Introdução às Probabilidades

tradução do Jay Davore

índice

Funções conjuntamente distribuidas

Trabalhamos até agora com funções de probabilidade de uma única v.a. Agora, definiremos para quaisquer variáveis aleatórias X e Y a função de distribuição de probabilidade cumulativa conjunta de X e Y como

$$F(a, b) = P\{X \leq a, Y \leq b\} \quad -\infty < a, b < \infty$$

Funções conjuntamente distribuidas

Trabalhamos até agora com funções de probabilidade de uma única v.a. Agora, definiremos para quaisquer variáveis aleatórias X e Y a função de distribuição de probabilidade cumulativa conjunta de X e Y como

$$F(a, b) = P\{X \leq a, Y \leq b\} \quad -\infty < a, b < \infty$$

Funções conjuntamente distribuidas

A função de distribuição de X (marginal) pode ser obtida a partir da conjunta de X e Y da seguinte forma:

$$\begin{aligned} F_X(a) &= P(X \leq a) \\ &= P(X \leq a, Y < \infty) \\ &= P(\lim_{b \to \infty} \{X \leq a, Y \leq b\}) \\ &= F(a, \infty) \end{aligned}$$

Marginais

$$
\begin{aligned}
P(X > a, Y > b) &= 1 - P(X > a, Y > b)^c \\
&= 1 - P\left((X > a)^c \cup (Y > b)^c\right) \\
&= 1 - P\left((X \leq a) \cup (Y \leq b)\right) \\
&= 1 - \left[P(X \leq a) + P(Y \leq b) - P\left(X \leq a, Y \leq b\right)\right] \\
&= 1 - F_X(a) - F_Y(b) + F(a, b)
\end{aligned}
$$

Função Massa de probabilidade conjunta

Sejam X e Y duas v.a. discretas definidas no espaço amostral de um experimento. A função massa de probabilidade conjunta $p(x, y)$ é definida para cada par de números (x, y) por

$$p(x, y) = P(X = x \text{ e } Y = y)$$

Seja A o conjunto que contém os valores dos pares (x, y), então:

$$P\left[(X, Y) \in A\right] = \sum_{(x,y) \in A} \sum p(x, y)$$

Definição

A função massa de probabilidade conjunta de X e Y, denotada por $p_X(x)$ e $p_Y(y)$, são dados respectivamente por:

$$p_X(x) = \sum_y p(x, y)$$

$$p_Y(y) = \sum_x p(x, y)$$

Exemplo

Suponha que três bolas sejam sorteadas de uma urna contendo 3 bolas vermelhas, 4 bolas brancas e 5 bolas azuis. Se X e Y representam respectivamente, o número de bolas vermelhas e brancas escolhidas, então a função de probabilidade conjunta de X e Y, $p(i,j) = P(X = i, Y = j)$ é dada por:

exemplo

$$p(0,0) = \frac{\binom{5}{3}}{\binom{12}{3}} = \frac{10}{220}$$

$$p(0,1) = \frac{\binom{4}{1}\binom{5}{2}}{\binom{12}{3}} = \frac{40}{220}$$

$$... = ...$$

$$p(1,1) = \frac{\binom{3}{1}\binom{4}{1}\binom{5}{1}}{\binom{12}{3}} = \frac{60}{220}$$

....

exemplo

representação das probabilidades numa tabela de dupla entrada:

i	j				
	0	1	2	3	soma linha
0	$\frac{10}{220}$	$\frac{40}{220}$	$\frac{30}{220}$	$\frac{40}{220}$	$\frac{84}{220}$
1	$\frac{30}{220}$	$\frac{60}{220}$	$\frac{18}{220}$	0	$\frac{108}{220}$
2	$\frac{15}{220}$	$\frac{12}{220}$	0	0	$\frac{27}{220}$
3	$\frac{1}{220}$	0	0	0	$\frac{1}{220}$
Soma Coluna	$\frac{56}{220}$	$\frac{112}{220}$	$\frac{48}{220}$	$\frac{4}{220}$	

Definição

Sejam X e Y v.a.contínuas. Então $f(x, y)$ é uma função densidade de probabilidade conjunta para X e Y se para qualquer conjunto bi-dimensional A,

