## 1. **Vetores Aleatórios**

Seja $(\Omega, \Im, P)$ um espaço de probabilidades. Nosso objetivo é definir uma função $\mathbf{X} = (X_1, X_2, ..., X_n)$ do espaço amostral $\Omega$ no conjunto dos números reais $\Re^n$, de dimensão $n$. Na realidade construimos um outro espaço de probabilidades $(\Re^n, \Im_{\mathbf{X}}, P_{\mathbf{X}})$, onde $\Im_{\mathbf{X}}$ é a $\sigma$-álgebra de Borel, $\beta^n$, em $\Re^n$ (no caso discreto, $\beta^n$ poderia ser substituida pelo conjunto das partes da imagem da aplicação $\mathbf{X}, \mathbf{X}(\Omega)$).

Se assumimos que a imagem inversa de qualquer conjunto $A$ em $\Im_{\mathbf{X}}$ pertence a $\Im$, isto é, $\mathbf{X}^{-1}(A) \in \Im, \quad \forall A \in \Im_{\mathbf{X}}$, podemos definir

$$P_{\mathbf{X}}(A) = P(\mathbf{X}^{-1}(A)), \quad \forall A \in \Im_{\mathbf{X}}.$$

Observe que:

$P_{\mathbf{X}}(\Re^n) = P(\mathbf{X}^{-1}(\Re^n)) = P(\Omega) = 1$, e que se $(A_k)_{k \geq 1}$ é uma sequência de eventos em $\Im_{\mathbf{X}}$, dois a dois disjuntos

$$P_{\mathbf{X}}(\cup_{k=1}^{\infty} A_k) = P(\mathbf{X}^{-1}(\cup_{k=1}^{\infty} A_k)) = P(\cup_{k=1}^{\infty} \mathbf{X}^{-1}(A_k)) =$$

$$\sum_{k=1}^{\infty} P\mathbf{X}^{-1}(A_k)) = \sum_{k=1}^{\infty} P_{\mathbf{X}}(A_k).$$

Portanto os axiomas de Kolmogorov estão satisfeitos e o espaço de probabilidades $(\Re^n, \Im_{\mathbf{X}}, P_{\mathbf{X}})$ esta bem definido e procedemos na formalização do conceito:

**Definição 1.1.** Seja $(\Omega, \Im, P)$ um espaço de probabilidade e $\mathbf{X}$ uma aplicação de $\Omega$ em $\Re^k$. $\mathbf{X}$ é uma variável aleatória se $\mathbf{X}^{-1}(A) \in \Im, \quad \forall A \in \Im_{\mathbf{X}}$.

Denominamos $(\Re^k, \Im_{\mathbf{X}}, P_{\mathbf{X}})$como o espaço de probabilidade induzido por $\mathbf{X}$.

*Observação* 1.2. Observe que um conjunto básico de $\beta^n$ é um retângulo da forma $A = A_1 X A_2 X ... X A_n$, onde $A_j \subseteq \beta$ e que se $w \in \Omega$

$$\mathbf{X}(w) = (X_1(w), X_2(w), ..., X_n(w)) \in A \Leftrightarrow$$

$$X_1(w) \in A_1, X_2(w) \in A_2, ..., X_n(w) \in A_n \Leftrightarrow$$

$$w \in X_1^{-1}(A_1), w \in X_2^{-1}(A_2), ..., w \in X_n^{-1}(A_n) \Leftrightarrow w \in \cap_{j=1}^{n} w \in X_j^{-1}(A_j).$$

Portanto se $X_j \quad 1 \leq j \leq n$ são variáveis aleatórias unidimensionais, $X_j^{-1}(A_j) \in \Im$ e $\mathbf{X}$ é um vetor aleatório.

**Vetores aleatórios discretos**

No caso discreto, a configuração de um conjunto $A$ de $\Im_{\mathbf{X}}$ é $A = \{\mathbf{x}_{i_1}, ..., \mathbf{x}_{i_n}\}$, onde $\mathbf{x}_{i_j} = \mathbf{X}(w_{i_j})$ para algum $w_{i_j} \in \Omega$. Portanto $A$

pode ser escrito como a união disjunta dos conjuntos unitários $\{\mathbf{x}_{i_j}\}$, $A = \cup_{j=1}^{n}\{\mathbf{x}_{i_j}\}$ e

$$\mathbf{X}^{-1}(A) = \mathbf{X}^{-1}(\cup_{j=1}^{n}\{\mathbf{x}_{i_j}\}) = \cup_{j=1}^{n}\mathbf{X}^{-1}(\{\mathbf{x}_{i_j}\}))$$

com

$$P_{\mathbf{X}}(A) = P(\mathbf{X} \in A) = P(\mathbf{X}^{-1}(A)) = P(\cup_{j=1}^{k}\mathbf{X}^{-1}(\{\mathbf{x}_{i_j}\})) =$$

$$\Sigma_{j=1}^{k} P(\mathbf{X}^{-1}(\{\mathbf{x}_{i_j}\})) = \Sigma_{j=1}^{k} P_X(\{\mathbf{x}_{i_j}\}).$$

Consequentemente o espaço de probabilidade induzido por $\mathbf{X}$ fica completamente caracterizado pelos valores $\mathbf{x}_i$ que $\mathbf{X}$ assume e suas probabilidades $P(\mathbf{X} = \mathbf{x}_i) = P(X_1 = x_{i_1}, X_2 = x_{i_2}, ..., X_n = x_{i_n})$ com $0 \leq P(\mathbf{X} = \mathbf{x}_i) \leq 1$ e $\Sigma P(\mathbf{X} = \mathbf{x}_i) = 1$.

**Definição 1.3.** A função que associa o valor da vetor aleatório à sua probabilidade é denominada função de probabilidade conjunta (distribuição de probabilidade conjunta).

| $x$ | $\mathbf{x}_1$ | $\mathbf{x}_2$ | ... | $\mathbf{x}_n$ |
|---|---|---|---|---|
| $P(\mathbf{X} = \mathbf{x})$ | $P(\mathbf{X} = \mathbf{x}_1)$ | $P(\mathbf{X} = \mathbf{x}_2)$ | ... | $P(\mathbf{X} = \mathbf{x}_n)$ |

*Observação* 1.4. Em resumo, se $X$ é uma variável aleatória assumindo valores $x_1, x_2, ...$ com probabilidades

$$P_X(\{x_k\}) = P(X = x_k) = P(w \in \Omega : X(w) = x_k)$$

e $Y$ é uma variável aleatória assumindo valores $y_1, y_2, ...$ com probabilidades

$$P_Y(\{y_k\}) = P(Y = y_k) = P(w \in \Omega : Y(w) = y_k),$$

então $(X, Y)$ é um vetor aleatório assumindo valores $(x_j, y_k)$ com probabilidades conjunta

$$P_{(X,Y)}((x_j, y_k)) = P((X, Y) = (x_j, y_k)) = P(w \in \Omega : X(w) = x_j, Y(w) = y_k).$$

