SAKUYA AOKI HONDA
Instrutora do Departamento de Matemática
IME - USP

SÔBRE O CONCEITO DE MARTINGAL E UM RECÍPROCO DO TEOREMA DA CONVERGÊNCIA DOMINADA

Trabalho apresentado ao Instituto de Matemática e Estatística da U.S.P. para mestrado em Probabilidade e Estatística Teórica.

Abril 1971

PREFÁCIO

Neste trabalho descrevemos a teoria dos martingais em L^1 (Capítulo I, § 4) dando ênfase à técnica de amostragem opcional, que pode interessar também aos analistas que não se ocupam da teoria das probabilidades.

Como aplicação, descrevemos resultados (Capítulo II) extraídos do trabalho de Blackwell e Dubins sôbre um recíproco do teorema da convergência dominada.

Desejamos ressaltar o papel essencial que cabe, nas aplicações acima referidas, às propriedades do rearranjo decrescente de uma variável aleatória deduzidas por Hardy e Littlewood, e parece-nos que o procedimento apresentado pode também ser aplicado em várias outras situações, tendo em vista que substituimos, sem perda das propriedades essenciais, uma função de comportamento não inteiramente definido por uma função contínua estritamente decrescente.

O Capítulo 0, em que apresentamos aspectos gerais da teoria das probabilidades, é perfeitamente dispensável para o leitor que conhece os fundamentos da teoria das probabilidades; sua presença se justifica sòmente pelo nosso desejo de estabelecer uma linguagem e dar referências de resultados usados nos capítulos subseqüentes.

No Capítulo I (§ 1, § 2 e § 3) procuramos apresentar os tópicos essenciais da teoria das probabilidades como a independência estocástica, a probabilidade e esperança condicionais à luz dos conceitos de Análise e Análise Funcional. Seguimos nesta parte do trabalho a orientação dada pelo professor Jacques Neveu, da Universidade de Paris, cuja presença oportuna no Departamento de Estatística do IME, durante o mês de agôsto de 1969, nos possibilitou ampliar os horizontes em curto espaço de tempo. De grande valia na elaboração dêste Capítulo foram os cursos de Análise e Análise Funcional ministrados pelo professor Chaim S. Hönig no IME em 1968 e 1969.

Êste trabalho foi-nos sugerido e orientado pelo professor Carlos Alberto Barbosa Dantas, a quem devemos inclusive nosso preparo na teoria das probabilidades, através dos cursos de Probabilidade Avançada, Processos Estacionários e Convergência de Medidas ,

por êle ministrados no IME no segundo semestre de 1969 e durante o ano de 1970.

Pelo muito que lhes devemos, desejamos expressar-lhes os nossos sinceros agradecimentos.

São Paulo, abril de 1971

Sakuya Aoki Honda

ÍNDICE

CAPÍTULO 0 - ESPAÇOS DE PROBABILIDADE

0.1 - Probabilidade em um espaço probabilizável

Na construção da teoria da probabilidade partimos de dois conceitos primitivos, a serem interpretados intuitivamente: pontos amostrais, que representam os resultados, considerados a priori possíveis, de uma experiência aleatória; espaço amostral, que é o conjunto Ω de todos os pontos amostrais.

Evento é um conjunto de pontos amostrais e a classe dos eventos é uma parte $\mathcal{J}$ de $\mathcal{P}(\Omega)$.

A cada evento $A \in \mathcal{J}$ associamos a probabilidade $P(A)$ de ocorrer o evento A, onde P é uma função de conjunto definida em $\mathcal{J}$ com valores reais, sendo no mínimo igual a zero e no máximo igual a 1.

Espaço amostral, evento e probabilidade estão relacionados com a experiência aleatória da seguinte maneira:

Qualquer resultado imaginável da experiência é descrito por um único ponto amostral ω e todo ponto amostral é um resultado possível de experiência. O conjunto de todos os pontos amostrais em que ocorre o evento A descreve completamente o evento A, isto é, A ocorre ou não ocorre conforme o resultado da experiência seja ou não descrito por um ponto amostral pertencente a A.