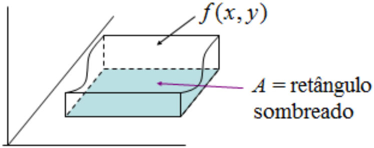
$$P\left[(X, Y) \in A\right] = \int_A \int f(x, y)dxdy$$

Se A é um retângulo bidimensional:

$$\{(x, y) : a \leq x \leq b, c \leq y \leq d\},$$

$$P\left[(X, Y) \in A\right] = \int_a^b \int_c^d f(x, y)dydx$$

exemplo



$$P\left[(X,Y)\in A\right]$$

= Volúmen abaixo da superficie acima de A

Figura: Representação de uma v.a. bidimensional

Definição

A função densidade de probabilidade marginal de X e Y, denotada por $f_X(x)$ e $f_Y(y)$, são dados respectivamente por:

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y)dy \quad para - \infty < x < \infty$$
$$f_Y(y) = \int_{-\infty}^{\infty} f(x, y)dx \quad para - \infty < y < \infty$$

Exemplo

A função de densidade conjunta de X e Y é dada por:

$$f(x,y) = \begin{cases} 2e^{-x}e^{-2y} & 0 < x < \infty,\ 0 < y < \infty \\ 0 & \text{caso contrário} \end{cases}$$

calcule:

1. $P(X > 1, Y < 1)$
2. $P(X < Y)$
3. $P(X < a)$

Exemplo

Solução a):

$$\begin{aligned} P(X > 1, Y < 1) &= \int_0^1 \int_1^\infty 2e^{-x} e^{-2y} \, dx dy \\ &= \int_0^1 2e^{-2y} (-e^{-x}|_1^\infty) \, dy \\ &= e^{-1}(1 - e^{-2}) \end{aligned}$$

Exemplo

Solução b)

$$
\begin{aligned}
P(X < Y) &= \int_{(x,y):x<y} \int 2e^{-x}e^{-2y}\,dxdy \\
&= \int_0^\infty \int_0^y 2e^{-x}e^{-2y}\,dxdy \\
&= \int_0^\infty 2e^{-2y}(1 - e^{-y})\,dy \\
&= \int_0^\infty 2e^{-2y}\,dy - \int_0^\infty 2e^{-3y}\,dy \\
&= 1 - \frac{2}{3} = \frac{1}{3}
\end{aligned}
$$

exemplo

Solução c)

$$\begin{aligned} P(X < a) &= \int_0^a \int_1^\infty 2e^{-x}e^{-2y}\,dydx \\ &= \int_0^a e^{-x}\,dx \\ &= 1 - e^{-a} \end{aligned}$$

definição

Duas variáveis X e Y são independentes se para cada par de valores x e y,

$$p(x, y) = p_X(x).p_y(y)$$

Quando X e Y são discretos, ou

$$f(x, y) = f_X(x).f_Y(y)$$

quando X e Y são contínuos.
Se as condições não são satisfeitas para todo (x, y), então X e Y são dependentes.

Exemplo

Um homem e uma mulher decidem se encontrar em certo lugar. Se cada um deles chega independentemente em um tempo uniformemente distribuido entre 12:00 e 13:00, determine a probabilidade de que o primeiro a chegar tenha que esperar mais de 10 minutos

exemplo

Se X e Y são os tempos de chegada (após meio-dia) do homem e da mulher, então X e Y são v.a. independentes,com distribuição uniforme no intervalo $(0, 60)$.
Queremos calcular:

$$P(X + 10 < Y) + P(Y + 10 < X)$$

o que é igual (por simetria) a $2P(X + 10 < Y)$

exemplo

Se X e Y são os tempos de chegada (após meio-dia) do homem e da mulher, então X e Y são v.a. independentes,com distribuição uniforme no intervalo $(0, 60)$.
Queremos calcular:

$$P(X + 10 < Y) + P(Y + 10 < X)$$

o que é igual (por simetria) a $2P(X + 10 < Y)$

exemplo

$$\begin{aligned}
2P(X+10< Y) &= 2\iint_{x+10<y} f(x,y)\,dxdy \\
&= 2\iint_{x+10<y} f_X(x)f_Y(y)\,dxdy \quad \text{pela independência} \\
&= 2\int_{10}^{60}\int_{0}^{y-10}\left(\frac{1}{60}\right)^2 dxdy \\
&= \frac{2}{(60)^2}\int_{10}^{60}(y-10)\,dy \\
&= \frac{25}{36}
\end{aligned}$$

