TIPOS DE DEPENDÊNCIA

ENTRE VARIÁVEIS ALEATÓRIAS

FLÁVIO WAGNER RODRIGUES

TESE APRESENTADA AO

INSTITUTO DE MATEMÁTICA E ESTATÍSTICA

DA

UNIVERSIDADE DE SÃO PAULO

PARA O CONCURSO DE

LIVRE DOCÊNCIA

NO

DEPARTAMENTO DE ESTATÍSTICA

- São Paulo, agosto de 1981 -

ÍNDICE

INTRODUÇÃO

O interesse por formas de dependência de variáveis aleatórias é relativamente recente. O artigo de Lehmann (1966) é um marco importante não apenas pelas idéias que introduz mas, também pelas sugestões de aplicações desses conceitos à outras áreas da Probabilidade e da Estatística.

A partir daí, estimulada pelo sucesso das aplicações à Teoria da Confiabilidade, desenvolve-se uma atividade intensa de pesquisa, especialmente em dependência positiva. São desse período os resultados sobre associação encontrados, por exemplo, em Esary & Proschan (1968,1972). Os trabalhos sobre dependência negativa são ainda mais recentes e os mais importantes, de Karlin & Rinnot e Block, Savits & Shakhed são de 1980.

O nosso interesse pelo assunto foi fortemente estimulado pelos seminários, dirigidos pelo Prof. Henry W. Block, da Universidade de Pittsburgh, quando de sua estada como Professor Visitante do IME-USP, no primeiro semestre de 1980.

Durante esse período obtivemos alguns resultados relativos à associação de vetores normalmente distribuídos e à geração de variáveis aleatórias que satisfaziam condições fortes de dependência. Posteriormente, conseguimos mostrar também que a condição de Karlin-Rinnot para que uma densidade normal multivariada fosse $SMRR_2$, era equivalente a impormos que ela fosse a densidade de uma distribuição condicional, gerada por vetores com componentes independentes. Com isso fica estabelecida uma conexão, que acreditamos ser importante, entre os resultados de Karlin & Rinnot e os de Block, Savits & Shakhed.

Este trabalho se divide basicamente em dois capítulos nos quais são tratados aspectos de dependência positiva e negativa. O terceiro e último capítulo complementa os anteriores com comentários e sugestões sobre alguns problemas em aberto.

Não houve intenção de se fazer aqui um apanhado geral e completo sobre todas as condições de dependência existentes na literatura. Existem diversos conceitos de dependência, alguns até importantes, que não são sequer mencionados neste trabalho. O que se procurou fazer foi uma análise das idéias de dependência positiva e negativa, com ênfase nas condições que mais se relacionassem com os nossos resultados.

Não se chega ao final de um trabalho como este sem contrair débitos de gratidão com pessoas que de uma forma ou outra contribuiram para sua realização. Gostaríamos, portanto de deixar registrados aqui nossos reconhecimento e gratidão a:

Prof. Henry W. Block pelo entusiasmo que conseguiu nos transmitir e pelas valiosas sugestões dadas no decorrer deste trabalho.

Wagner de Souza Borges e Carlos Alberto de Bragança Pereira, meus colegas no Seminário, pelas estimulantes discussões sobre o assunto.

Carlos Alberto Barbosa Dantas que sempre se interessou pelo andamento do trabalho e nunca faltou com seu estímulo e encorajamento nos momentos mais difíceis.

João Baptista Esteves de Oliveira que mesmo com a premencia do tempo conseguiu uma vez mais realizar um excelente trabalho de datilografia.

Finalmente uma menção especial para minha esposa Maria Regina e meus filhos Paulo e Ana Lúcia, pela paciência e compreensão com que suportaram minhas freqüentes crises de mau humor nos últimos doze meses.

- ⊟ -

CAPÍTULO 1

1.1 - DEPENDÊNCIA POSITIVA NO CASO MULTIVARIADO

Uma análise sucinta e razoavelmente completa sobre dependência positiva entre duas variáveis aleatórias X e Y pode ser encontrada em Barlow & Proschan (1975, pp. 142 a 146). No caso multivariado, as noções existentes são muito mais numerosas e as relações entre elas bem mais complicadas. Isto se deve principalmente ao fato de que as generalizações sugeridas pelo caso bivariado não tem o mesmo relacionamento direto que tinham naquela situação. Esse estado de coisas levou diversos autores a introduzirem novos conceitos de dependência positiva, sempre com o objetivo de obterem critérios, de fácil verificação, que conduzissem à formas de dependência importantes nas aplicações práticas.

Não pretendemos fazer aqui uma análise de todos os critérios de dependência positiva conhecidos. Discutiremos apenas os conceitos, a nosso ver, mais importantes do ponto de vista prático, com ênfase especial naqueles que serão utilizados nas próximas seções.

Antes de entrarmos propriamente no assunto, vale a pena observar que embora dependência seja uma propriedade da medida induzida e portanto da distribuição conjunta, já está consagrado o abuso de linguagem que consiste em falarmos de dependência indiferentemente como uma propriedade das variáveis ou de sua distribuição.

Iniciaremos com a generalização do conceito de dependência quadrante positiva introduzida em Lehmann (1966).

DEFINIÇÃO 1.1.1 - Diremos que a dependência positiva entre as variáveis aleatórias $X_1,\ldots,X_n$ é *superior* se para toda n-upla, $(x_1,\ldots,x_n)$, de números reais tivermos:

$$(1.1.1) \qquad P[X_1>x_1,\ldots,X_n>x_n] \geq \prod_{i=1}^{n} P[X_i>x_i].$$

DEFINIÇÃO 1.1.2 - Diremos que a dependência positiva entre X_1, $\ldots,X_n$ é *inferior* se para toda n-upla $(x_1,\ldots,x_n)$ de números reais, tivermos:

$$(1.1.2) \qquad P[X_1\leq x_1,\ldots,X_n\leq x_n] \geq \prod_{i=1}^{n} P[X_i\leq x_i]$$

OBSERVAÇÕES 1 - Por uma questão de facilidade, quando nos referirmos à variáveis que satisfazem as condições (1.1.1) ou (1.1.2), diremos apenas que essas variáveis são DPS ou DPI respectivamente.

2 - No caso particular $n = 2$, as condições (1.1.1) e (1.1.2) são equivalentes e qualquer uma delas pode ser usada para definir

dependência quadrante positiva. O exemplo sequinte mostra que essa equivalência se perde quando n é maior que 2.

EXEMPLO 1.1.3 - Seja $\underset{\sim}{X} = (X_1, X_2, X_3)$ um vetor aleatório que assume valores no conjunto $\{(1,1,1),(0,1,0),(0,0,1)\}$ com probabilidades iguais a 1/3 para cada ponto. É fácil mostrar que $\underset{\sim}{X}$ satisfaz a condição 1.1.1. No entanto se fizermos $x_1 = x_2 = x_3 = 0$, teremos:

$$P[X_1 \leq 0, X_2 \leq 0, X_3 \leq 0] = 0$$

e

$$P[X_1 \leq 0]P[X_2 \leq 0]P[X_3 \leq 0] = \frac{2}{27}$$

o que mostra que $\underset{\sim}{X}$ não satisfaz (1.1.2). Uma das aplicações práticas das noções de DPS e DPI diz respeito à possibilidade de obtermos estimativas aproximadas de probabilidades que dependem da distribuição conjunta, (e que porisso nem sempre são fáceis de serem calculadas), conhecendo apenas as distribuições marginais. As determinações de coeficientes de confiança, de regiões críticas de testes, da confiabilidade de um sistema, são alguns exemplos de situações nas quais o conhecimento, de que as distribuições envolvidas são DPS ou DPI, pode ser bastante útil.

ASSOCIAÇÃO

A medida de associação mais conhecida e mais utilizada é sem dúvida a covariância entre duas variáveis aleatórias X e Y. Quando ela é introduzida nos textos elementares de Estatística, em geral é mencionado que um valor positivo da covariância é uma indicação de que X e Y tendem a variar num mesmo sentido, enquanto que um valor negativo indicaria que X e Y tendem a variar em sentidos opostos. Dada a pouca informação contida na covariância (afinal ela nada mais é do que um dos momentos da distribuição conjunta), seria demais esperarmos que ela funcionasse bem como medida de dependência. Vamos lembrar, por exemplo, que ela vale zero sempre que X e Y forem independentes e também que ela pode valer zero até mesmo em situações nas quais uma das variáveis é função da outra.

Dada a facilidade de se tratar matematicamente com covariâncias, era natural que se procurasse utilizar a mesma idéia básica para definir medidas mais fortes de dependência. Uma possibilidade é dizermos que X e Y são associadas quando a covariância entre f(X) e f(Y) é não negativa, para todo par de funções crescentes f e g, para o qual a covariância exista. Essa condição conduz a um critério bem mais forte de dependência, cujo único defeito é só poder ser utilizado no caso bivariado.

O conceito de associação, que introduziremos abaixo, é mais forte ainda e tem a vantagem adicional de se aplicar a um número finito qualquer de variáveis aleatórias.

Inicialmente vamos relembrar algumas definições e introduzir a notação necessária.

1. Se $\underset{\sim}{x} = (x_1,\ldots,x_n)$ e $\underset{\sim}{y} = (y_1,\ldots,y_n)$ são elementos de R^n, diremos que $\underset{\sim}{x} \le \underset{\sim}{y}$ se $x_j \le y_j$ para $j = 1,\ldots,n$. Escreveremos $\underset{\sim}{x} < \underset{\sim}{y}$ se $\underset{\sim}{x} \le \underset{\sim}{y}$ e existe ao menos um j, para o qual $x_j < y_j$.
2. Uma função f de R^n em R é crescente se $f(\underset{\sim}{y}) \ge f(\underset{\sim}{x})$ sempre que $\underset{\sim}{y} > \underset{\sim}{x}$.
3. Uma função f de R^n em R é binária se $f(\underset{\sim}{x})$ for igual a 0 ou 1 para todo $\underset{\sim}{x} \in R^n$.

DEFINIÇÃO 1.1.4 - Diremos que as variáveis aleatórias $X_1, X_2, \ldots X_n$ são *associadas* se para todo par de funções binárias crescentes, γ e δ, tivermos:

$$\text{(1.1.3)} \qquad \mathrm{Cov}(\gamma(X_1,\ldots,X_n),\ \delta(X_1,\ldots,X_n)) \ge 0.$$

OBSERVAÇÃO - O fato de considerarmos apenas funções binárias pode dar a impressão de que seria possível fortalecer ainda mais a condição (1.1.3). No entanto, uma desigualdade de Hoeffding (ver Lehmann, 1966) nos permite mostrar que se um vetor aleatório $\underset{\sim}{X} = (X_1,\ldots,X_n)$ satisfaz (1.1.3) então $\mathrm{Cov}[f(\underset{\sim}{X}), g(\underset{\sim}{X})] \ge 0$ para todo par de funções crescentes para o qual a covariância exista.

Damos a seguir algumas propriedades importantes da associação.

A.1 - O conjunto formado por uma única variável aleatória X é associado.

A.2 - Qualquer subconjunto de um conjunto de variáveis associadas é associado.

A.3 - Funções crescentes de variáveis aleatórias associadas são associadas.

A.4 - Seja $\underset{\sim}{X} = (X_1, \ldots, X_n)$ um vetor associado e $\underset{\sim}{Y} = (Y_1, \ldots, Y_m)$ um outro vetor associado, independente de $\underset{\sim}{X}$. O vetor $\underset{\sim}{Z} = (X_1, \ldots, X_n, Y_1, \ldots, Y_m)$ será também um vetor associado.

Para as demonstrações ver Esary, Proschan & Walkup (1967).

Das propriedades A.1 e A.4 segue-se que qualquer conjunto finito de variáveis aleatórias independentes é associado.

PROPOSIÇÃO 1.1.5 - Seja $\underset{\sim}{X} = (X_1, \ldots, X_n)$ um vetor associado e $\underset{\sim}{Y} = (Y_1, \ldots, Y_n)$ um outro vetor associado independente de $\underset{\sim}{X}$. Nessas condições o vetor $\underset{\sim}{Z} = (X_1+Y_1, X_2+Y_2, \ldots, X_n+Y_n)$ será também associado.

DEMONSTRAÇÃO - Imediata usando propriedades A.4 e A.3.

EXEMPLO 1.1.6 - Seja $\underset{\sim}{X} = (X_1, X_2)$ um vetor aleatório com distribuição conjunta normal com médias iguais a zero, variâncias iguais a 1 e suponha que $Cov(X_1, X_2) = \rho > 0$. Então $\underset{\sim}{X}$ é associado. De fato, seja $\underset{\sim}{Y}$ um vetor aleatório formado por duas variáveis aleatórias normais, independentes, com médias iguais a zero e variâncias iguais a 1. Vamos definir $\underset{\sim}{Z} = (Z_1, Z_2)$ através de:

$$Z_1 = Y_1$$

$$Z_2 = \rho Y_1 + \sqrt{1-\rho^2}\,Y_2$$

Como $\rho > 0$ e $\underset{\sim}{Y}$ é associado segue-se de A.3 que $\underset{\sim}{Z}$ é associado. Como a distribuição de $\underset{\sim}{Z}$ é a mesma de $\underset{\sim}{X}$ segue-se que $\underset{\sim}{X}$ é associado.

A proposição seguinte contém dois resultados sobre associação entre variáveis aleatórias binárias isto é, sobre associação entre indicadores de eventos.

