1. CAPÍTULO 7

Distribuições conjuntas.

Neste capítulo atribuimos a um ponto amostral valores de várias variáveis aleatórias e analisamos a sua distribuição conjunta. Por praticidade desenvolvemos a teoria para vetores de variáveis aleatórias bidimensionais mas os resultados estendem-se para o caso multidimensional de dimensão finita.

1.1. **Distribuições conjuntas discretas.** Começamos com um exemplo:

Exemplo 1.1. Um investidor simula uma sequência de sucessos, ou fracassos, anuais de suas aplicações por um período de três anos. Para isso supõe que a probabilidade de sucesso em determinado ano é p, $0 < p < 1$, independente dos resultados nos outros anos. Considera três variáveis de interesse:

A variável aleatória X, que indica se houve sucesso ou fracasso no primeiro ano. O sucesso (S) indicado pelo algarismo 1 e o fracasso (F) pelo algarismo 0;

a variável aleatória Y que indica o número de sucessos nos três anos e a variável aleatória

Z, indicando o número de mudanças (SF ou FS) que ocorreram durante os três anos.

A configuração dos resultados possíveis com as respectivas probabilidades é a distribuição conjunta tridimensional que apresenta os valores (x, y, z) do vetor aleatório (X, Y, Z) com suas respectivas probabilidades

$$P(X = x, Y = y, Z = z) = P(\{X = x\} \cap \{Y = y\} \cap \{Z = z\}),$$

por exemplo $P(X = 1, Y = 2, Z = 2) = p^2(1 - p)$.

Tabela 7.1- Distribuição tridimensional de (X, Y, Z)

Realização	(x,y,z)	$P(X=x,Y=y,Z=z)$
(S,S,S)	$(1,3,0)$	p^3
(S,S,F)	$(1,2,1)$	$p^2(1-p)$
(S,F,S)	$(1,2,2)$	$p^2(1-p)$
(F,S,S)	$(0,2,1)$	$p^2(1-p)$
(S,F,F)	$(1,1,1)$	$p(1-p)^2$
(F,S,F)	$(0,1,2)$	$p(1-p)^2$
(F,F,S)	$(0,1,1)$	$p(1-p)^2$
(F,F,F)	$(0,0,0)$	$(1-p)^3$

As distribuições unidimensionais das variáveis X, Y e Z são obtidas fixando o valor da variável de interesse e somando sobre os valores das outras variáveis.Analiticamente temos

$$P(X=x)=\sum_y\sum_z P(X=x,Y=y,Z=z).$$

Por exemplo

$$P(X=1)=P(X=1,Y=3,Z=0)+P(X=1,Y=2,Z=1)+$$
$$P(X=1,Y=2,Z=2)+P(X=1,Y=1,Z=1)=$$
$$p^3+2p^2(1-p)+p(1-p)^2=p[p^2+2p(1-p)+(1-p)^2]=p[p+(1-p)]^2=p.$$

A probabilidade do evento complementar é $P(X=0)=1-p$. O resultado era esperado desde que X é uma variável aleatória de Bernoulli que tem média $E[X]=p$ e variância $Var(X)=p(1-p)$.

Obtemos a distribuição da variável Y de maneira semelhante e a sua função de probabilidade é

y	0	1	2	3
$P(Y=y)$	$(1-p)^3$	$3p(1-p)^2$	$3p^2(1-p)$	p^3

Resumindo, $P(Y=k)=\binom{3}{k}p^k(1-p)^{3-k}$ para $k\in\{0,1,2,3\}$. Isto é, o número de sucessos em 3 ensaios de Bernoulli, independentes e identicamente distribuidos, com probabilidade de sucesso igual a p, tem distribuição binomial com média $E[Y]=3p$ e variância $Var(Y)=3p(1-p)$.

A variável Z assume os valores $0,1,2$ com probabilidades
$P(Z=0)=p^3+(1-p)^3$;
$P(z=1)=2p^2(1-p)+2p(1-p)^2=2p(1-p)$;
$P(Z=2)=p^2(1-p)+p(1-p)^2$.

A sua esperança é

$$E[Z]=2p(1-p)+2p^2(1-p)+2p(1-p)^2=4p(1-p)$$

e sua variância

$Var(Z) = E[Z^2] - E[Z]^2 = 6p(1-p) - 16p^2(1-p)^2 = 6p - 22p^2 + 32p^3 - 16p^4.$

A distribuição conjunta de (X, Z) pode ser representada por uma tabela de dupla entrada

Tabela 7.2- Distribuição bidimensional de (X, Z)

X, Z	0	1	2	total
0	$(1-p)^3$	$p(1-p)$	$p(1-p)^2$	1-p
1	p^3	$p(1-p)$	$p^2(1-p)$	p
total	$p^3 + (1-p)^3$	$2p(1-p)$	$pr(1-p) + p(1-p)^2$	1

No seu interior a tabela nos fornece a distribuição conjunta das variáveis (X, Z), isto é, os valores

$$P(X = x, Z = z) = P(\{X = x\} \cap \{Z = z\})$$

para todos os valores de X e de Z representados por x e z respectivemente. As suas margens fornecem as distribuições (marginais) de X e de Z.

Em uma primeira simulação o investidor imaginou o menos pior, ou seja, considerou a probabilidade de sucesso igual a $\frac{1}{2}$. A tabela torna-se

Tabela 7.3- Distribuição bidimensional de (X, Z), $p = \frac{1}{2}$

X, Z	0	1	2	total
0	$\frac{1}{8}$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{2}$
1	$\frac{1}{8}$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{2}$
total	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$	1

Em seguida, para um cenário de sucesso na carreira considerou $p = \frac{2}{3}$, projetando

Tabela 7.4- Distribuição bidimensional de (X, Z), $p = \frac{2}{3}$

X, Z	0	1	2	total
0	$\frac{1}{27}$	$\frac{6}{27}$	$\frac{2}{27}$	$\frac{1}{3}$
1	$\frac{8}{27}$	$\frac{6}{27}$	$\frac{4}{27}$	$\frac{2}{3}$
total	$\frac{3}{9}$	$\frac{4}{9}$	$\frac{2}{9}$	1

Quando estudamos a distribuição conjunta de variáveis aleatórias desejamos conhecer se, de alguma maneira, uma variável esta associada às outras. Um conceito essencial para uma análise neste sentido é o de distribuição condicional fundamentada no conceito de eventos condicionais: Se A e B são eventos com $P(B) > 0$, a probabilidade condicional do evento A dado a ocorrência de B é $P(A|B) = \frac{P(A \cap B)}{P(B)}$.

