# Teorema Limite Central

Quando os $X_i$'s são distribuídos normalmente, $\bar{X}$ também é para cada tamanho de amostra $n$. Porém, os dois últimos experimentos de simulação sugerem que, quando o tamanho da amostra aumenta, independentemente da forma da distribuição da população (

(consideramos uniforme e binomial), a distribuição amostral de $\bar{X}$ aproxima-se cada vez mais de uma distribuição normal. Esse resultado é conhecido como Teorema Limite central (TLC).

Teorema fundamental na teoria de inferência estatística.

Teorema Limite central

Para amostras aleatórias simples $X_1, \ldots, X_n$, retiradas de uma população com média (valor esperado) $\mu$ e variância $\sigma^2$ finita, a distribuição amostral da média $\bar{X}$ aproxima-se, para $n$ grande, de uma distribuição normal, com média $\mu$ e variância $\sigma^2/n$.

Além disso, $T_0$ também tem aproximadamente uma distribuição normal, com média $n\mu$ e variância $n\sigma^2$.

quanto maior o valor de n, melhor a apro-
ximação.

A rapidez dessa convergência depende da distri-
buição da população da qual a amostra é
retirada. Se a população original tem uma
distribuição próxima da normal, a conver-
gência é rápida; se a população original
se afasta muito de uma normal, a convergên-
cia é mais lenta, ou seja, necessitamos de
uma amostra maior para que $\bar{X}$ tenha
uma distribuição aproximadamente normal.
Para amostras da ordem de 30 ou 50 elementos
a aproximação pode ser considerada boa
(razoável).

<u>Exemplo</u>: considere uma máquina que
enche pacotes cujos pesos seguem uma
distribuição $N(500, 100)$. colhendo-se uma
amostra de 100 pacotes e pesando-os,
pelo que foi dito acima, $\bar{X}$ terá uma
distribuição normal com média 500 e
variância $100/100 = 1$. Logo, se a máquina
estiver regulada, a probabilidade de

encontrarmos a média de 100 pacotes diferindo de 500 g de menos de 2 gramas será

$$P(|\bar{X}-500| < 2) = P(498 < \bar{X} < 502) =$$

$$P(-2 < Z < 2) \approx 95\%$$

ou seja, dificilmente 100 pacotes terão uma média fora do intervalo (498, 502). caso 100 pacotes apresentem uma média fora desse intervalo, podemos considerar como um evento raro, e será razoável supor que a máquina esteja desregulada.

<u>corolário</u>: se $X_1, \dots, X_n$ for uma amostra aleatória simples da população $X$, com média $\mu$ e variância $\sigma^2$ finita, e $\bar{X} = \frac{X_1 + \cdots + X_n}{n}$,

então

$$Z = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1).$$

observe também que podemos escrever da seguinte forma

$$Z = \frac{\sqrt{n}(\bar{X}-\mu)}{\sigma} \sim N(0,1).$$

Chamemos de "e" a variável aleatória que mede a diferença entre a estatística $\bar{X}$ e o parâmetro $\mu$, isto é, $e = \bar{X} - \mu$

e é chamado de erro amostral da média. Então, temos o seguinte corolário.

corolário A distribuição de e aproxima-se de uma distribuição normal com média 0 e variância $\sigma^2/n$, isto é,

$$\frac{\sqrt{n}\, e}{\sigma} \sim N(0,1) .$$

Outras aplicações do Teorema Limite Central

DISTRIBUIÇÃO AMOSTRAL de uma proporção

Vamos considerar uma população em que a proporção de elementos portadores de certa característica é $p$.

Logo, podemos definir uma variável aleatória $X$, da seguinte maneira:

$$X = \begin{cases} 1, & \text{se o indivíduo for portador da característica} \\ 0, & \text{se o indivíduo não for portador da característica} \end{cases}$$

Assim, $\mu = E(X) = p$ e $\sigma^2 = Var(X) = p(1-p)$.

Retirada uma amostra aleatória simples $X_1, \dots, X_n$ dessa população, e indicando por $Y$ o total de indivíduos portadores da característica na amostra, ou seja, $Y = X_1 + \dots + X_n$, onde cada $X_i$ tem distribuição de Bernoulli, com média $\mu = p$, variância $\sigma^2 = p(1-p)$ e são independentes. Podemos escrever

$$Y = n\bar{X} \sim \text{binomial}(n, p).$$

Considere $X_1, \ldots, X_n$ i.i.d $X \sim$ Bernoulli $(p)$. Mostre que $T_0 = X_1 + \cdots + X_n = \sum_{i=1}^{n} X_i \sim$ Binomial$(n, p)$.

PROVA:

Sabemos que se $X \sim$ Bernoulli $(p)$ então

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

Além disso,

$$m_{T_0}(t) = m_{\sum_{i=1}^{n} X_i}(t) = \left[ m_X(t) \right]^n,$$

pois $X_1, \ldots, X_n$ são v.a.'s i.i.d. Assim,

$$m_{T_0}(t) = \left(1 - p + e^t p\right)^n$$ // cqd

Vimos em sala de aula que a fgm para uma v.a. que tem distribuição Binomial$(n, p)$ é dada por

$$(1 - p + e^t p)^n.$$

Portanto, $T_0 \sim$ Binomial$(n, p)$.

Vamos definir por $\hat{p}$ a proporção de indivíduos portadores da característica na amostra, isto é,

$$\hat{p} = Y/n$$

Então,

$$P(Y = k) = P(Y/n = k/n) = P(\hat{p} = k/n),$$

ou seja, a distribuição amostral de $\hat{p}$ é obtida da distribuição de $Y$.

Note que $Y = n\bar{X}$, mas pelo TLC, $\bar{X}$ terá distribuição aproximadamente normal, com média $p$ e variância $\frac{p(1-p)}{n}$, ou seja,

$\mu$ $\qquad$ $\sigma^2/n$

$$\bar{X} \sim N\left(p, \frac{p(1-p)}{n}\right).$$

Logo, a transformação $Y = n\bar{X}$ terá a distribuição

$$Y \sim N(np, np(1-p)).$$

((fgm) ou função característica)

observe que $\bar{X}$ é a própria variável $\hat{p}$ e, desse modo, para $n$ grande podemos considerar a distribuição amostral de $\hat{p}$ como aproximadamente normal:

$$\hat{p} \sim N\left( p, \frac{p(1-p)}{n} \right).$$

<u>Exemplo</u>: Suponha que $p = 30\%$ dos estudantes de uma escola sejam mulheres. Colhemos uma aas de $n = 10$ estudantes e calculamos $\hat{p}$ = proporção de mulheres na amostra. Qual a probabilidade de que $\hat{p}$ difira de $p$ em menos de $0,01$? Temos que esta probabilidade é dada por

$$P(|\hat{p} - p| < 0,01) = P(-0,01 < \hat{p} - p < 0,01),$$

mas $\hat{p} - p \sim N\left(0, \frac{p(1-p)}{n}\right)$. Como $p = 0,3$, temos que

$$Var(\hat{p}) = \frac{p(1-p)}{n} = \frac{0,3 \times 0,7}{10} = 0,021.$$

Portanto, a probabilidade pedida é 

igual a

$$P\left(\frac{-0{,}01}{\sqrt{0{,}021}} < Z < \frac{0{,}01}{\sqrt{0{,}021}}\right) = P\left(-0{,}07 < Z < 0{,}07\right) =$$

$0{,}056$, onde

$Z \sim N(0,1)$.