PROPOSIÇÃO 1.1.7 - a) Se A e B são dois eventos quaisquer, uma condição necessária e suficiente para que I_A e I_B sejam associadas é que $P(A \cap B) \geq P(A) \cdot P(B)$.

b) Se $X_1, X_2, \ldots, X_n$ são variáveis aleatórias aleatórias binárias associadas, então também serão associadas as variáveis aleatórias:

$$1 - X_1,\ 1 - X_2, \ldots,\ 1 - X_n.$$

DEMONSTRAÇÃO - a) As funções binárias crescentes de I_A e I_B podem ser determinadas explicitamente e são as seguintes:

$$f_1 \equiv 0;\ f_2 = \min(I_A, I_B);\ f_3 = I_A;$$
$$f_4 = I_B;\ f_5 = \max(I_A, I_B);\ f_6 \equiv 1.$$

É fácil ver que:

$$f_1 \leq f_2 \leq f_3 \leq f_5 \leq f_6$$

e que:

$$f_1 \leq f_2 \leq f_4 \leq f_5 \leq f_6.$$

Para mostrar que I_A e I_B são associadas nós teríamos que mostrar que $E(f_j f_k) \geq E(f_j)E(f_k)$, para todo $j \neq k$, $j=1, 2,\ldots,6$, $k=1,2,\ldots,6$. É fácil verificar que essa condição estará trivialmente satisfeita sempre que $f_j \leq f_k$. O único par que não satisfaz essa relação de ordem é o par (f_3,f_4). Conclue-se portanto que I_A e I_B serão associadas se e somente se

$$E(f_3f_4) \geq E(f_3)E(f_4).$$

Da definição de f_3 e f_4 temos:

$$E[f_3f_4] - E(f_3)E(f_4) = P[I_A=1 \text{ e } I_B=1] -$$
$$- P[I_A=1]\cdot P[I_B=1] = P(A\cap B)-P(A)\cdot P(B).$$

b) Sejam γ' e δ' funções binárias crescentes de $(1-X_1, 1-X_2,\ldots,1-X_n)$. Vamos definir:

$$\gamma(X_1,\ldots,X_n) = 1-\gamma'[1-X_1,\ldots,1-X_n]$$
$$\delta(X_1,\ldots,X_n) = 1-\delta'[1-X_1,\ldots,1-X_n].$$

Verifica-se facilmente que γ e δ são funções binárias crescentes de $(X_1,\ldots,X_n)$. Como esse vetor é associado temos:

$$E(\gamma\delta) \geq E(\gamma)E(\delta).$$

Segue-se então que $E(\gamma'\delta') \geq E(\gamma')\cdot E(\delta')$ o que conclui a demonstração.

PROPOSIÇÃO 1.1.8 - Seja $\underset{\sim}{X}$ um vetor aleatório associado. Então $\underset{\sim}{X}$ satisfaz (1.1.1) e (1.1.2) ou seja, $\underset{\sim}{X}$ é DPS e DPI.

DEMONSTRAÇÃO - Seja $\underset{\sim}{X} = (X_1,\ldots,X_n)$ e vamos definir para cada $j = 1,\ldots,n$,

$$Y_j = \begin{cases} 0 \text{ se } X_j \leq x_j \\ 1 \text{ se } X_j > x_j. \end{cases}$$

Sejam agora:

$$f(X_1,\ldots,X_n) = Y_1\cdot Y_2,\ldots,Y_{n-1}$$

$$g(X_1,\ldots,X_n) = Y_n.$$

É fácil ver que f e g são funções binárias crescentes, e como $\underset{\sim}{X}$ é associado, temos:

$$E[f\cdot g] \geq E(f)\cdot E(g)$$

ou seja:

$$P[X_1>x_1,\ldots,X_{n-1}>x_{n-1},X_n>x_n] \geq$$

$$\geq P[X_1>x_1,\ldots,X_{n-1}>x_{n-1}]P[X_n>x_n].$$

Segue-se por indução que $\underline{X}$ é DPS. Para mostrar que é DPI basta observar pela parte b) da Proposição 1.1.7 que $1-Y_1,\ldots,1-Y_n$ são variáveis associadas.

1.2 - DEPENDÊNCIA POSITIVA E ORDENAÇÃO ESTOCÁSTICA

Vamos considerar agora dois conceitos de dependência positiva que estão relacionados com a idéia de ordenação estocástica de variáveis e de vetores aleatórios. Inicialmente discutiremos as principais definições e propriedades da ordenação estocástica.

DEFINIÇÃO 1.2.1 - Diremos que uma variável aleatória X é estocásticamente maior do que uma variável aleatória Y se para todo t real, tivermos: $P[X>t] \geq P[Y>t]$.

OBSERVAÇÃO - Se X e Y tem a mesma distribuição nós diremos que X é estocasticamente igual a Y [indica-se: $X \overset{st}{=} Y$]. Se X for estocasticamente maior do que Y escreveremos $X \overset{st}{\geq} Y$.

As propriedades que apresentamos a seguir, decorrem imediatamente da definição.

O_1 - Se $X \overset{st}{=} Y$ e $\underline{a}$ é uma constante positiva, então $X+a \overset{st}{\geq} Y$. Em particular, $X+a \overset{st}{\geq} X$.

O_2 - Se $P[Y\geq 0]=1$ então $X+Y \overset{st}{\geq} X$ para todo X.

O_3 - Se $P[X\geq 0]=1$ e $\underline{a}$ for uma constante maior do que 1, então $aX \overset{st}{\geq} X$.

OBSERVAÇÃO - O conceito de ordenação estocástica depende apenas das distribuições marginais de X e Y e não de sua distribuição conjunta. Como veremos nos exemplos, em algumas situações, a verificação de que $X \overset{st}{\geq} Y$ fica mais simples quando supomos que X e Y tem uma distribuição conjunta conveniente.

EXEMPLO 1.2.2 - Suponha que X e Y tem distribuições normais com variâncias iguais. Então $X \overset{st}{\geq} Y$ se e somente se $E(X) \geq E(Y)$. De fato, suponha que $a = E(X)-E(Y) \geq 0$. Das hipóteses feitas segue-se que:

$$X + E(Y) - E(X) = X - a \overset{st}{=} Y.$$

De O_1 segue-se que $X \overset{st}{\geq} Y$. Reciprocamente, suponha $X \overset{st}{\geq} Y$.

$$P[X>t] = P[X-E(X)>t-E(X)] = P[Y-E(Y)>t-E(X)] =$$

$$= P[Y>t+E(Y)-E(X)] \geq P[Y>t].$$

Segue-se necessariamente que $E(Y) - E(X) \leq 0$.

EXEMPLO 1.2.3 - Suponha que X e Y tem distribuições de Poisson, com médias λ_1 e λ_2 com $\lambda_1 < \lambda_2$. Então $Y \overset{st}{\geq} X$. De acordo com a observação feita logo após as propriedades da ordenação estocástica, não há perda de generalidade se supusermos que X e Y são independentes. Nessas condições, $Z = Y-X$ tem distribuição de Poisson e portanto, $P[Z\geq 0] = 1$. Segue-se então, de O_3 que $X+Z = X+Y-X = Y$ é estocasticamente maior do que X.

A generalização do conceito de ordenação estocástica, para vetores aleatórios $\underset{\sim}{X}$ e $\underset{\sim}{Y}$, é feita com base numa idéia semelhante a que foi utilizada na generalização da covariância.

DEFINIÇÃO 1.2.4 - Diremos que o vetor aleatório $\underset{\sim}{X} = (X_1,\ldots,X_n)$ é estocasticamente maior do que $\underset{\sim}{Y} = (Y_1,\ldots,Y_n)$ se $g(\underset{\sim}{X}) \overset{st}{\geq} g(\underset{\sim}{Y})$ para todo função g mensurável, crescente de R^n em R. Como em toda generalização desse tipo, a principal dificuldade é de ordem prática, isto é, o problema é como verificar se um dado vetor $\underset{\sim}{X}$ é ou não estocasticamente maior do que um outro vetor $\underset{\sim}{Y}$. Nesse sentido, a definição e a proposição que apresentaremos a seguir são bastante úteis.

DEFINIÇÃO 1.2.5 - Um subconjunto $S \subset R^n$ se denominará *superior* se da hipótese $\underset{\sim}{x} \in S$ decorrer que $\underset{\sim}{y} \in S$ para todo $\underset{\sim}{y} \geq \underset{\sim}{x}$.

PROPOSIÇÃO 1.2.6 - Uma condição necessária e suficiente para que $\underset{\sim}{X} \overset{st}{\geq} \underset{\sim}{Y}$ é que para todo conjunto mensurável superior S de R^n tenhamos:

(1.2.1) $$P[\underset{\sim}{X} \in S] \geq P[\underset{\sim}{Y} \in S].$$

DEMONSTRAÇÃO - Suponha que os vetores aleatórios $\underset{\sim}{X}$ e $\underset{\sim}{Y}$ satisfazem (1.2.1) para todo conjunto superior mensurável S de R^n. Seja g uma função mensurável crescente de R^n em R. É fácil mostrar que, para todo $t \in R$, o conjunto

$$S_t = \{\underset{\sim}{x} : g(\underset{\sim}{x}) > t\}$$

é um conjunto superior de R^n. Temos então:

$$P[g(\underset{\sim}{X})>t] = P[\underset{\sim}{X}\in S_t] \geq P[\underset{\sim}{Y}\in S_t] = P[g(\underset{\sim}{Y})>t]$$

o que mostra que $\underset{\sim}{X} \overset{st}{\geq} \underset{\sim}{Y}$.

Reciprocamente, se $\underset{\sim}{X} \overset{st}{\geq} \underset{\sim}{Y}$ e S é qualquer conjunto superior mensurável de R^n, para mostrar que vale (1.2.1) basta observar que o indicador de S é uma função mensurável crescente de R^n em R.

OBSERVAÇÃO - Da Proposição 1.2.6 segue-se que as propriedades O_1, O_2 e O_3 se estendem a vetores aleatórios com valores em R^n.

As noções de dependência que foram discutidas até agora estão bastante relacionadas com a idéia de ordenação estocástica. Assim, por exemplo, dado um vetor $\underset{\sim}{X} = (X_1,\ldots, X_n)$ vamos considerar o vetor $\underset{\sim}{Y} = (Y_1,\ldots,Y_n)$ como componentes independentes e cujas distribuições marginais unidimensionais são iguais as de $\underset{\sim}{X}$, isto é, $X_j \overset{st}{=} Y_j$, para $j=1,\ldots,n$. Nessas condições se $\underset{\sim}{X} \overset{st}{\geq} \underset{\sim}{Y}$, $\underset{\sim}{X}$ será DPS. No que diz respeito à associação, vamos considerar a medida $P_{\underset{\sim}{X}}$, induzida em R^n, pelo vetor $\underset{\sim}{X}$. Dizer que $\underset{\sim}{X}$ é associado é equivalente a dizer que

$$P_{\underset{\sim}{X}}(U_1 \cap U_2) \geq P_{\underset{\sim}{X}}(U_1)P_{\underset{\sim}{X}}(U_2)$$

para todo par (U_1,U_2) de conjuntos superiores mensuráveis de R^n.

Block, Savits e Shaked, (1980), introduziram e utilizaram com sucesso o conceito de variáveis aleatórias negativamente dependentes em seqüência. Esse conceito tem um correspondente natural, no caso da dependência positiva, que até hoje parece não ter sido considerado na literatura.

DEFINIÇÃO 1.2.7 - O vetor $\underset{\sim}{X} = (X_1,\ldots,X_n)$ é dito positivamente dependente em seqüência (PDS) se para todo $i=2,\ldots,n$, a distribuição condicional de $(X_1,X_2,\ldots,X_{i-1})$ dado $X_i = x_i$ for estocasticamente crescente em x_i.

A condição acima pode ser expressa na forma:

$$(1.2.2) \qquad (X_1,\ldots,X_{i-1}|X_i=x_i') \overset{st}{\geq} (X_1,\ldots,X_{i-1}|X_i=x_i)$$

sempre que $x_i' \geq x_i$, $i=2,\ldots,n$.

Em algumas situações, em lugar de (1.2.2), é mais fácil verificar que o vetor $\underset{\sim}{X}$ satisfaz:

$$(X_1,\ldots,X_{i-1},X_{i+1},\ldots,X_n|X_i=x_i)$$

é estocasticamente crescente em x_im para $i=1,2,\ldots,n$.

Para mostrar que essa última condição implica em (1.2.2), basta notar que se $\underset{\sim}{Z} \overset{st}{\geq} \underset{\sim}{Y}$, então qualquer vetor cujas componentes constituem um sub-conjunto de $\underset{\sim}{Z}$ será estocasticamente maior que o vetor formado pelas componentes correspondentes de $\underset{\sim}{Y}$.

O lema que provaremos a seguir, nos permitirá mostrar que todo vetor que é positivamente dependente em seqüencia é DPS.

LEMA - Suponha que $\underset{\sim}{X}$ é um vetor aleatório que satisfaz (1.2.2). Então para i=2,...,n e y' > y:

$$P[X_1>y_1,\ldots,X_{i-1}>y_{i-1}|X_i>y'] \geq P[X_1>y_1,\ldots,X_{i-1}>y_{i-1}|X_i>y].$$

DEMONSTRAÇÃO - De (1.2.2) segue-se que:

$$(1.2.3) \qquad P[X_1>y_1,\ldots,X_{i-1}>y_{i-1}|X_i=y'] \geq P[X_1>y_1,\ldots,X_{i-1}>y_{i-1}|X_i=y]$$

sempre que y' for maior do que y.

Por uma questão de facilidade de notação, vamos representar $P[X_1>y_1,\ldots,X_{i-1}>y_{i-1}|X_i]$ como uma função G do valor da variável X_i. Isto posto, (1.2.3) diz simplismente que G é uma função crescente, isto é:

$$G(y') \geq G(y) \text{ se } y' > y.$$

Se $F_i(\cdot)$ for a distribuição marginal de X_i, podemos dizer que (1.2.3) é equivalente a:

$$\begin{vmatrix} G(y')dF_i(y') & G(y)dF_i(y) \\ & \\ dF_i(y') & dF_i(y) \end{vmatrix} \geq 0$$

Como esse determinante é não negativo sempre que y' for maior do que y, nós podemos fixar $Z_1 < Z_2$ e integrar a primeira coluna entre Z_2 e $+\infty$ e a segunda entre Z_1 e Z_2 sem alterar o sinal do determinante. Temos então:

$$\begin{vmatrix} \int_{Z_2}^{\infty} G(y)\,dF_i(y) & \int_{Z_1}^{Z_2} G(y)\,dF_i(y) \\ & \\ \int_{Z_2}^{\infty} dF_i(y) & \int_{Z_1}^{Z_2} dF_i(y) \end{vmatrix} \geq 0.$$

Substituindo a segunda coluna por uma soma dela com a primeira, temos:

$$\begin{vmatrix} \int_{Z_2}^{\infty} G(y)\,dF_i(y) & \int_{Z_1}^{\infty} G(y)\,dF_i(y) \\ & \\ \int_{Z_2}^{\infty} dF_i(y) & \int_{Z_1}^{\infty} dF_i(y) \end{vmatrix} \geq 0.$$

Essa última desigualdade é equivalente a

$$P[X_1>y_1,\ldots,X_{i-1}>y_{i-1}|X_i>Z_2] \geq P[X_1>y_1,\ldots,X_{i-1}>y_{i-1}|X_i>Z_1]$$

sempre que $Z_2 > Z_1$, o que completa a demonstração.