Definição 1.2. Sejam X e Y variáveis aleatórias discretas que assumem valores $x_1, ..., x_n$ e $y_1, ..., y_m$ respectivamente. Se $P(X = x_i) > 0$, a probabilidade condicional de $\{Y = y_j\}$ dado $\{X = x_i\}$, denotada por $P(Y = y_j|X = x_i)$ é definida por

$$P(Y = y_j|X = x_i) = \frac{P(X = x_i, Y = y_j)}{P(X = x_i)}, \quad 1 \leq j \leq m.$$

Podemos observar que para x_i fixado com $P(X = x_i) > 0$, os pares $(y_j, P(Y = y_j|X = x_i))$ $\ 1 \leq j \leq m$ caracterizam a distribuição de probabilidade da variável aleatória condicional $(Y|X = x_i)$. Observe que

$$\sum_{j=1}^{m} P(Y = y_j|X = x_i) = \sum_{j=1}^{m} \frac{P(X = x_i, Y = y_j)}{P(X = x_i)} =$$

$$\frac{1}{P(X = x_i)} \sum_{j=1}^{m} P(X = x_i, Y = y_j) = \frac{P(X = x_i)}{P(X = x_i)} = 1.$$

Na suposição de que $p = \frac{2}{3}$, a probabilidade condicional de termos uma única mudança, condicionada que se obteve sucesso no primeiro ano é:

$P(Z = 1|X = 1) = \frac{P(X=1,Z=1)}{P(X=1)} = \frac{\frac{6}{27}}{\frac{2}{3}} = \frac{6}{18}$.

A distribuição condicional da variável Z condicionada ao valor da variável $X = 1$ tem função de probabilidade:

$Z\|X = 1$	0	1	2
$P(Z = z\|X = 1)$	$\frac{8}{18}$	$\frac{6}{18}$	$\frac{4}{18}$

Observe que $\sum_{k=0}^{2} P(Z = k|X = 1) = 1$. A esperança da variável $(Z|X = 1)$ é

$$E[Z|X = 1] = \frac{6}{18} + \frac{8}{18} = \frac{14}{18}$$

e a variância da variável $(Z|X = 1)$ é

$$Var(Z|X = 1) = \frac{22}{18} - \frac{196}{324} = \frac{200}{324},$$

pois $E[Z^2|X = 1] = \frac{6}{18} + \frac{16}{18} = \frac{22}{18}$.

Com argumento análogo concluimos que a distribuição condicional da variável Z condicionada ao valor da variável $X = 0$ tem função de probabilidade:

$Z\|X = 0$	0	1	2
$P(Z = z\|X = 0)$	$\frac{1}{9}$	$\frac{6}{9}$	$\frac{2}{9}$

com
$E[Z|X=0]=\frac{10}{9}$, $E[Z^2|X=0]=\frac{14}{9}$ e $Var(Z|X=0)=\frac{56}{81}$.

Observe que uma variável Y quando condicionada aos valores de uma variável X é função de tais valores e como função de X é uma variável aleatória. Esta variável aleatória denotada por $\varphi(X)=E[Y|X]$ assume valores $E[Y|X=x]$ com respectivas probabilidades P(X=x).

$E[Y\|X]$	$E[Y\|X=x_1]$	...	$E[Y\|X=x_n]$
$P(E[Y\|X]=E[Y\|X=x])$	$P(X=x_1)$	...	$P(X=x_n)$

Portanto ao calcular $E\{E[Y|X]\}$ temos

$$E\{E[Y|X]\}=\sum_{j=1}^{m}E[Y|X=x_j]P(X=x_j)=$$

$$\sum_{j=1}^{m}[\sum_{i=1}^{n}y_iP(Y=y_i|X=x_j)]P(X=x_j)+\sum_{j=1}^{m}[\sum_{i=1}^{n}y_i\frac{P(Y=y_i,X=x_j)}{P(X=x_j)}]P(X=x_j)=$$

$$\sum_{j=1}^{m}\sum_{i=1}^{n}y_iP(Y=y_i,X=x_j)=\sum_{i=1}^{n}y_iP(Y=y_i)=E[Y].$$

Pode-se provar tambem que

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

Podemos utilizar os cálculos anteriores para, no caso em que $p=\frac{2}{3}$, exemplificarmos:

$$E\{E[Z|X]\}=E[Z|X=0]P(X=0)+E[Z|X=1]P(X=1)=$$

$$\frac{10}{9}\frac{1}{3}+\frac{14}{18}\frac{2}{3}=\frac{20+28}{54}=\frac{8}{9}=E[Z].$$

Um tópico importante na análise das distribuições conjuntas de variáveis aleatórias é o estudo da independência das mesmas.

Definição 1.3. Duas variáveis aleatórias X e Y assumindo valores nos conjuntos $\{x_1,...,x_n\}$ e $\{y_1,...,y_m\}$, respectivamente, são independentes se, e somente se,

$$P(X=x_i,Y=y_j)=P(X=x_i)P(Y=y_j)$$

para quaiquer possiveis pares de valores (x_i,y_j), $1\leq i\leq n$ $1\leq j\leq m$.

As variáveis aleatórias X e Z acima, quando consideramos $p = \frac{1}{2}$, tem distribuição conjunta representada por

Tabela 7.5- Distribuição bidimensional de (X, Z), $p = \frac{1}{2}$

X, Z	0	1	2	total
0	$\frac{1}{8}$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{2}$
1	$\frac{1}{8}$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{2}$
total	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$	1

E podemos notar que a probabilidade conjunta é igual ao produto de suas marginais para todos os valores possíveis, por exemplo

$$P(X = 1, Z = 1) = \frac{2}{8} = \frac{1}{2}\frac{1}{2} = P(X = 1)P(Z = 1).$$

Portanto as variáveis X e Z são independentes.

