

BIOESTATÍSTICA

Parte 3 – Variáveis Aleatórias

Aulas Teóricas de 29/03/2011 a 26/04/2011

3.1. Conceito de Variável Aleatória. Função de Distribuição

Variáveis aleatórias

Uma variável aleatória pode ser entendida como uma variável quantitativa [numérica], cujo valor depende de factores aleatórios, isto é, cujo valor resulta de um processo em que intervém o acaso.

Exemplos:

- Número obtido no lançamento de um dado;
- Número de comprimidos defeituosos numa amostra retirada de um lote e número de indivíduos infectados pelo vírus da gripe numa dada região, num certo período de tempo;
- Volume de água perdido, por dia, num sistema de abastecimento;
- Tempo de vida de um ser vivo de certa espécie.

A uma experiência aleatória pode-se associar um modelo de probabilidade que pressupõe a construção de um espaço de resultados e a atribuição de uma probabilidade a cada um dos acontecimentos elementares. Ora, nem sempre os resultados de uma experiência aleatória são numéricos (e.g. lançamento de uma moeda: cara ou coroa). [neste caso atribui-se valores a cara e a coroa, transformando a variável em numérica].

Além disso, frequentemente é preferível sintetizar os aspectos mais significativos dos resultados das experiências aleatórias através da definição de uma variável.

Ex: lançamento de uma moeda equilibrada

$\Omega = \{\text{cara, coroa}\}$

X – variável aleatória indicatriz do acontecimento $A = \{\text{coroa}\}$

A variável indicatriz indica se o acontecimento se realizou. $X=0$ se A se realiza e $X=1$ se B se realizou, etc.

Então: $P(X=0) = P(\text{sair cara}) = \frac{1}{2}$ e $P(X=1) = P(\text{sair coroa}) = \frac{1}{2}$

Ex: São escolhidos ao acaso 100 comprimidos de um grande lote e observa-se se cada comprimido é defeituoso ou não defeituoso.

O espaço de resultados Ω é constituído por 2^{100} elementos! É mais conveniente definir a variável X – número de comprimidos defeituosos (ou não defeituosos) em 100. Para representar variáveis aleatórias usam-se letras maiúsculas do final do

alfabeto: X, Y, Z, W ... Se X for uma variável aleatória (v.a.) representamos por x (minúsculo) um valor observado dessa v.a., ou seja, um valor que a v.a. pode assumir

[note-se que se usam letras maiúsculas do início do alfabeto para representar acontecimentos]

Variável aleatória $X \Rightarrow$ valor observado x
[variável aleatória $Y \Rightarrow$ valor observado y]

As variáveis aleatórias podem ser classificadas como discretas ou como contínuas.

Uma v.a. diz-se discreta se apenas assume um número finito ou infinito numerável de valores distintos. [associado a contagens]

Exemplos:

- Número obtido no lançamento de um dado;
- Número de comprimidos defeituosos numa amostra retirada de um lote.

Uma v.a. diz-se contínua se pode assumir qualquer valor de um intervalo da recta real, sendo nula a probabilidade de assumir um valor específico. [associado a medições]

Exemplos:

- Nível de glicose no sangue de um indivíduo numa dada população;
- Tempo até ser atingida a concentração máxima de um fármaco na circulação sistémica.

Se X é uma via contínua então:

$P(X = a) = 0$ para qualquer real a

[De forma não rigorosa, pode-se pensar: $P(x=a) = \frac{\text{casos favoráveis}}{\text{casos possíveis}} = \frac{1}{\infty} = 0$]

[Para variáveis contínuas calcula-se a probabilidade de intervalos de valores, por exemplo, $P(X > a)$, $P(X \leq a)$, $P(a < X < b)$]

Função de Distribuição

Uma forma de definir uma variável aleatória, discreta ou indiscreta, consiste na utilização de uma função real de variável real $[\mathbb{R} \rightarrow \mathbb{R}]$ designada por função de distribuição.

Função de distribuição (f.d.) de uma variável aleatória X (discreta ou contínua) é a função definida por:

$$\begin{array}{c} F(x) = P(X \leq x) \text{ para qualquer } x \in \mathbb{R} \\ \downarrow \\ \text{domínio} \end{array}$$

A função de distribuição é cumulativa, representa a probabilidade acumulada até x . Notemos que $0 \leq F(x) \leq 1$.

[Apesar de ser possível calcular f.d. para variáveis contínuas e discretas, o seu cálculo é diferente para umas e outras. Esta função é mais interessante para variáveis contínuas, pois em alguns destes casos assume uma expressão analítica. Para variáveis discretas soma-se a probabilidade até ao ponto x (semelhante às frequências acumuladas) e para variáveis contínuas calculam-se através de integrais – interpretação através de áreas]

Seja X uma v.a. com f.d. $F(X)$. Então, dados os pontos a e b , tais que $a < b$, tem-se que:

$$P(a < X \leq b) = F(b) - F(a)$$

O conhecimento da função de distribuição possibilita, de forma expedita, o cálculo da probabilidade de uma v.a. assumir valores num dado intervalo. Notemos que, se a v.a. for contínua:

$$P(a < X \leq b) = P(a < X < b) = P(a \leq X \leq b) = P(a \leq X < b) = F(b) - F(a)$$

Se a v.a. X for contínua estas probabilidades são todas iguais! [o que não é verdade para variáveis discretas].

3.2. Variáveis Aleatórias Discretas. Função Massa de Probabilidade

É natural definir a probabilidade de uma variável aleatória assumir um determinado valor como sendo a probabilidade do acontecimento que fez com que a v.a. tivesse esse valor.

Chama-se função massa da probabilidade (f.m.p.) à função que, a cada valor numérico que a v.a. assume, faz corresponder a probabilidade associada a esse valor.

Uma v.a. X discreta fica perfeitamente identificada pela sua f.m.p., ie, pelos valores de x que a v.a. pode tomar e pelas probabilidades com que assume cada um desses valores.

$$X = \begin{cases} x_i \\ P(i) = P(X=x_i) \end{cases}$$

Atendendo à definição de probabilidade, é imediato que qualquer f.m.p., goza das seguintes probabilidades:

1. $P_i \geq 0$
2. $\sum_i P_i = 1$

$$X = \begin{cases} \begin{array}{|c|c|c|c|} \hline -1 & 0 & 1 & 2 \\ \hline 0,3 & 0,1 & a & 0,4 \\ \hline \end{array} \end{cases} \Rightarrow a = 0,2$$

[Ex: lançamento de uma moeda:

$$X = \begin{cases} 0 & 1 \\ \frac{1}{2} & \frac{1}{2} \end{cases} \quad \underline{\text{ou}} \quad \begin{array}{|c|c|c|} \hline x & 0 & 1 \\ \hline P(X=x) & \frac{1}{2} & \frac{1}{2} \\ \hline \end{array}$$

Nota: só se inserem nas tabelas os valores com probabilidade não nula]

Ex: Uma urna que contém 4 bolas numeradas de 1 a 4. Extraem-se ao acaso 2 bolas de uma urna. Defina a v.a. que representa a soma dos números observados nas bolas extraídas.