As distribuições marginais de $X$ e de $Y$ são obtidas fixando o valor de interesse da variável e somando sobre os valores da outra variável:

$$P(X = x_j) = \Sigma_{y_k} P((X, Y) = (x_j, y_k));$$

$$P(Y = y_k) = \Sigma_{x_j} P((X, Y) = (x_j, y_k)).$$

**Exemplo 1.5.** A alocação de 2 bolas em 3 urnas resulta na seguinte configuração

$$\begin{array}{lll} \{ab|\ |\ \} & \{\ a|\ b|\ \ \} & \{\ b|\ a|\ \ \} \\ \{\ |ab|\ \} & \{\ a|\ \ |\ b\} & \{\ b|\ \ |\ a\} \\ \{\ |\ |ab\} & \{\ \ |\ a|\ b\} & \{\ \ |\ b|\ a\} \end{array}$$

Sejam as variáveis aleatórias definidas por :
$N$: O número de urnas ocupadas;
$X_i$: O número de bolas no i-ésimo compartimento, $1 \leq i \leq 3$.
As distribuições conjuntas de $(N, X_1)$ e de $(X_1, X_2)$ bem como suas marginais estão nas tabelas

Tabela 1- Distribuição conjunta de $(N, X_1)$

| $N, X_1$ | 0 | 1 | 2 | total |
|---|---|---|---|---|
| 1 | 0,22 | 0 | 0,12 | 0,34 |
| 2 | 0,22 | 0,44 | 0 | 0,66 |
| total | 0,44 | 0,44 | 0,12 | 1 |

A distribuição conjunta de $(N, X)$ está no corpo da tabela e as distribuições marginais estão nas marginais.
Portanto, $E[N] = 1.0,34+2.0,66 = 1,66$, $E[N^2] = 1.0,34+4.0,66 = 2,98$ e $Var(N) = 2,98 - (1,66)^2 = 0,23$.

$(X_1, X_2)$ tem distribuição de probabilidade conjunta

Tabela 2- Distribuição conjunta de $(X_1, X_2)$

| $X_2, X_1$ | 0 | 1 | 2 | total |
|---|---|---|---|---|
| 0 | 0,11 | 0,22 | 0,11 | 0,44 |
| 1 | 0,22 | 0,22 | 0 | 0,44 |
| 2 | 0,11 | 0 | 0 | 0,12 |
| total | 0,44 | 0,44 | 0,12 | 1 |

Observe que as distribuições marginais de $X_1$ e $X_2$ são iguais e dizemos que $X_1$ e $X_2$ são identicamente distribuidas.

**Exemplo 1.6.** Suponha que uma amostra de tamanho $n$, com reposição, seja retirada de uma população com várias subpopulações. Seja $X_j$ o número de elementos da amostra que pertence à $j$-ésima subpopulação. Então $(X_1, X_2, X_3)$ tem distribuição multinomial caracterizada completamente pela distribuição de probabilidade conjunta $P(X_1 = x_1, X_2 = x_2, X_3 = x_3) =$

$$\frac{n!}{x_1!x_2!x_3!(n - x_1 - x_2 - x_3)!} p_1^{x_1} p_2^{x_2} p_3^{x_3} (1 - p_1 - p_2 - p_3)^{(n - x_1 - x_2 - x_3)},$$

onde $x_1, x_2, x_3$ são inteiros não negativos tais que $x_1 + x_2 + x_3 \leq n$, $0 < p_i < 1, \quad i = 1, 2, 3$ e $p_i$ é a probabilidade de que um elemento pertence à $i$-ésima subpopulação.

A distribuição de probabilidade marginal de $(X_1, X_2)$ é

$$P(X_1 = x_1, X_2 = x_2) = \sum_{x_3=0}^{n-x_1-x_2} P(X_1 = x_1, X_2 = x_2, X_3 = x_3) =$$

$$\sum_{x_3=0}^{n-x_1-x_2} \frac{n!}{x_1!x_2!x_3!(n - x_1 - x_2 - x_3)!} p_1^{x_1} p_2^{x_2} p_3^{x_3} (1-p_1-p_2-p_3)^{(n-x_1-x_2-x_3)} =$$

$$\frac{n! p_1^{x_1} p_2^{x_2}}{x_1!x_2!(n - x_1 - x_2)!} \sum_{x_3=0}^{(n-x_1-x_2)} \frac{(n - x_1 - x_2)!}{x_3!(n - x_1 - x_2 - x_3)!} p_3^{x_3} (1-p_1-p_2-p_3)^{(n-x_1-x_2-x_3)} =$$

$$\frac{n!}{x_1!x_2!(n - x_1 - x_2)!} p_1^{x_1} p_2^{x_2} (1 - p_1 - p_2)^{n-x_1-x_2}.$$

A distribuição de probabilidade marginal de $X_1$ é a distribuição binomial.

$$P(X_1 = x_1) = \binom{n}{x_1} p_1^{x_1} (1 - p_1)^{n-x_1}.$$

**Variáveis aleatórias discretas condicionais**

No Capítulo 1 definimos o espaço de probabilidade condicional $(\Omega_B, \Im_B, P(.|B))$ onde

$$\forall A \in \Im_B, P(A|B) = \frac{P(A \cap B)}{P(B)}, \quad P(B) > 0$$

satisfazendo o teorema de Kolmogorov.

No momento utilizamos tal definição para estudarmos as distribuições de probabilidades das variáveis aleatórias condicionais. Se $(X, Y)$ é um vetor aleatório que assume valores $(x_j, y_k)$ com probabilidade conjunta

$$P_{(X,Y)}((x_j, y_k)) = P((X, Y) = (x_j, y_k)) = P(w \in \Omega : X(w) = x_j, Y(w) = y_k),$$

definimos , para cada valor $y_k$ fixado com $P(Y = y_k) > 0$, da variável aleatória $Y$, a variável aleatória condicional $(X|Y = y_k)$ que assume os valores de $X$ com probabilidades

$$P_{\{Y=y_k\}}(X = x_j) = P(X = x_j|Y = y_k) = \frac{P(X = x_j, Y = y_k)}{P(Y = y_k)}$$

obtida considerando os eventos $A = \{X = x_j\}$ e $B = \{Y = y_k\}$, no espaço de proabilidade condicional (se $P(Y = y_k) = 0$, definimos $P_{\{Y=y_k\}}(X = x_j) = 0$.)

A variável aleatória condicional $(X|Y = y_k)$ fica completamente caracterizada pelos valores $x_i$ que assume e suas probabilidades $P(X =$

$x_j|Y = y_k)$ com $0 \leq P(X = x_j|Y = y_k) \leq 1$ e $\Sigma P(X = x_j|Y = y_k) = 1$. Sua função de probabilidade é

| $(X\|Y = y_k)$ | $x_1$ | ... | $x_n$ |
|---|---|---|---|
| $P(X = x\|Y = y_k)$ | $P(X = x_1\|Y = y_k)$ | ... | $P(X = x_n\|Y = y_k)$ |

**Exemplo 1.7.** Recordando o exemplo 1.4, observamos que a variável aleatória condicional $(N|X_1 = 0)$ assume os valores 1 e 2 comm probabilidades:

$$P(N = 1|X_1 = 0) = \frac{P(N = 1, X_1 = 0)}{P(X_1 = 0)} = \frac{0,22}{0,44} = 0,5$$

$$P(N = 2|X_1 = 0) = \frac{P(N = 2, X_1 = 0)}{P(X_1 = 0)} = \frac{0,22}{0,44} = 0,5.$$

Portanto $E[N|X_1 = 0] = 1.0,5 + 2.0,5 = 1,5$,
$E[N^2|X_1 = 0)] = 1.0,5 + 4.0,5 = 2,5$ e
$Var((N|X_1 = 0)) = 2,5 - (1,5)^2 = 0,25$. Por outro lado, a variável

aleatória condicional $(N|X_1 = 1)$ assume os valores 1 e 2 comm probabilidades:

$$P(N = 1|X_1 = 1) = \frac{P(N = 1, X_1 = 1)}{P(X_1 = 1)} = \frac{0}{0,44} = 0$$

$$P(N = 2|X_1 = 1) = \frac{P(N = 2, X_1 = 1)}{P(X_1 = 1)} = \frac{0,44}{0,44} = 1.$$

Portanto $(N|X_1 = 1)$ tem distribuição degenerada no valor 2 e $E[N|X_1 = 1] = 2$ e $Var((N|X_1 = 1)) = 0$.