Mais de duas variáveis

Se $X_1, X_2, ..., X_n$ são todas v.a. discretas. A f.p.conjunta das v.a. é a função:

$$p(x_1, ..., x_n) = P(X_1 = x_1, ..., X_n = x_n)$$

Se as variáveis são contínuas, a f.d.p. conjunta é a função tal que para qualquer dos n intervalos $[a_1, b_1], ...[a_n, b_n]$

$$P(a_1 \leq X_1 \leq b_1, ..., a_n \leq X_n \leq b_n) = \int_{a_1}^{b_1} ... \int_{a_n}^{b_n} f(x_1, ..., x_n) dx_n ... dx_1$$

Independência de mais de duas v.a.

As v.a. $X_1, ..., X_n$ são independentes se para cada subconjunto $X_{i_1}, X_{i_2}, ..., X_{i_n}$ das variáveis, a f.m.p (f.p.) conjunta ou f.d.p. conjunta do subconjunto é igual ao produto das f.m.p marginais ou f.d.p. marginais.

Exemplo

Se a função de densidade conjunta de X e Y é

$$f(x,y) = \begin{cases} 6e^{-2x}e^{-3y} & \text{na região } 0 < x < \infty, 0 < y < \infty \\ 0 & \text{caso contrário} \end{cases}$$

verifique se as variáveis aleatórias são independentes

Exemplo

Solução:
Se fizermos X e Y variáveis aleatórias exponenciais, teremos:

$$f(x) = \begin{cases} 2e^{-2x} & \text{na região } 0 < x < \infty \\ 0 & \text{caso contrário} \end{cases}$$

$$f(y) = \begin{cases} 3e^{-3y} & \text{na região } 0 < y < \infty \\ 0 & \text{caso contrário} \end{cases}$$

Definição

Sejam X e Y duas v.a. continuas com f.d.p conjunta $f(x,y)$ e f.d.p marginal de X $f_X(x)$. Então, para qualquer valor x de X, para o qual $f_X(x) > 0$, a função densidade de probabilidade condicional (f.d.p. condicional) de Y dado $X = x$ é:

$$f_{Y|X}(y|x) = \frac{f(x,y)}{f_X(x)} \qquad -\infty < y < \infty$$

Se X e Y são discretos, substituindo f.d.p't́s por f.m.p.t́s resultarão na função massa de probabilidade condicional de Y cuando $X = x$

Funções de uma variável aleatória

Suponha que uma v.a. X tem f.p. f (discreta) e que outra v.a. $Y = r(X)$ é definida como função de X. Então, a f.p. g de Y pode ser derivada de f de uma forma direta como segue:

para qualquer valor possível y de Y,

$$\begin{aligned} g(y) &= P(Y = y) = P(r(X) = y) \\ &= \sum_{x:r(x)=y} f(x) \end{aligned}$$

Funções de uma variável aleatória

Se X tem f.d.p. f (contínua), e que uma v.a. $Y = r(X)$ é definida. a função de distribuição acumulada $G(y)$ de Y pode ser derivado como segue:

$$G(y) = P(Y \leq y) = P(r(X) \leq y)$$
$$= \int_{\{x: r(x) \leq y\}} f(x)\, dx$$

Função de uma variável aleatória

Se a v.a. Y tem uma função de distribuição contínua, sua f.d.p. g pode ser obtida da relação:

$$g(y) = \frac{dG(y)}{dy}$$

Exemplo

Suponha que X tem uma distribuição no intervalo $(-1, 1)$, tal que

$$f(x) = \begin{cases} 1/2 & \text{para} \quad -1 < x < 1 \\ 0 & \text{caso contrário} \end{cases}$$

determinar a f.d.p. da v.a. $Y = X^2$

Exemplo

Como $Y = X^2$, então Y deve pertencer ao intervalo $0 \leq Y \leq 1$. Assim, para qualquer valor de Y tal que $0 < Y < 1$, a f.de distribuição acumulada $G(y)$ de Y é

$$\begin{aligned} G(Y) &= P(Y \leq y) = P(X^2 \leq y) \\ &= P(-y^{1/2} \leq X \leq y^{1/2}) \\ &= \int_{-y^{1/2}}^{y^{1/2}} f(x)\,dx = y^{1/2} \end{aligned}$$