Se consideramos $p = \frac{2}{3}$ temos

$$P(X = 1, Z = 1) = \frac{6}{27} \neq \frac{2}{3}\frac{4}{9} = P(X = 1)P(Z = 1)$$

e podemos afirmar que, neste caso, X e Z não são independentes.

Podemos generalizar a definição de independência para um vetor de variáveis aleatórias com dimensões maiores. Vejamos o caso $n = 3$.

Definição 1.4. As variáveis aleatórias X_1, X_2 e X_3 são independentes se, e sómente se

$$P(X_1 = x_1, X_2 = x_2, X_3 = x_3) = P(X_1 = x_1)P(X_2 = x_2)P(X_3 = x_3),$$
$$P(X_1 = x_1, X_2 = x_2) = P(X_1 = x_1)P(X_2 = x_2),$$
$$P(X_1 = x_1, X_3 = x_3) = P(X_1 = x_1)P(X_3 = x_3)$$

e

$$P(X_2 = x_2, X_3 = x_3) = P(X_2 = x_2)P(X_3 = x_3).$$

Operação entre 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:

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

No que segue estudaremos algumas destas operações:

Se, no exemplo consideramos $p = \frac{2}{3}$, a distribuição conjunta de (X, Y) é dada por

Tabela 7.6- Distribuição bidimensional de (X, Y), $p = \frac{2}{3}$

X, Y	0	1	2	3	total
0	$\frac{1}{27}$	$\frac{4}{27}$	$\frac{4}{27}$	0	$\frac{1}{3}$
1	0	$\frac{2}{27}$	$\frac{8}{27}$	$\frac{8}{27}$	$\frac{2}{3}$
total	$\frac{1}{27}$	$\frac{6}{27}$	$\frac{12}{27}$	$\frac{8}{27}$	1

A distribuição da variável aleatória $X + Y$ assume os valores $x + y$, para todos os valores x e y de X e Y, respectivamente. A função de probabilidade de $X + Y$ é:

$X + Y$	0	1	2	3	4
$P(X + Y = x + y)$	$\frac{1}{27}$	$\frac{4}{27}$	$\frac{6}{27}$	$\frac{8}{27}$	$\frac{8}{27}$

de maneira que

$$E[X + Y] = \frac{8}{3} = \frac{2}{3} + 2 = E[X] + E[Y].$$

Este resultado sempre é verdadeiro, isto é, a esperança da soma de variáveis aleatórias é a soma de suas esperanças. O fato é uma consequência do teorema que segue que aceitamos sem prova. Sua prova pode ser encontrada em literatura mais especializada.

Teorema 1.5. *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 o Teorema 7.5 podemos provar o corolário:

Corolário 1.6. *Sejam X e Y variáveis aleatórias discretas que assumem valores $x_1, ..., x_n$ e $y_1, ..., y_m$ respectivamente, com probabilidade conjunta $P(X = x_i, Y = y_j)$. Então*

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

Prova

Considerando, no Teorema 7.5, $g(x,y)=x+y$, obtemos

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

$$\sum_{i=1}^{n}\sum_{j=1}^{m}x_iP(X=x_i,Y=y_j)+\sum_{i=1}^{n}\sum_{j=1}^{m}y_jP(X=x_i,Y=y_j)=$$

$$\sum_{i=1}^{n}x_i\sum_{j=1}^{m}P(X=x_i,Y=y_j)+\sum_{j=1}^{m}y_j\sum_{i=1}^{n}P(X=x_i,Y=y_j)=$$

$$\sum_{i=1}^{n}x_iP(X=x_i)+\sum_{j=1}^{m}y_jP(Y=y_j)=E[X]+E[Y].$$

Se consideramos a transformação produto, isto é, a variável aleatória XY que assume valores xy para todos os pares (x,y) do vetor aleatório (X,Y) com função de probabilidade

XY	0	1	2	3
$P(XY=xy)$	$\frac{9}{27}$	$\frac{2}{27}$	$\frac{8}{27}$	$\frac{8}{27}$

concluimos que $E[XY]=\frac{17}{9}\neq\frac{2}{3}.2=E[X].E[Y]$.

Contudo, se consideramos a distribuição conjunta das variáveis aleatórias independentes X e Z, quando $p=\frac{1}{2}$ obtemos

Tabela 7.7- Distribuição bidimensional de (X,Z), $p=\frac{1}{2}$

X,Z	0	1	2	total
0	$\frac{1}{8}$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{2}$
1	$\frac{1}{8}$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{2}$
total	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$	1

e a função de probabilidade de XZ é

y	0	1	2
$P(XZ=xz)$	$\frac{5}{8}$	$\frac{2}{8}$	$\frac{1}{8}$

com $E[XZ]=\frac{1}{2}=\frac{1}{2}.1=E[X].E[Z]$.

Utilizando o Teorema 7.5 podemos provar que esta propriedade é verdadeira

Corolário 1.7. *Sejam X e Y variáveis aleatórias discretas independentes, que assumem valores $x_1, ..., x_n$ e $y_1, ..., y_m$ respectivamente, com probabilidade conjunta $P(X = x, Y = y) = P(X = x).P(Y = y)$. Então*

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

Prova

Considerando, no Teorema 7.5, $g(x, y) = x.y$, obtemos

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

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

Observamos que o corolário 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 1.8. Se (X, Y) tem distribuição de probabilidade conjunta

Tabela 7.8- Distribuição conjunta de (X, Y)

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 corolários 7.6 e 7.7 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

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

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 1.9. 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 1.10. 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 utilizando o Teorema 9.5 que

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

Claramente, usando o Corolário 7.7 , se X e Y são variáveis aleatórias independentes $Cov(X,Y) = 0$. Observamos tambem que, como no exemplo 7.8, a $Cov(X,Y)$ pode ser igual a zero quando X e Y são variáveis aleatórias dependentes.