(i) Supondo que a extracção é feita com reposição:

$$\Omega = \{(1,1), (1,2), (1,3), (1,4), (2,1), (2,2), (2,3), (2,4), (3,1), (3,2), (3,3), (3,4), (4,1), (4,2), (4,3), (4,4)\}$$

$$X = \begin{cases} 2 & 3 & 4 & 5 & 6 & 7 & 8 \\ \frac{1}{16} & \frac{1}{8} & \frac{3}{4} & \frac{1}{4} & \frac{3}{16} & \frac{1}{8} & \frac{1}{16} \end{cases}$$

(ii) Supondo que a extracção é feita sem reposição:

$$\Omega = \{(1,2), (1,3), (1,4), (2,1), (2,3), (2,4), (3,1), (3,2), (3,4), (4,1), (4,2), (4,3)\}$$

$$X = \begin{cases} 3 & 4 & 5 & 6 & 7 \\ \frac{1}{6} & \frac{1}{6} & \frac{1}{3} & \frac{1}{6} & \frac{1}{6} \end{cases}$$

3.3. Variáveis Aleatórias Contínuas. Função Densidade de Probabilidade

Define-se **função densidade de probabilidade (f.d.p.)** de uma variável aleatória X contínua e representa-se por $f(x)$ como sendo a derivada, se existir, da função de distribuição $F(x)$. [note-se que f.d. representa-se por letras maiúsculas e f.d.p. por letras minúsculas]

$$f(x) = \frac{d F(x)}{dx}$$

O que é equivalente a:

$$F(x) = \int_{-\infty}^x f(x)dt$$

visto que $F(-\infty) = 0$

Notação que significa limite

[O domínio $]-\infty, x]$ é explicado pois $F(x) = P(X \leq x) \Rightarrow P(x \in]-\infty, x])$

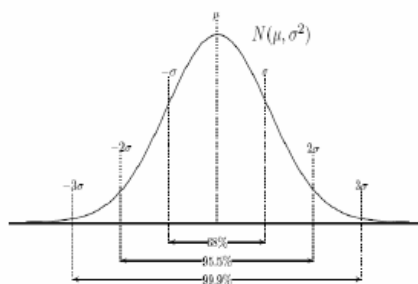
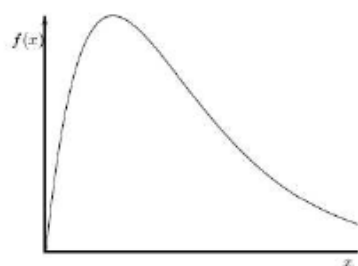
f.m.p. da função discreta é a correspondente à f.d.p. da variável aleatória contínua]

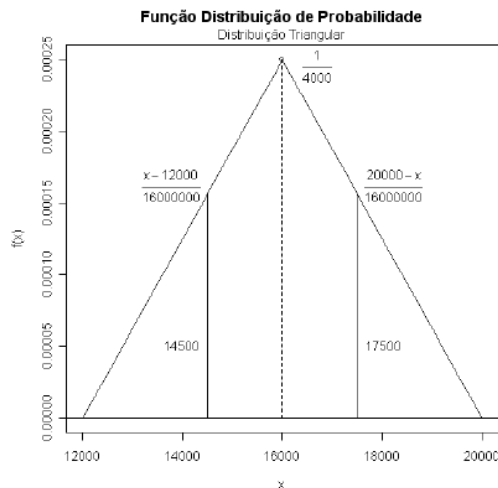
Nos pontos em que não existe derivada de f.d., é habitual considerar que $f(x) = 0$. Quando a v.a. assume valores apenas num subconjunto de \mathbb{R} , admite-se que $f(x) = 0$ no exterior daquele subconjunto.

Propriedades de f.d.p.:

- 1) $f(x) \geq 0$;
- 2) $\int_{-\infty}^{+\infty} f(x)dx = 1$ [$\Rightarrow A = 1$]

[A função densidade de probabilidade está sempre acima dos eixos dos xx. Possíveis gráficos:

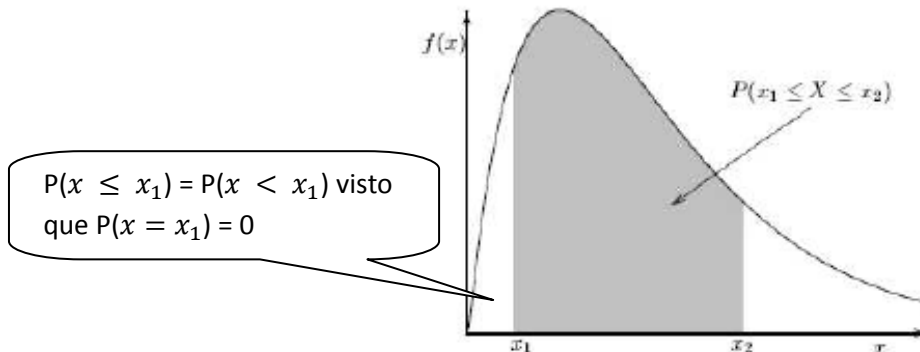




Notemos que para $a < b$:

$$\int_a^b f(x)dx = F(b) - F(a) = P(a < X \leq b)$$

O cálculo do integral anterior é equivalente a calcular o valor da área compreendida entre o eixo das abcissas, o gráfico da f.d.p. e as rectas $x = a$ e $x = b$.



3.4. Características Populacionais

Vamos agora definir medidas de localização e dispersão que permitem identificar muitas das características fundamentais de uma população representada pela v.a. X . São os chamados parâmetros populacionais, dos quais depende a distribuição de X .

Parâmetro: quantidade numérica fixa, embora por vezes seja desconhecida

- [- nunca são aleatórios;
- representados com frequência por letras gregas]
- por vezes são desconhecidos;

Valor Médio

[corresponde à média]

Define-se valor médio ou valor esperado de X e representa-se por $E(X)$ ou μ_X como sendo:

- Seja X uma v.a. discreta que assume valores x_i com probabilidade p_i , $i=1,2,\dots$

Então

$$E(X) = \sum_i x_i p_i$$

[é uma espécie de média ponderada]

$$\text{quando } \sum_i |x_i| p_i < \infty$$

[ou seja, quando a série é convergente]

O valor esperado de uma v.a. discreta não é necessariamente um valor que a v.a. pode assumir com probabilidade não nula.