A variável aleatória condicional $(N|X_1 = 2)$ tem distribuição degenerada no valor 1 e $E[N|X_1 = 2] = 1$, com $Var((N|X_1 = 2)) = 0$.

Observamos, pelo exemplo anterior, que as distribuições das variáveis aleatórias condicionais $(X|Y = y_k)$ depende do valor fixado $y_k$ da variável aleatória $Y$ e que $E(X|Y = y)$ é uma função $\varphi(Y)$ de $Y$ $(E(X|Y = y) = \varphi(y))$ . Como a função de uma variável aleatória é uma variável aleatória, podemos calcular sua esperança e variância.

**Teorema 1.8.** *Se $(X, Y)$ é uma variável aleatória bidimensional, então*

$$E\{E[(X|Y)]\} = E[X];$$

$$Var(X) = E\{Var((X|Y))\} + Var(E[(X|Y)]\}.$$

**Prova**

*Como* $E(X|Y) = \varphi(Y)$ *é uma função de* $Y$, $\varphi(y_k) = E(X|Y = y_k) = \Sigma_{x_j} x_j P(X = x_j|Y = y_k)$. *A esperança de* $\varphi(Y)$ *é*

$$E[\varphi(Y) = \Sigma_{y_k}\varphi(y_k)P(Y = y_k) = \Sigma_{y_k}\Sigma_{x_j} x_j P(X = x_j|Y = y_k)P(Y = y_k) =$$
$$\Sigma_{y_k}\Sigma_{x_j} P(X = x_j, Y = y_k) = \Sigma_{x_j} x_j P(X = x_j) = E[X].$$

$$E\{Var((X|Y))\} = E\{E[(X^2|Y)]\} - E\{E[(X|Y)]^2\} =$$
$$E[X^2] - E[X^2] - E\{E[(X|Y)]^2\} + E[X]^2 =$$
$$Var(X) - E\{E[(X|Y)]^2\} + \{E[E(X|Y)]\}^2 = Var(X) - Var(E(X|Y)).$$

**Exemplo 1.9.** Continuando o exemplo 1.6, $E\{E[(N|X_1)]\} =$

$$E[(N|X_1 = 0)]P(X_1 = 0) + E[(N|X_1 = 1)]P(X_1 = 1)+$$
$$E[(N|X_1 = 2)]P(X_1 = 2) = 1,5.0,44 + 2.0,44 + 1.0,12 = 1,66 = EN.$$

Com respeito à variância, $Var(N) = E\{Var((N|X_1))\} + Var(E[(N|X_1)]\}$. Contudo

$$Var(E[(N|X_1)]) = E\{E[(N|X_1)]^2\} - E\{E[(N|X_1)\}^2 =$$
$$E\{E[(N|X_1)]^2\} - (1,66)^2 = E\{E[(N|X_1)]^2\} - 2,75$$

e $E\{E[(N|X_1)]^2\} =$

$$E[(N|X_1 = 0)]^2 P(X_1 = 0) + E[(N|X_1 = 1)]^2 P(X_1 = 1)+$$
$$E[(N|X_1 = 2)]^2 P(X_1 = 2) = (1,5)^2.0,44 + 2^2.0,44 + 1^2.0,12 = 2,87.$$

Portanto $Var(E[(N|X_1)]) = 2,87 - 2,75 = 0,12$.
Por outro lado $E\{Var((N|X_1))\} =$

$$Var((N|X_1 = 0))P(X_1 = 0) + Var((N|X_1 = 1))P(X_1 = 1)+$$
$$Var((N|X_1 = 2))P(X_1 = 2) = 0,25.0,44 + 0.0,44 + 0.0,12 = 0,11.$$

Concluimos que $E\{Var((N|X_1))\} + Var(E[(N|X_1)]\} = 0,11 + 0,12 = 0,23 = Var(N)$.

**Exemplo 1.10.** Suponha que uma amostra de tamanho 2, seja retirada de uma urna com 3 bolas pretas, 4 bola brancas e 2 bolas vermelhas, com reposição. Seja $X_j$ a cor da bola retirada no $j$-ésima ensaio,assumindo valores 1, 2 e 3 correspondentes à bola retirada ser preta, branca ou vermelha, respectivamente.. Então os $(X_j$ so identicamente distribuidos desde que a urna é reconstituida com a reposição da bola retirada.

$(X_1, X_2)$ tem distribuição de probabilidade conjunta

| $X_2, X_1$ | 1 | 2 | 3 | total |
|---|---|---|---|---|
| 1 | $\frac{9}{81}$ | $\frac{12}{81}$ | $\frac{6}{81}$ | $\frac{27}{81}$ |
| 2 | $\frac{12}{81}$ | $\frac{16}{81}$ | $\frac{8}{81}$ | $\frac{36}{81}$ |
| 3 | $\frac{6}{81}$ | $\frac{8}{81}$ | $\frac{4}{81}$ | $\frac{18}{81}$ |
| total | $\frac{27}{81}$ | $\frac{36}{81}$ | $\frac{18}{81}$ | 1 |

Observe que

$$P(X_1 = 1, X_2 = 1) = \frac{9}{81} = \frac{27}{81}.\frac{27}{81} = P(X_1 = 1).P(X_2 = 1),$$

$$P(X_1 = 1, X_2 = 3) = \frac{6}{81} = \frac{27}{81}.\frac{36}{81} = P(X_1 = 1).P(X_2 = 2),$$

e

$$P(X_1 = x_1, X_2 = x_2) = P(X_1 = x_1).P(X_2 = x_2), \quad \forall x_1, x_2 \in \{1, 2, 3\}.$$

Concluimos que

$$P(X_1 = x_1 | X_2 = x_2) = \frac{P(X_1 = x_1, X_2 = x_2)}{P(X_2 = x_2)} =$$

$$\frac{P(X_1 = x_1)P(X_2 = x_2)}{P(X_2 = x_2)} = P(X_1 = x_1)$$

e que, equivalentemente,

$$P(X_1 = x_1, X_2 = x_2) = P(X_1 = x_1).P(X_2 = x_2), \quad \forall x_1, x_2 \in \{1, 2, 3\}.$$