Para o vetor aleatório (X,Y) com distribuição conjunta

Tabela 7.9- Distribuição conjunta de (X,Y)

X,Y	0	1	2	3	total
0	$\frac{1}{27}$	$\frac{4}{27}$	$\frac{4}{27}$	0	$\frac{1}{3}$
1	0	$\frac{2}{27}$	$\frac{8}{27}$	$\frac{8}{27}$	$\frac{2}{3}$
total	$\frac{1}{27}$	$\frac{6}{27}$	$\frac{12}{27}$	$\frac{8}{27}$	1

temos que

$$Cov(X,Y) = E[XY] - E[X].E[Y] = \frac{14}{9} - \frac{2}{3}.2 = \frac{14}{9} - \frac{12}{9} = \frac{2}{9}.$$

No Corolário 7.6 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 corolário

Corolário 1.11. *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 1.12. Para o vetor aleatório (X,Y) com distribuição conjunta

Tabela 7.10- Distribuição conjunta de (X,Y)

X,Y	0	1	2	3	total
0	$\frac{1}{27}$	$\frac{4}{27}$	$\frac{4}{27}$	0	$\frac{1}{3}$
1	0	$\frac{2}{27}$	$\frac{8}{27}$	$\frac{8}{27}$	$\frac{2}{3}$
total	$\frac{1}{27}$	$\frac{6}{27}$	$\frac{12}{27}$	$\frac{8}{27}$	1

temos que

$Var(X) = \frac{2}{9}$, $Var(Y) = \frac{2}{3}$ e $Cov(X,Y) = \frac{14}{9} - \frac{2}{3}.2 = \frac{2}{9}$. Portanto

$$Var(X+Y) = \frac{2}{9} + \frac{6}{9} + 2.\frac{2}{9} = \frac{12}{9}.$$

Exemplo 1.13. A distribuição Binomial com parâmetros n e p, $0 < p < 1$ pode ser interpretada como 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

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

Em relaço à distribuição de transformações de variáveis aleatórias independentes, a função geradora de momentos torna-se uma ferramenta importante:

Teorema 1.14. *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 1.15. Utilizando o teorema acima podemos afirmar que a distribui¸cão da soma de n variáveis aleatórias, de Bernoulli, digamos $X_1, X_2, ..., X_n$, independentes e identicamente distribuidas com probabilidade de sucesso p, $0 < p < 1$ é a distribuição binomial de parâmetros n e p. A distribuição de Bernoulli tem função geradora de momentos

$$M_{X_1}(t) = E[e^{tX_1}] = pe^{1.t} + (1-p)e^{0.t} = (1-p) + pe^t.$$

Portanto

$$M_{X_1+...+X_n}(t) = M_{X_1}(t).....M_{X_n}(t) =$$
$$[(1-p) + pe^t].....[(1-p) + pe^t] = [(1-p) + pe^t]^n.$$

que é a função geradora de momentos da distribuição binomial.

Um outro exemplo seria o da soma de distribuições de Poisson independentes. Se X e Y são variáveis aleatórias independentes com parâmetros λ_1 e λ_2, respectivamente, então as suas funções geradoras de momentos so $M_X(t) = \exp[\lambda_1(e^t - 1)]$ e $M_Y(t) = \exp[\lambda_2(e^t - 1)]$, de maneira que

$$M_X(t).M_Y(t) = \exp[(\lambda_1 + \lambda_2)(e^t - 1)]$$

que é a função geradora de momentos de uma distribuição de Poisson com par ametro $\lambda_1 + \lambda_2$.

Observe que a função geradora de momentos da distribuição de Poisson é:

$$M_X(t) = E[e^{tX}] = \sum_{k=0}^{\infty} e^{tk} \frac{e^{-\lambda_1}(\lambda_1)^k}{k!} = e^{-\lambda_1} \sum_{k=0}^{\infty} \frac{(\lambda_. e^t 1)^k}{k!} =$$
$$e^{-\lambda_1} e^{\lambda_1 e^t} = e^{\lambda_1(e^t - 1)}.$$

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 1.16. 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 1.17. O coeficiente de correlação linear entre as variáveis aleatórias X e Y do exemplo é

$$\rho = \frac{Cov(X, Y)}{DP(X).DP(Y)} = \frac{\frac{2}{9}}{\sqrt{\frac{2}{9}.\frac{6}{9}}} = \frac{2}{\sqrt{8}} = 0,71.$$

Proposição 1.18. *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

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

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

Teorema 1.19. *Se X e Y são variáveis aleatórias com variâncias finitas, então $|\rho|\leq 1$. Em adição, se vale a igualdade, existem núemros 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, $Y=t.X+(1-t).E[X]$. Definindo $a=t$ e $b=(1-t).E[X]$ temos $Y=a.X+b$.

1.2. **Distribuições conjuntas contínuas.** Como no caso das distribuições contínuas unidimensionais para definirmos a probabilidade induzida pelo vetor aleatório $(X,Y):\Omega\to\Re^2$ devemos considerar $\beta^2=\beta X\beta$, a classe de subconjuntos de $\Re^2$ obtida através das operações de reunião, intersecção, complementar, em número finito ou infinito de subconjuntos na forma $(-\infty,s]\cap(-\infty,t],\forall s,t\in\Re$.

Definição 1.20. Seja $(\Omega,\Im,P)$ um espaço de probabilidade e (X,Y) uma aplicação de Ω em $\Re^2$. (X,Y) é um vetor aleatório contínuo se $X^{-1}((0,s])\cap Y^{-1}((0,t])\in\Im,\ \ \forall s,t\in\Re$. Denominamos $(\Re^2,\beta^2,P_{(X,Y)})$ como o espaço de probabilidade induzido por (X,Y).

A medida de probabilidade induzida pelo vetor aleatório

$$(X, Y) : \Omega \to \Re^2.$$

é definida por

$$P_{(X,Y)}(B_1 X B_2) = P(X^{-1}(B_1) \cap Y^{-1}(B_2)), \quad \forall B_1 X B_2 \in \beta^2.$$

Observe que $P_{(X,Y)}$ esta bem definida e

$$P_{(X,Y)}((-\infty, s] \cap (-\infty, t]) = P(X^{-1}((-\infty, s]) \cap Y^{-1}((-\infty, \{t])) =$$

$$P(\{w : X(w) \in (-\infty, t]\} \cap \{w : X(w) \in (-\infty, t]) =$$

$$P(X \leq s, Y \leq t) = F_{(X,Y)}(s, t)$$

é a função de distribuição do vetor aleatório (X, Y) que , neste caso, é uma função contínua.