A seguir vamos descrever propriedades da classe $\mathcal{J}$ onde vamos definir a probabilidade P:

Como todo ponto amostral ω pode estar em A ou não estar em A, para cada evento A o seu complementar A^c é ainda um evento, isto é:

$$\text{se } A \in \mathcal{J}, \text{ então } A^c \in \mathcal{J}.$$

Em particular, como todo ponto amostral ω está em Ω, Ω é chamado evento certo, a êle atribuimos a probabilidade 1 $(P(\Omega) = 1)$; o complementar de Ω, o ϕ, é chamado evento vazio ou impossível, e temos sempre:

$$\Omega, \ \phi \in \mathcal{J}$$

Como um ponto amostral pode estar simultâneamente em dois eventos, o conjunto de todos os pontos amostrais pertencentes a ambos os eventos A e B é o evento intersecção de A e B, que designamos por $A \cap B$ ou AB, e $A \cap B \in \mathcal{J}$, para todo $A, B \in \mathcal{J}$.

Em particular, se A e B se excluem mùtuamente, isto é, se a ocorrência de um impede a ocorrência do outro, dizemos que A e B são disjuntos, e indicamos:

$$A \cap B = \phi \quad ,$$

Se $A, B \in \mathcal{J}$, então $A^c, B^c \in \mathcal{J}$ logo $A^c \cap B^c \in \mathcal{J}$ e $(A^c \cap B^c)^c = A \cup B \in \mathcal{J}$ sempre que $A, B \in \mathcal{J}$.

Dados dois eventos A e B, a reunião $A \cup B$ corresponde à "ocorrência de A, ou de B, ou de ambos".

Se, em particular, A e B são disjuntos, indicamos a reunião $A \cup B$ como soma: $A + B$. O evento $A + B$ ocorre se, e sòmente se ocorre um e apenas um dos eventos A e B, sendo que a probabilidade de ocorrer $A + B$ será a soma da probabilidade de ocorrer A com a probabilidade de ocorrer B:

$$P(A + B) = P(A) + P(B) \qquad \text{(aditividade de } P)$$

Como

$$A + A^c = \Omega \quad ,$$

temos

$$P(A^c) = 1 - P(A) \quad .$$

Em particular,

$$P(\phi) = 1 - P(\Omega) = 0$$

Ainda, se $A, B \in \mathcal{J}$, $A - B = A \cap B^c$ é o conjunto dos pontos amostrais de A que não estão em B e a diferença simétrica $A \Delta B = (A - B) \cup (B - A)$ é o conjunto dos pontos amostrais que estão em um dos eventos, mas não no outro.

É claro que, se $A, B \in \mathcal{J}$, então $A - B, A \Delta B \in \mathcal{J}$.

Na realidade, vamos atribuir à classe $\mathcal{J}$ dos eventos uma estrutura mais rica, que passamos a descrever:

Definição 0.1 - Dada uma classe $\mathcal{A}$ de subconjuntos de Ω contendo Ω e satisfazendo:

A_1) Se $A \in \mathcal{A}$, então $A^c \in \mathcal{A}$;

então, se $\mathcal{A}$ satisfaz a uma das condições:

A_2) Se $A, B \in \mathcal{A}$ então $A \cap B \in \mathcal{A}$;

A_2') Se $A, B \in \mathcal{A}$ então $A \cup B \in \mathcal{A}$;

a outra também está satisfeita, e neste caso dizemos que $\mathcal{A}$ é uma álgebra sôbre Ω.

Em outras palavras, uma álgebra é uma classe de eventos contendo Ω e ϕ, estável pelas operações de complementação e intersecção finita (portanto também estável por reunião finita).

Definição 0.2 - Dada uma classe $\mathcal{J}$ de subconjuntos de Ω contendo Ω e satisfazendo:

$\mathcal{J}_1)$ Se $A \in \mathcal{J}$ então $A^c \in \mathcal{J}$,

se além disso $\mathcal{J}$ satisfaz a uma das condições:

$\mathcal{J}_2)$ Se $A_n \in \mathcal{J}$ $(n \geq 1)$, então $\bigcap_{n \geq 1} A_n \in \mathcal{J}$;

$\mathcal{J}_2')$ Se $A_n \in \mathcal{J}$ $(n \geq 1)$, então $\bigcup_{n \geq 1} A_n \in \mathcal{J}$;

a outra também está satisfeita, e neste caso dizemos que $\mathcal{J}$ é uma σ - álgebra sôbre Ω.