PROPOSIÇÃO 1.2.8 - Seja X um vetor aleatório que satisfaz (1.2.2). Então $\underset{\sim}{X}$ satisfaz (1.1.1), isto é, $\underset{\sim}{X}$ é DPS.

DEMONSTRAÇÃO - Do lema segue-se que para k=2,...,n, temos:

$$P[X_1>y_1,\dots,X_{k-1}>y_{k-1}|X_k>y_k] \geq P[X_1>y_1,\dots,X_{k-1}>y_{k-1}|X_k>-\infty] =$$

$$= P[X_1>y_1,\dots,X_{k-1}>y_{k-1}].$$

Portanto, para $k=2,\dots,n$, temos:

$$P[X_1>y_1,\dots,X_k>y_k] \geq P[X_1>y_1,\dots,X_{k-1}>y_{k-1}]P[X_k>y_k].$$

Multiplicando essas n-1 desigualdades e efetuando os cancelamentos possíveis, obtemos

$$P[X_1>y_1,\dots,X_n>y_n] \geq \prod_{j=1}^{n} P[X_j>y_j]$$

isto é, $\underset{\sim}{X}$ é DPS.

DEFINIÇÃO 1.2.9 - O vetor $\underset{\sim}{X}$ é dito condicionalmente crescente em seqüência (CCS) se para $i=2,\dots,n$ a distribuição condicional de $(X_1=x_i, X_2=x_2,\dots,X_{i-1}=x_{i-1})$ for estocasticamente crescente em $(x_1,\dots,x_{i-1})$.

OBSERVAÇÃO - Embora aparentemente os conceitos de CCS e PDS tenham alguma semelhança, não há nenhuma relação de implicação entre as duas noções. Mostraremos esse fato através dos exemplos desta e da próxima seção.

O Lema seguinte nos permitirá provar que se $\underset{\sim}{X}$ é CCS então $\underset{\sim}{X}$ é associado.

LEMA - Sejam $\underset{\sim}{X}$ um vetor aleatório e Y uma variável aleatória tais que a distribuição condicional de $Y/X = \underset{\sim}{x}$ cresça estocasticamente com $\underset{\sim}{x}$. Nessas condições, existe uma função crescente $h(\underset{\sim}{u},\underset{\sim}{x})$ de R^{n+1} em R e uma variável aleatória U, independente de $\underset{\sim}{X}$ tais que a distribuição de $(Y,\underset{\sim}{X})$ coincide com a distribuição de $(h(U,\underset{\sim}{X}),\underset{\sim}{X})$.

DEMONSTRAÇÃO - Vamos usar o fato bastante conhecido de que se G é uma função distribuição então a v.a. $G^{-1}(U)$ tem distribuição G se U tiver distribuição uniforme em $[0,1]$ e

$$G^{-1}(u) = \inf\{t:G(t)\geq u\}.$$

Seja $F_{\underset{\sim}{x}}(z) = P[Y\leq z|\underset{\sim}{X}=\underset{\sim}{x}]$, isto é, $F_{\underset{\sim}{x}}(z)$ é a função distribuição de $Y/\underset{\sim}{X} = \underset{\sim}{x}$. Defina:

$$h(u,\underset{\sim}{x}) = F_{\underset{\sim}{x}}^{-1}(u) = \inf\{t:F_{\underset{\sim}{x}}(t)\geq u\}.$$

Da própria definição segue-se que h é uma função crescente de u. Como $Y/\underset{\sim}{X} = \underset{\sim}{x}$ cresce estocasticamente com $\underset{\sim}{x}$ temos:

$$F_{\underset{\sim}{x}}(z) \geq F_{\underset{\sim}{x}'}(z)$$

sempre que $\underset{\sim}{x}' > \underset{\sim}{x}$. Segue-se então que h é crescente em $\underset{\sim}{x}$.

Seja agora U uma variável aleatória uniforme em $[0,1]$ e independente de $\underset{\sim}{X}$. Pela observação feita no início da demonstração temos que $h(U,\underset{\sim}{x})$ tem distribuição $F_{\underset{\sim}{x}}$ e portanto:

$$h(U,\underset{\sim}{x}) \overset{st}{=} Y/\underset{\sim}{X} = \underset{\sim}{x}.$$

Como U é independente de $\underset{\sim}{X}$ temos:

$$h(U,\underset{\sim}{x})/\underset{\sim}{X} = \underset{\sim}{x} \overset{st}{=} h(U,\underset{\sim}{x}) \overset{st}{=} h(U,\underset{\sim}{X})/\underset{\sim}{X} = \underset{\sim}{x}.$$

Segue-se que para todo $\underset{\sim}{x}$:

$$Y/\underset{\sim}{X} = \underset{\sim}{x} \overset{st}{=} h(U,\underset{\sim}{X})/\underset{\sim}{X} = \underset{\sim}{x}$$

e portanto que:

$$(Y,\underset{\sim}{x}) \overset{st}{=} (h(U,\underset{\sim}{X}),\underset{\sim}{X}).$$

PROPOSIÇÃO 1.2.10 - Seja $\underset{\sim}{X} = (X_1,\ldots,X_n)$ um vetor condicionalmente crescente em seqüência. Então $\underset{\sim}{X}$ é associado.

DEMONSTRAÇÃO - Vamos considerar inicialmente o caso $n = 2$. Seja $\underset{\sim}{X} = (X_1,X_2)$ e suponha que a distribuição de $X_2/X_1 = x_1$ cresce estocasticamente com x_1. Pelo lema, existe uma função crescente h e uma variável U; independente de X_1 de tal modo que

$$(X_2,X_1) \overset{st}{=} (h(U,X_1),X_1).$$

Como U e X_1 são independentes, o vetor (U,X_1) é associado e conseqüentemente, pela propriedade A_3 da associação, (X_2, X_1) é associado.

Suponha agora que $\underset{\sim}{X} = (X_1,\ldots,X_k)$, $(X_1,\ldots,X_{k-1})$ é associado e que a distribuição condicional de

$$X_k/X_1 = x_1,\ldots,X_{k-1} = x_{k-1}$$

cresce estocasticamente com $(x_1,\ldots,x_{k-1})$. O lema nos permite então dizer que:

$$(X_k,X_1,\ldots,X_{k-1}) \overset{st}{=} (h(U,X_1,\ldots,X_{k-1}),X_1,\ldots,X_{k-1})$$

onde h é crescente e U é independente de $(X_1,\ldots,X_{k-1})$. Segue-se então que $\underset{\sim}{X}$ é associado e a proposição fica demonstrada por indução.

1.3 - POSITIVIDADE TOTAL MULTIVARIADA

Nas situações práticas, quando consideramos um número finito de variáveis aleatórias, a informação de que dispomos é geralmente a função distribuição conjunta ou a função densidade conjunta, quando esta última existe. A partir deste tipo de informação, nem sempre é fácil verificarmos se o vetor $\underset{\sim}{X} = (X_1,\ldots,X_n)$ satisfaz ou não as condições de dependência que discutimos até aqui. Assim, por exemplo, a não ser em casos muito particulares, é praticamente impossível verificarmos diretamente se $\underset{\sim}{X}$ é ou não associado. Seria importante, portanto, se dispusessemos de condições, de fácil verificação, sobre a função densidade (ou sobre a distribuição de probabilidade no caso discreto), que uma vez satisfeitas implicassem na validade das condições de dependência já discutidas.

Neste sentido, a noção de função totalmente positiva de segunda ordem, é fundamental. Vamos relembrar a definição no caso $n = 2$ e em seguida discutir as possíveis generalizações para mais de duas dimensões.

DEFINIÇÃO 1.3.1 - Sejam A e B dois conjuntos mensuráveis da reta. Uma função $f(x,y)$ definida em $A \times B$, é totalmente positiva de ordem $2(TP_2)$ se para todo $x < x'$ em A e $y < y'$ em B, tivermos

(1.3.1) $$f(x,y)f(x',y') \geq f(x,y')f(x',y).$$

OBSERVAÇÃO - Sejam A e B respectivamente, os conjuntos dos valores assumidos por duas variáveis aleatórias X e Y. Diremos que o vetor $\underset{\sim}{X} = (X,Y)$ é TP_2 se a sua função densidade (ou a distribuição no caso discreto) for TP_2 em $A \times B$.

Barlow & Proschan (1975) p. 146, sumarizam num diagrama todas as implicações existentes entre as condições de dependência positiva bivariada. Podemos ver então que, excetuada a independência (que satisfaz trivialmente todas as condições), TP_2 é a condição mais forte de dependência positiva entre duas variáveis X e Y.

A questão agora é como generalizar (1.3.1) para funções de mais de duas variáveis. A maneira mais natural é a que apresentamos a seguir.

DEFINIÇÃO 1.3.2 - Sejam $A_1, A_2, \ldots, A_n$ conjuntos mensuráveis da reta e seja $f(x_1, x_2, \ldots, x_n)$ uma função com valores em R, definida em

$$\prod_{j=1}^{n} A_j.$$

Diremos que f é TP_2 em pares se para cada par (i,j), $i \neq j$, f for TP_2 em x_i e x_j (isto é, em $A_i \times A_j$), quando as demais variáveis são mantidas fixas.

Karlin & Rinott em [1980] preferem trabalhar com as chamadas funções MTP_2 que, a não ser em casos especiais, são equivalentes às funções TP_2 em pares. Para definir o conceito de MTP_2 vamos introduzir a notação: se $\underset{\sim}{x} = (x_1,\ldots,x_n)$ e $\underset{\sim}{y} = (y_1,\ldots,y_n)$ são dois elementos quaisquer de R^n, colocamos:

$$(\underset{\sim}{x} \vee \underset{\sim}{y}) = (\max(x_1,x_2),\ldots,\max(x_n,y_n))$$

$$(\underset{\sim}{x} \wedge \underset{\sim}{y}) = (\min(x_1,y_1),\ldots,\min(x_n,y_n)).$$

DEFINIÇÃO 1.3.3 - Sejam $A_1,\ldots,A_n$ conjuntos mensuráveis da reta e seja f uma função, com valores em R, definida em

$$\prod_{j=1}^{n} A_j .$$

Diremos que f é multivariada totalmente positiva de ordem 2(MTP_2) se quaisquer que sejam $\underset{\sim}{x}$ e $\underset{\sim}{y}$ em

$$\prod_{j=1}^{n} A_j$$

tivermos:

$$f(\underset{\sim}{x})f(\underset{\sim}{y}) \le f(\underset{\sim}{x} \wedge \underset{\sim}{y})f(\underset{\sim}{x} \vee \underset{\sim}{y}). \tag{1.3.2}$$

A verificação de que uma função MTP_2 é TP_2 em pares é trivial. Kemperman (1977) mostrou que, se $f(x) > 0$ para todo

$$\underset{\sim}{x} \in \prod_{j=1}^{n} A_j ,$$

então a recíproca é verdadeira.

Outros conceitos do mesmo tipo encontrados na literatura são m*-dependência (Allan & Wallenius,(1976)) e TPD (Ahmed et al.(1978)). Block & Ting (1981) mostraram que, sob a hipótese $f(\underset{\sim}{x}) > 0$, todos esses quatro conceitos são equivalentes.

O conceito de positividade total desempenha um papel bastante importante em diversas áreas da Matemática e da Estatística. Karlin, nos dois volumes de seu livro, Total Positivity (1968) trata estensivamente da teoria e das aplicações desse conceito. No que se segue nós nos limitamos a apresentar as propriedades que serão utilizadas no decorrer desse trabalho e que são casos particulares (para funções densidade), das propriedades gerais das funções totalmente positivas.

PROPOSIÇÃO 1.3.4 - Seja $\underset{\sim}{X} = (X_1,\dots,X_n)$ um vetor aleatório cuja função densidade é MTP_2. Nessas condições, para $2 \le k < n$, a densidade marginal de k componentes quaisquer de $\underset{\sim}{X}$ será também MTP_2.

DEMONSTRAÇÃO - Se mostrarmos que a redução de uma dimensão não altera o caráter MTP_2 da densidade, a proposição ficará demonstrada por recorrência. Basta portanto, mostrar que a densidade de $(X_1,\dots,X_{n-1})$ é MTP_2. Vamos indicar por f a densidade de $\underset{\sim}{X}$ e sejam $\underset{\sim}{x}$ e $\underset{\sim}{y}$ pontos de R^{n-1}. Sejam A e B dados por:

$$A = \left[\int f(\underset{\sim}{x},z)\,dz\right]\left[\int f(\underset{\sim}{y},z)\,dz\right]$$

e

$$B = \left[\int f(\underset{\sim}{x} \vee \underset{\sim}{y}, z)\,dz\right]\left[\int f(\underset{\sim}{x} \wedge \underset{\sim}{y}, z)\,dz\right].$$

O nosso objetivo é mostrar que $A \leq B$. Vamos reescrever A e B utilizando integrais duplas, da seguinte maneira:

$$A = \int_{z_1 < z_2}\int \left[f(\underset{\sim}{x}, z_1) f(\underset{\sim}{y}, z_2) + f(\underset{\sim}{x}, z_2) f(\underset{\sim}{y}, z_1)\right] dz_1 dz_2$$

e

$$B = \int_{z_1 < z_2}\int \left[f(\underset{\sim}{x} \vee \underset{\sim}{y}, z_1) f(\underset{\sim}{x} \wedge \underset{\sim}{y}, z_2) + f(\underset{\sim}{x} \vee \underset{\sim}{y}, z_2) f(\underset{\sim}{x} \wedge \underset{\sim}{y}, z_1)\right] dz_1 dz_2 .$$

Para simplificar um pouco a notação vamos representar os integrandos da seguinte maneira:

$$a = f(\underset{\sim}{x}, z_1) f(\underset{\sim}{y}, z_2) \qquad b = f(\underset{\sim}{x}, z_2) f(\underset{\sim}{y}, z_1)$$

$$c = f(\underset{\sim}{x} \vee \underset{\sim}{y}, z_1) f(\underset{\sim}{x} \wedge \underset{\sim}{y}, z_2) \qquad d = f(\underset{\sim}{x} \vee \underset{\sim}{y}, z_2) f(\underset{\sim}{x} \wedge \underset{\sim}{y}, z_1)$$

Como a região de integração é a mesma nos dois casos, bastará mostrar que:

$$a + b \leq c + d.$$

Por hipótese f é MTP_2 e $z_1 < z_2$. Logo da definição de MTP_2 vem que:

$$a \leq d \qquad (1.3.3)$$

$$b \leq d.$$

Novamente, o fato de f ser MTP_2 nos permite concluir que:

$$f(\underset{\sim}{x},z_1)f(\underset{\sim}{y},z_1) \le f(\underset{\sim}{x}\wedge\underset{\sim}{y},z_1)f(\underset{\sim}{x}\vee\underset{\sim}{y},z_1)$$

$$f(\underset{\sim}{x},z_2)f(\underset{\sim}{y},z_2) \le f(\underset{\sim}{x}\wedge\underset{\sim}{y},z_2)f(\underset{\sim}{x}\vee\underset{\sim}{y},z_1).$$

Multiplicando essas duas desigualdades obtemos:

(1.3.4) $$ab \le cd.$$

Observe agora que:

$$c + d - a - b = \frac{1}{d}[(d-a)(d-b) + (cd-ab)].$$

Segue-se de (1.3.3) e (1.3.4) que:

$$c + d \ge a + b$$

o que conclui a demonstração.