**Definição 1.11.** As variáveis aleatórias $X_1, X_2, ..., X_n$ , definidas em $(\Omega, \Im, P)$ so estocásticamente independentes se, e somente a distribuição conjunta de qualquer subconjunto $\{X_{i_1}, ..., X_{i_k})\}$ de $\{X_1, X_2, ..., X_n\}$ é o produto de suas marginais, isto é

$$P(\mathbf{X} = \mathbf{x}_i) = P(X_{i_1} = x_{i_1}, X_{i_2} = x_{i_2}, ..., X_{i_k} = x_{i_k}) =$$

$$P(X_{i_1} = x_{i_1}).P(X_{i_2} = x_{i_2})....P(X_{i_k} = x_{i_k}), \quad k = 2, 3, ..., n.$$

Um conjunto básico de $\Re^n$ é um retângulo da forma $A = A_1 X A_2 X ... X A_n$, onde $A_j \subseteq \beta$. Então pode-se provar que as variáveis aleatórias $X_1, X_2, ..., X_n$ são estocásticamente independentes se, e somente se,

$$P(X_1 \in A_1, X_2 \in A_2, ..., X_n \in A_n) = P(X_1 \in A_1)P(X_2 \in A_2)...P(X_n \in A_n),$$

para quaisquer $A = A_1 X A_2 X ... X A_n$ em $\beta^n$. No caso particular, discreto, em que $n = 2$, temos

$$P(X_1 \in A_1, X_2 \in A_2) = \Sigma_{x_1 \in A_1} \Sigma_{x_2 \in A_2} P(X_1 = x_1, X_2 = x_2) =$$

$$\Sigma_{x_1 \in A_1} \Sigma_{x_2 \in A_2} P(X_1 = x_1)P(X_2 = x_2) =$$

$$\Sigma_{x_1 \in A_1} P(X_1 = x_1) \Sigma_{x_2 \in A_2} P(X_2 = x_2) = P(X_1 \in A_1)P(X_2 \in A_2).$$

### Operações com variáveis aleatórias

Operações com variáveis aleatórias resultam em variáveis aleatórias. Portanto, se $X$ e $Y$ são variáveis aleatórias, as operações $\sqrt{X}$, $\ln Y$, $X+Y$, $X.Y$ são variáveis aleatórias e como tais, cada uma tem sua função de distribuição, sua média, sua variância e outras medidas.

A função de probabilidade induzida pela variável aleatória $g(X,Y)$ é caracterizada por:

$$P_{g(X,Y)}(k) = P(g(X,Y)=k) = \sum_{\{(x_i,y_j):g(x_i,y_j)=k\}} P(X=x_i, Y=y_j).$$

**Exemplo 1.12.** Sejam $X$ e $Y$ variáveis aleatórias independentes com distribuições geométricas de parâmetros $p_1$ e $p_2$ respectivamente. Então $P(X=x) = p_1(1-p_1)^{x-1}, \quad x=1,2,...$ e $P(Y=y) = p_2(1-p_2)^{y-1}, \quad y=1,2,...$ e

A) a distribuição de $Z=\max\{X,Y\}$ é:

$$P(Z \leq z) = P(\max\{X,Y\} \leq z) = P(X \leq z, Y \leq z) =$$
$$P(X \leq z)P(Y \leq z) = [1-(1-p_1)^z][1-(1-p_2)^z];$$

B) a probabilidade $P(\max\{X,Y\}=X)$ é

$$P(Y \leq X) = \sum_{y=1}^{\infty} P(Y \leq X|Y=y)P(Y=y) =$$

$$\sum_{y=1}^{\infty} P(y-1 < X|Y=y)P(Y=y) = \sum_{y=1}^{\infty} P(X > y-1)P(Y=y) =$$

$$\sum_{x=1}^{\infty}(1-p_1)^{y-1}p_2(1-p_2)^{y-1} =$$

$$p_2\sum_{y=1}^{\infty}[(1-p_1)(1-p_2)]^{y-1} = \frac{p_2}{p_1+p_2-p_1.p_2}.$$

Observe que, na primeira igualdade utilizamos a regra da probabilidade total.

C) a probabilidade $P(S=X+Y=k)$ é

$$P(S=k) = \sum_{x=1}^{k-1} P(X=x, Y=k-x) = \sum_{x=1}^{k-1} P(X=x)P(Y=k-x) =$$

$$p_1p_2\sum_{x=1}^{k-1}(1-p_1)^{x-1}(1-p_2)^{k-x-1} = p_1p_2(1-p_2)^{k-2}\sum_{x=1}^{k-1}[\frac{(1-p_1)}{(1-p_2)}]^{x-1} =$$

$$\frac{p_1p_2[(1-p_2)^{k-1}-(1-p_1)^{k-1}]}{p_1-p_2}.$$

Em geral se $X$ e $Y$ são variáveis aleatórias independentes, assumindo valores em $\{x_1, x_2, ..., x_k, ..\}$ e $\{y_1, y_2, ..., y_k, ..\}$ respectivamente, a soma $S = X + Y$ é uma variável aleatória assumindo valores $x_i + y_j, i = 1, 2, ..., k, ..., j = 1, 2, ..., k, ...$, com probabilidades

$$P(X+Y=k) = \sum_{x_i} P(X=x_i, Y=k-x_i) = \sum_{x_i} P(X=x_i)P(Y=k-x_i).$$

A função de probabilidade induzida pela variável aleatória $g(X, Y)$, com sua respectiva função de probabilidade permite definir:

**Definição 1.13.** Sejam $X$ e $Y$ variáveis aleatórias discretas que assumem valores $x_1, ..., x_n$ e $y_1, ..., y_m$, respectivamente, com probabilidade conjunta $P(X = x_i, Y = y_j)$. Se $g(x, y)$ é uma função a valores reais, limitada, então

$$E[g(X,Y)] = \sum_{i=1}^{n}\sum_{j=1}^{m} g(x_i, y_j)P(X=x_i, Y=y_j).$$

Utilizando a definição acima podemos provar o teorema:

**Teorema 1.14.** *Sejam $X$ e $Y$ variáveis aleatórias discretas que assumem valores $x_1, ..., x_n, ...$ e $y_1, ..., y_m, ...$ respectivamente, com probabilidade conjunta $P(X = x_i, Y = y_j)$. Se $E[X]$ e $E[Y]$ existem, então $E[X + Y]$ existe e*

$$E[X+Y] = E[X] + E[Y].$$

*Prova*

*Considerando, no Teorema 7.13, $g(x, y) = x + y$, obtemos*

$$E[|X+Y|] = \sum_{x_i}\sum_{y_j} |x_i + y_j|)P(X=x_i, Y=y_j) \leq$$

$$\sum_{x_i}\sum_{y_j} |x_i|P(X=x_i, Y=y_j) + \sum_{x_i}\sum_{y_j} |y_j|P(X=x_i, Y=y_j) =$$

$$\sum_{x_I} |x_i| \sum_{y_j} P(X=x_i, Y=y_j) + \sum_{y_j} |y_j| \sum_{x_i} P(X=x_i, Y=y_j) =$$

$$\sum_{x_i} |x_i| P(X = x_i) + \sum_{y_j}^{m} |y_j| P(Y = y_j) = E[|X|] + E[|Y|].$$

*Portanto* $E[|X+Y|] \leq E[|X|] + E[|Y|] < \infty$ *e* $E[X+Y]$ *existe.*

*Contudo*

$$E[X+Y] = \sum_{x_i}\sum_{y_j}(x_i + y_j)P(X = x_i, Y = y_j) =$$

$$\sum_{x_i}\sum_{y_j} x_i P(X = x_i, Y = y_j) + \sum_{x_i}\sum_{y_j} y_j P(X = x_i, Y = y_j) =$$

$$\sum_{x_i} x_i \sum_{y_j} P(X = x_i, Y = y_j) + \sum_{y_j} y_j \sum_{x_i} P(X = x_i, Y = y_j) =$$

$$\sum_{x_i} x_i P(X = x_i) + \sum_{y_j} y_j P(Y = y_j) = E[X] + E[Y].$$

**Exemplo 1.15.** No exemplo 1.5, onde alocamos duas bolas em três urnas , consideramos as variáveis aleatórias

$N$: O número de urnas ocupadas;

$X_1$: O número de bolas no primeiro compartimento, .