Exemplo

Para $0 < y < 1$, segue-se que a f.d.p. $g(y)$ de Y é

$$g(y) = \frac{dG(y)}{dy} = \frac{1}{2y^{1/2}}$$

Funções de duas ou mais variáveis aleatórias

Suponha que n v.a. $X_1, ..., X_n$ tem f.d.p.conjunta discreta f, e que m funções $Y_1, ..., Y_m$ de estas n variáveis são definidas como segue:

$$
\begin{aligned}
Y_1 &= r_1(X_1, ..., X_n) \\
Y_2 &= r_2(X_1, ..., X_n) \\
&\ldots\ldots \\
Y_m &= r_m(X_1, ..., X_n)
\end{aligned}
$$

Funções de duas ou mais variáveis aleatórias

Para valores dados $y_1, ..., y_m$ das m v.a. $Y_1, ..., Y_m$, seja A o conjunto de todos os pontos $(x_1, ..., x_n)$ tal que

$$r_1(x_1, ..., x_n) = y_1$$
$$r_2(x_1, ..., x_n) = y_2$$
$$......$$
$$r_m(x_1, ..., x_n) = y_m$$

Então a f.p.conjunta g de $Y_1, ..., Y_m$ é especificado no ponto $(y_1, ..., y_m)$ pela relação:

$$g(y_1, ..., y_m) = \sum_{(x_1, ..., x_n) \in A} f(x_1, ..., x_n)$$

Funções de duas ou mais variáveis aleatórias

Se a função de distribuição conjunta de $X_1, ..., X_n$ (continua) é $f(x_1, ..., x_n)$ e se $Y = r(X_1, ..., X_n)$, então a função de distribuição acumulada $G(y)$ de Y pode ser determinada da seguinte forma:
Para qualquer valor dado de y $(-\infty < y < \infty)$, seja A_y um subconjunto de R^n que contém os pontos $(x_1, ..., x_n)$ tal que $r(x_1, ..., x_n) \leq y$, então:

$$\begin{aligned} G(y) &= P(Y \leq y) = P(r(X_1, ..., X_n) \leq y) \\ &= \int_{A_y} .... \int f(x_1, ..., x_n) dx_1 ... dx_n \end{aligned}$$

Distribuição do máximo e o mínimo em uma amostra aleatória

Suponha que as v.a. $X_1, ..., X_n$ formam uma a.a. de tamanho n de uma distribuição para a qual a f.d.p. é f e a f.de distribuição acumulada é F. Considere as v.a. Y_n e Y_1 definidas como

$$Y_n = max\{X_1, ..., X_n\}$$

e

$$Y_1 = min\{X_1, ..., X_n\}$$

determine as f.d.p. de cada uma delas.

Distribuição do máximo e o mínimo em uma amostra aleatória

Para o caso Y_n,

$$\begin{aligned} G_n(y) &= P(Y_n \leq y) = P(X_1 \leq y, ..., X_n \leq y) \\ &= P(X_1 \leq y)...P(X_n \leq y) \\ &= F(y)...F(y) = [F(y)]^n \end{aligned}$$

a f.d.p de Y_n pode ser determinado derivando:

$$g_n(y) = n\,[F(y)]^{n-1} f(y), \quad \text{para} \quad -\infty < y < \infty$$

Distribuição do máximo e o mínimo em uma amostra aleatória

para o caso Y_1, para $(-\infty < y < \infty)$

$$\begin{aligned} G_1(y) &= P(Y_1 \leq y) = 1 - P(Y_1 > y) \\ &= 1 - P(X_1 > y, ..., X_n > y) \\ &= 1 - P(X_1 > y)...P(X_n > y) \\ &= 1 - [1 - F(y)] \ldots [1 - F(y)] \\ &= 1 - [1 - F(y)]^n \end{aligned}$$

a f.d.p de Y_1 pode ser determinado derivando:

$$g_1(y) = n\,[1 - F(y)]^{n-1} f(y), \quad \text{para} \quad -\infty < y < \infty$$