A proposição e o corolário abaixo são conseqüências imediatas da definição.

PROPOSIÇÃO 1.3.5 - Se f e g são funções MTP_2 definidas em R^n então o produto fg é também MTP_2.

COROLÁRIO - A densidade de um vetor $\underset{\sim}{X}$ com componentes indepentes é MTP_2.

DEFINIÇÃO 1.3.6 - Seja $f(x)$ a função densidade de uma variável aleatória X. Diremos que f é PF_2 (função de Polya de ordem 2) se $g(x,y) = f(x-y)$ for TP_2.

PROPOSIÇÃO 1.3.7 - Seja $\underset{\sim}{X} = (X_1,\ldots,X_n)$ um vetor aleatório MTP_2. Seja $\underset{\sim}{Y} = (Y_1,\ldots,Y_n)$ um vetor com componentes independentes e suponha que a densidade de Y_i é PF_2 para $i=1,2,\ldots,n$. Se além disso, $\underset{\sim}{Y}$ for independente de $\underset{\sim}{X}$ então o vetor $\underset{\sim}{Z} = \underset{\sim}{X}+\underset{\sim}{Y}$ será MTP_2.

DEMONSTRAÇÃO - Seja $f(x_1,\ldots,x_n)$ a densidade de $\underset{\sim}{X}$ e seja $g_i(y_i)$ a densidade de Y_i para $i=1,2,\ldots,n$. Da independência entre $\underset{\sim}{X}$ e $\underset{\sim}{Y}$ segue-se que a densidade conjunta será dada por:

$$f(x_1,\ldots,x_n)\prod_{i=1}^{n} g_i(y_i).$$

Portanto, a densidade de $\underset{\sim}{Z}$ será:

$$h(z_1,\ldots,z_n) = \int\cdots\int \prod_{i=1}^{n} g_i(z_i-x_i)f(x_1,\ldots,x_n)dx_1\ldots dx_n.$$

Das hipóteses feitas e da Proposição 1.3.5 segue-se que o integrando é MTP_2 em $(x_1,\ldots,x_n,z_1,\ldots,z_n)$. O mesmo argumento usado na demonstração da Proposição 1.3.4 mostra que h é MTP_2.

COROLÁRIO - Seja $\underset{\sim}{Y}$ um vetor aleatório satisfazendo as condições da Proposição anterior e seja X_0 uma variável aleatória. Se definirmos: $Z_i = Y_i + X_0$, o vetor $\underset{\sim}{Z}$ será MTP_2.

Vamos agora mostrar que se $\underset{\sim}{X}$ é um vetor aleatório MTP_2 então $\underset{\sim}{X}$ é CCS e conseqüentemente associado.

PROPOSIÇÃO 1.3.8 - Seja $\underset{\sim}{X} = (X_1,\ldots,X_n)$ um vetor aleatório cuja função densidade é MTP_2. Então $\underset{\sim}{X}$ é condicionalmente crescente em seqüência.

De acordo com a Definição 1.2.9 nós temos que mostrar que para $j=2,\ldots,n$, a distribuição condicional de X_j, dados $X_1,\ldots,X_{j-1}$, cresce estocasticamente com $X_1,\ldots,X_{j-1}$.

Para $j=2$, a distribuição marginal de (X_1,X_2) é TP_2 (a Proposição 1.3.4) e portanto, das relações no caso bivariado (ver Barlow & Proschan (1975),p. 146) segue-se que a distribuição condicional de $X_2/X_1 = x_1$ cresce estocasticamente com x_1.

Para $j=3$, novamente a Proposição 1.3.4 nos permite dizer que a distribuição de (X_1,X_2,X_3) é MTP_2. O mesmo argumento utilizado no caso $j=2$, nos permite afirmar que para $X_2 = x_2$, fixado, a distribuição condicional de X_3 cresce estocasticamente com X_1.

Analogamente, para $X_1 = x_1$, fixado, a distribuição condicional de X_3 cresce estocasticamente com X_2. Para $x_1 < x_1'$ e $x_2 < x_2'$ teremos então:

$$P[X_3>t\,|\,X_1=x_1',X_2=x_2'] \geq P[X_3>t\,|\,X_1=x_1',X_2=x_2'] \geq$$

$$\geq P[X_3>t\,|\,X_1=x_1,X_2=x_2]$$

o que mostra que X_3 cresce estocasticamente com (X_1,X_2). Repetições desse argumento nos permitem mostrar que $\underset{\sim}{X}$ é CCS.

COROLÁRIO - Se a função densidade do vetor aleatório $\underset{\sim}{X}$ for MTP_2 então $\underset{\sim}{X}$ é associado.

DEMONSTRAÇÃO - Esse resultado é conseqüência da proposição anterior e da Proposição 1.2.10.

1.4 - DEPENDÊNCIA POSITIVA DA DISTRIBUIÇÃO NORMAL MULTIVARIADA

Quando consideramos duas variáveis aleatórias X e Y, com distribuição conjunta normal, o fato da covariância ser não negativa é suficiente para concluirmos que a densidade conjunta, $f(x,y)$ é TP_2 e que portanto, o vetor (X,Y) satisfaz a todos os critérios de dependência positiva conhecidos. A situação não é tão simples assim quando consideramos mais de duas variáveis aleatórias. Existem inúmeros exemplos de vetores aleatórios normais com covariâncias não negativas cujas densidades não são MTP_2.

Nesta seção, nós apresentamos uma condição necessária e suficiente para que a densidade de um vetor aleatório normal seja MTP_2 e descrevemos algumas técnicas que nos permitem gerar distribuições normais que satisfazem essa condição. Além disso, analisamos quais as propriedades de dependência positiva que são preservadas mesmo quando a condição não está satisfeita.

A proposição abaixo contém a condição mencionada no início do parágrafo anterior. Essa condição foi utilizada pela primeira vez por Sarkar (1969) e aparece também em Barlow & Proschan (1975).

PROPOSIÇÃO 1.4.1 - Seja $\underset{\sim}{X} = (X_1, \ldots, X_n)$ um vetor aleatório com distribuição $N(\underset{\sim}{\mu}, \Sigma)$. Seja $B = \| b_{ij} \|$, $i,j = 1, \ldots, n$, a matriz inversa da matriz de variância-covariância Σ. Nestas condições, a função densidade, $f(x_1, \ldots, x_n)$, de $\underset{\sim}{X}$ será MTP_2 se e somente se $b_{ij} \leq 0$ para todo $i \neq j$.

DEMONSTRAÇÃO - A densidade conjunta do vetor $\underset{\sim}{X}$ é dada por:

$$f(\underset{\sim}{x}) = (2\pi)^{-n/2} |\Sigma|^{1/2} \exp[-\tfrac{1}{2}(\underset{\sim}{x}-\underset{\sim}{\mu})' B (\underset{\sim}{x}-\underset{\sim}{\mu})].$$

Como $f(\underset{\sim}{x})$ é estritamente positiva, ela será MTP_2 se, para todo $i \neq j$, $f(\underset{\sim}{x})$ for TP_2 em x_i e x_j quando as demais variáveis são mantidas fixas. Portanto, $f(\underset{\sim}{x})$ será MTP_2 se e somente se

$$e^{-\frac{1}{2} b_{ij} x_i x_j}$$

for TP_2 para todo $i \neq j$. É fácil ver no entanto que essa última função será TP_2 se e só se $b_{ij} \leq 0$.

EXEMPLO 1.4.2 - Vamos considerar um vetor $(X_1, \ldots, X_n)$ para o qual temos:

$$E(X_i) = 0, \; E(X_i^2) = 1 \text{ para } i=1, \ldots, n$$

e

$$E(X_i, X_j) = \rho \geq 0 \text{ para todo } i \neq j.$$

É claro que devemos ter necessariamente $\rho < 1$. Nesse caso a matriz de variância-covariância será:

$$\Sigma = \begin{Vmatrix} 1 & \rho & \rho & \dots & \rho \\ \rho & 1 & \rho & \dots & \rho \\ \vdots & \vdots & \vdots & & \vdots \\ \rho & \rho & \rho & \dots & 1 \end{Vmatrix}.$$

Dado o tipo especial da matriz Σ, fica fácil mostrar que a matriz inversa B é do mesmo tipo e que portanto pode ser obtida diretamente. Segue-se então que os elementos da matriz $B = \|b_{ij}\|$ serão dados por:

$$b_{ii} = \frac{1 + (n-2)\rho}{[1+(n-1)\rho](1-\rho)}$$

$$b_{ij} = \frac{-\rho}{[1+(n-1)\rho](1-\rho)}$$

Como $0 \le \rho < 1$, segue-se que $b_{ij} \le 0$ e portanto que a densidade de $\underset{\sim}{X}$ é MTP_2.

EXEMPLO 1.4.3 - Considere um vetor (X_1,X_2,X_3) cuja matriz de variância-covariância é:

$$\Sigma = \begin{Vmatrix} 1 & 1/5 & 1/2 \\ 1/5 & 1 & 1/2 \\ 1/2 & 1/2 & 1 \end{Vmatrix}.$$

É fácil ver que b_{12} é estritamente positivo e portanto que a densidade de $\underset{\sim}{X}$ não é MTP_2.

Vamos considerar agora algumas operações que podem ser realizadas com um vetor $\underset{\sim}{X}$, cuja densidade é normal MTP_2 e que preservam essa característica da densidade. Algumas dessas técnicas decorrem diretamente das propriedades gerais discutidas na seção 1.3 e portanto, para elas, a a preservação da característica MTP_2 não depende das hipóteses de normalidade feitas. Assim, por exemplo, se $\underset{\sim}{X} = (X_1,\ldots,X_n)$ é um vetor aleatório aleatório com densidade normal MTP_2 e Z é uma variável normalmente distribuída e independente de $\underset{\sim}{X}$ então a densidade do vetor aleatório $(X_1+Z,\ldots,$ (X_n+Z) é normal MTP_2.

Outro exemplo do mesmo tipo pode ser obtido se tomarmos $\underset{\sim}{X}$ nas mesmas condições acima e considerarmos um outro vetor $\underset{\sim}{Y} = (Y_1,\ldots,Y_n)$, independente de $\underset{\sim}{X}$, cujas componentes são normalmente distribuídas e independentes entre si. Segue-se então, das propriedades da seção 1.3, que a densidade de $(X_1+Y_1,\ldots,X_n+Y_n)$ é normal MTP_2.

OBSERVAÇÃO - Este exemplo nos dá uma informação que pode ser útil na identificação de densidades normais MTP_2.

Quando interpretado em termos da matriz de variância-covariância, ele nos diz que se partirmos da matriz de uma densidade normal MTP_2 e aumentarmos arbitrariamente as variâncias, sem mexer nas covariâncias, a matriz resultante estará também associada a uma densidade normal MTP_2.

Vamos analisar agora o que ocorre quando consideramos somas parciais das componentes de um vetor normal MTP_2. Em caso particular importante está contido na proposição que apresentamos a seguir.

PROPOSIÇÃO 1.4.4 - Seja $\underset{\sim}{X} = (X_1,\ldots,X_n)$ um vetor aleatório cuja distribuição é $N(\underset{\sim}{0},\Sigma)$ onde $\Sigma = \|r_{ij}\|$ é tal que $r_{ii} = 1$, para $i = 1,\ldots,n$ e $r_{ij} = \rho$ para todo $i \neq j$, $0 \leq \rho < 1$. Considere o vetor $\underset{\sim}{Y} = (Y_1,\ldots,Y_n)$ cujas componentes são dadas pelas relações: $Y_1 = X_1$ e $Y_j = \sum_{i=1}^{j} X_i$ para $2 \leq j \leq n$. Nessas condições, a densidade de $\underset{\sim}{Y}$ é normal MTP_2.

DEMONSTRAÇÃO - Seja A a matriz da transformação que define $\underset{\sim}{Y}$ como função de $\underset{\sim}{X}$. A matriz de variância-covariância de $\underset{\sim}{Y}$ será então dada por:

$$B = A\Sigma A^T.$$

Da definição de $\underset{\sim}{Y}$ segue-se que $X_1 = Y_1$ e que $X_j = Y_j - Y_{j-1}$, para todo $j \geq 2$ e portanto que a matriz inversa de A é:

$$A^{-1} = \begin{Vmatrix} 1 & 0 & 0 & \ldots & 0 & 0 & 0 \\ -1 & 1 & 0 & \ldots & 0 & 0 & 0 \\ 0 & -1 & 1 & \ldots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \ldots & -1 & 1 & 0 \\ 0 & 0 & 0 & \ldots & 0 & -1 & 1 \end{Vmatrix}.$$

A matriz Σ é a mesma que aparece no Exemplo 1.4.2 e portanto Σ^{-1} é do mesmo tipo de Σ podendo ser representada na forma:

$$\Sigma^{-1} = \begin{Vmatrix} a & b & b & \cdots & b & b \\ b & a & b & \cdots & b & b \\ b & b & a & \cdots & b & b \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ b & b & b & \cdots & a & b \\ b & b & b & \cdots & b & a \end{Vmatrix}$$

onde:

$$a = \frac{1 + (n-2)\rho}{[1+(n-1)\rho](1-\rho)}$$

(1.4.1)

$$b = \frac{-\rho}{[1+(n-1)\rho](1-\rho)}$$

O produto $[A^{-1}]^T\Sigma$ será então igual a

$$\begin{Vmatrix} a-b & b-a & 0 & \cdots & 0 & 0 \\ 0 & a-b & b-a & \cdots & 0 & 0 \\ 0 & 0 & a-b & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & a-b & b-a \\ b & b & b & \cdots & b & a \end{Vmatrix}$$

Portanto, a matriz inversa da matriz de variância-covariância de $\underset{\sim}{Y}$ será:

$$B^{-1} = (A^{-1})^T \Sigma^{-1} A^{-1} = \left\| \begin{matrix} 2(a-b) & b-a & 0 & \dots & 0 & 0 \\ b-a & 2(a-b) & b-a & \dots & 0 & 0 \\ 0 & b-a & 2(a-b) & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 2(a-b) & b-a \\ 0 & 0 & 0 & \dots & b-a & a \end{matrix} \right\|$$

Vê-se portanto que os elementos fora da diagonal principal de B^{-1} ou são nulos ou iguais a b-a. Porém de (1.4.1) vem que:

$$b-a = \frac{-\rho-[1+(n-2)\rho]}{[1+(n-1)\rho](1-\rho)} = \frac{-1}{1-\rho},$$

donde se conclui que a densidade de $\underset{\sim}{Y}$ é MTP_2.