Se a função de distribuição de (X, Y), for diferenciável em cada variável com $\frac{\delta^2 F_{(X,Y)}(s,t)}{\delta_s \delta_t} = f_{(X,Y)}(s, t)$, definimos

Definição 1.21. Se (X, Y) é uma vetor aleatório contínuo com função de distribuição

$$F_{X,Y)}(s, t) = \int_{-\infty}^{s} \int_{-\infty}^{t} f_{(X,Y)}(x, y) dy dx,$$

dizemos que $f_{(X,Y)}(s, t)$ é a função densidade de probabilidade de (X, Y) e que é (absolutamente) contínua.

Observação 1.22. É evidente que a função densidade de probabilidade é positiva, isto é, $f_{(X,Y)}(s, t) \geq 0, \forall s, t$ em seu domínio de definição e que

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{(X,Y)}(x, y) dy dx = 1.$$

Tambem calculamos

$$P(a < X \leq b, c < Y \leq d) = \int_{a}^{b} \int_{c}^{d} f_{(X,Y)}(x, y) dy dx$$

Exemplo 1.23. Considere que o tempo entre acidentes e o custo do sinistro correspondente são modelados pelo vetor aleatório (X, Y) com função densidade de probabilidade conjunta

$$f(x, y) = \left\{ \begin{array}{rcr} 0 & : & x < 0 \\ c. \exp[-(x + y)] & : & 0 \leq x < y < \infty \end{array} \right. .$$

onde c é um número real positivo. Qual o valor da constante c?

Para que $f(x,y)$ seja uma função densidade de probabilidade conjunta devemos ter:
a) $f(x,y) \geq 0$ e
b) $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{(X,Y)}(x,y)dydx = 1$.
Como $\exp[-(x+y)] > 0$ o item a) é obvio se $c > 0$.
Contudo, para que o item b seja verdadeiro devemos ter

$$c.\int_0^{\infty}\int_x^{\infty}\exp[-(x+y)]dydx = c.\int_0^{\infty}\exp[-x]\int_x^{\infty}\exp[-y]dydx =$$

$$\int_0^{\infty}\exp[-2x]dx = \frac{c}{2} = 1$$

e concluimos que $c = 2$. Para calcular $P(-\infty < X \leq 3, 2 < Y \leq 5)$

procedemos com

$$P(-\infty < X \leq 3, 2 < Y \leq 5) = 2.\int_0^3\int_{\max\{x,2\}}^5 \exp[-(x+y)]dydx =$$

$$2.\int_0^2\exp[-x]\int_2^5\exp[-y]dydx + 2.\int_2^3\exp[-x]\int_x^5\exp[-y]dydx =$$

$$2.(\exp[-2]-\exp[-5])\int_0^2\exp[-x]dx + 2.\int_2^3\exp[-x](\exp[-x]-\exp[-5])dx =$$

$$2.(\exp[-2]-\exp[-5]).(1-\exp[-2]+2.(\exp[-4]-$$
$$\exp[-6]) - 2.\exp[-5].(\exp[-2]-\exp[-3]).$$

Como no caso discreto podemos definir as funçõe densidades de probabilidades marginais de X e de Y:

Definição 1.24. Se (X,Y) é um vetor aleatório, as funções densidades de probabilidades marginais de X, $f_X(x)$, e de Y, $f_Y(y)$ são definidas, para x e y fixados, por

$$f_X(x) = \int_{-\infty}^{\infty} f_{(X,Y)}(x,y)dy$$

e

$$f_Y(y) = \int_{-\infty}^{\infty} f_{(X,Y)}(x,y)dx,$$

respectivamente.

Em continuação ao exemplo acima temos

Exemplo 1.25. Se

$$f(x,y) = \begin{cases} 0 & : \quad x < 0 \\ 2.\exp[-(x+y)] & : \quad 0 \leq x < y < \infty \end{cases}$$

temos que, para x fixado

$$f_X(x) = \int_x^{\infty} 2.\exp[-(x+y)]dy = 2.\exp[-2x], \quad 0 \leq x < \infty,$$

a função densidade de probabilidade de uma variável aleatória exponencial de parâmetro 2.

Para y fixado

$$f_Y(y) = 2.\int_0^y .\exp[-(x+y)]dx = 2.\exp[-y].(1-\exp[-y]), \ \ 0 \leq y < \infty.$$

Para o cálculo da esperança de uma transformação do vetor aleatório (X,Y) utilizamos um teorema análogo ao Teorema para o caso bivaria do discreto.

Teorema 1.26. *Se (X,Y) é um vetor aleatório e com função densidade de probabilidade $f_{(X,Y)}(x,y)$ e se $g(x,y)$ é uma função a valores reais, limitada, então*

$$E[g(X,Y)] = \int_{\infty}^{\infty} \int_{-\infty}^{\infty} g(x,y)dydx.$$

Em continuação ao exemplo acima temos

Exemplo 1.27. Se

$$f(x,y) = \begin{cases} 0 & : \quad x < 0 \\ 2.\exp[-(x+y)] & : \quad 0 \leq x < y < \infty \end{cases}$$

A transformação $X.Y$ tem média

$$E[X.Y] = \int_0^{\infty} \int_x^{\infty} x.y2.\exp[-(x+y)]dydx = \int_0^{\infty} 2.x\exp[-x] \int_x^{\infty} y.\exp[-y]dydx =$$

$$\int_0^{\infty} 2.x\exp[-x](x+1)\exp[-x]dx = 1.$$

Como X é uma variável aleatória exponencial de parâmetro 2, $E[X] = \frac{1}{2}$ e $Var(X) = \frac{1}{4}$.

Com respeito à variável Y temos

$$E[Y] = \int_0^{\infty} 2.y.\exp[-y](1-\exp[-y])dy = \frac{3}{2}$$

$$E[Y^2] = \int_0^\infty 2.y^2.\exp[-y](1-\exp[-y])dy = \frac{7}{2}$$

e $Var(Y) = \frac{5}{4}$.