Em outras palavras, uma σ - álgebra é uma classe de eventos contendo Ω e ϕ, estável por complementação e por intersecção enumerável (portanto também estável por reunião enumerável).

Definição 0.3 - Dada uma σ - álgebra $\mathcal{J}$ sôbre Ω, uma probabilidade P sôbre $\mathcal{J}$ é uma função definida sôbre $\mathcal{J}$ com valores em $[0,1]$, tal que:

$P_1)$ $P(\Omega) = 1$;

$P_2)$ (σ -aditividade): Dada uma seqüência $(A_n)_{n \geq 1}$ de eventos de $\mathcal{J}$ dois a dois disjuntos, vale:

$$P\left(\sum_{n \geq 1} A_n\right) = \sum_{n \geq 1} P(A_n) .$$

Assim, uma probabilidade sôbre uma σ -álgebra $\mathcal{J}$ sôbre Ω nada mais é do que uma medida sôbre $\mathcal{J}$ normalizada $(P(\Omega) = 1)$, e é claro que no estudo da Probabilidade vamos usar muitos resultados da Teoria da Medida, o que explica o ponto de vista de muitos autores, que consideram a Probabilidade como uma parte ou mera aplicação da Teoria da Medida. Por outro lado, a Probabilidade tem desenvolvido métodos e raciocínios a ela peculiares, de modo que, muitas vêzes se justifica o estudo da Teoria da Médidaorientada no sentido da Probabilidade.

Definição 0.4 - Um par $(\Omega, \mathcal{J})$ formado por um conjunto não vazio Ω e uma σ -álgebra $\mathcal{J}$ de partes de Ω é chamado um espaço probabilizável.

<u>Definição 0.5</u> - Se P é uma probabilidade sôbre a σ -álgebra $\mathcal{F}$ de um espaço probabilizável $(\Omega, \mathcal{F})$, então a tripla $(\Omega, \mathcal{F}, P)$ é um <u>espaço probabilizado</u> ou um <u>espaço de probabilidade</u>.

As seguintes propriedades são conseqüências da σ -aditividade de P:

$$P(\lim_{n\to\infty} \downarrow Bn) = \lim_{n\to\infty} \downarrow P(Bn) , \qquad (0.6)$$

$$P(\lim_{n\to\infty} \uparrow Bn) = \lim_{n\to\infty} \uparrow P(Bn) \qquad (0.7)$$

Qualquer das igualdades (0.6) e (0.7) exprime que a <u>probabilidade</u> P goza da propriedade monótona sequencial.

<u>Proposição</u> 0.8: <u>Se uma função P definida na σ -álgebra $\mathcal{F}$ sôbre Ω com valores no intervalo $[0,1]$ possui as propriedades</u>:

a) $P(\Omega) = 1$;

b) (aditividade): para tôda soma finita $A_1 + \ldots + A_n$ de eventos de $\mathcal{F}$:

$$P(A_1 + \ldots + A_n) = P(A_1) + \ldots + P(A_n) ;$$

c) propriedade monótona sequencial, <u>então</u> P <u>é uma probabilidade</u>.

<u>Definição 0.9</u> - Dada uma classe $\mathcal{D}$ de subconjuntos de Ω, a menor álgebra (σ -álgebra) contendo $\mathcal{D}$ é chamada <u>álgebra</u> (σ-álgebra) gerada por $\mathcal{D}$.

Indicamos por $\mathcal{Q}(\mathcal{D})$ e $\mathcal{F}(\mathcal{D})$, respectivamente, a álgebra e σ -álgebra geradas por $\mathcal{D}$; a álgebra $\mathcal{Q}(\mathcal{D})$ (σ -álgebra $\mathcal{F}(\mathcal{D})$) é a intersecção de tôdas as álgebras (σ -álgebras) que contêm $\mathcal{D}$.