OBSERVAÇÃO - Em caso particular importante é aquele no qual as componentes de $\underset{\sim}{X}$ são independentes e tem todas distribuição $N(0,1)$. Nesse caso a demonstração da Proposição 1.4.4 fica bem mais simples pois, $\Sigma = \Sigma^{-1} = I$.

A Proposição 1.4.4 nada mais faz do que confirmar a idéia intuitiva, de que a formação de somas parciais só poderia contribuir para aumentar a dependência que, por hipótese, já existia entre as componentes do vetor $\underset{\sim}{X}$. Vale a pena observar no entanto, que esse tipo de intuição nem sempre conduz a conclusões verdadeiras.

Considere, por exemplo, dois vetores $\underset{\sim}{X}$ e $\underset{\sim}{Y}$ independentes entre si e tendo ambos densidade normal MTP_2. Pode-se então afirmar que a soma $\underset{\sim}{X}+\underset{\sim}{Y}$ terá densidade normal MTP_2? Vimos anteriormente, que a resposta a essa questão é afirmativa quando as componentes de um dos vetores são independentes entre si. O mesmo tipo de intuição nos levaria então a pensar que a resposta deveria continar sendo afirmativa quando as componentes de $\underset{\sim}{Y}$ e de $\underset{\sim}{X}$ são positivamente dependentes. O exemplo que apresentamos a seguir mostra que essa conclusão é falsa.

EXEMPLO 1.4.5 - Sejam $\underset{\sim}{X}$ e $\underset{\sim}{Y}$ independentes com distribuições $N(\underset{\sim}{0},\Sigma_1)$ e $N(\underset{\sim}{0},\Sigma_2)$ respectivamente, onde:

$$\Sigma_1 = \begin{Vmatrix} 1 & 1/4 & 1/4 \\ 1/4 & 1 & 1/4 \\ 1/4 & 1/4 & 1 \end{Vmatrix} \quad \text{e} \quad \Sigma_2 = \begin{Vmatrix} 1 & 0 & 0 \\ 0 & 1 & 2 \\ 0 & 2 & 5 \end{Vmatrix}$$

Σ_1 é um caso particular da matriz do Exemplo 1.4.2 e portanto a densidade de $\underset{\sim}{X}$ é MTP_2. É fácil verificar que o mesmo acontece com a densidade de $\underset{\sim}{Y}$. Como $\underset{\sim}{X}$ e $\underset{\sim}{Y}$ são independentes, a matriz associada a $\underset{\sim}{X}+\underset{\sim}{Y}$ será:

$$\Sigma_1 + \Sigma_2 = \begin{Vmatrix} 2 & 1/4 & 1/4 \\ 1/4 & 2 & 9/4 \\ 1/4 & 9/4 & 6 \end{Vmatrix} .$$

Um cálculo direto mostra que o elemento b_{31} da matriz inversa de $\Sigma_1 + \Sigma_2$ é estritamente positivo seguindo-se portanto que a densidade de $\underset{\sim}{X} + \underset{\sim}{Y}$ não é MTP_2.

Vamos agora analisar o problema dos vetores com covariâncias não-negativas cujas densidades não são MTP_2.

PROPOSIÇÃO 1.4.6 - Seja $\underset{\sim}{X}$ um vetor aleatório normal com covariâncias não negativas. Então $\underset{\sim}{X}$ satisfaz (1.2.2) isto é $\underset{\sim}{X}$ é PDS e conseqüentemente DPS.

DEMONSTRAÇÃO - Vamos mostrar que para todo $i = 1,2,\ldots,n$, a distribuição condicional de $(X_1,\ldots,X_{i-1},X_{i+1},\ldots,X_n)$ dado $X_i=x_i$ cresce estocasticamente com x_i. Como foi observado anteriormente, essa condição implica em 1.2.2.

Seja $\underset{\sim}{Y}$ um vetor com n-1 componentes cuja distribuição coincide com a distribuição condicional acima. Das propriedades da normal multivariada segue-se que:

$$E(\underset{\sim}{Y}) = \begin{pmatrix} \rho_{i1}x_i \\ \rho_{i2}x_i \\ \vdots \\ \rho_{in}x_i \end{pmatrix}$$

$$V(\underset{\sim}{Y}) = (\rho_{i1},\ldots,\rho_{in})\Sigma_1^{-1} \begin{pmatrix} \rho_{i1} \\ \vdots \\ \rho_{in} \end{pmatrix}$$

onde, Σ_1 é a matriz de variância-covariância de $(X_1,\ldots,X_{i-1}, X_{i+1},\ldots,X_n)$.

Seja agora $x_i' > x_i$ e $\underset{\sim}{Y}$ um vetor cuja distribuição coincide com a distribuição condicional de $(X_1,\ldots,X_n)/X_i = x_i'$. Segue-se que:

$$\underset{\sim}{Y}' \overset{st}{=} \underset{\sim}{Y} + \begin{pmatrix} \rho_{i1}(x_i'-x_i) \\ \vdots \\ \rho_{in}(x_i'-x_i) \end{pmatrix}.$$

Como as covariâncias são todas não-negativas e $x_i' > x_i$ segue-se que:

$$\underset{\sim}{Y}' \overset{st}{=} \underset{\sim}{Y} + \underset{\sim}{A}$$

onde $\underset{\sim}{A} \geq 0$. Portanto, $\underset{\sim}{Y}' \overset{st}{\geq} \underset{\sim}{Y}$ o que equivale a dizer que $\underset{\sim}{X}$ é PDS.

Segue-se então da Proposição 1.2.8 que $\underset{\sim}{X}$ é DPS.

A questão seguinte é o que se pode dizer sobre associação, quando a densidade não é MTP_2. Há fortes razões para suspeitarmos que todo vetor aleatório normal, com covariâncias não-negativas, é associado. No entanto, não encontramos na literatura uma demonstração dessa hipótese e nem um contra-exemplo que mostrasse ser ela não verdadeira. No trabalho de Karlin (1980) ele usa vias indiretas para concluir que um dado vetor com três componentes é associado. A proposição seguinte mostra que para três componentes esse resultado é sempre verdadeiro.

PROPOSIÇÃO 1.4.7 - Seja $\underset{\sim}{X} = (X_1,X_2,X_3)$ um vetor aleatório com distribuição normal trivariada, com covariâncias não negativas. Então, $\underset{\sim}{X}$ é associado.

DEMONSTRAÇÃO - Não há perda de generalidade em supormos que $E(X_i) = 0$, $V(X_i) = 1$, para $i = 1,2,3$. Portanto, a matriz de variância-covariância de $\underset{\sim}{X}$ pode ser representada na forma:

$$\Sigma = \begin{Vmatrix} 1 & \rho_{12} & \rho_{13} \\ \rho_{12} & 1 & \rho_{23} \\ \rho_{13} & \rho_{23} & 1 \end{Vmatrix}$$

onde os ρ_{ij} são tais que $0 \leq \rho_{ij} < 1$ e Σ é positiva definida.

Vamos considerar as diferenças:

$$a_1 = \rho_{12}\rho_{13} - \rho_{23}$$

$$a_2 = \rho_{12}\rho_{23} - \rho_{13}$$

$$a_3 = \rho_{13}\rho_{23} - \rho_{12}$$

A menos do produto por uma constante positiva, os a_j são os elementos que aparecem fora da diagonal principal da matriz Σ^{-1}. Se para $j = 1,2,3$ tivermos $a_j \leq 0$ segue-se da Proposição 1.4.1 que a densidade de $\underset{\sim}{X}$ é MTP_2 e portanto que $\underset{\sim}{X}$ é associado. O problema da existência ou não de associação só se coloca quando ao menos um dos a_j's for estritamente positivo. Por outro lado, é fácil ver que no máximo um deles pode ser estritamente positivo. De fato, suponha por exemplo que $a_1 > 0$ e

$a_2 > 0$. Temos então:

$$\text{(1.4.2)} \qquad \begin{aligned} \rho_{12}\rho_{13} &> \rho_{23} \\ \rho_{12}\rho_{23} &> \rho_{13} \end{aligned}$$

Multiplicando essas duas desigualdades teremos:

$$\rho_{12}^2\rho_{13}\rho_{23} > \rho_{13}\rho_{23}.$$

De (1.4.2) segue-se que $\rho_{13}\rho_{23} \neq 0$ e portanto teremos $\rho_{12}^2 > 1$ o que é um absurdo. Analogamente pode-se mostrar que não podemos ter simultaneamente $a_1 > 0$, $a_3 > 0$ e nem $a_2 > 0$, $a_3 > 0$.

Vamos supor que $a_1 \leq 0$, isto é que $\rho_{23} \geq \rho_{12}\rho_{13}$. Defina as variáveis aleatórias:

$$\text{(1.4.3)} \qquad \begin{aligned} Y_1 &= X_1 \\ Y_2 &= X_2 - \rho_{12}X_1 \\ Y_3 &= X_3 - \rho_{13}X_1 \end{aligned}$$

Um cálculo simples mostra que $\mathrm{cov}(Y_2,Y_3) \geq 0$ e que Y_1 é independente de (Y_2,Y_3). Das propriedades A.1 e A.4 da associação segue-se que o vetor (Y_1,Y_2,Y_3) é associado. De (1.4.3) segue-se que:

$$\begin{aligned} X_1 &= Y_1 \\ X_2 &= \rho_{12}Y_1 + Y_2 \\ X_3 &= \rho_{13}Y_1 + Y_3 \end{aligned}$$

o que mostra que as componentes de $\underset{\sim}{X}$ são funções crescentes das componentes de um vetor associado e portanto segue-se da propriedade A.3 que $\underset{\sim}{X}$ é associado.

OBSERVAÇÃO - Note que para provar a proposição acima nós só precisamos que um dos elementos a_j seja menor ou igual a zero. O fato de que pelo menos dois deles satisfazem essa condição nos permite mostrar um resultado um pouco mais forte. Na realidade é possível mostrar que existe uma ordenação das componentes de $\underset{\sim}{X}$, tal que o vetor resultante e condicionalmente crescente em seqüência. A associação sairia então como conseqüência da Proposição 1.2.10.

A Proposição 1.4.7 sugere que se considere a possibilidade de demonstrarmos o resultado geral por um processo de recorrência. Os comentários que faremos a seguir sobre o caso $n = 4$ mostram que esse processo não é muito simples de ser obtido.

Seja $\underset{\sim}{X} = (X_1, X_2, X_3, X_4)$ um vetor aleatório normal com variâncias iguais a 1 e covariâncias não negativas. Suponhamos ainda que exista uma linha da matriz de variância covariância, digamos por exemplo a primeira, que satisfaça as condições:

$$(1.4.4) \qquad \begin{aligned} \rho_{12}\rho_{13} &\leq \rho_{23} \\ \rho_{12}\rho_{14} &\leq \rho_{24} \\ \rho_{13}\rho_{14} &\leq \rho_{34} \end{aligned}$$

Nesse caso, para $j = 2,3,4$ as variáveis aleatórias $Y_j = X_j - \rho_{1j} X_1$ teriam covariâncias não negativas e seriam portanto associadas pela proposição anterior. Como todas elas são independentes de X_1 o argumento utilizado na demonstração da proposição mostra que $\underset{\sim}{X}$ é associado.

O exemplo seguinte mostra que existem vetores que não satisfazem condições do tipo dado em (1.4.4).

EXEMPLO 1.4.8 - Considere a matriz Σ dada por:

$$\Sigma = \begin{Vmatrix} 1 & \rho & \rho & 0 \\ \rho & 1 & 0 & \rho \\ \rho & 0 & 1 & \rho \\ 0 & \rho & \rho & 1 \end{Vmatrix}$$

Se $0 \leq \rho < 1/2$, Σ será positiva definida e portanto pode ser pensada como a matriz de variância covariância de um vetor aleatório normal. No entanto se ρ for diferente de zero nenhuma linha da matriz Σ satisfará condições do tipo de (1.4.4).

Vale a pena observar que Σ não é um contra-exemplo, uma vez que por meios indiretos é possível mostrarmos que um vetor que tem Σ como matriz de variância covariância é necessariamente associado.

CAPÍTULO 2

2.1 - DEPENDÊNCIA NEGATIVA MULTIVARIADA

Mesmo durante o período em que se desenvolveu uma razoável atividade de pesquisa em dependência positiva, pouca ou quase nenhuma atenção foi dada à dependência negativa. A explicação para essa aparente falta de interesse estava na impressão, mais ou menos generalizada, de que existia uma analogia perfeita entre os dois tipos de dependência. Acreditava-se que bastaria inverter o sentido de desigualdade em alguns casos e substituir funções concordantes por discordantes em outros, para que as formas de dependência negativa fossem obtidas a partir das positivas correspondentes.

No caso bivariado essa analogia funciona bastante bem e de um ponto de vista intuitivo, a dependência negativa entre duas variáveis aleatórias significa que o crescimento de uma delas provavelmente estará associado a uma diminuição da outra.