Ex: Seja X a v.a. que representa o número de pontos obtidos no lançamento de um dado equilibrado.

$$E(X) = 1 \times 1/6 + 2 \times 1/6 + 3 \times 1/6 + 4 \times 1/6 + 5 \times 1/6 + 6 \times 1/6 = 3.5$$

[$E(X)$ é um valor não obtido no lançamento do dado, mas não se arredonda]

- Seja X uma v.a. contínua com f.d.p. $f(x)$. Então

$$E(X) = \int_{-\infty}^{+\infty} x f(x) dx$$

quando

$$\int_{-\infty}^{+\infty} |x| f(x) dx < \infty$$

O valor médio corresponde ao centro de gravidade da distribuição de probabilidades. É um parâmetro de localização do centro da distribuição.

Valor médio de uma função de uma variável aleatória

Seja X uma v.a. e $Y = \psi(X)$ uma v.a. que é função de X . Então

- Se X é uma v.a. discreta que assume valores x_i com probabilidade p_i , $i=1,2,\dots$

$$E(Y) = E(\psi(X)) = \sum_i \psi(x_i) p_i$$

- Se X é uma v.a. contínua com f.d.p. $f(x)$ e supondo que $\psi(X)$ é ainda uma v.a. contínua

$$E(Y) = E(\psi(X)) = \int_{-\infty}^{+\infty} \psi(x) f(x) dx$$

Supomos em ambos os casos que o valor médio de $Y = \psi(X)$ existe. Notemos que a existência de $E(X)$ não implica a existência de $E[\psi(X)]$ e inversamente.

Propriedades do valor médio:

1. Dadas as v.a. X e Y com valores médios $E(X)$ e $E(Y)$, respectivamente, então

$$E(X + Y) = E(X) + E(Y)$$

2. Dada a v.a. X e as constantes a e b , tem-se

$$E(aX + b) = a E(X) + b \quad \text{[visto que } E(b) = \text{]}$$

3. Generalizando as propriedades anteriores, se existir $E(X_j)$, $j=1,\dots, n$, então

$$E\left(\sum_{j=1}^n a_j X_j\right) = \sum_{j=1}^n a_j E(X_j)$$

Quantil de probabilidade p ($0 < p < 1$)

Vamos considerar apenas o caso de uma v.a. X contínua, visto que são estes os quantis que têm mais interesse do ponto de vista prático.

Define-se quantil de probabilidade p de uma v.a. X contínua como sendo o número real x tal que

$$F(x) = p \Leftrightarrow P(X \leq x) = p$$

O quantil de probabilidade p representa-se genericamente por χ_p

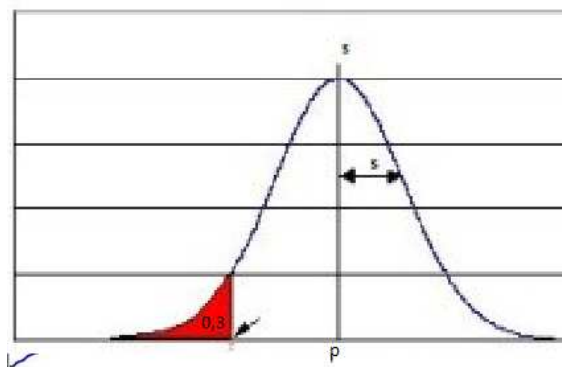
É também uma medida de localização.

Quando $p = \frac{1}{2}$ obtem-se a mediana de X

$p = \frac{1}{4}$ obtem-se o 1º quantil e para $p = \frac{3}{4}$ obtem-se o 3º quantil de X

Quando $p = 0.1, \dots, 0.9$ obtêm-se os decis; para $p = 0.01, \dots, 0.99$ obtêm-se os percentis

[Exemplo:



Nota: gráfico representa f.d.p.

Quantil de probabilidade 0,3 = b

$$p = F(b) = 0,3]$$

Variância (populacional)

Define-se variância da v.a. X e representa-se por $\text{var}(X)$ ou σ^2 como sendo

$$\text{var}(X) = E[(X - E(X))^2]$$

Facilmente daqui se deduz uma expressão mais simples para o cálculo da variância:

$$\text{var}(X) = E(X^2) - E^2(X)$$

Propriedades da variância:

1. $\text{var}(X) \geq 0$, sendo que a variância de uma constante é zero
2. Dadas as constantes a e b , tem-se que $\text{var}(aX+b) = a^2 \text{var}(X)$

A variância de X é uma medida da dispersão da distribuição de X relativamente ao seu valor médio.

Desvio Padrão (populacional)

É por vezes mais conveniente usar como parâmetro de dispersão a raiz quadrada da variância, visto que assim estaremos a usar uma medida que se exprime nas mesmas unidades da variável.

Então, define-se desvio padrão da v.a. X como

$$\sigma = \sqrt{\text{var}(X)}$$

O desvio padrão mede a variabilidade da população relativamente ao seu valor médio.

Coefficiente de Variação (populacional)

O coeficiente de variação da v.a. X é igual, se existir, ao quociente entre o desvio padrão e o valor médio de X

$$CV = \sigma/\mu$$

O coeficiente de variação é uma grandeza que não depende de qualquer unidade de medida e é, portanto, muito utilizada para comparar a dispersão relativa de variáveis aleatórias.

Exemplo do cálculo do valor médio e variância de uma variável aleatória discreta

X - nº de pacientes hipertensos, em 4, cuja tensão arterial fica controlada após a administração de um certo fármaco

k	0	1	2	3	4
P(X=k)	0.008	0.076	0.265	0.411	0.240

$$E(X) = 0 \times 0.008 + 1 \times 0.076 + 2 \times 0.265 + 3 \times 0.411 + 4 \times 0.24 = 2.799 \approx 2.8$$

$$E(X^2) = 0^2 \times 0.008 + 1^2 \times 0.076 + 2^2 \times 0.265 + 3^2 \times 0.411 + 4^2 \times 0.24 = 8.675$$

$$\text{var}(X) = E(X^2) - E^2(X) = 8.675 - 2.799^2 = 0.841$$

$$\sigma_X = \sqrt{\text{Var}(X)} = \sqrt{0.841} = 0.917$$

Par aleatório Discreto

Por vezes, é necessário descrever o resultado de uma experiência aleatória recorrendo a mais do que uma variável. Vamos considerar o caso de **duas variáveis aleatórias discretas** e vamos estudar o comportamento conjunto dessas variáveis.