Portanto $Cov(X,Y) = 1 - \frac{3}{4} = \frac{1}{4}$ e $\rho(X,Y) = \frac{\frac{1}{4}}{\sqrt{\frac{1}{4}\cdot\frac{5}{4}}} = 0,45$.

Como no caso discreto podemos definir as probabilidades condicionais de um evento do espaço amostral induzido por X, condicionado a um valor particular y de Y. Devemos ser cuidadosos pois sabemos que, como a variável Y é contínua, $P(Y=y)=0$. Tais probabilidades são calculadas através das densidades de probabilidades condicionais:

Definição 1.28. Se (X,Y) é um vetor aleatório com função densidade de probabilidade conjunta $f_{(X,Y)}(x,y)$, a densidade condicional de X, dado que $Y=y$ é definida por

$$f_{X|Y}(x|y) = \frac{f_{(X,Y)}(x,y)}{f_Y(y)}, \quad f_Y(y) > 0,$$

e densidade condicional de Y, dado que $X=x$ é definida por

$$f_{Y|X}(y|x) = \frac{f_{(X,Y)}(x,y)}{f_X(x)}, \quad f_X(x) > 0.$$

Em continuação ao exemplo acima temos

Exemplo 1.29. Se

$$f(x,y) = \begin{cases} 0 & : \quad x < 0 \\ 2.\exp[-(x+y)] & : \quad 0 \le x < y < \infty \end{cases}$$

temos

$f_X(x) = 2.\exp[-2x], \quad 0 \le x < \infty$ e

$f_Y(y) = 2.\exp[-y].(1-\exp[-y]), \quad 0 \le y < \infty$.

Portanto

$$f_{X|Y}(x|y) = \frac{2.\exp[-(x+y)]}{2.\exp[-y].(1-\exp[-y])} = \frac{\exp[-x]}{1-\exp[-y]}, \quad 0 \le x < y.$$

$$f_{Y|X}(y|x) = \frac{2.\exp[-(x+y)]}{2.\exp[-2x]} = \frac{\exp[-y]}{\exp[-x]}, \quad x \le y < \infty.$$

e podemos calcular

$$E[X|Y=y] = \int_0^y x.\frac{\exp[-x]}{1-\exp[-y]} = \frac{1}{1-\exp[-y]}[1-\exp[-y]-y\exp[-y]] = \varphi(y).$$

Assim $\varphi(Y)$ é uma variável aleatória, com média

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

$$\int_0^\infty \frac{1}{1-\exp[-y]}[1-\exp[-y]-y\exp[-y]].2.\exp[-y].(1-\exp[-y])dy =$$

$$\int_0^\infty (2.\exp[-y] - 2.\exp[-2y] - 2.y.\exp[-2y])dy = \frac{1}{2} = E[X].$$

Fato é que, como no caso discreto, as propriedades

$$E[X] = E\{E[X|Y]\}$$

$$Var(X) = E[Var(X|Y)] + Var(E[Y|X])$$

são verdadeiras.

Quando a função de densidade condicional é igual à densidade marginal dizemos que as variáveis são independentes.

Definição 1.30. As variáveis aleatórias X e Y do vetor aleatório (X, Y) são independentes se, e somente se,

$$f_{(X,Y)(x,y)=f_X(x).f_Y(y)}$$

para todo par (x, y) de (X, Y).

Exemplo 1.31. Se

$$f(x,y) = \begin{cases} 0 & : \quad x < 0 \\ \exp[-(x+y)] & : \quad 0 \leq x < \infty, 0 \leq y < \infty \end{cases}$$

temos

$$f_X(x) = \int_0^\infty \exp[-(x+y)]dy = \exp[-x],$$

$$f_Y(y) = \int_0^\infty \exp[-(x+y)]dx = \exp[-y]$$

de maneira que

$$f_{(X,Y)}(x,y) = \exp[-(x+y)] = \exp[-x].\exp[-y] = f_X(x).f_Y(y)$$

e concluimos que X e Y são independentes.

Observação 1.32. **Distribuição normal bivariada** Um vetor aleatório bidimensional (X, Y), representando duas características de uma população, tem distribuição normal bivariada, $N(\mu_X, \mu_Y, \sigma_X^2, \sigma_Y^2, \rho_{XY})$, com $E[X] = \mu_X$, $E[Y] = \mu_Y$, $Var(X) = \sigma_X^2$, $Var(Y) = \sigma_Y^2$ e $CORR(X, Y) = \rho_{XY} = \rho$ se, e somente se,

(X, Y) tem função densidade de probabilidade

$$f_{XY}(x, y) = [\frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} \exp\{-K(x, y)\}.$$

onde

$$K(x, y) = \frac{1}{2(1-\rho^2)}[(\frac{x-\mu_X}{\sigma_X})^2 + (\frac{y-\mu_Y}{\sigma_Y})^2 - \frac{2\rho(x-\mu_X)(y-\mu_Y)}{\sigma_X\sigma_Y}].$$

O vetor aleatório bidimensional normal tem algumas propriedades impares, tais como:

As marginais X e Y são independentes se, e somente se, $COV(X, Y) = 0$, pois neste caso a densidade conjunta é o produto de suas marginais as quais tambem tem distribuições normais.

$$f_{XY}(x, y) = \frac{1}{\sqrt{2\pi}\sigma_X} \exp[(\frac{x-\mu_X}{\sigma_X})^2]. \frac{1}{\sqrt{2\pi}\sigma_Y} \exp[(\frac{y-\mu_Y}{\sigma_Y})^2] = f_X(x).f_Y(y),$$

A distribuição condicional de Y dado X, $(Y|X = x)$ (de X dado Y, $(X|Y = y)$) tambem tem distribuição normal, em particular

$$(Y|X = x) \sim N(\mu_Y + \frac{\rho\sigma_Y}{\sigma_X}(x - \mu_X); \sigma_Y^2(1-\rho^2))$$

Portanto

$$E[Y|X = x] = \mu_Y + \frac{\rho\sigma_Y}{\sigma_X}(x - \mu_X) = (\mu_Y - \frac{\rho\sigma_Y}{\sigma_X}) + \frac{\rho\sigma_Y}{\sigma_X}x$$

é uma função linear de x de parâmetros $\mu_Y - \frac{\rho\sigma_Y}{\sigma_X}$ e $\frac{\rho\sigma_Y}{\sigma_X}$.