É fácil ver que essa interpretação intuitiva não funciona em mais de duas dimensões. De fato, dadas três variáveis aleatórias, X_1, X_2 e X_3 e considerados os pares (X_1,X_2),(X_2,X_3)

e (X_1,X_3), se supusermos que as componentes de dois desses pares são negativamente dependentes, então entre as componentes do terceiro par o que deve existir é alguma forma de dependência positiva.

É necessário portanto, repensar essa interpretação intuitiva de modo que ela funcione para vetores aleatórios com mais de duas componentes. A forma mais adequada nos parece ser a seguinte: as componentes de um vetor $\underset{\sim}{X} = (X_1,\ldots,X_n)$ serão negativamente dependentes quando para toda partição de $\underset{\sim}{X}$ em dois blocos disjuntos, um com k e o outro com n-k componentes ($1\leq k<n$), o crescimento de um dos blocos estiver provavelmente associado a uma diminuição do outro. É claro que quando n=2 essa interpretação coincide com a que foi dada anteriormente. Finalmente, para tornar mais precisa essa idéia intuitiva de variação em sentidos opostos, vale a pena considerá-la em termos de probabilidades condicionais. Em outras palavras, o que es está afirmando é que o crescimento de um dos blocos aumenta a probabilidade condicional de que o outro diminua.

Estabelecidas as bases intuitivas, vamos passar a descrever as principais formas de dependência negativa, começando pelas formas análogas a (1.1.1) e (1.1.2).

DEFINIÇÃO 2.1.1 - Diremos que a dependência negativa entre as variáveis aleatórias $X_1,X_2,\ldots,X_n$ é *superior* se para toda n-upla, $(x_1,\ldots,x_n)$, de números reais, tivermos:

$$(2.1.1) \qquad P[X_1>x_1,\ldots,X_n>x_n] \geq \prod_{i=1}^{n} P(X_i>x_i).$$

DEFINIÇÃO 2.1.2 - Diremos que a dependência negativa entre as variáveis aleatórias $X_1,\ldots,X_n$ é *inferior* se para toda n-upla, $(x_1,\ldots,x_n)$, de números reais, tivermos:

$$(2.1.2) \qquad P[X_1\le x_1,\ldots,X_n\le x_n] \le \prod_{i=1}^{n} P[X_i\le x_i].$$

Assim como no caso positivo, variáveis que satisfazem as condições (2.1.1) ou (2.1.2) serão designadas por DNS ou DNI respectivamente. É fácil ver também que, para n = 2, (2.1.1) e (2.1.2) são equivalentes.

O exemplo seguinte, devido a Ebrahimi & Ghosh (1980) mostra que essa equivalência não vale para $n > 3$.

EXEMPLO 2.1.3 - Seja $\underset{\sim}{X} = (X_1,X_2,X_3)$ um vetor aleatório que assume valores no conjunto $\{(1,1,0),(1,0,1),(0,1,1),(0,0,0)\}$ com probabilidades iguais a 1/4 para cada ponto. Teremos então:

$$P[X_1>0,X_2>0,X_3>0] = 0$$

e

$$P[X_1>0]\cdot P[X_2>0]\cdot P[X_3>0] = 1/8$$

o que mostra que $\underset{\sim}{X}$ satisfaz (2.1.1).

Por outro lado, temos:

$$P[X_1\le 0,X_2\le 0,X_3\le 0] = 1/4$$

e

$$P[X_1 \leq 0] \cdot P[X_2 \leq 0] \cdot P[X_3 \leq 0] = 1/8$$

e portanto $\underset{\sim}{X}$ não satisfaz (2.1.2).

EXEMPLO 2.1.4 - É fácil mostrar que se o par (X,Y) satisfaz a condição (1.1.1) (e conseqüentemente (1.1.2)) então o par (X,-Y) satisfará as condições (2.1.1) e (2.1.2). Em particular, se X e Y tem distribuição conjunta normal com covariância negativa então o vetor (X,Y) será DNS e DNI. De fato, o par (X,-Y) tem covariância positiva e portanto satisfaz (1.1.1).

Se pretendessemos manter nesse capítulo a mesma seqüência adotada no Capítulo 1, seria agora o momento de falarmos de associação. Infelizmente, apesar das tentativas feitas por diversos autores, ninguém conseguiu até agora produzir uma definição satisfatória de associação negativa. Entende-se por satisfatória, uma noção que além de satisfazer à outras condições, ocupe o lugar correspondente ao que a associação ocupava, nas interrelações entre as formas de dependência positiva. Block (comunicação pessoal) acredita ter razões suficientes para duvidar da existência de uma definição adequada. Voltaremos ao assunto no final deste trabalho.

Vamos considerar a seguir os conceitos de dependência negativa relacionados com a ordenação estocástica.

DEFINIÇÃO 2.1.5 - O vetor $\underset{\sim}{X} = (X_1,...,X_n)$ é dito negativamente dependente em seqüência (NDS) se para todo i=2,...,n a distribuição condicional de $(X_1,\ldots,X_{i-1})$ dado $X_i = x_i$, for estocasticamente decrescente em x_i.

(2.1.3) $\quad (X_1,\ldots,X_{i-1}|X_i=x_i') \overset{st}{\leq} (X_1,\ldots,X_{i-1}|X_i=x_i)$

sempre que $x_i' \geq x_i$, $i = 2,\ldots,n$.

Note que (2.1.3) é idêntica a (1.2.2) com o sentido da desigualdade trocado.

PROPOSIÇÃO 2.1.6 - Seja $\underset{\sim}{X}$ um vetor aleatório NDS. Então $\underset{\sim}{X}$ satisfaz (2.1.1) e (2.1.2) isto é, $\underset{\sim}{X}$ é DNS e DNI.

DEMONSTRAÇÃO - Idêntica à demonstração da Proposição 1.2.8 e do lema que a precede.

A definição 2.1.5 e a Proposição 2.1.6 mostram uma perfeita analogia entre PDS e NDS. Veremos agora que isso não acontece com a noção de condicionalmente crescente em seqüência (Definição 1.2.9).

DEFINIÇÃO 2.1.7 - O vetor $\underset{\sim}{X} = (X_1,\ldots,X_n)$ é dito condicionalmente decrescente em seqüência (CDS) se para $i = 2,\ldots,n$ a distribuição condicional de X_i dado $X_1 = x_1,\ldots,X_{i-1} = x_{i-1}$ for estocasticamente decrescente em $(x_1,\ldots,x_{i-1})$.

Vimos no caso da dependência positiva que se $\underset{\sim}{X}$ era condicionalmente crescente em seqüência (CCS) então $\underset{\sim}{X}$ satisfazia as condições (1.1.1) e (1.1.2) isto é, $\underset{\sim}{X}$ era DPS e DPI. A demonstração lá era feita indiretamente através de uma passagem intermediária pela associação (ver Proposições 1.2.10 e 1.1.8).

Como no presente contexto nós não dispomos de conceito de associação, teríamos que verificar diretamente se um vetor que é CDS satisfaz ou não as condições (2.1.1) e (2.1.2). O exemplo seguinte, devido a Block, Savits & Shaked (1980) mostra que $\underset{\sim}{X}$ pode ser CDS sem ser DNS ou DNI.

EXEMPLO 2.1.8 - Suponha que a distribuição conjunta de X_1 e X_2 é dada na tabela abaixo:

$X_2 \backslash X_1$	1	2
1	0,1	0,7
2	0,1	0,1

Considere agora uma variável X_3 que fica completamente determinada quando são dados X_1 e X_2, isto é:

$$X_3 = 11 \text{ se } (X_1,X_2) = (1,1)$$

$$X_3 = 10 \text{ se } (X_1,X_2) = (2,1)$$

$$X_3 = 1 \text{ se } (X_1,X_2) = (1,2)$$

$$X_3 = 1 \text{ se } (X_1,X_2) = (2,2).$$

Da própria construção da variável X_3 segue-se que (X_1,X_2,X_3) é condicionalmente decrescente em seqüência. No entanto:

$$P[X_1>1, X_3>1] = 0,7$$

e

$$P[X_1>1]\cdot P[X_3>1] = 0,8 \times 0,8 = 0,64.$$

Portanto o para $(X_1, X_3$ não é DNS e conseqüentemente não é DNI já que para duas variáveis esses conceitos são equivalentes. Segue-se então que (X_1, X_2, X_3) não é DNS nem DNI.

O exemplo acima mostra por que as relações entre as formas de dependência negativa são um pouco mais complicadas do que as relações correspondentes no caso positivo. Na realidade, a vista desse exemplo, seria até razoável perguntarmos se CDS merece ser considerada como condição de dependência negativa. A resposta a essa questão é afirmativa uma vez que se $\underset{\sim}{X}$ é CDS não é difícil mostrar que $\text{cov}(X_i, X_j) \leq 0$ para todo $i \neq j$. Logo CDS verifica a condição mínima que se exige de um vetor negativamente dependente. Seria lamentável se isso não ocorresse, uma vez que CDS (assim como CCS no caso positivo) é uma condição importante em si, dado o seu relacionamento natural com a análise de regressão.

No entanto, ao contrário do que ocorria no caso positivo, se estivermos interessados na estimação de probabilidades ligadas à distribuição conjunta, NDS é que é a condição que deve ser verificada na prática, ou admitida como hipótese num problema teórico. Na maioria dos casos, ela é uma condição cuja validade é facilmente verificada e sempre implica em (2.1.1) e (2.1.2).

Em resumo, o papel atribuído a CCS no caso positivo tem agora que ser repartido entre a noção correspondente (CDS) e a condição negativamente dependente em seqüência.

Veremos nas próximas seções que várias distribuições importantes satisfazem uma condição que é suficiente para garantirmos que elas são simultaneamente CDS e NDS.

2.2 - DEPENDÊNCIA NEGATIVA E A MEDIDA INDUZIDA

Pelas mesmas razões expostas no início da seção 1.3, seria importante dispormos de condições sobre a função densidade (ou sobre a distribuição de probabilidades no caso discreto) que uma vez satisfeitas garantissem a validade de uma ou mais das condições de dependência negativa discutidas na seção anterior. Vamos começar de maneira natural invertendo o sentido das desigualdades nas Definições 1.3.1 e 1.3.3.

DEFINIÇÃO 2.2.1 - Sejam A e B dois conjuntos mensuráveis da reta. Uma função $f(x,y)$ é reversa regular de ordem 2 (RR_2) se para todo $x < x'$ em A e $y < y'$ em B, tivermos:

(2.2.1) $$f(x,y)f(x',y') \le f(x,y')f(x',y).$$

Diremos que um vetor aleatório (X,Y) é RR_2 quando sua função densidade for RR_2.

DEFINIÇÃO 2.2.2 - Sejam $A_1,\ldots,A_n$ conjuntos mensuráveis da reta e seja f uma função, definida em

$$\prod_{j=1}^{n} A_j,$$

com valores em R. Diremos que f é multivariada reversa regular de ordem 2 (MRR_2) se quaisquer que sejam $\underset{\sim}{x}$ e $\underset{\sim}{y}$ em

$$\prod_{j=1}^{n} A_j,$$

tivermos:

(2.2.2) $$f(\underset{\sim}{x})f(\underset{\sim}{y}) \geq f(\underset{\sim}{x}\wedge\underset{\sim}{y})f(\underset{\sim}{x}\vee\underset{\sim}{y}).$$

A definição de função RR_2 em pares é análoga à definição 1.3.2. Pode-se mostrar também, sem nenhuma dificuldade, que sob condições bastante gerais RR_2 em pares é equivalente a MRR_2.

A analogia com a dependência positiva termina aqui. O exemplo seguinte mostra que uma função pode ser RR_2 em pares sem que as marginais sejam. Em outras palavras não temos aqui um resultado análogo ao da Proposição 1.3.4.

EXEMPLO 2.2.3 - Seja $f(x_1,x_2,x_3)$ igual a $e^{-x_3(x_1+x_2)}$, $x_1 > 0$, $x_2 > 0$, $x_3 > 0$. É fácil verificar que f é RR_2 em pares. No entanto, a função:

$$g(x_1,x_2) = \int_0^{\infty} e^{-x_3(x_1+x_2)}\,dx_3 = \frac{1}{x_1+x_2}$$

não é RR_2 mas, sim TP_2.

Outros exemplos envolvendo funções densidade de estatísticas de ordem podem ser encontrados em Ebrahimi & Ghosh (1980).

Sempre que as componentes de um vetor $\underset{\sim}{X}$ satisfazem a uma determinada condição de dependência, todo vetor formado por um subconjunto das componentes de $\underset{\sim}{X}$ deverá satisfazer a mesma condição. Segue-se portanto que, ao contrário de MTP_2, MRR_2 não é uma condição suficiente para que sejam válidas as condições de dependência negativa que vimos até aqui.

Uma possível solução para o problema seria supormos a condição adicional de que todas as marginais sejam também MRR_2. Essa solução foi adotada em Ebrahimi & Ghosh (1980) e com ela fica fácil mostrar que $\underset{\sim}{X}$ é CDS. Ao que tudo indica no entanto, ela ainda não é suficiente para garantir a validade de (2.1.1) e (2.1.2).

Uma condição bem mais forte, denominada $SMRR_2$, foi proposta por Karlin & Rinnot (1980).

DEFINIÇÃO 2.2.4 - Suponha que a função densidade, $f(x_1,\ldots,x_n)$, de um vetor $\underset{\sim}{X}$ é MRR_2. Diremos que f é $SMRR_2$ se qualquer que seja a partição do conjunto de índices $\{1,\ldots,n\}$ em dois subconjuntos $\{j_1,\ldots,j_k\}$ e $\{i_1,\ldots,i_{n-k}\}$ $(k\geq 2)$ e quaisquer que sejam as funções PF_2, $\phi_1,\ldots,\phi_{n-k}$, a função $g(x_{j_1},\ldots,x_{j_k})$ dada por:

$$\int \ldots \int f(x_1,\ldots,x_n)\phi_1(x_{i_1})\ldots\phi_{n-k}(x_{i_{n-k}})dx_{i_1}\ldots dx_{i_{n-k}}$$

for MRR_2 nas variáveis $x_{j_1},\ldots,x_{j_k}$.

CONSEQÜÊNCIAS DA DEFINIÇÃO - Se a densidade f do vetor $\underset{\sim}{X}$ é $SMRR_2$, teremos:

1 - Para todo $k \geq 2$, a distribuição marginal de quaisquer k componentes de $\underset{\sim}{X}$ é MRR_2. De fato, para obtermos esse resultado basta que tomemos $\phi_1,\ldots,\phi_{-k}$ como funções constantes iguais a 1, na definição 2.2.4.