Sejam X e Y as duas v.a.'s discretas de interesse. Um par aleatório discreto (X, Y) fica perfeitamente identificado pela sua **função massa de probabilidade conjunta**, i.e., pelos valores (x_i, y_j) que o par de v.a.'s assume e pelas probabilidades de assumir esses valores.

$$p_{ij} = P(X = x_i \cap Y = y_j) = P(X = x_i, Y = y_j) \quad i = 1, \dots, k; j = 1, \dots, m$$

A f.m.p. do par aleatório (X, Y) costuma representar-se numa tabela:

$X \backslash Y$	y_1	\dots	y_j	\dots	y_m	
x_1	p_{11}		p_{1j}		p_{1m}	$p_{1.}$
\dots		\ddots				\dots
x_i	p_{i1}		p_{ij}		p_{im}	$p_{i.}$
\dots			\ddots			\dots
x_k	p_{k1}		p_{kj}		p_{km}	$p_{k.}$
	$p_{.1}$	\dots	$p_{.j}$	\dots	$p_{.m}$	1

f.m.p. marginal de Y

f.m.p. marginal de X

Nota: A célula p_{ij} está rotulada como $x_i \cap y_j$.

A f.m.p. do par (X, Y) goza das seguintes propriedades:

1) $p_{ij} \geq 0 \quad \forall i, j.$

2) $\sum_i \sum_j p_{ij} = 1$ [a soma das probabilidades é zero]

Exemplo:

Uma farmácia tem duas marcas (A e B) de certo medicamento para tratamento de uma doença pouco frequente. Considere-se o par aleatório (X, Y) em que

X - nº de caixas da marca A vendidas por semana

Y - nº de caixas da marca B vendidas por semana

$\begin{matrix} Y \\ X \end{matrix}$	1	2	3	$p_{i.}$
0	0.10	0.15	0.10	0.35
1	0.15	0.20	0.10	0.45
2	0.15	0.05	0	0.20
$p_{.j}$	0.40	0.40	0.20	1

Qual a probabilidade de se venderem 2 caixas de A?

Qual a probabilidade de se vender 1 caixa de B?

Qual a probabilidade de se venderem 2 caixas de A e 1 de B?

Variáveis Aleatórias Independentes:

X e Y são variáveis aleatórias (discretas) independentes se e só se

[, = ∩]

$$P(X = x_i, Y = y_j) = P(X = x_i) \times P(Y = y_j) \quad \forall (x_i, y_j)$$

ou seja, para todo o par de valores (x_i, y_j) em que (X, Y) está definido.

[Nota: A e B são independentes se e só se $P(A \cap B) = P(A) \cdot P(B)$]

[se existir apenas um par de valores em que a condição não se verifica as variáveis são dependentes;

Variáveis cuja intersecção de um par de valores é zero não são independentes]

No exemplo anterior, X e Y não são variáveis aleatórias independentes.

Por exemplo: $P(X=2, Y=3) \neq P(X=2)P(Y=3) \Leftrightarrow 0 \neq 0.2 \times 0.2$

Covariância

Dadas duas v.a. X e Y, existe uma medida do grau de intensidade com que as v.a.'s se associam (linearmente) que se designa por covariância entre X e Y

$$\text{cov}(X, Y) = E[(X - E(X))(Y - E(Y))]$$

Pelas propriedades do valor médio, obtém-se uma expressão equivalente:

$$\text{cov}(X, Y) = E(XY) - E(X)E(Y)$$

1. Para quaisquer v.a.'s X e Y tem-se que

$$\text{var}(X \pm Y) = \text{var}(X) + \text{var}(Y) \pm 2\text{cov}(X, Y)$$

2. Se as v.a.'s X e Y são independentes, então $\text{cov}(X, Y) = 0$.

Se X e Y são v.a.'s independentes, então

$$E(XY) = E(X) E(Y)$$

$$\text{var}(X + Y) = \text{var}(X) + \text{var}(Y)$$

$$\text{var}(X - Y) = \text{var}(X) + \text{var}(Y)$$

[$\text{cov}(X, Y) = 0$ não implica que as v.a. sejam independentes, mas $\text{cov}(X, Y) \neq 0$ implica que as variáveis não sejam independentes]

Quando X e Y não são variáveis aleatórias independentes, pode ainda acontecer que $\text{cov}(X, Y) = 0$

Ex:

$X \backslash Y$	-1	0	1
0	0.1	0.1	0.1
2	0.1	0.2	0.1
4	0.1	0.1	0.1

$$E(XY) = \sum_i \sum_j x_i y_j p_{ij} = 0 + 0 + 0 - 0.2 + 0 + 0.2 - 0.4 + 0 + 0.4 = 0$$

$$\text{Cov}(X, Y) = E(XY) - E(X) E(Y) = 0 - 2 \times 0 = 0$$

Coefficiente de Correlação

A covariância depende das unidades em que se exprimem as v.a. X e Y . Assim, é conveniente introduzir um novo parâmetro, não dependente dessas unidades, para medir o grau de associação linear entre X e Y .

O coeficiente de correlação entre as v.a. X e Y é dado por

$$\rho = \frac{\text{cov}(X, Y)}{\sqrt{\text{var}(X) \text{var}(Y)}} = \frac{\text{cov}(X, Y)}{\sigma_X \sigma_Y}$$

Pode demonstrar-se que ρ toma valores no intervalo $[-1,1]$. Além disso,

- um valor de ρ próximo de 1 ou de -1 significa uma forte associação linear entre X e Y
- um valor de ρ próximo de zero significa que a associação linear entre X e Y não existe ou é muito fraca. Neste caso as v.a. X e Y podem estar correlacionadas não linearmente.

Retomemos o exemplo da venda do medicamento:

X \ Y	1	2	3	$P_{i.}$
0	0.10	0.15	0.10	0.35
1	0.15	0.20	0.10	0.45
2	0.15	0.05	0	0.20
$P_{.j}$	0.40	0.40	0.20	1

X e Y não são independentes
por exemplo, $P(X=2, Y=1)=0.15$
que é diferente de
 $P(X=2)P(Y=1) = 0.20 \times 0.40 = 0.08$

Cálculo da covariância e do coeficiente de correlação:

$$E(X) = 0 \times 0.35 + 1 \times 0.45 + 2 \times 0.20 = 0.85$$

$$E(X^2) = 0^2 \times 0.35 + 1^2 \times 0.45 + 2^2 \times 0.20 = 1.25$$

$$\text{Var}(X) = E(X^2) - E^2(X) = 1.25 - 0.85^2 = 0.5275$$

$$\sigma_X = 0.726$$

$$E(XY) = 0 \times 1 \times 0.1 + 0 \times 2 \times 0.15 + \dots + 2 \times 3 \times 0 = 1.35$$

$$E(Y) = 1.8 \quad \sigma_Y = 0.748$$

$$\text{cov}(X, Y) = 1.35 - 0.85 \times 1.8 = -0.18$$

$$\rho = \frac{-0.18}{0.726 \times 0.748} = -0.33$$

Soma de Variáveis Aleatórias

Dadas duas v.a.'s X e Y podemos construir uma nova v.a. X+Y.