Valor esperado

Sejam X e Y duas v.a. conjuntamente distribuidos com f.m.p conjunta $p(x, y)$ ou f.d.p. conjunta $f(x, y)$ dependendo em que as variáveis são discretas ou continuas. Então, o valor esperade de uma função $h(X, Y)$ denotado por $E\left[h(X, Y)\right]$ ou $\mu_{h(X,Y)}$ é

$$E\left[h(X, Y)\right] = \begin{cases} \sum_x \sum_y h(x, y).p(x, y) & \text{no caso discreto} \\ \\ \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(x, y).f(x, y)\,dxdy & \text{no caso contínuo} \end{cases}$$

exemplo

Considere a distribuição conjunta de X e Y:

	X				
Y	0	1	2	3	$p(y)$
O	1/8	2/8	1/8	0	1/2
1	0	1/8	2/8	1/8	1/2
$p(x)$	1/8	3/8	3/8	1/8	1

Encontre a distribuição de

1. a soma $X+Y$, o valor esperado $E(X+Y)$
2. o produto XY

Exemplo

Solução
a distribuição da soma é:

$x+y$	0	1	2	3	4
$p(x+y)$	1/8	2/8	2/8	2/8	1/8

O valor esperado

$$E(x+y) = 0 \times 1/8 + 1 \times 2/8 + ... + 4 \times 1/8 = 2$$

Exemplo

Solução
a distribuição do produto é:

xy	0	1	2	3
$p(xy)$	$4/8$	$1/8$	$2/8$	$1/8$

definição

A covariância entre duas v.a.t's X e Y é

$$Cov(X, Y) = E\left[(X - \mu_X)(Y - \mu_y)\right] =$$

$$\begin{cases} \sum_x \sum_y (x - \mu_X)(y - \mu_Y)p(x, y) & \text{caso discreto} \\ \\ \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \mu_X)(y - \mu_Y)f(x, y)\,dxdy & \text{caso contínuo} \end{cases}$$

Forma curta para o cálculo da covariância

$$Cov(X, Y) = E(X, Y) - \mu_X . \mu_Y$$

definição

O coeficiente de correlação de X e Y, denotado por $Corr(X, Y)$, $\rho_{X,Y}$, ou apenas ρ, é definida por :

$$\rho_{X,Y} = \frac{Cov(X, Y)}{\sigma_X . \sigma_Y}$$

Proposições para a Correlação

- 1. Se a e c são ambos positivos ou ambos negativos,

$$Corr(aX + b, cY + d) = Corr(X, Y)$$

- 2. Para duas v.a.'s X e Y quaisquer:

$$-1 \leq Corr(X, Y) \leq 1$$

Proposições para a Correlação

- 3. Se X e Y são independentes, então

$$\rho = 0$$

mas, $\rho = 0$ não implica independência.

- 4. $\rho = 1$ ou $\rho = -1$ se e somente se

$$Y = aX + b$$

para alguns números a e b com $a \neq 0$

Definição

Uma estatística é qualquer quantidade cujo valor pode ser calculado dos dados amostrais (função de valores amostrais).

Antes de obter os dados (amostra), existe incerteza sobre qual será o valor de uma particular estatística

Uma estatística é uma v.a. denotada por uma letra maiúscula; uma letra minúscula representará o valor observado da estatística.

Estatísticas comúns

$$\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n} \quad \text{: média amostral}$$

$$S^2 = \frac{\sum_{i=1}^{n} (X_i - \bar{X})^2}{n-1} \quad \text{: variância amostral}$$

$$X_{(1)} = min\{X_1, ..., X_n\} \quad \text{: mínimo da amostra}$$

$$X_{(n)} = max\{X_1, ..., X_n\} \quad \text{: maximo da amostra}$$

Definição de amostra aleatória

As v.a. $X_1, ..., X_n$ formam uma amostra aleatória simples de tamanho n, se:

- 1. Os X_i's são v.a. independentes.
- 2. Cada X_i tem a mesma distribuição de probabilidade.

Experimentos simulados

As seguintes características devem ser especificadas:

- 1. A estatística de interesse
- 2. A distribuição da população
- 3. O tamanho da amostra n
- 4. O número de repetições k

Distribuições amostrais

O propósito da inferência estatística é fazer afirmações sobre parâmetros populacionais, através de dados amostrais.

Usamos uma a.a.s. de n elementos da população, e a decisão será baseada na estatística T, que será função da amostra $(X_1, ..., X_n)$, i.e.,

$$T = f(X_1, ..., X_n)$$

Distribuições amostrais

O propósito da inferência estatística é fazer afirmações sobre parâmetros populacionais, através de dados amostrais.

