostral proveniente de uma amostra aleatória retirada de uma população com distribuição Poisson, assume os valores

$0, 1/n, 2/n, \ldots$, com as respectivas probabilidades

(55)

$$\frac{e^{-n\lambda}(n\lambda)^0}{0!}, \quad \frac{e^{-n\lambda}(n\lambda)^1}{1!}, \quad \frac{e^{-n\lambda}(n\lambda)^2}{2!}, \ldots$$

A distribuição amostral de $\bar{X}$ é obtida da distribuição de $\sum_{i=1}^{n} X_i$.