Vamos ilustrar o procedimento, usando o par aleatório do exemplo da venda do medicamento

X \ Y	1	2	3	$P_{i.}$
0	0.10	0.15	0.10	0.35
1	0.15	0.20	0.10	0.45
2	0.15	0.05	0	0.20
$P_{.j}$	0.40	0.40	0.20	1

Qual a probabilidade de serem vendidas 2 caixas do medicamento por semana, isto é, qual a probabilidade de que $X+Y=2$?

$$P(X+Y=2) =$$

$$= P(X=0, Y=2) + P(X=1, Y=1)$$

$$= 0.15 + 0.15 = 0.30$$

Verifique que

X+Y	1	2	3	4
Prob.	0.10	0.30	0.45	0.15

$$\begin{aligned}E(X+Y) &= 2.65 = E(X) + E(Y) \\Var(X+Y) &= 0.7275 \neq Var(X) + Var(Y) \\ \sigma_{X+Y} &= 0.853\end{aligned}$$

3.5. Modelos Discretos: binomial, Poisson e hipergeométrica

Vamos agora estudar alguns modelos probabilísticos com grande aplicação na prática.

Modelo Binomial

Consideremos uma experiência aleatória com as seguintes características:

- A experiência é constituída por n provas, entendendo-se por prova uma repetição em condições idênticas;
- São provas independentes;
- Em cada prova ocorre apenas um de dois resultados possíveis [estuda-se apenas um acontecimento], designados por sucesso ou insucesso. A probabilidade de sucesso mantém-se constante de prova para prova e representa-se por p [letra minúscula].

[sucesso \Rightarrow ocorreu o acontecimento estudado;

Insucesso \Rightarrow não ocorreu o acontecimento estudado]

Uma sucessão de provas com estas características designa-se por sucessão de provas de Bernoulli:

Exemplos:

- Lançamento de uma moeda: cara (sucesso), coroa (insucesso) [quer seja a moeda equilibrada ou não];
- Observação de comprimidos de um lote: comprimido defeituoso (sucesso), comprimido não defeituoso (insucesso);
- Resultado de um tratamento para uma certa doença aplicado a pacientes em condições bem definidas: paciente curado (sucesso), paciente não curado (insucesso) [a independência resulta da escolha dos pacientes ser aleatória].

Seja X a v.a. que representa o número de sucessos em n provas de Bernoulli e em que a probabilidade de sucesso é p . Então a f.m.p. é:

$$P(X = K) = \binom{n}{K} p^K (1-p)^{n-K}$$

[equivalente a C_K^n]

$$K = 0, 1, 2, \dots, n$$

[Número de sucessos]

X é uma v.a. binomial de parâmetros n e p e representa-se por:

$$X \sim \text{Bi}(n, p)$$



[não é uma intersecção;

Lê-se X tem distribuição binomial de parâmetros n e p]

[Exemplo:

Foram feitos 3 lançamentos de uma moeda não equilibrada:

$$P(\text{cara} = F) = \frac{2}{3}$$

$$P(\text{coroa} = C) = \frac{1}{3}$$

X – número de coroas que saem em 3 lançamentos

$$P(X=0) = \left(\frac{2}{3}\right)^3 = \frac{8}{27}$$

$$P(X=2) = 3 \times \frac{2}{3} \times \left(\frac{1}{3}\right)^2 = \frac{6}{27}$$

$$P(X=1) = 3 \times \left(\frac{2}{3}\right)^2 \times \frac{1}{3} = \frac{12}{27}$$

$$P(X=3) = \left(\frac{1}{3}\right)^3 = \frac{1}{27}$$

Recordar que:

$$C_k^n = \binom{n}{k} = \frac{n!}{k! (n-k)!}$$

$$0! = 1$$

Mostra-se que:

$$E(X) = np \quad \text{e} \quad \text{var}(X) = np(1-p)$$

[Número médio de sucessos em n provas]

Caso particular: quando $n=1$, X é designado por v.a. de Bernoulli

[é a variável indicatriz]

Amostragem com reposição:

Consideremos o processo de amostragem em que, para cada indivíduo recolhido aleatoriamente na população, se verifica se tem ou não uma determinada característica, repondo o elemento antes de proceder a uma nova extracção. Então, estamos em condições de aplicar o modelo Binomial quando se quer estudar a v.a. que representa o número de indivíduos da amostra assim obtida que têm a dita característica.

[mesmo que não se faça reposição em determinadas condições pode-se utilizar o modelo binomial, se a dimensão da amostra for muito superior à probabilidade considerada]

Soma de v.a. independentes com distribuição binomial:

Dadas as v.a. x_1, \dots, x_k independentes com distribuição binomial

$$X_i \sim \text{Bi}(n_i, p), \quad i=1, \dots, k$$

A soma $S_k = X_1 + \dots + X_k$ também tem distribuição binomial

$$S_i \sim \text{Bi}\left(\sum_{i=1}^k n_i, p\right) \quad i = 1, \dots, k$$

[para isto ser válido o parâmetro p tem de ser igual para todas as v.a.]

Exemplo: lançamento de uma moeda. Seja:

X_1 – número de caras em 5 lançamentos;

X_2 – número de caras em 19 lançamentos;

X_3 – número de caras em 3 lançamentos.

Uma vez que as variáveis têm a mesma probabilidade p de sucesso, a variável soma também é binomial]

Distribuição Geométrica

Uma variável aleatória X diz-se geométrica de parâmetros p se representar o número de provas necessárias até à obtenção do primeiro sucesso (inclusive), numa sucessão de provas de Bernoulli em que a probabilidade de sucesso é p .

[É um caso particular de uma outra distribuição – Distribuição Binomial Negativa – em que se considera o número de provas necessárias até à obtenção do n -ésimo sucesso (terceiro, quinto, décimo ...)]

A f.m.p. de X é dada por:

$$P(X = k) = p(1-p)^{k-1}$$

$$k = 1, 2, \dots \quad [k \in [1, \infty[]$$

Mostra-se que:

$$E(X) = \frac{1}{p}$$

$$\text{Var}(X) = \frac{1-p}{p^2}$$

[Exemplo: Numa urna estão 20 bolas, das quais 5 são brancas. Qual a probabilidade de a bola branca sair pela primeira vez à terceira extração?

$$P(X=3) = \frac{1}{4} \times \left(1 - \frac{1}{4}\right)^{3-1} = \frac{9}{64}$$

Neste exemplo, $E(X) = 4$

Distribuição Poisson

Vamos agora considerar um modelo discreto que se aplica em situações em que interessa estudar o número de ocorrências de um certo acontecimento, num determinado intervalo de tempo ou região do espaço.