<u>Observação</u> 0.10 - Construimos $\mathcal{Q}(\mathcal{D})$ segundo as etapas sucessivas seguintes:

1º - Seja $\mathcal{C}_1$ a classe formada por ϕ, Ω e todos os eventos $A \subset \Omega$ tais que A ou $A^c \in \mathcal{D}$.

2º - Seja $\mathcal{C}_2$ a classe das intersecções finitas dos eventos de $\mathcal{C}_1$.

3º - $\mathcal{A}(\mathcal{D})$ é a classe das reuniões finitas disjuntas duas a duas, dos eventos de $\mathcal{C}_2$.

<u>Proposição</u> 0.11 - Para que uma álgebra $\mathcal{A}$ seja uma σ-álgebra é necessário e suficiente que ela seja estável para as operações de limite crescente ($\lim \uparrow$) e limite decrescente ($\lim \downarrow$).

Vamos citar um resultado que é útil para definir probabilidade em um espaço $(\Omega, \mathcal{F})$, quando $\mathcal{F}$ é a σ-álgebra gerada por uma classe C de subconjuntos de Ω estável por intersecção.

<u>Proposição</u> 0.12 - <u>Seja</u> $(\Omega, \mathcal{F})$ <u>espaço probabilizável e</u> C <u>uma classe de subconjuntos de</u> Ω <u>estável pela operação de intersecção, tal que</u> $\mathcal{F} = \mathcal{F}(C)$ ($\mathcal{F}$ <u>é a</u> σ <u>-álgebra gerada</u> por C). <u>Se duas probabilidades</u> Q <u>e</u> Q' <u>sôbre</u> $(\Omega, \mathcal{F})$ <u>coincidem nos conjuntos de</u> C, <u>então</u> Q <u>e</u> Q' <u>coincidem em todo o</u> $\mathcal{F}$.

A demonstração desta proposição pode ser encontrada em Breiman (pg. 26) ou Neveu ([1], pg. 3).

0.2 - <u>Distribuição de Probabilidade</u>

Na reta $R^{(1)}$ temos a σ-álgebra dos borelianos ou de Borel, que é a σ-álgebra gerada pelos intervalos da reta (bastaria tomar só intervalos abertos, ou fechados, ou semi-abertos). Vamos indicar com $\mathcal{B}_1$ a σ-álgebra de Borel em $R^{(1)}$. Então $(R^{(1)}, \mathcal{B}_1)$ é um espaço probabilizável e, podemos definir uma probabilidade P em $(R^{(1)}, \mathcal{B}_1)$ através dos valores de tal função na classe dos intervalos da forma $(-\infty, a)$, pois, tal classe é estável por intersecção e gera a σ-álgebra de Borel da reta.

<u>Definição</u> 0.13 - Dado um espaço de probabilidade $(\Omega, \mathcal{F}, P)$; uma <u>variável aleatória</u> sôbre $(\Omega, \mathcal{F}, P)$ é uma função

$$X : \Omega \longrightarrow R^{(1)}$$

tal que

para todo $B \in \mathcal{B}_1$ o conjunto $[X \in B] = X^{-1}(B) =$
$= \{\omega \in \Omega ; X(\omega) \in B\}$ está em $\mathcal{F}$. (0.14)

A condição (0.14) exprime que X é uma função mensurável de $(\Omega, \mathcal{F})$ em $(R^{(1)}, \mathcal{B}_1)$.

<u>Observação</u> 0.15: Se $\mathcal{F}_1 \in \mathcal{F}$ é uma sub-σ-álgebra de $\mathcal{F}$, tôda variável aleatória sôbre $(\Omega, \mathcal{F}_1)$ é

também uma variável aleatória de $(\Omega, \mathcal{F})$, mas não reciprocamente ; para indicarmos uma variável aleatória X sôbre $(\Omega, \mathcal{F}_1)$ basta dizer então X é uma função $\mathcal{F}_1$ - mensurável .