Transformações de variáveis aleatórias

Suponha que desejamos obter a função densidade de probabilidade conjunta de um vetor aleatório (Y_1, Y_2), $f_{(Y_2,Y_2)}(y_1, y_2)$ obtido através de transfomações do vetor aleatório (X_1, X_2), com função densidade de probabilidade conjunta $f_{(X_2,X_2)}(x_1, x_2)$, onde $Y_1 = g_1(X_1, X_2)$ e $Y_2 = g_2(X_1, X_2)$ definem transformações bijetivas de $\Re_{(X_1,X_2)} = \{(x_1, x_2) : f_{(X_1,X_2)}(x_1, x_2) > 0\}$ em $\Re_{(Y_1,Y_2)} = \{(y_1, y_2) : f_{(Y_2,Y_2)}(y_1, y_2) > 0\}$. A solução esta no teorema seguinte que admitiremos sem prova a qual é encontrada em literatura especifica.

Teorema 1.33. *Seja* (X_1, X_2) *um vetor aleatório com função densidade de probabilidade conjunta* $f_{(X_2,X_2)}(x_1, x_2)$. *Se*

a) $Y_1 = g_1(X_1, X_2)$ *e* $Y_2 = g_2(X_1, X_2)$ *definem transformações bijetivas e contínuas de* $\Re_{(X_1,X_2)}$ *em* $\Re_{(Y_1,Y_2)}$*;*

b) as derivadas parciais de $x_1 = g_1^{-1}(y_1, y_2)$ *e* $x_2 = g_2^{-1}(y_1, y_2)$ *são contínuas em* $\Re_{(Y_1,Y_2)}$*;*

c) O Jacobiano $J = \frac{\delta x_1}{\delta y_1}.\frac{\delta x_2}{\delta y_2} - \frac{\delta x_1}{\delta y_2}.\frac{\delta x_2}{\delta y_1}$ *é diferente de zero,* 0, *em* $\Re_{(Y_1,Y_2)}$,

então a função densidade de probabilidade conjunta de (Y_1, Y_2) *é*

$$f_{(Y_2,Y_2)}(y_1, y_2) = |J|.f_{(X_1,X_2)}(g_1^{-1}(y_1, y_2), g_2^{-1}(y_1, y_2)), \;\; (y_1, y_2) \in \Re_{(Y_1,Y_2)}.$$

Exemplo 1.34. Suponha que X_1 e X_2 são variáveis aleatórias independentes e identicamente distribuidas com distribuição uniforme no intervalo $[0, 1]$. Então $\Re_{(X_1,X_2)} = \{(x_1, x_2) : 0 \leq x_1 \leq 1, \;\; 0 \leq x_2 \leq 1\}$.

Sejam $Y_1 = g_1(X_1, X_2) = X_1 + X_2$ e e $Y_2 = g_2(X_1, X_2) = X_2 - X_1$. Então $x_1 = g_1^{-1}(y_1, y_2) = \frac{y_1 - y_2}{2}$ e $x_2 = g_2^{-1}(y_1, y_2) = \frac{y_1 + y_2}{2}$ e

$$J = \frac{\delta x_1}{\delta y_1}.\frac{\delta x_2}{\delta y_2} - \frac{\delta x_1}{\delta y_2}.\frac{\delta x_2}{\delta y_1} =$$

$$\frac{1}{2}.\frac{1}{2} + \frac{1}{2}.\frac{1}{2} = \frac{1}{2}.$$

Os domínios $\Re_{(X_1,X_2)}$ e $\Re_{(Y_1,Y_2)}$ estão na Figura 8.1.

Figura 8.1 - $\Re_{(X_1,X_2)}$ *e* $\Re_{(Y_1,Y_2)}$

A fronteira $\{(x_1, x_2) : x_1 = 0, \quad 0 \leq x_2 \leq 1\}$ de $\Re_{(X_1,X_2)}$ é transformada na fronteira $\{(y_1, y_2) : 0 \leq y_1 \leq 1, \quad y_2 = y_1\}$ de $\Re_{(Y_1,Y_2)}$;

A fronteira $\{(x_1, x_2) : 0 \leq x_1 \leq 1, \quad x_2 = 0\}$ de $\Re_{(X_1,X_2)}$ é transformada na fronteira $\{(y_1, y_2) : 0 \leq y_1 \leq 1, \quad y_2 = -y_1\}$ de $\Re_{(Y_1,Y_2)}$;

A fronteira $\{(x_1, x_2) : 0 \leq x_1 \leq 1, \quad x_2 = 1\}$ de $\Re_{(X_1,X_2)}$ é transformada na fronteira $\{(y_1, y_2) : 1 \leq y_1 \leq 2, \quad y_2 = 2 - y_1\}$ de $\Re_{(Y_1,Y_2)}$;

A fronteira $\{(x_1, x_2) : x_1 = 1 \;\; 0 \leq x_2 \leq 1\}$ de $\Re_{(X_1,X_2)}$ é transformada na fronteira $\{(y_1, y_2) : 1 \leq y_1 \leq 2, \quad y_2 = y_1 - 2\}$ de $\Re_{(Y_1,Y_2)}$;

A função densidade de probabilidade conjunta de (Y_1, Y_2) é

$$f_{(Y_2,Y_2)}(y_1, y_2) = \frac{1}{2}.1_{[0,1]}(\frac{y_1 - y_2}{2}).1_{[0,1]}(\frac{y_1 + y_2}{2}) = \frac{1}{2}, \;\; (y_1, y_2) \in \Re_{(Y_1,Y_2)}.$$