Usamos uma a.a.s. de n elementos da população, e a decisão será baseada na estatística T, que será função da amostra $(X_1, ..., X_n)$, i.e.,

$$T = f(X_1, ..., X_n)$$

Colhida a amostra, teremos observado um valor particular de T (t_0). Baseados naquele valor é que faremos uma afirmação sobre θ.

Distribuições amostrais

O propósito da inferência estatística é fazer afirmações sobre parâmetros populacionais, através de dados amostrais.

Usamos uma a.a.s. de n elementos da população, e a decisão será baseada na estatística T, que será função da amostra $(X_1, ..., X_n)$, i.e.,

$$T = f(X_1, ..., X_n)$$

Colhida a amostra, teremos observado um valor particular de T (t_0). Baseados naquele valor é que faremos uma afirmação sobre θ.

Distribuições amostrais

A distribuição amostral da estatística T é obtida quando retiram-se todas as amostras da população segundo o plano amostral adotado.

Esquematicamente temos:

1. uma população X com parâmetro de interesse θ

Distribuições amostrais

A distribuição amostral da estatística T é obtida quando retiram-se todas as amostras da população segundo o plano amostral adotado.

Esquematicamente temos:

1. uma população X com parâmetro de interesse θ
2. todas as amostras retiradas da população (de acordo com o plano amostral)

Distribuições amostrais

A distribuição amostral da estatística T é obtida quando retiram-se todas as amostras da população segundo o plano amostral adotado.

Esquematicamente temos:

1. uma população X com parâmetro de interesse θ
2. todas as amostras retiradas da população (de acordo com o plano amostral)
3. para cada amostra, calcula-se o valor t da estatística T

Distribuições amostrais

A distribuição amostral da estatística T é obtida quando retiram-se todas as amostras da população segundo o plano amostral adotado.

Esquematicamente temos:

1. uma população X com parâmetro de interesse θ
2. todas as amostras retiradas da população (de acordo com o plano amostral)
3. para cada amostra, calcula-se o valor t da estatística T
4. os valores de t formam uma nova população, cuja distribuição é a distribuição amostral de T.

Distribuições amostrais

A distribuição amostral da estatística T é obtida quando retiram-se todas as amostras da população segundo o plano amostral adotado.

Esquematicamente temos:

1. uma população X com parâmetro de interesse θ
2. todas as amostras retiradas da população (de acordo com o plano amostral)
3. para cada amostra, calcula-se o valor t da estatística T
4. os valores de t formam uma nova população, cuja distribuição é a distribuição amostral de T.

Exemplo

Considere a população $\{1, 3, 5, 5, 7\}$

1. Construa a distribuição conjunta bidimensional (X_1, X_2)
2. calcule a distribuição da estatística $\bar{X} = \frac{X_1+X_2}{2}$.
3. calcule a distribuição da estatística $S^2 = \frac{\sum_{i=1}^{n}(X_i-\bar{X})^2}{n-1}$

Solução

a distribuição do par (X_1, X_2) é:

X_2	X_1				total
	1	3	5	7	
1	1/25	1/25	2/25	1/25	1/5
3	1/25	1/25	2/25	1/25	1/5
5	2/25	2/25	4/25	2/25	2/5
7	1/25	1/25	2/25	1/25	1/5
Total	1/5	1/5	2/5	1/5	1

Solução

Para cada par (X_1, X_2), calculamos a estatística $\bar{X}$ e a distribuição será:

$\bar{X}$	1	2	3	4	5	6	7	total
$p(\bar{X})$	1/25	2/25	5/25	6/25	6/25	4/25	1/25	1

Solução

Para cada par (X_1, X_2), calculamos a estatística S^2 e a distribuição será:

s^2	0	2	8	18	total
$p(S^2 = s^2)$	7/25	10/25	6/25	2/25	1

Usando a média amostral

Seja $X_1, ..., X_n$ uma a.a. de uma distribuição com média μ e desvio padrão σ. Então:

1. $E(\bar{X}) = \mu_{\bar{X}} = \mu$
2. $V(\bar{X}) = \sigma^2_{\bar{X}} = \frac{\sigma^2}{n}$

Em adição: Com $T_0 = X_1 + ... + X_n$,

$$E(T_0) = n\mu$$
$$V(T_0) = n\sigma^2, \quad \text{e}$$
$$\sigma_{T_0} = \sqrt{n}\sigma$$