[ex: número de chamadas recebidas num call center em 15 minutos]

Suponham-se que se verificam as seguintes hipóteses (processo de Poisson):

- A ocorrência do acontecimento num determinado intervalo é independente da ocorrência do acontecimento em qualquer intervalo distinto;
- A probabilidade de exactamente uma ocorrência do acontecimento em qualquer intervalo de amplitude h arbitrariamente pequeno é λh ;
- A probabilidade de duas ou mais ocorrências do acontecimento em qualquer intervalo de amplitude h é aproximadamente 0.

Seja X a variável aleatória que representa o número de ocorrências do acontecimento num intervalo unitário [seja 15 minutos, 1 cm³, ou outro intervalo considerado]. Então dizemos que λ tem distribuição de Poisson de parâmetro λ com f.m.p. dada por:

$$P(X = k) = e^{-\lambda} \frac{\lambda^k}{k!} \quad k = 0, 1, 2, \dots \quad \lambda > 0$$

Representa-se por $X \sim P(\lambda)$

O parâmetro λ é o número médio de ocorrências por intervalo de tempo:

$$E(X) = \lambda \quad \text{e} \quad \text{var}(X) = \lambda$$

[no exemplo acima λ é o número médio de chamadas recebidas em 15 minutos]

Exemplos de aplicação da distribuição de Poisson:

- Número de novos casos de hepatite numa certa região num mês;
- Número de clientes que chegam a um balcão de atendimento num certo intervalo de tempo;
- Número de erros de impressão por página de livro;
- Número de chamadas telefônicas recebidas num *call center* numa hora;
- Número de bactérias num determinado volume.

Cada intervalo unitário é dividido em h subintervalos, de modo que cada subintervalo tenha uma amplitude arbitrariamente pequena. Então em cada um desses subintervalos:

- A probabilidade de exactamente uma ocorrência do acontecimento é $\frac{\lambda}{h}$;
- A probabilidade de 2 ou mais ocorrências do acontecimento é aproximadamente 0;

Então, a probabilidade de k ocorrências do acontecimento no intervalo é para k fixo.

[ou seja, ou ocorre o acontecimento ou não ocorre nada, como na distribuição binomial.

$$P(X=k) \approx \binom{n}{k} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k}$$

Se $n \rightarrow \infty$, tem-se

$$\lim_{n \rightarrow \infty} \frac{n(n-1) \dots (n-k+1)}{n^k} \frac{\lambda^k}{k!} \left(1 - \frac{\lambda}{n}\right)^n \left(1 - \frac{\lambda}{n}\right)^{-k} = 1 \cdot e^{-\lambda} \cdot 1$$

A distribuição Binomial $Bi(n,p)$ converge para a distribuição de Poisson $P(\lambda)$ quando:

- $n \rightarrow \infty$ (o número de provas aumenta) [a partir de 20 ou 30 provas]
- $p \rightarrow 0$ (a probabilidade de sucesso tende para zero)
- $np = \lambda$ (o número médio de sucesso mantém-se constante de prova para prova)

O modelo de Poisson surge assim como limite do modelo binomial, nas condições anteriores, por isso é designado por Lei dos Acontecimentos Raros.

Soma de v.a. independentes com distribuição poisson:

Dadas as v.a. x_1, \dots, x_n independentes com distribuição de Poisson

$$X_i \cap P(\lambda), \quad i=1, \dots, n$$

A soma $S_k = X_1 + \dots + X_n$ também tem distribuição de Poisson

$$S_n \cap P\left(\sum_{i=1}^k \lambda_i\right) \quad i = 1, \dots, n$$

Distribuição Hipergeométrica

Quando se procede a extracções sem reposição numa população finita, o que podemos dizer quanto à variável aleatória que representa o número de elementos da amostra assim obtida que possuem determinada característica?

Consideremos uma população de N elementos dos quais M possuem determinada característica (sucesso), enquanto que os restantes $N - M$ não a têm. Seja n a dimensão de uma amostra extraída da população (sem reposição) e X a v.a. que representa o número de elementos da amostra que possuem a dita característica. Então a v.a. X tem distribuição hipergeométrica dada por:

$$P(X = k) = \frac{\binom{M}{k} \binom{N-M}{n-k}}{\binom{N}{n}}$$

$$K = \text{Max. } (0, n(N-M)), \dots, \min(M, n)$$

[Exemplo: num lote de 12 caixas de comprimidos sabe-se que 3 estão danificadas e 9 estão normais. Retiraram-se 4 caixas sem reposição. Qual a probabilidade de 2 dessas caixas serem danificadas?

X – número de caixas danificadas em 4 retiradas aleatoriamente sem reposição

The diagram illustrates the calculation of the probability $P(X = 2)$ for a hypergeometric distribution. The formula is shown as $P(X = 2) = \frac{\binom{3}{2} \binom{9}{2}}{\binom{12}{4}} = 0,218$. Three callout boxes explain the terms:

- Top-left callout:** "De quantas maneiras posso retirar 2 caixas em 3 danificadas" (How many ways can I withdraw 2 boxes from 3 damaged ones). This corresponds to the numerator term $\binom{3}{2}$.
- Top-right callout:** "De quantas maneiras posso retirar 2 caixas em 9 normais" (How many ways can I withdraw 2 boxes from 9 normal ones). This corresponds to the numerator term $\binom{9}{2}$.
- Bottom callout:** "De quantas maneiras posso retirar 4 caixas em 12" (How many ways can I withdraw 4 boxes from 12). This corresponds to the denominator term $\binom{12}{4}$.

Nota: é sempre necessário conhecer a distribuição da população]

Os parâmetros são N , n e p — (que é a probabilidade de que um elemento extraído ao acaso da população possua a característica).

$$E(X) = np$$

$$\text{Var}(X) = np(1-p)$$

O modelo hipergeométrico é adequado quando se procede a extracções sem reposição:

- as extracções não são independentes;
- a probabilidade de sucesso varia de extracção para extracção.

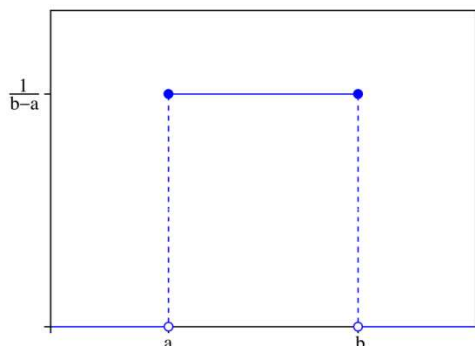
No entanto, quando se procede a extracções sem reposição, se a dimensão N da população for muito maior que a dimensão n da amostra, o modelo binomial pode ser usado para representar o número de elementos da amostra que possuem determinada característica. De facto, neste caso, as sucessivas extracções podem ser consideradas independentes e a probabilidade de sucesso não se altera substancialmente de extracção para extracção.