As distribuições conjuntas de $(N, X_1)$ bem como suas marginais são

| $N, X_1$ | 0 | 1 | 2 | total |
|---|---|---|---|---|
| 1 | $0,22$ | $0$ | $0,12$ | $0,34$ |
| 2 | $0,22$ | $0,44$ | $0$ | $0,66$ |
| total | $0,44$ | $0,44$ | $0,12$ | $1$ |

Portanto, $E[N] = 1.0,34 + 2.0,66 = 1,66$, $E[N^2] = 1.0,34 + 4.0,66 = 2,98$ e $Var(N) = 2,98 - (1,66)^2 = 0,23$.

$E[X_1] = 0,44 + 2.0,12 = 0,68$, $E[X_1^2] = 0,44 + 4.0,12 = 0,92$ e $Var(X_1) = 0,92 - (0,68)^2 = 0,46$.

A função de probabilidade de $S = N + X_1$ é

| $s$ | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| $P(S = s$ | $0,22$ | $0,22$ | $0,56$ | $0$ |

Portanto

$$E[N+X_1] = 0,22 + 2.0,22 + 3.0,56 = 2,34 = 1,66 + 0,68 = E[N] + E[X_1].$$

$$E[(N+X_1)^2] = 0,22 + 4.0,22 + 9.0,56 = 6,14$$

$$Var(N+X_1) = E[(N+X_1)^2] - (E[N+X_1])^2 = 6,14 - (2,34)^2 =$$

$$0,67 \neq 0,23 + 0,46 = Var(N) + Var(X_1).$$

**Exemplo 1.16.** No exemplo 1.10 em que observamos a variável aleatória $X_j$, a cor da bola retirada no $j$-ésimo ensaio, a distribuição conjunta de $(X_1, X_2)$ é

| $X_2, X_1$ | 1 | 2 | 3 | total |
|---|---|---|---|---|
| 1 | $\frac{9}{81}$ | $\frac{12}{81}$ | $\frac{6}{81}$ | $\frac{27}{81}$ |
| 2 | $\frac{12}{81}$ | $\frac{16}{81}$ | $\frac{8}{81}$ | $\frac{36}{81}$ |
| 3 | $\frac{6}{81}$ | $\frac{8}{81}$ | $\frac{4}{81}$ | $\frac{18}{81}$ |
| total | $\frac{27}{81}$ | $\frac{36}{81}$ | $\frac{18}{81}$ | 1 |

indicando que $X_1$ e $X_2$ são variáveis aleatórias independentes e identicamente distribuidas com
$E[X_1] = E[X_2] = \frac{27}{81} + \frac{72}{81} = \frac{54}{81} = \frac{153}{81}$
$E[X_1^2] = E[X_1^2] = \frac{27}{81} = \frac{144}{81} + \frac{162}{81} = \frac{333}{81}$ e
$Var(X_1) = Var(X_2) = \frac{333}{81} - (\frac{153}{81})^2 = \frac{44}{81}$.
A função de probabilidade de $S = X_1 + X_2$ é

| $s$ | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|
| $P(S = s)$ | $\frac{9}{81}$ | $\frac{24}{81}$ | $\frac{28}{81}$ | $\frac{16}{81}$ | $\frac{4}{81}$ |

Portanto

$$E[X_1 + X_2] = \frac{9}{81} + \frac{72}{81} + \frac{112}{81} + \frac{80}{81} + \frac{24}{81} =$$

$$\frac{306}{81} = \frac{153}{81} + \frac{153}{81} + E[X_1] + E[X_2].$$

$E[(X_1 + X_2)^2] = \frac{36}{81} + \frac{216}{81} + \frac{448}{81} + \frac{400}{81} + \frac{144}{81} = \frac{1244}{81}$ e

$$Var(X_1+X_2) = \frac{1244}{81} - (\frac{306}{81})^2 = \frac{88}{81} = \frac{44}{81} + \frac{44}{81} = Var(X_1) + Var(X_2).$$

**Teorema 1.17.** *Sejam $X$ e $Y$ variáveis aleatórias discretas independentes que assumem valores $x_1, ..., x_n, ...$ e $y_1, ..., y_m, ...$ respectivamente, com probabilidade conjunta $P(X = x_i, Y = y_j)$. Se $E[X]$ e $E[Y]$ existem, então $E[X.Y]$ existe e*

$$E[X.Y] = E[X].E[Y].$$

*Prova*

*Considerando a função $g(x, y) = x.y$, obtemos*

$$E[|X.Y|] = \sum_{x_i} \sum_{y_j} |x_i.y_j|)P(X = x_i, Y = y_j) =$$

$$\sum_{x_i} \sum_{y_j} |x_i|P(X = x_i, Y = y_j). \sum_{x_i} \sum_{y_j} |y_j|P(X = x_i, Y = y_j) =$$

$$\sum_{x_i} |x_i| \sum_{y_j} P(X = x_i, Y = y_j). \sum_{y_j} |y_j| \sum_{x_i} P(X = x_i, Y = y_j) =$$

$$\sum_{x_i} |x_i| P(X = x_i). \sum_{y_j}^{m} |y_j| P(Y = y_j) = E[|X|].E[|Y|].$$

*Portanto* $E[|X.Y|] = E[|X|].E[|Y|] < \infty$ *e* $E[X.Y]$ *existe.*

*Contudo*

$$E[X.Y] = \sum_{x_i} \sum_{y_j} (x_i.y_j)) P(X = x_i, Y = y_j) =$$

$$\sum_{x_i} \sum_{y_j} x_i P(X = x_i, Y = y_j). \sum_{x_i} \sum_{y_j} y_j P(X = x_i, Y = y_j) =$$

$$\sum_{x_i} x_i \sum_{y_j} P(X = x_i, Y = y_j). \sum_{y_j} y_j \sum_{x_i} P(X = x_i, Y = y_j) =$$

$$\sum_{x_i} x_i P(X = x_i). \sum_{y_j} y_j P(Y = y_j) = E[X].E[Y].$$

**Exemplo 1.18.** No exemplo 1.10 as variáveis $X_1$e $X_2$ são independentes com $E[X_1] = E[X_2] = \frac{153}{81}$

A função de probabilidade de $Z = X_1.X_2$ é

| $z$ | 1 | 2 | 3 | 4 | 6 | 9 |
|---|---|---|---|---|---|---|
| $P(Z = z)$ | $\frac{9}{81}$ | $\frac{24}{81}$ | $\frac{12}{81}$ | $\frac{16}{81}$ | $\frac{16}{81}$ | $\frac{4}{81}$ |