Provas

$$E(\bar{X}) = E\left(\frac{\sum_{i=1}^{n} X_i}{n}\right) = \frac{1}{n}\left(E(X_1) + \ldots + E(X_n)\right)$$
$$= \frac{1}{n}\left(\mu + \ldots + \mu\right) = \mu$$

$$Var(\bar{X}) = Var\left(\frac{\sum_{i=1}^{n} X_i}{n}\right) = \frac{1}{n^2}\left(Var(X_1) + \ldots + Var(X_n)\right)$$
$$= \frac{1}{n^2}\left(\sigma^2 + \ldots + \sigma^2\right) = \frac{\sigma^2}{n}$$

Provas

$$E(\bar{X}) = E\left(\frac{\sum_{i=1}^{n} X_i}{n}\right) = \frac{1}{n}\left(E(X_1) + ... + E(X_n)\right)$$
$$= \frac{1}{n}\left(\mu + ... + \mu\right) = \mu$$

$$Var(\bar{X}) = Var\left(\frac{\sum_{i=1}^{n} X_i}{n}\right) = \frac{1}{n^2}\left(Var(X_1) + ... + Var(X_n)\right)$$
$$= \frac{1}{n^2}\left(\sigma^2 + ... + \sigma^2\right) = \frac{\sigma^2}{n}$$

O teorema central do limite.

Seja $X_1, ..., X_n$ uma a.a. de uma distribuição com média μ e variância σ^2. Então, se n é suficientemente grande, $\bar{X}$ terá aproximadamente uma distribuição normal com $\mu_{\bar{X}} = \mu$ e $\sigma^2_{\bar{X}} = \frac{\sigma^2}{n}$, e T_0 também tem uma distribuição aproximadamente normal, com parâmetros

$$\mu_{T_0} = n\mu,$$

$$\sigma_{T_0} = n\sigma^2$$

Quanto maior for o valor de n, melhor a aproximação.

Simulação do T.L.C

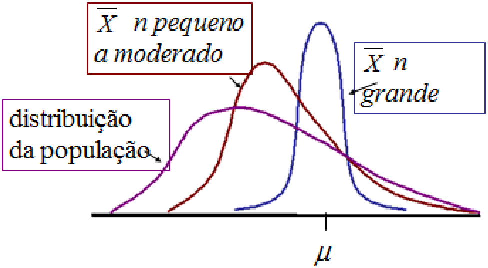


Figura: O Teorema do Limite Central

Regra de Ouro

Se $n > 30$, o T.L.C. pode ser utilizado.

Exemplo

Uma máquina enche pacotes de café, com distribuição normal com média 500 e variancia 100. Colhe-se uma amostra de $n = 100$ pacotes, os que são pesados. Qual a probabilidade de encontrarmos a média dos 100 pacotes diferindo de 500 em menos de dois gramas.

Solução

Tomamos $X \sim N(500, 100)$,pelo T.L.C, teremos que $X \sim N(500, 100/100 = 1)$. Portanto,

$$P(|\bar{X} - 500| < 2) = P(498 < \bar{X} < 502) = P(-2 < Z < 2) \approx 95\%$$

corolários

Corolário 1: Se $(X_1, ..., X_n)$ é uma a.a.s. da população X com média μ e variância σ^2 finita, então:

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

Corolário 2: se $e = \bar{X} - \mu$ então

$$e \sim N(0, \sigma^2/n)$$

corolários

Corolário 1: Se $(X_1, ..., X_n)$ é uma a.a.s. da população X com média μ e variância σ^2 finita, então:

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

Corolário 2: se $e = \bar{X} - \mu$ então

$$e \sim N(0, \sigma^2/n)$$

Distribuição amostral da proporção

Considere uma população em que a proporção de elementos portadores de certa característica é p. A v.a X é definidaa da seguinte maneira:

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

Temos que $\mu = E(X) = p$, $\sigma^2 = Var(X) = p(1-p)$

Distribuição amostral da proporção

Se retirarmos uma a.a.s. e indicando Y_n o total de indivíduos na amostra, portadores da característica, teremos que:

$$Y_n \sim bin(n, p)$$

Seja $\hat{p}$ a proporção de indivíduos na amostra, portadores da característica:

$$\hat{p} = \frac{Y_n}{n}$$

Então $P(Y_n = k) = P(\frac{Y_n}{n} = \frac{k}{n}) = P(\hat{p} = \frac{k}{n})$

Distribuição amostral da proporção (aproximação)