2 - A distribuição condicional de $(X_{j_1},\ldots,X_{j_k})$ dado que

$$(X_{i_1},\ldots,X_{i_{n-k}}) \in \prod_{\ell=1}^{n-k} I_\ell,$$

onde para todo ℓ, I_ℓ é um intervalo na reta, é MRR_2. Para verificar esse resultado tome ϕ_ℓ como o indicador de I_ℓ, o que é possível uma vez que o indicador de um intervalo é uma função PF_2.

Embora várias distribuições multivariadas importantes satisfaçam as condições da Definição 2.2.4, ela é por demais restritiva para os nossos propósitos. Um possível abrandamento dessas restrições pode ser obtido quando consideramos a medida induzida em R^n pela função densidade f. Vamos mostrar como isso pode ser feito, começando pelo caso $n = 2$.

Dada uma medida μ nos subconjuntos de Borel de R^2, vamos considerar a restrição de μ aos produtos cartesianos de intervalos de reta $I_1 \times I_2$ e utilizar a notação:

$$\mu(I_1, I_2) = \mu(I_1 \times I_2).$$

Dessa forma. a restrição de μ aos produtos cartesianos de intervalos passa a ser considerada como uma função de conjuntos de duas variáveis, I_1 e I_2.

Dados dois intervalos na reta, I_1 e I_1', diremos que $I_1 < I_1'$, se do fato de supormos que $x \in I_1$ e $y \in I_1'$ decorrer que $x < y$. Em outras palavras, $I_1 < I_1'$ significa que I_1 está totalmente à esquerda de I_1'.

DEFINIÇÃO 2.2.5 - Diremos que uma medida μ nos subconjuntos de Borel de R^2 é RR_2 se a restrição de μ aos produtos cartesianos de intervalos da reta satisfaz:

$$\mu(I_1, I_2) \cdot \mu(I_1', I_2') \leq \mu(I_1, I_2') \cdot \mu(I_1', I_2)$$

sempre que $I_1 < I_1'$, $I_2 < I_2'$.

De maneira análoga, dada uma medida μ nos subconjuntos de Borel de R^n, nóa podemos considerar a restrição de μ aos produtos cartesianos de intervalos,

$$\prod_{j=1}^{n} I_j ,$$

como uma função de conjuntos nas n variáveis, $I_1,\dots,I_n$, e escrever:

$$\mu(I_1,\dots I_n) = \mu\left(\prod_{j=1}^{n} I_j\right).$$

DEFINIÇÃO 2.2.6 - Diremos que uma medida μ nos subconjuntos de Borel de R^n é RR_2 em pares, se a restrição de μ aos produtos cartesianos de intervalos for RR_2 em I_j, I_k, para todo $j \neq k$, quando as demais variáveis são mantidas fixas.

Vamos considerar agora uma função densidade f e a medida μ_f por ela induzida nos subconjuntos de Borel de R^n. Não é difícil mostrar que, sob condições bastante gerais, se μ_f for RR_2 em pares, f será também RR_2 em pares.

No caso $n = 2$ vale a recíproca porém para $n > 2$, o exemplo 2.2.3 mostra que f pode ser RR_2 em pares, sem que μ_f satisfaça as condições da definição 2.2.6. Essa situação justifica a definição que daremos a seguir.

DEFINIÇÃO 2.2.7 - Diremos que a função densidade f de um vetor aleatório $\underset{\sim}{X} = (X_1,\dots,X_n)$ é fortemente RR_2 em pares, se a medida μ_f for RR_2 em pares no sentido da Definição 2.2.6.

OBSERVAÇÕES 1 - Do que foi dito acima segue-se que $f(x_1,\dots,x_n)$ é fortemente RR_2 em pares se quando integrada com relação a quaisquer $n-2$ variáveis sobre $n-2$ intervalos, a função resultante for RR_2 nas variáveis não integradas. Em termos de variáveis aleatórias essa condição significa que para todo $j \neq k$, a distribuição condicional de (X_j, X_k) dado

$$\prod_{i \neq j,k} \{X_i \in I_i\}$$

é RR_2.

2 - É claro que uma função $SMRR_2$ (Definição 2.2.4) é fortemente RR_2 em pares. Para mostrar esse fato basta tomarmos os ϕ_i da Definição 2.2.4 como indicadores de intervalos da reta. É claro também que se f for uma densidade fortemente RR_2 em pares essa propriedade será herdada pela densidade marginal de quaisquer k $(k \geq 2)$ componentes do vetor $\underset{\sim}{X}$.

Vamos agora mostrar que se a função densidade de um vetor $\underset{\sim}{X}$ satisfaz as condições da Definição 2.2.7 (ou da Definição 2.2.4) então $\underset{\sim}{X}$ será CDS DNS e DNI.

PROPOSIÇÃO 2.2.8 - Se a função densidade f de um vetor aleatório $\underset{\sim}{X}$ é fortemente RR_2 em pares, então $\underset{\sim}{X}$ é CDS.

DEMONSTRAÇÃO - Ebrahimi & Ghosh (1980) mostraram que, sob a hipótese das marginais de f serem todas RR_2 em pares, $\underset{\sim}{X}$ seria CDS. Portanto, o resultado se segue da $2^{\underline{a}}$ observação feita após a Definição 2.2.7.

LEMA - Seja X um vetor aleatório cuja função densidade f é fortemente RR_2 em pares. Então a função distribuição conjunta

$$F(t_1,\ldots,t_n) = P[X_1 \leq t_1,\ldots,X_n \leq t_n]$$

e a função de sobrevivência

$$\bar{F}(t_1,\ldots,t_n) = P[X_1>t_1,\ldots,X_n>t_n]$$

satisfazem (2.2.2), isto é, são MRR_2.

DEMONSTRAÇÃO - Se para $1 \le \ell \le n$, indicarmos por I_ℓ o intervalo $(-\infty,t_\ell]$ teremos:

$$F(t_1,t_2,\ldots,t_n) = \mu_f(I_1,I_2,\ldots,I_n).$$

Vamos fixar as variáveis t_ℓ e os intervalos I_ℓ correspondentes, para todo ℓ diferente de j e de k e para facilitar a notação indicaremos a função F e a medida μ_f, quando essas variáveis estão fixas, por $G(t_j,t_k)$ e $\nu(I_j,I_k)$ respectivamente.

Em correspondência a $t_j' > t_j$ e $t_k' > t_k$ vamos considerar os intervalos:

$$I_j' = (t_j,t_j'] \quad e \quad I_k' = (t_k,t_k'].$$

Teremos então:

$$\nu(I_j,I_k) = G(t_j,t_k)$$

$$\nu(I_j',I_k') = G(t_j',t_k') - G(t_j,t_k') - G(t_j',t_k) + G(t_j,t_k)$$

$$\nu(I_j,I_k') = G(t_j,t_k') - G(t_j,t_k)$$

$$\nu(I_j',I_k) = G(t_j',t_k) - G(t_j,t_k).$$

Como, por construção $I_j < I_j'$ e $I_k < I_k'$ e por hipótese

μ_f é RR_2 em pares, temos:

$$\nu(I_j,I_k)\cdot\nu(I_j',I_k') \le \nu(I_j,I_k')\cdot\nu(I_j',I_k).$$

Efetuando os cálculos indicados e as simplificações possíveis obtemos:

$$G(t_j,t_k)\cdot G(t_j',t_k') \le G(t_j,t_k')\cdot G(t_j',t_k)$$

o que mostra que F é MRR_2. A demonstração de que $\bar{F}$ é MRR_2 é similar.

PROPOSIÇÃO 2.2.9 - Seja $\underset{\sim}{X}$ um vetor aleatório cuja função densidade f é fortemente RR_2 em pares. Então $\underset{\sim}{X}$ satisfaz (2.1.1) e (2.1.2), isto é $\underset{\sim}{X}$ é DNS e DNI.

DEMONSTRAÇÃO - Do lema, segue-se que a função distribuição F do vetor $\underset{\sim}{X}$ é MRR_2. Portanto, quaisquer que sejam os pontos $\underset{\sim}{x}$ e $\underset{\sim}{y}$ de R^n, temos:

(2.2.3) $$F(\underset{\sim}{x}\wedge\underset{\sim}{y})F(\underset{\sim}{x}\vee\underset{\sim}{y}) \le F(\underset{\sim}{x})F(\underset{\sim}{y}).$$

Vamos considerar os pontos, $\underset{\sim}{x} = (t_1,+\infty,\ldots,+\infty)$ e $\underset{\sim}{y} = (+\infty,t_2+\infty,\ldots,+\infty)$. Teremos então:

$$F(t_1,t_2,+\infty,\ldots,+\infty)\cdot 1 \le F(t_1,+\infty,\ldots,+\infty)F(+\infty,t_2,\ldots,+\infty)$$

ou equivalentemente:

$$P[X_1\le t_1,X_2\le t_2] \le P[X_1\le t_1]P[X_2\le t_2].$$

Se considerarmos agora, os pontos:

$$\underset{\sim}{x} = (t_1, t_2, \ldots, t_{n-1}, +\infty)$$

e

$$\underset{\sim}{y} = (+\infty, +\infty, \ldots, +\infty, t_n)$$

teremos:

$$P[X_1 \leq t_1, \ldots, X_n \leq t_n] \leq P[X_1 \leq t_1, \ldots, X_{n-1} \leq t_{n-1}] P[X_n \leq t_n].$$

Segue-se então que $\underset{\sim}{X}$ é DNI.

A demonstração de que $\underset{\sim}{X}$ é DNS segue-se analogamente do fato de que $\bar{F}$ é MRR_2.

OBSERVAÇÃO - Os resultados contidos nas Proposições 2.2.8 e 2.2.9 não dependem da existência de uma densidade.Basta substituir a hipótese sobre μ_f por uma hipótese que envolva a medida induzida em R^n pelo vetor aleatório $\underset{\sim}{X}$.

As Proposições 2.2.8 e 2.2.9 mostram que $SMRR_2$ e fortemente TP_2 em pares são conceitos adequados para o estudo da dependência negativa. No entanto, ao contrário do caso positivo, onde a verificação da condição MTP_2 podia ser complicada mas não era impossível, no caso presente as condições sobre a densidade pouco ou quase nada ajudam do ponto de vista prático.

De fato, a não ser em casos muito particulares ,a verificação direta de que uma dada função satisfaz essas condições é impraticável.

Mallows (1968) mostrou diretamente que a distribuição multinomial satisfazia duas desigualdades que correspondiam exatamente aos conceitos de DNS e DNI. Nesse trabalho ele utilizou o fato da distribuição multinomial poder ser pensada como a distribuição condicional de um vetor com componentes independentes, dada a soma dessas componentes. Essa idéia se encaixa bem dentro do espírito da discussão feita no início da seção 2.1. Observe que a fixação do valor da soma funciona como um vínculo, que vai forçar que o crescimento de um bloco de componentes seja necessariamente acompanhado pela diminuição do bloco complementar.

Block, Savits & Shaked (1980) inspiraram-se nessa idéia de Mallows para provar a proposição abaixo, que nos permitirá identificar facilmente diversas distribuições multivariadas que são fortemente RR_2 em pares.

PROPOSIÇÃO 2.2.10 -Seja $\underset{\sim}{X} = (X_1, \ldots, X_n)$ um vetor aleatório com função densidade de f e seja $\underset{\sim}{Y} = (Y_0, Y_1, \ldots, Y_n)$ um vetor com componentes independentes e densidades que são PF_2. Seja $S = Y_0 + \ldots + Y_n$ e suponha que:

$$\underset{\sim}{X} \overset{st}{=} (Y_1, \ldots, Y_n / S=s).$$

Nessas condições, f é fortemente RR_2 em pares e conseqüentemente $\underset{\sim}{X}$ é CDS, DNS e DNI.

DEMONSTRAÇÃO - Vamos indicar por $f_j(y_j)$ a densidade de Y_j, para $j = 0,\ldots,n$. A densidade condicional de $(Y_1,\ldots,Y_n/S=s)$ será então dada por

$$(2.2.4) \qquad k\cdot f_1(y_1)f_2(y_2)\ldots f_n(y_n)f_0(s-y_1-y_2-\ldots-y_n)$$

onde k é uma constante normalizadora. Vamos mostrar que se essa função for integrada com relação a $y_3,\ldots,y_n$ sobre os intervalos $A_3,\ldots,A_n$, a função resultante será RR_2 em y_1 e y_2. O mesmo argumento poderá ser aplicado a qualquer par (y_j,y_k) o que nos permitirá concluir que f é fortemente RR_2 em pares.

Considere a função:

$$g(z) = \int_{A_3}\ldots\int_{A_n} f_3(y_3)\ldots f_n(y_n)f_0(z-y_3\ldots-y_n)dy_3\ldots dy_n.$$

Essa função é a convolução das funções PF_2

$$f_3 I_{A_3},\ldots,f_n I_{A_n}$$

com a função PF_2, f_0. Segue-se então que g é PF_2.

A integral de (2.2.4) com relação a $y_3,\ldots,y_n$ sobre $A_3,\ldots,A_n$ é igual a

$$(2.2.5) \qquad k\cdot f_1(y_1)f_2(y_2)g(s-y_1-y_2).$$

Observe que se h é uma função PF_2 então h(u-v-w) é

RR_2 em v e w. Segue-se então que (2.2.5) é RR_2 em y_1 e y_2 o que demonstra a proposição.

EXEMPLO 2.2.11 (Distribuição Multinomial) - Seja $\underset{\sim}{X} = (X_1,\ldots,X_n)$ um vetor aleatório com distribuição multinomial com parametros n, p_1, $p_2,\ldots,$ p_k. Temos então:

$$(2.2.6) \quad P[X_1=x_1,\ldots,X_k=x_k] = \frac{n!}{x_1!\ldots x_k!x_{k+1}!}\, p_1^{x_1}\ldots p_k^{x_k}p_{k+1}^{x_{k+1}}$$

onde

$$x_{k+1} = n - \sum_{j=1}^{k} x_j \quad \text{e} \quad p_{k+1} = 1 - \sum_{j=1}^{k} p_j .$$

n, $x_1,\ldots,x_k$, x_{k+1} são inteiros não negativos e $p_j > 0$, para $j = 1,\ldots,k+1$. Suponha agora que $Y_1,Y_2,\ldots,Y_{k+1}$ são variáveis aleatórias independentes e que para cada $j = 1,\ldots,k+1$, Y_j tem distribuição de Poisson com média $\lambda_j = p_j$. É fácil verificar que a distribuição condicional de

$$\left\{Y_1,\ldots Y_k \middle/ \sum_{j=1}^{k+1} Y_j\right\}$$

é dada por (2.2.6). Por outro lado verifica-se diretamente que a função

$$f(x) = \frac{e^{-\lambda}\lambda^x}{x!},$$

definida no conjunto dos inteiros não negativos, é PF_2. Segue-se que a distribuição multinomial satisfaz as condições da Proposição 2.2.10 e portanto é fortemente RR_2 em pares, CDS, DNS e DNI.