Portanto

$$E[X_1.X_2] = \frac{9}{81} + \frac{48}{81} + \frac{36}{81} + \frac{64}{81} + \frac{96}{81} + \frac{36}{81} = \frac{289}{81} = \frac{153}{81}.\frac{153}{81} = E[X_1].E[X_2].$$

Observamos que o teorema prova que a independência de $X$ e $Y$ é condição necessária para que $E[X.Y] = E[X].E[Y]$. A condição não é suficiente:

**Exemplo 1.19.** Se $(X, Y)$ tem distribuição de probabilidade conjunta

| $X, Y$ | $-1$ | $0$ | $1$ | total |
|---|---|---|---|---|
| $-1$ | 0 | $\frac{1}{4}$ | 0 | $\frac{1}{4}$ |
| 0 | $\frac{1}{4}$ | 0 | $\frac{1}{4}$ | $\frac{1}{2}$ |
| 1 | 0 | $\frac{1}{4}$ | 0 | $\frac{1}{4}$ |
| total | $\frac{1}{4}$ | $\frac{1}{2}$ | $\frac{1}{4}$ | 1 |

temos $E[X.Y] = 0 = 0.0 = E[X].E[Y]$ mas $P(X = 0, Y = 0) \neq \frac{1}{2}.\frac{1}{2} = P(X = 0).P(Y = 0)$.

As propriedades nos teoremas 1.14 e 1.17 se estendem para um número finito de variáveis aleatórias. Se $(X_1, X_2, ..., X_n)$ é um vetor de variáveis aleatórias, então

$$E[X_1 + X_2 + ... + X_n] = E[X_1] + E[X_2] + ... + E[X_n].$$

Se, em adição as variáveis aleatórias são independentes

$$E[X_1.X_2.....X_n] = E[X_1].E[X_2]....E[X_n].$$

Em relação à distribuição de soma de variáveis aleatórias independentes, a função geradora de probabilidades torna-se uma ferramenta importante:

Em particular, se $X$ e $Y$ assumem valores nos números naturais, com funções geradoras de probabilidades $\varphi_X(s)$ e $\varphi_Y(s)$ respectivamente, temos

$$\varphi_{X+Y}(s) = E[s^{X+Y}] = E[s^X]E[s^y] = \varphi_X(s)\varphi_Y(s)$$

**Exemplo 1.20.** Se consideramos $n$ variáveis aleatórias, $X_1, X_2, ..., X_n$ independentes e identicamente distribuidas com distribuição de Bernoulli de parâmetro $p$ temos $\varphi_{X_i}(s) = sp + (1-p)$ e

$$\varphi_{\Sigma_{i=1}^n X_i}(s) = \pi_{i=1}^n \varphi_{X_i}(s) = [sp + (1-p)]^n$$

Desde que a função geradora de probabilidade caracteriza completamente sua distribuição concluimos que $\Sigma_{i=1}^n X_i$ tem distribuição binomial de parâmetros $n$ e $p$. Por outro lado, como $E[X_i] = p$, $E[\Sigma_{i=1}^n X_i] = \Sigma_{i=1}^n E[X_i] = np$.

**Exemplo 1.21.** Se consideramos $n$ variáveis aleatórias, $X_1, X_2, ..., X_n$ independentes, onde $X_i$ tem distribuição de Poisson com parâmetro $\lambda_i, \quad i = 1, .., n$, temos

$$\varphi_{\Sigma_{i=1}^n X_i}(s) = \pi_{i=1}^n \varphi_{X_i}(s) = \pi_{i=1}^n e^{-\lambda_i(1-s)} = e^{-\Sigma_{i=1}^n \lambda_i(1-s)},$$

e concluimos que $\Sigma_{i=1}^n X_i$ tem distribuição de Poisson de parâmetro $\Sigma_{i=1}^n \lambda_i$. Assim $E[\Sigma_{i=1}^n X_i] = \Sigma_{i=1}^n E[X_i] = \Sigma_{i=1}^n \lambda_i$.

**Exemplo 1.22.** Seja $(X_n)_{n\geq 1}$ uma sequência de variaáveis aleatórias de Bernoulli, independentes e identicamente distribuidas com parâmetro $p$ e seja $N$ uma variável aleatória comm distribuição de Poisson de parâmetro $\lambda$, independente dos $X_i$'s. A função de probabilidade de $S_N = \Sigma_{i=1}^N X_i$ é

$$P(\Sigma_{i=1}^N X_i = k) = \sum_{n=k}^{\infty} P(\Sigma_{i=1}^N X_i = k, N = n) =$$

$$\sum_{n=k}^{\infty} P(\Sigma_{i=1}^{N} X_i = k | N = n) P(N = n) =$$

$$\sum_{n=k}^{\infty} P(\Sigma_{i=1}^{n} X_i = k | N = n) \frac{e^{-\lambda} \lambda^n}{n!} =$$

$$\sum_{n=k}^{\infty} P(\Sigma_{i=1}^{n} X_i = k) \frac{e^{-\lambda} \lambda^n}{n!} =$$

$$\sum_{n=k}^{\infty} \binom{n}{k} p^k (1-p)^{n-k} \frac{e^{-\lambda} \lambda^n}{n!} =$$

$$\frac{(\lambda p)^k e^{-\lambda}}{k!} \sum_{n=k}^{\infty} \frac{[(1-p)\lambda]^{n-k}}{(n-k)!} =$$

$$\frac{(\lambda p)^k e^{-\lambda}}{k!} \sum_{m=0}^{\infty} \frac{[(1-p)\lambda]^{m}}{m!} =$$

$$\frac{(\lambda p)^k e^{-\lambda}}{k!} e^{[(1-p)\lambda]} = \frac{(\lambda p)^k e^{-\lambda p}}{k!}$$

e concluimos que, neste caso particular, a soma aleatória das variáveis alaatórias tem distribuição de Poisson de parâmetro $\lambda p$.