Seja a v.a $Y_n = X_1 + X_2 + ... + X_n$ onde cada X_i tem distribuição Bernoulli com média $\mu = p$ e variância $\sigma^2 = p(1-p)$ e são independentes entre si. Podemos escrever:

$$Y_n = n\bar{X}$$

pelo T.L.C, $\bar{X}$ tem distribuição normal com média p e variância $\frac{p(1-p)}{n}$, ou seja:

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

Distribuição amostral da proporção

Logo

$$Y_n = n\bar{X} \sim N(np, np(1-p))$$

observe que $\bar{X}$ é a própria $\hat{p}$ e, para n grande,

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

Determinação do tamanho da amostra

Suponha que desejamos estimar a média μ e para tanto usamos a média amostral $\bar{X}$, baseada numa amostra de tamanho n.
Suponha que desejamos determinar o valor de n tal que:

$$P(|\bar{X} - \mu| \leq \epsilon) \geq \gamma$$

com $0 < \gamma < 1$ e ϵ é o erro amostral máximo permissível (ambos dados).

Determinação do tamanho da amostra

Como $\bar{X} \sim N(\mu, \sigma^2/n)$, logo $\bar{X} - \mu \sim N(0, \sigma^2/n)$ e pode ser escrita como:

$$P(-\epsilon \leq \bar{X} - \mu \leq \epsilon) = P\left(\frac{-\sqrt{n}\epsilon}{\sigma} \leq Z \leq \frac{\sqrt{n}\epsilon}{\sigma}\right) \approx \gamma$$

Com $Z = (\bar{X} - \mu)\sqrt{n}/\sigma$. Dado γ, obtemos z_γ da $N(0,1)$, tal que $P(-z_\gamma < Z < z_\gamma) = \gamma$ de modo que

$$\frac{\sqrt{n}\epsilon}{\sigma} = z_\gamma$$

do que se obtêm finalmente $n = \frac{\sigma^2 z_\gamma^2}{\epsilon^2}$

Definição

Seja $X_1, ..., X_n$ uma a.a. de uma distribuição para a qual apenas são possíveis os valores positivos $[P(X_i > 0 = 1]$.
Então, para n suficientemente grande, o produto $Y = X_1.X_2...X_n$ tem aproximadamente uma distribuição log-normal

C.L.

Dada uma coleção de n v.a. $X_1, ..., X_n$ e n constantes numéricas $a_1, ..., a_n$, a v.a.

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

Isto é chamado de combinação linear dos X_i's

Esperança de uma C.L.

Seja $X_1, ..., X_n$ com médias $\mu_1, \mu_2, ..., \mu_n$ e variâncias de $\sigma_1^2, \sigma_2, 2, ..., \sigma_n^2$, respectivamente.
Ainda que os X_i's não sejam independentes,

$$\begin{aligned} E(a_1X_1 + ... + a_nX_n) &= a_1E(X_1) + ... + a_nE(X_n) \\ &= a_1\mu_1 + ... + a_nE(X_n) \end{aligned}$$

Variância de uma C.L.

Se $X_1, \dots, X_n$ são independentes,

$$\begin{aligned} V(a_1 X_1 + \dots + a_n X_n) &= a_1^2 V(X_1) + \dots + a_n^2 V(X_n) \\ &= a_1^2 \sigma_1^2 + \dots + a_n^2 \sigma_n^2 \end{aligned}$$

e

$$\sigma_{a_1 X_1 + \dots + a_n X_n} = \sqrt{a_1^2 \sigma_1^2 + \dots + a_n^2 \sigma_n^2}$$

Variância de uma C.L.

Para qualquer $X_1, ..., X_n$,

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

definição

$$E(X_1 - X_2) = E(X_1) - E(X_2)$$

e se X_1 e X_2 são independentes,

$$V(X_1 - X_2) = V(X_1) + V(X_2)$$

Diferença

Se $X_1, ..., X_n$ são v.a. independentes e normalmente distribuidas, então qualquer combinação linear dos X_i‘s também têm uma distribuição normal.
A diferença $X_1 - X_2$ entre duas v.a. independentes e normalmente distribuídas, é também normalmente distribuída.