3.6. Modelos Contínuos: uniforme, exponencial, o modelo gaussiano e o Teorema Limite Central

Distribuição Uniforme

X tem distribuição uniforme no intervalo (a, b) se a sua f.m.p. for dada por:

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{se } a < x < b \\ 0 & \text{se } x \leq a \text{ ou } x \geq b \end{cases}$$



[Demonstração:

Função uniforme $\Rightarrow f(x) = k$ se

$f(x) = 0$ se

$$\Leftrightarrow (b-a)k = 1 \Leftrightarrow k = \frac{1}{b-a}]$$

Simbolicamente, representa-se por $X \sim U(a,b)$.

A função de distribuição é dada por:

$$F(x) = \begin{cases} 0 & \text{se } x < a \\ \frac{x-a}{b-a} & \text{se } a \leq x \leq b \\ 1 & \text{se } x \geq b \end{cases}$$

Então, se $X \cap U(a,b)$, temos que:

$$E(X) = \frac{a+b}{2} \Rightarrow [\text{se a distribuição for uniforme o valor médio é o ponto médio}]$$

$$\text{Var}(X) = \frac{(b-a)^2}{12}$$

Distribuição Normal (ou Gaussiana)

A distribuição normal é a mais conhecida das distribuições contínuas e é uma das mais importantes.

Do ponto de vista das aplicações empíricas, tem-se comprovado que muitas das características observáveis de determinada população são bem representadas por variáveis aleatórias que seguem uma distribuição normal.

Exemplo: distribuição da altura ou do peso dos indivíduos em populações razoavelmente homogêneas.

Por outro lado, prova-se que a distribuição da soma (ou da média) de um número suficientemente grande de variáveis aleatórias s.i.d.d. [independentes e identicamente distribuídas, ou seja, com o mesmo parâmetro] é bem aproximada à distribuição normal, o que permite justificar a sua utilização em muitas situações. Tem também um papel importante na inferência estatística.

Uma v.a. X tem distribuição normal de parâmetro μ e σ se a sua f.m.p. é dada por:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right]$$

$$x \in \mathbb{R};$$

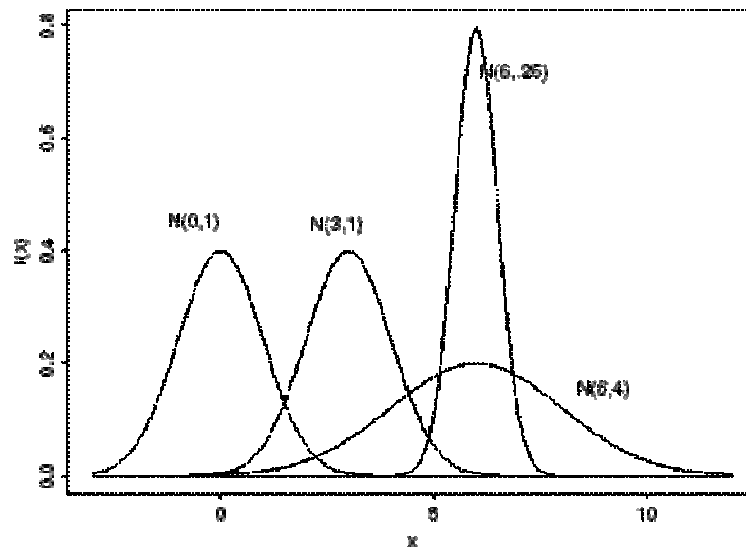
$$\mu \in \mathbb{R};$$

$$\sigma \in \mathbb{R}^+$$

Se $X \cap N(\mu, \sigma)$, então $E(X)=\mu$ e $\text{var}(X) = \sigma^2$
▼
Desvio padrão

Vejamos algumas propriedades da f.m.p. da normal relativamente à sua representação gráfica:

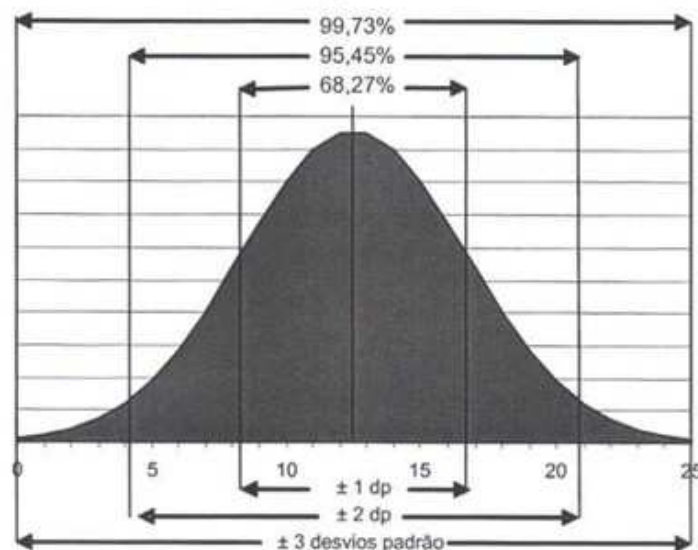
1. É simétrica relativamente ao seu valor médio μ , de modo que 2 curvas correspondentes a 2 distribuições com o mesmo desvio padrão σ têm a mesma forma, diferindo apenas na localização.
2. É tanto mais achatada quanto maior o valor de σ , de modo que 2 curvas correspondentes a 2 distribuições com o mesmo valor médio são simétricas relativamente ao mesmo ponto μ , diferindo apenas quanto ao seu grau de achatamento.



Para qualquer v.a. $X \sim N(\mu, \sigma)$:

- $P(\mu - \sigma < x < \mu + \sigma) = 0,6826$
- $P(\mu - 2\sigma < x < \mu + 2\sigma) = 0,9544$
- $P(\mu - 3\sigma < x < \mu + 3\sigma) = 0,9973$

Distribuição Normal



A distribuição normal com valor médio $\mu=0$ e desvio padrão $\sigma=1$ é designada por distribuição normal “standard” ou padrão e representa-se por $N(0,1)$. Se $X \sim N(\mu, \sigma)$, então:

$$z = \frac{X - \mu}{\sigma} \sim N(0,1)$$

A função de distribuição da normal “standard” tem uma notação especial:

$$z \sim N(0,1) \Rightarrow P(Z \leq z) = \Phi(z)$$

$$X \sim N(\mu, \sigma)$$

$$P(X \leq b) = P\left(\frac{X - \mu}{\sigma} \leq \frac{b - \mu}{\sigma}\right) = P\left(b \leq \frac{b - \mu}{\sigma}\right) = \Phi\left(\frac{b - \mu}{\sigma}\right)$$

$$P(a \leq X \leq b) = P\left(\frac{a - \mu}{\sigma} \leq \frac{X - \mu}{\sigma} \leq \frac{b - \mu}{\sigma}\right) = \Phi\left(\frac{b - \mu}{\sigma}\right) - \Phi\left(\frac{a - \mu}{\sigma}\right)$$