As condições a, b e c do Teorema são de certa forma restritivas. As transformações (Y_1, Y_2) de $\Re_{(X_1,X_2)}$ podem não ser injetoras, mas o domínio $\Re_{(X_1,X_2)}$ pode ser particionado em várias partes, digamos $\Re^1_{(X_1,X_2)}, \Re^2_{(X_1,X_2)}, ...\Re^m_{(X_1,X_2)}$, de maneira que a restrição da transformação (Y_1, Y_2) de $\Re^i_{(X_1,X_2)}$ em $\Re_{(Y_1,Y_2)}$, $1 \leq i \leq m$, seja injetora. Neste caso, definindo $x^i_1 = g_1^{i^{-1}}(y_1, y_2)$ e $x^i_2 = g_2^{i^{-1}}(y_1, y_2)$ em $\Re^i_{(X_1,X_2)}$ e $J^i = \frac{\delta x^i_1}{\delta y_1}.\frac{\delta x^i_2}{\delta y_2} - \frac{\delta x^i_1}{\delta y_2}.\frac{\delta x^i_2}{\delta y_1}$ podemos aplicar o Teorema em cada domínio da partição e temos:

Corolário 1.35. *Seja (X_1, X_2) um vetor aleatório com função densidade de probabilidade conjunta $f_{(X_2,X_2)}(x_1, x_2)$. Se*

a) $Y_1 = g^i_1(X_1, X_2)$ e $Y_2 = g^i_2(X_1, X_2)$ definem transformações bijetivas e contínuas de $\Re^i_{(X_1,X_2)}$ em $\Re_{(Y_1,Y_2)}$;

b) as derivadas parciais de $x^i_1 = g_1^{i^{-1}}(y_1, y_2)$ e $x^i_2 = g_2^{i^{-1}}(y_1, y_2)$ são contínuas em $\Re_{(Y_1,Y_2)}$;

c) O Jacobiano J^i é diferente de zero, 0, em $\Re_{(Y_1,Y_2)}$,

então a função densidade de probabilidade conjunta de (Y_1, Y_2) é

$$f_{(Y_2,Y_2)}(y_1, y_2) = |J^i|.f_{(X_1,X_2)}(g_1^{i^{-1}}(y_1, y_2), g_1^{i^{-1}}(y_1, y_2)), \;\; (y_1, y_2) \in \Re_{(Y_1,Y_2)}.$$

Teorema 1.36. *Sejam $Z \sim N(0, 1)$ uma variável aleatória com distribuição normal padrão e $Y \sim \chi^2_n$ uma variável aleatória com distribuição Qui-quadrado com n graus de liberdade, independentes. Então a variável aleatória*

$$T = \frac{Z}{\sqrt{\frac{y}{n}}},$$

tem função densidade de probabilidade

$$f(t)\frac{\Gamma(\frac{n+1}{2})}{\Gamma(\frac{n}{2})\sqrt{n\pi}}(1+\frac{t^2}{n})^{-\frac{v+1}{2}}, \quad -\infty < t < \infty.$$

A variável aleatória obtida no teorema anterior tem distribuição de Student com n graus de liberdade e é denoda por T_n. n é o parâmetro da distribuição e pode-se provar que

$$E[T_n] = 0, \quad Var(T_n) = \frac{n}{n-2}, \quad n > 2.$$

Uma tabela da distribuição de Student esta disponível no fim do livro. É importante notar que a distribuição de T_n aproxima-se de uma distribuição normal padrão para n suficientemente grande (n ¿ 120).

Teorema 1.37. *Sejam X e Y duas varáveis aleatórias independentes com distribuições qui-quadrado com n e m graus de liberdades, respectivamente. Então a variável aleatória*

$$F = \frac{\frac{X}{n}}{\frac{Y}{m}}$$

tem função densidade de probabilidade

$$f(x) = \frac{\Gamma(\frac{n+m}{2})}{\Gamma(\frac{n}{2})\Gamma(\frac{m}{2})}(\frac{n}{m})^{\frac{n}{2}}\frac{x^{\frac{n-2}{2}}}{(1+\frac{nx}{m})^{\frac{n+m}{2}}}, \quad x > 0.$$

Diremos que a variável aleatória F, do teorema anterior tem distribuição F com n e m graus de liberdade, nessa ordem, o que denotamos por $F \sim F_{n,m}$. n e m são os parâmetros da distribuição e que $F_{n,m}$ é diferente de $F_{m,n}$. Pode-se provar que

$$E[F_{n,m}] = \frac{m}{m-2}, \quad Var(F_{n,m}) = \frac{2m^2(n+m-2)}{n(m-2)^2(m-4)}.$$

Uma tabela da distribuição $F_{n,m}$ esta disponível no fim do livro.

Como no caso univariado a função geradora de momentos bivariada caracteriza completamente o modelo probabilístico.

Definição 1.38. A função geradora de momentos de um vetor aleatório (X, Y) é definida por

$$M_{(X,Y)}(t_1, t_2) = E[\exp[t_1.X + t_2.Y].$$

Observação 1.39. Segue da definição e da caracterização da distribuição da variável aleatória pela função geradora de momentos que, se X e Y são variáveis aleatórias independentes se, e somente se,

$M_{(X,Y)}(t_1, t_2) = E[\exp[t_1.X+t_2.Y] = E[\exp[t_1.X]]E[\exp[t_1.Y]] = M_X(t_1)M_Y(t_2).$

Se (X, Y) tem função densidade de probabilidade $f_{(X,Y)}(x, y)$,

$$M_{(X,Y)}(t_1, t_2) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \exp[t_1.x + t_2.y].f_{(X,Y)}(x, y)dydx.$$

Exemplo 1.40. Se (X, Y) tem distribuição normal bivariada, $N(\mu_X, \mu_Y, \sigma_X^2, \sigma_Y^2, \rho_{XY})$, então, pode-se provar que a função geradora de momentos bivariada é

$$M_{(X,Y)}(t_1, t_2) = \exp[t_1.\mu_X + t_2.\mu_Y + \frac{1}{2}(t_1^2\sigma_X^2 + 2.\rho.\sigma_X.\sigma_Y + t_2^2\sigma_Y^2].$$

No caso em que $\rho = 0$ temos

$$M_{(X,Y)}(t_1, t_2) = \exp[t_1.\mu_X + t_2.\mu_Y + \frac{1}{2}(t_1^2\sigma_X^2 + t_2^2\sigma_Y^2] =$$
$$\exp[t_1.\mu_X + \frac{1}{2}t_1^2\sigma_X^2]\exp[t_2.\mu_Y + \frac{1}{2}t_1^2\sigma_Y^2] = M_X(t_1)M_Y(t_2),$$

e as variáveis aleatórias X e Y são independentes.