EXEMPLO 2.2.12 (Normal Multivariada) - Seja $\underset{\sim}{X} = (X_1,\dots,X_k)$ um vetor aleatório com distribuição normal multivariada com médias iguais a zero e matriz de variância-covariância dada por:

$$\Sigma = \begin{Vmatrix} 1 & \rho & \rho & \dots & \rho \\ \rho & 1 & \rho & \dots & \rho \\ \vdots & \vdots & \vdots & & \vdots \\ \rho & \rho & \rho & \dots & 1 \end{Vmatrix}.$$

Suponhamos ainda que $\rho < 0$. Como

$$E\left(\sum_{j=1}^{k} X_j\right)^2 \geq 0$$

teremos necessariamente $\rho > \frac{-1}{k-1}$. Suponha agora que $Y_1,\dots,Y_k$ são variáveis normais, independentes com médias iguais a zero e variâncias iguais a $1-\rho$. Seja Y_{k+1} uma variável aleatória normal independente dos Y_j, com média zero e variância igual a $-\frac{(1-\rho)(1+(k+1)\rho)}{\rho}$. É fácil verificar que

$$\underset{\sim}{X} \overset{st}{=} \left(Y_1,\dots,Y_k \Big/ \sum_{j=1}^{k+1} Y_j = 0\right).$$

Como toda densidade normal é PF_2 segue-se da Proposição 2.2.10 que $\underset{\sim}{X}$ é fortemente RR_2 em pares, CDS, DNS e DNI.

CAPÍTULO 3

3.1 - DEPENDÊNCIA NEGATIVA DA DISTRIBUIÇÃO NORMAL MULTIVARIADA

Neste capítulo, apresentaremos uma análise da situação atual das pesquisas sobre dependência negativa em comparação com os resultados já conhecidos no caso positivo. Faremos também alguns comentários sobre as limitações da dependência negativa e daremos algumas sugestões sobre possíveis caminhos, que eventualmente poderão conduzir à solução de problemas em aberto.

Vamos iniciar com alguns comentários sobre a distribuição normal multivariada. Se $\underset{\sim}{X}$ for um vetor aleatório normal, com covariâncias negativas, prova-se de maneira análoga ao caso positivo (ver Proposição 1.4.6) que $\underset{\sim}{X}$ é NDS. É possível mostrar também que $\underset{\sim}{X}$ satisfaz (2.1.1) e (2.1.2), isto é, que $\underset{\sim}{X}$ é DNS e DNI.

No que diz respeito à condições sobre a densidade, a situação é um pouco mais complicada. A validade de uma condição análoga à da Proposição 1.4.1 não nos ajuda muito uma vez que MRR_2 não é uma condição forte de dependência negativa.

Não se conhece uma condição necessária e suficiente para que um vetor aleatório normal seja $SMRR_2$. Karlin & Rinnot (1980-b) dão uma condição sobre a matriz de variância-covariância que é suficiente para garantir que a densidade é $SMRR_2$. Essencialmente essa condição diz que se $\underset{\sim}{X}$ é um vetor $N(\underset{\sim}{0},\Sigma)$, $\underset{\sim}{X}$ será $SMRR_2$ se Σ puder ser colocada na forma:

$$\Sigma = D - \| \alpha_i \alpha_j \|_{i,j=1}^{n},$$

onde, $D = (d_1,\ldots,d_n)$ é uma matriz diagonal positiva definida, isto é, $d_j > 0$, para $j = 1,\ldots,n$, e os α_j são não negativos e tais que

$$\sum_{j=1}^{n} \frac{\alpha_j^2}{d_j} < 1.$$

Vimos no Exemplo 2.2.12 que se as covariâncias são iguais então f é $SMRR_2$. Esse é um caso particular de (3.1.1) quando tomamos $d_j = 1-\rho$, $\alpha_j = \sqrt{-\rho}$ para $-\frac{1}{n-1} < \rho \leq 0$. Nesse caso,

$$\sum \frac{\alpha_j^2}{d_j} = \frac{-n\rho}{1-\rho}$$

que é menor do que 1 uma vez que $1+(n-1)\rho > 0$.

Na proposição abaixo nós caracterizamos os vetores aleatórios normais que satisfazem a condição (3.1.1) de Karlin-Rinnot.

PROPOSIÇÃO 3.1.1 - Uma matriz Σ satisfaz a condição (3.1.1) se e somente se ela for a matriz de variância-covariância da distribuição condicional de $(Y_1,\ldots,Y_n)$ dada a soma $\sum_{j=1}^{n+1} \beta_j Y_j$,

onde $(Y_1,\ldots,Y_n,Y_{n+1})$ é um vetor aleatório normal com componentes independentes e os β_j são tais que, $\beta_{n+1} > 0$ e $\beta_j \geq 0$ para $j = 1,\ldots,n$.

DEMONSTRAÇÃO - a) SUFICIÊNCIA - A demonstração é imediata uma vez que se Σ for a matriz de variância-covariância de uma distribuição condicional do tipo descrito acima, as propriedades da distribuição normal multivariada nos permitem escrever:

$$(3.1.2)\qquad \Sigma = \begin{Vmatrix} \sigma_1^2 & 0 & \ldots & 0 \\ 0 & \sigma_2^2 & \ldots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \ldots & \sigma_n^2 \end{Vmatrix} - \frac{1}{\sigma^2} \begin{Vmatrix} \beta_1^2\sigma_1^4 & \ldots\ldots\ldots\ldots & \beta_1\beta_n\sigma_1^2\sigma_2^2 \\ \vdots & & \vdots \\ \beta_i\beta_n\sigma_1^2\sigma_n^2 & \ldots\ldots\ldots\ldots & \beta_n^2\sigma_n^2 \end{Vmatrix}$$

onde

$$\sigma_j^2 = E(Y_j^2) \quad e \quad \sigma^2 = \sum_{j=1}^{n+1} \beta_j^2\sigma_j^2 .$$

É claro que Σ está na forma dada em (3.1.1) com

$$d_j = \sigma_j^2 \quad e \quad \alpha_j = \frac{\beta_j\sigma_j^2}{\sigma}.$$

Observe ainda que:

$$\sum \frac{\sigma_j^2}{d_j} = \frac{\sum_{j=1}^{n} \beta_j^2 \sigma_j^2}{\sum_{j=1}^{n+1} \beta_j^2 \sigma_j^2}$$

o que é menor do que 1 uma vez que por hipótese $\beta_{n+1} > 0$.

b) NECESSIDADE - Suponha agora que Σ é uma matriz que satisfaz a condição (3.1.1). Defina:

$$\sigma^2 = \frac{1}{1 - \sum_{j=1}^{n} \frac{\sigma_j^2}{d_j}}$$

$$\beta_j = \frac{\alpha_j \sigma}{d_j} \text{ para } j=1,\dots,n \tag{3.1.3}$$

e

$$\beta_{n+1} = 1.$$

De (3.1.1) segue-se que $\sigma^2 > 0$ e portanto os β_j estão bem definidos e são não-negativos.

Considere um vetor aleatório normal $(Y_1,\dots,Y_{n+1})$ com componentes independentes, $V(Y_{n+1}) = 1$ e $V(Y_j) = d_j$ para $j = 1,\dots,n$. Segue-se então que:

$$V(Z) = \sum_{j=1}^{n} \frac{\alpha_j^2 \sigma_j^2}{d_j^2} d_j + 1 = 1 + \sigma^2 \sum \frac{\alpha_j^2}{d_j} = \sigma^2$$

Segue-se então que a matriz de variância-covariância da distribuição condicional de $(Y_1,\dots,Y_n)/Z$ será:

$$D - \left\| \frac{\beta_i \beta_j d_i d_j}{\sigma^2} \right\|_{i,j=1}^{n}.$$

Como $\beta_j = \frac{\alpha_j \sigma}{d_j}$ segue-se que a matriz de variância - covariância será:

$$\Sigma = D - \| \alpha_i \alpha_j \|_{i,j=1}^{n}$$

o que completa a demonstração.

OBSERVAÇÃO - O resultado de Block, Savits & Shakhed (Proposição 2.2.10) quando aplicado à normal multivariada nos dá o caso particular no qual $\beta_j = 1$ para $j = 1,\dots,n+1$.

3.2 - SUGESTÕES E COMENTÁRIOS FINAIS

As dificuldades encontradas na análise da dependência negativa explicam-se em parte, pelas restrições que tem que ser impostas à distribuição conjunta, para que as componentes do vetor se comportem de acordo com a idéia intuitiva discutida no início do Capítulo 2.

Para dar uma idéia sobre a natureza dessas restrições vamos considerar um vetor $\underset{\sim}{X} = (X_1,\dots,X_n)$ que satisfaz a mais fraca das condições de dependência negativa conhecidas, isto é, suponhamos que $Cov(X_i,X_j) \leq 0$ para todo $i \neq j$. Suponhamos ainda que $E(X_j) = 0$ e $V(X_j) = 1$, para $j = 1,\ldots,n$. Segue-se então que:

$$E\left(\sum_{j=1}^{n} X_j^2\right) = n + \sum_{\substack{i \neq j \\ i,j=1}}^{n} \mathrm{Cov}(X_i, X_j) \geq 0.$$

Dessa desigualdade segue-se que se considerarmos a média aritmética $\bar{\rho}$ das $\binom{n}{2}$ covariâncias teremos necessariamente:

$$\frac{-1}{n-1} \leq \bar{\rho} \leq 0$$

Portanto, em valores absolutos, o fato de algumas covariâncias serem moderadamente grandes forças as demais a serem bastante pequenas, limitação essa que não existia no caso positivo.

A nosso ver existem pelo menos dois problemas importantes ainda não resolvidos sobre dependência negativa. O primeiro tem caráter prático e diz respeito à determinação de condições, facilmente verificáveis, que nos permitam decidir se uma dada função densidade é ou não $SMRR_2$. O resultado de Block & Shakhed (Proposição 2.2.10) é um importante passo nessa direção mas, não é suficiente para se resolver completamente nem mesmo o caso particular da normal multivariada.

Uma questão que valeria a pena ser considerada é se toda densidade $SMRR_2$ pode ser pensada como uma distribuição condicional. Ao que tudo indica, uma resposta a essa questão teria que ser procurada no contexto das medidas induzidas pelas densidades e envolveria uma análise do comportamento de derivadas de Radon Nykodim. Sobre esse assunto existem alguns resultados esparsos na literatura mas nenhum deles responde completamente a questão.

Do ponto de vista teórico, o problema mais importante seria a obtenção de um conceito satisfatório de associação negativa. As modificações que foram tentadas na definição de associação positiva tem fracassado, seja por conduzirem a resultados contraditórios ou por resultarem em condições inúteis do ponto de vista prático. No momento, seguindo uma sugestão do Prof. Henry Block estamos analisando, juntamente com o Prof. Wagner de Souza Borges, o problema de associação negativa de variáveis binárias. A esperança é que, nesse contexto mais simples, possamos obter uma condição suscetível de ser estendida ao caso geral ou que, pelo menos, consigamos reunir elementos suficientes para que possamos garantir que efetivamente não existe uma definição adequada. Os poucos resultados que obtivemos até agora não permitem ainda uma conclusão definitiva sobre o assunto.

BIBLIOGRAFIA

AHMED,A.N. et alii. (1978) *Two concepts of positive dependence with applications in multivariate analysis* Tallahassee, University Press (Technical Report, M486. Dept.Statist.)

ALAM,K & WALLENIUS,K.T. (1976) Positive dependence and monotonicity in conditional distributions *Comm.Statist.-Theory Methods*, A5(6):525-534.

BARLOW,R.E. & PROSCHAN,F. (c1975) *Statistical theory of reliability and life testing: probability models.* New York, Holt. 290p. (International Series in Decision Processes).

BLOCK,H.W; SAVITS,T.H.; SHAKED,M. (1980) *Some concepts of negative dependence.* Pittsburgh, University Press. (Technical Report. Dept. Math. Statist.)

BLOCK,H.W. & TING,M.L. (1981) Some concepts of multivariate dependence. *Comm.Statist.-Theory Methods*, A10(8):749-762.

EBRAHIMI,N, & GHOSH,M. (1980) *Multivariate negative dependence.* Ames, Iowa University Press. (Technical Report.Dept. Statist.)

ESARY,J.D. & PROSCHAN,F. (1972). Relationships among some concepts of bivariate dependence. *Ann. Math. Statist.*,43(2): 651-655.

ESARY,J.D.; PROSCHAN,F.; WALKUP,D.W. (1967). Association of Random variables with applications. *Ann. Math. Statist.*, 38(5):1466-1474.

KARLIN,S. (1968) *Total positivity* Stanford, University Press. v.1.

KARLIN,S.& RINNOT,Y. (1980). *Classes of orderings of measures and related correlation inequalities: I-multivariate totally positive distributions.* Stanford,University Press. (Technical Report. Dept. Math.)

KARLIN,S. & RINNOT,Y (1980b). *Classes of orderings of measures and related correlation inequalities: II - multivariate reverse rule distributions.* Stanford,University Press. (Technical Report. Dept. Math.)

KEMPERMAN,J.H.B. (1977) On the FKG-inequality for measures on a partially ordered space. *Nederl. Akad. Wetensch. Proc. Ser. A, 80*(4):313-331.

LEHMANN,E.L. (1966) Some concepts of dependence. *Ann. Math. Statist., 37*(5):1137-1153.

MALLOWS,C.L. (1968). An inequality involving multinomial probabilities. *Biometrika,* 55(2):422-424.

SARKAR,T.K. (1969). *Some lower bounds of reliability.* Stanford, University Press. (Technical Report).