Uma propriedade importante da f.d. da Normal(0,1):

pela simetria da f.d.p. relativamente ao valor médio 0 deduz-se que

$$\Phi(-x) = 1 - \Phi(x)$$

[note-se que o gráfico da normal é simétrico e que $A=1$]

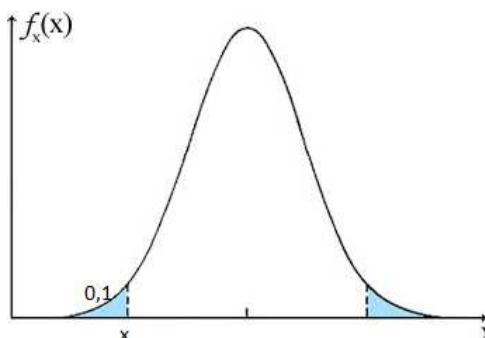
Existem tabelas para a f.d. da Normal(0,1). Pela propriedade acima referida, constata-se que basta haver tabelas para valores positivos ou negativos.

O quantil de probabilidade α da $N(0,1)$ representa-se por z_α e é tal que

$$\Phi(z_\alpha) = \alpha \Leftrightarrow P(Z \leq z_\alpha) = \alpha$$

$$\text{Note-se que } z_\alpha = -z_{1-\alpha}$$

O quantil de probabilidade α é igual ao simétrico do quantil de probabilidade $1-\alpha$



[x é o quantil do ordem 0,1 $\Rightarrow F(x) = 0,1$

o quantil de ordem 0,5 corresponde à mediana, que é igual à média, porque a distribuição normal é simétrica;

qualquer quantil de ordem menor que 0,5 é negativo (numa $N(0,1)$)]

Soma de variáveis aleatórias independentes com distribuição normal

Dadas as v.a.'s X_1, \dots, X_n independentes com distribuição normal

$$X_i \sim N(\mu_i, \sigma_i^2) \quad i = 1, \dots, n$$

a soma $S_n = X_1 + \dots + X_n$ também tem distribuição normal

$$S_n \sim N\left(\sum_{i=1}^n \mu_i, \sqrt{\sum_{i=1}^n \sigma_i^2}\right) \quad i = 1, \dots, n$$

Soma dos
valores médios

Se as variáveis forem independentes
a variância (σ^2) da soma é a soma das
variâncias

Propriedade mais geral:

Qualquer combinação linear de v.a.'s independentes com distribuição normal tem ainda distribuição normal

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

[combinação linear = multiplicação por constantes]

Valor Médio e a Variância da Média

Quando as variáveis aleatórias X_1, \dots, X_n são independentes e identicamente distribuídas (i.i.d.) diz-se que (X_1, \dots, X_n) é uma amostra aleatória.

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

[i.i.d. implica que as variáveis têm os mesmos parâmetros]

Dada uma amostra aleatória (X_1, \dots, X_n) proveniente de uma população X de valor médio μ e desvio padrão σ , então

$$E(\bar{X}) = \mu \quad \text{var}(\bar{X}) = \frac{\sigma^2}{n}$$

Dem.:

$$E(X_i) = \mu \quad \text{var}(X_i) = \sigma^2 \quad i = 1, \dots, n$$

$$E(\bar{X}) = \frac{1}{n} \sum_{i=1}^n E(X_i) = \frac{n\mu}{n} = \mu \quad \text{var}(\bar{X}) = \frac{1}{n^2} \sum_{i=1}^n \text{var}(X_i) = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}$$

Distribuição da Média para Populações Normais

Já dissemos que qualquer combinação linear de v.a.'s independentes com distribuição normal ainda tem distribuição normal. Como a média é uma combinação linear vem que:

Seja (X_1, \dots, X_n) uma amostra aleatória em que $X_i \sim N(\mu, \sigma), i = 1, \dots, n$.
Então $\bar{X} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right) \Leftrightarrow Z = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}} \sim N(0,1)$

O que é sempre verdade seja σ desconhecido ou não.

Quando o parâmetro σ é desconhecido, se substituirmos σ por S dado por

$$S = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2}, \quad \text{tem-se que} \quad T = \frac{\bar{X} - \mu}{S / \sqrt{n}} \sim t_{(n-1)}$$

T tem distribuição t-Student com $n-1$ graus de liberdade.

[A distribuição t-Student é extremamente parecida com a distribuição Normal: também tem forma de sino e é simétrica em relação a μ , mas as caudas são mais “pesadas”]

Teorema Limite Central

Distribuição da média para populações não normais

Quando a distribuição da população X não é normal, a distribuição da média não é em geral conhecida.

No entanto, um dos teoremas fundamentais da Teoria das Probabilidades fornece uma indicação sobre o comportamento da distribuição da média de um número suficientemente grande de v.a.'s i.i.d.

TEOREMA LIMITE CENTRAL

Se X_1, \dots, X_n são v.a.'s independentes e identicamente distribuídas, com valor médio μ e variância σ^2 finita, então a distribuição da soma $S_n = X_1 + \dots + X_n$ ou da média $(X_1 + \dots + X_n)/n$ tende a aproximar-se da distribuição normal para n suficientemente grande ($n \geq 30$).

$$P\left(\frac{S_n - n\mu}{\sigma\sqrt{n}} \leq z\right) \approx \Phi(z) \qquad P\left(\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \leq z\right) \approx \Phi(z)$$

Aplicações do Teorema Limite Central:

Aproximação da distribuição binomial pela distribuição normal

Uma v.a. X com distribuição $Bi(n,p)$ pode ser considerada como a soma de n v.a.'s independentes cada uma com distribuição $Bi(1,p)$. Então, pelo Teorema Limite Central, temos o seguinte resultado:

Se $X \cap Bi(n,p)$, então para n suficientemente grande

$$\frac{X - np}{\sqrt{np(1-p)}} \approx Z \cap N(0,1)$$

a distribuição da v.a. $X \cap Bi(n,p)$ pode ser aproximada pela distribuição $N(np, \sqrt{np(1-p)})$

Regra prática: considera-se uma aproximação razoável da distribuição binomial pela distribuição normal quando $np > 10$ e $n(1-p) > 10$

Aproximação da distribuição de Poisson pela distribuição normal

Uma v.a. X com distribuição $P(\lambda)$ pode ser considerada como a soma de n v.a.'s independentes cada uma com distribuição $P(\lambda/n)$. Então, pelo Teorema Limite Central, temos o seguinte resultado:

Se $X \cap P(\lambda)$, então para λ suficientemente grande

$$\frac{X - \lambda}{\sqrt{\lambda}} \approx Z \cap N(0,1)$$

Regra prática: considera-se uma aproximação razoável da distribuição de Poisson pela distribuição normal quando $\lambda > 20$.