No caso geral temos:

**Teorema 1.23.** *Sejam $X_1, X_2, ..., X_n$ variáveis aleatórias independentes e identicamente distribuidas, com valores nos números naturais e $N$ uma variável aleatória inteira, não negativa, independende dos $X_i's$. $S_N = \Sigma_{i=1}^{N} X_i$ $S_0 = 0$ é a soma aleatória de variáveis aleatórias que tambem assume valores nos naturais. Então*

$$\varphi_{S_N}(s) = \varphi_N(\varphi_{X_1}(s).$$

*Prova*

*Observe que condicionado a $\{N = n\}$ temos*

$$E[s^{\Sigma_{i=1}^{N} X_i} | N = n] = E[s^{\Sigma_{i=1}^{n} X_i} | N = n] = E[s^{\Sigma_{i=1}^{n} X_i}] = E[\pi_{i=1}^{n} s^{X_i}] =$$

$$\pi_{i=1}^{n} E[s^{X_i}] = \pi_{i=1}^{n} \varphi_{X_i}(s) = \pi_{i=1}^{n} \varphi_{X_1}(s) = (\varphi_{X_1}(s))^n.$$

*Portanto*

$$\varphi_{S_N}(s) = E\{E[\varphi_{S_N}(s) | N]\} = E\{E[s^{\Sigma_{i=1}^{N} X_i} | N]\} =$$

$$E\{(\varphi_{X_1}(s))^N\} = \varphi_N(\varphi_{X_1}(s).$$

**Exemplo 1.24.** Seja $(X_n)_{n\geq 1}$ uma sequência de variaáveis aleatórias de Bernoulli, independentes e identicamente distribuidas com parâmetro $p$, e função geradora de probabilidade é $\varphi_{X_1}(s) = ps + (1-p)$. Seja $N$ uma variável aleatória comm distribuição de Poisson de parâmetro $\lambda$ , com função geradora de probabilidade $\varphi_N(s) = e^{-\lambda(1-s)}$. Então

$$\varphi_{S_N}(s) = e^{-\lambda p(1-s)},$$

e concluimos que $S_N = \Sigma_{i=1}^N X_i$ tem distribuição de Poisson com parâmetro $\lambda p$.

Em relação à distribuição de soma de variáveis aleatórias independentes, mas que não assumem valores inteiros positivos, a função geradora de momentos, que também caracteriza completamente a distribuição de probabilidade, é mais indicada:

**Teorema 1.25.** *Se $X$ e $Y$ são variáveis aleatórias independentes com funções geradoras de Momentos $M_X(t)$ e $M_Y(t)$ respectively, then*

$$M_{X+Y}(t) = M_X(t).M_Y(t).$$

*Prova*

$$M_{X+Y}(t) = E[\exp[t(X+Y)]] = E[\exp[tX].\exp[tY]] =$$
$$E[\exp[tX]].E[\exp[tY]] = M_X(t).M_Y(t).$$

*Observação* 1.26. Utilizando o teorema acima podemos afirmar que a distribui¸cão da soma de $n$ variáveis aleatórias, digamos $X_1, X_2, ..., X_n$, independentes e identicamente distribuidas com distribuição geométrica de parâmetro $p$, $\quad 0 < p < 1$ é a distribuição binomial negativa de parâmetros $n$ e $p$:

Observe que

$$M_{X_1}(t) = E[e^{tX_1}] = \sum_{k=0}^{\infty} e^{tk} p(1-p)^k =$$

$$p\sum_{k=0}^{\infty}[(e^t(1-p)]^k = \frac{p}{1-(1-p)e^t}.$$

$$M_{X_1+...+X_n}(t) = M_{X_1}(t).....M_{X_n}(t) =$$

$$[(\frac{p}{1-(1-p)e^t}].....[(\frac{p}{1-(1-p)e^t}] = [(\frac{p}{1-(1-p)e^t}]^n.$$

que é a função geradora de momentos da distribuição binomiaL negativa.

Um tipo de dependência entre duas variáveis $X$ e $Y$ muito importante nas aplicações é a associação linear entre $X$ e $Y$. Esta medida de relação linear entre as variáveis é denominada covariância e denotada por $Cov(X, Y)$.

**Definição 1.27.** Sejam $X$ e $Y$ variáveis aleatórias. A covariância entre $X$ e $Y$ é definida pela esperança do produto dos desvios de $X$ e $Y$ em relação às suas respectivas médias, isto é

$$Cov(X, Y) = E[(X - E[X]).(Y - E[Y])].$$

*Observação* 1.28. De maneira mais fácil podemos escrever

$$Cov(X, Y) = E[(X - E[X]).(Y - E[Y])] =$$
$$E[XY - X.E[Y] - Y.E[X] + E[X].E[Y]] = E[XY] - E[X].E[Y].$$

Quando $X$ e $Y$ são variáveis aleatórias discretas que assumem valores $x_1, ..., x_n$ e $y_1, ..., y_n$ respectivamente, podemos escrever

$$Cov(X, Y) = \sum_{x_i} \sum_{x_j} (x_i - E[X]).(y_j - E[Y])P(X = x_i, Y = y_j).$$

Claramente, usando o Teorema 1.17 , se $X$ e $Y$ são variáveis aleatórias independentes $Cov(X, Y) = 0$. Observamos tambem que, como no exemplo 1.19, a $Cov(X, Y)$ pode ser igual a zero quando $X$ e $Y$ são variáveis aleatórias dependentes.

Para o vetor aleatório $(N, X_1)$ com distribuição conjunta

| $N, X_1$ | 0 | 1 | 2 | total |
|---|---|---|---|---|
| 1 | 0,22 | 0 | 0,12 | 0,34 |
| 2 | 0,22 | 0,44 | 0 | 0,66 |
| total | 0,44 | 0,44 | 0,12 | 1 |

temos $E[N] = 1,66$ e $Var(N) = 2,98 - (1,66)^2 = 0,23$.
$E[X_1] = 0,68$ e $Var(X_1) = 0,92 - (0,68)^2 = 0,46$.
A função de probabilidade de $Z = N.X_1$ é

| $z$ | 0 | 1 | 2 | 4 |
|---|---|---|---|---|
| $P(Z = z$ | 0,44 | 0 | 0 | 0,56 |

e $E[N.X_1] = 2.0,56 = 1,12$. Portanto

$Cov(N, X_1) = E[N.X_1] - E[N].E[X_1] = 1,12 - 0,68.1,66 = -0,02.$

No Teorema1.14 demonstramos que o valor esperado da soma de variáveis aleatórias é a soma dos valores esperados. O que podemos dizer sobre a variância da soma segue do teorema

**Teorema 1.29.** *Sejam $X$ e $Y$ variáveis aleatórias , então*

$$Var(X + Y) = Var(X) + Var(Y) + 2.Cov(X, Y).$$

*Se, em adição, $X$ e $Y$ forem independentes temos $Var(X + Y) = Var(X) + Var(Y)$.*

*Prova*

*A prova é imediata:*

$$Var(X+Y) = E\{[(X+Y)-E(X+Y)]^2\} = E\{[(X-E[X])+(Y-E[Y])]^2\} =$$
$$E[(X - E[X])^2 + E[(Y - E[Y])^2 + 2.E[(X - E[X]).(Y - E[Y])] =$$
$$Var(X) + VAR(Y) + 2.Cov(X, Y).$$

*Se $X$ e $Y$ são independentes, $Cov(X, Y) = 0$ e temos*

$$Var(X + Y) = Var(X) + Var(Y).$$

**Exemplo 1.30.** Para o vetor aleatório $(N, X_1)$ do exemplo anterior temos que

$Var(N) = 2,98 - (1,66)^2 = 0,23$, $Var(X_1) = 0,92 - (0,68)^2 = 0,46$ e $Cov(N, X_1) = -0,0088$.

portanto

$Var(N+X_1) = Var(N)+Var(X_1)+2.Cov(N, X_1) = 0,23+0,46-0,02 = 0,67$

*Observação* 1.31. A extensão do resultado do Teorema 1.28 para uma soma de variáveis aleatórias, $\Sigma_{i=1}^n X_i$ é:

$$Var(\Sigma_{i=1}^n X_i) = \Sigma_{i=1}^n Var(X_i) + 2\Sigma_i \Sigma_{j:i<j} cov(X_i, X_j).$$

**Exemplo 1.32.** A distribuição Binomial com parâmetros $n$ e $p$, $0 < p < 1$ é o número de sucessos quando realizamos $n$ ensaios de Bernoulli com probabilidade de sucesso $p$, independentes e identicamente distribuidos. A função de probabilidade da variável aleatória, X, de Bernoulli é

| $X$ | 0 | 1 |
|---|---|---|
| $P(X=x)$ | $1-p$ | $p$ |

Assim a média de $X$ é $E[X]=p$ e sua variância $Var(X)=p.(1-p)$.

Podemos interpretar a variável aleatória Binomial , $Y$, como a soma $Y=\sum_{i=1}^{n} X_i$ de variáveis aleatórias, $X_i$, $1\leq i\leq n$, de Bernoulli, independentes e identicamente distribuidas a $X$.

Portanto

$$E[Y]=E[\sum_{i=1}^{n} X_i]=\sum_{i=1}^{n} E[X_i]=\sum_{i=1}^{n} p=np$$

e

$$Var(Y)=Var(\sum_{i=1}^{n} X_i)=\sum_{i=1}^{n} Var(X_i)=\sum_{i=1}^{n} p,(1-p)=np(1-p).$$

**Exemplo 1.33.** Suponha que temos uma urna com $N$ bolas idênticas das quais $K$ tem a côr branca e $N-K$ preta. Retiramos da urna uma amostra sem reposição de $n$ bolas. Seja $S_n$ o número de bolas brancas na amostra. Observe que $S_n$ assume os valores $0,1,2,...,\min\{n,K\}$ com probabilidades

$$P(S_n=k)=\frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}}$$

e tem distribuição hipergeométrica.

Entretanto, podemos interpretar $S_n$ como $S_n=\sum_{i=1}^{n} X_i$ onde os $X_i$'s são variáveis aleatórias de Bernoulli:

$X_i=1$ se a i-ésimo ensaio resultar em bola branca;

$X_i=0$ se a i-ésimo ensaio resultar em bola preta.

Observe que $P(X_i=1)=\frac{K}{N}$, $k=1,2,...,\min\{n,K\}$ e portanto são identicamente distribuidas e

$$E[S_n]=E[\sum_{i=1}^{n} X_i]=\sum_{i=1}^{n} E[X_i]=\sum_{i=1}^{n} \frac{K}{N}=n\frac{K}{N}.$$

Contudo, como cada ensaio modifica a quantidade de bolas na urna, as variáveis aleatórias $X_i$'s não são independentes.

Note que $E[X_i]=P(X_i=1)=\frac{K}{N}$ e que $X_i^2=X_i$,produzindo $E[X_i^2]=\frac{K}{N}$ e

$$Var(X_i)=\frac{K}{N}-(\frac{K}{N})^2=\frac{K}{N}(1-\frac{K}{N}).$$

Por outro lado,

$$E[X_i.X_j] = P(X_i = 1, X_j = 1) = P(X_i = 1|X_j = 1)P(X_j = 1) = \frac{K}{N}(\frac{K-1}{N-1})$$

e

$$Cov(X_i, X_j) = \frac{K(K-1)}{N(N-1)} - (\frac{K}{N})^2 = \frac{K}{N}(\frac{K-1}{N-1} - \frac{K}{N}) = \frac{K}{N}(\frac{K-N}{N(N-1)})$$

Portanto

$$\begin{gathered} Var(\Sigma_{i=1}^n X_i) = \Sigma_{i=1}^n Var(X_i) + 2\Sigma_i \Sigma_{j:i<j} cov(X_i, X_j) = \\ n\frac{K}{N}(1 - \frac{K}{N}) + 2\binom{n}{2}\frac{K}{N}(\frac{K-N}{N(N-1)}) = \\ n\frac{K}{N}(1 - \frac{K}{N})(1 - \frac{n-1}{N-1}). \end{gathered}$$

Uma medida da relação linear entre duas variáveis aleatórias $X$ e $Y$ que não depende da unidade de medida é o coeficiente de correlação linear, denotado por $\rho = \rho(X, Y)$, que é a covariância padronizada pelos desvios padrões de $X$ e de $Y$:

**Definição 1.34.** O coeficiente de correlação linear entre as variáveis aleatórias $X$ e $Y$ é definido por

$$\rho = \rho(X, Y) = \frac{Cov(X, Y)}{DP(X).DP(Y)}.$$

**Exemplo 1.35.** O coeficiente de correlação linear entre as variáveis aleatórias $N$ e $X_1$ do exemplo é

$$\rho = \frac{Cov(N, X_1)}{DP(N).DP(X_1)} = \frac{-0,0088}{\sqrt{0,23.0,46}} = -0,03.$$

**Proposição 1.36.** *Se existem números reais $a$ e $b$ tais que $Y = aX + b$, isto é, $Y$ é uma função linear de $X$, então $|\rho| = 1$.*

*Prova:*

*Se $Y = aX + b$ temos $E[Y] = a.E[X] + b$, $Var(Y) = a^2.Var(X)$ e $E[XY] = E[X.(aX + b)] = a.E[X^2] + b.E[X] = a.Var(X) + a.E[X]^2 + b.E[X]$.*

*Portanto*

$$Cov(X, Y) = a.Var(X) + a.E[X]^2 + b.E[X] = a.Var(X)$$

*e*

$$\rho(X,Y) = \frac{a.Var(X)}{\sqrt{a^2 Var(X)^2}} = \frac{a}{|a|}.$$

**Teorema 1.37.** *Se $X$ e $Y$ são variáveis aleatórias com variâncias finitas, então $|\rho| \leq 1$. Em adição, se vale a igualdade, existem números reais $a$ e $b$ tais que, com probabilidade 1, $Y = a.X + b$. Prova:*

*Consideramos a função $f$ de $\Re$ em $\Re$ definida por*

$$f(t) = E\{[(Y - E[Y]) + t.(X - E[X])]^2\}$$

*que é maior ou igual a zero. Desenvolvendo o quadrado perfeito temos*

$$f(t) = Var(X).t^2 + 2.t.Cov(X,Y) + Var(Y) \geq 0.$$

*A função $f(t)$ é uma equação quadrática que é positiva se, e somente se, o seu discriminante $\triangle = 4.Cov(X,Y)^2 - 4.Var(X).Var(Y)$ é menor ou igual a zero. Portanto*

$$\frac{Cov(X,Y)^2}{Var(X).Var(Y)} \leq 1 \Leftrightarrow \rho^2 \leq 1 \Leftrightarrow |\rho| \leq 1.$$

*$|\rho| = 1$ se, e somente se, $E\{[(Y - E[Y]) + t.(X - E[X])]\} = 0$, o que implica, com probabilidade 1, $(Y - E[Y]) + t.(X - E[X]) = 0$, isto é, $Y = -t.X + t.E[X] + E[Y]$. Definindo $a = -t$ e $b = t.E[X] + E[Y]$ temos $Y = a.X + b$.*

*E-mail address*: `bueno@ime.usp.br`

Departamento de Estatística, Instituto de Matemática e Estatística, Universidade de São Paulo, Caixa Postal 66281, CEP 05311-970, São Paulo, Brazil