Das relações:

$$[X \in B^c] = [X \in B]^c \quad ; \quad [X \in \bigcup_{n \geq 1} B_n] = \bigcup_{n \geq 1} [X \in B_n] \qquad (0.16)$$

concluimos que uma função real X definida em Ω é uma variável aleatória se e só se a imagem inversa por X de todo intervalo $I \subset R^{(1)}$ está em $\mathcal{F}$.

<u>Definição</u> 0.17 - <u>Dada uma variável aleatória</u> X sôbre o espaço de probabilidade $(\Omega, \mathcal{F}, P)$, <u>a função</u> $\hat{P}$ <u>definida em</u> $(R^{(1)}, \mathcal{B}_1)$ <u>por</u>

$$\hat{P}(B) = P[X \in B]$$

<u>é uma probabilidade e</u> $\hat{P}$ <u>é chamada distribuição de probabilidade da variá</u>vel aleatória X .

A distribuição $\hat{P}$ de X é determinada quando conhecemos $\hat{P}(I) = P[X \in I]$, para todo intervalo I da forma $(-\infty, a)$.

As indagações referentes a uma experiência aleatória podem ser descritas por meio de variáveis aleatórias, e portanto o estudo dos problemas se reduz ao estudo das distribuições de probabilidade de variáveis aleatórias. Sob o ponto de vista da probabilidade não há diferença essencial entre as duas situações seguintes:

a) No lançamento de uma moeda perfeita, considerar $X(\omega) = 0$ se o resultado fôr "cara" e $X(\omega) = 1$ caso contrário.

b) No lançamento de um dado, tomar $Y(\omega) = 0$ se o resultado fôr um número par e $Y(\omega) = 1$ se o resultado fôr ímpar.

No caso a) o espaço amostral tem dois elementos e no caso b), seis elementos. Mas X e Y têm a mesma distribuição de probabilidade $\hat{P}$:

$$\hat{P}(B) = \begin{cases} \frac{1}{2} & \text{se } 0 \in B \text{ ou } 1 \in B \text{ (mas não ambos)} \\ 0 & \text{se } 0 \notin B \text{ nem } 1 \notin B \\ 1 & \text{se ambos } 0,1 \in B \end{cases}$$

A distribuição de probabilidade $\hat{P}$ <u>da variável aleatória</u> X <u>está univocamente determinada quando conhecemos a função</u> $F(x)$ <u>definida por</u>:

$$F(x) = P[X < x] = \hat{P}(-\infty, x) \quad , \tag{0.18}$$

<u>Se</u> $F(x)$ <u>é conhecida, a probabilidade do evento</u> $[x_1 \leq X < x_2]$ <u>é dada por</u>:

$$P[x_1 \leq X < x_2] = \hat{P}\{(-\infty, x_2) \cap (-\infty, x_1)^c\} = \tag{0.19}$$
$$= F(x_2) - F(x_1) = \Delta_{(x_1, x_2)} F(x) \ .$$

Em geral, para todo $B \in \mathcal{B}_1$, vale:

$$\hat{P}[B] = \int_B d\,F(x) \quad ,$$

onde a integral é a de Lebesgue-Stieltjes.

<u>Definição</u> 0.20 - A função $F(x)$ associada à distribuição de probabilidade $\hat{P}$ da variável aleatória X pelas relações (0.18) e (0.19) é denominada <u>função de distribuição da variável aleatória</u> X.

Das propriedades da probabilidade são deduzidas as propriedades da função de distribuição $F(x)$:

f_1) $\Delta_{[x_1, x_2)} F(x) \geq 0$ para todo intervalo finito $[x_1, x_2)$

(isto é, se $x_1 < x_2$ então $F(x_1) \leq F(x_2)$; $\quad$ (0.21)

F é não decrescente)

f_2) $\lim\limits_{x_k \uparrow x_0} f(x_k) = F(x_0)$ (continuidade pela esquerda) ;

f_3) $\lim\limits_{x_k \uparrow +\infty} F(x_k) = 1$, $\lim\limits_{x_k \downarrow -\infty} F(x_k) = 0$ (normalização) .

Em suma, dada uma variável aleatória X sôbre um espaço de probabilidade $(\Omega, \mathcal{F}, P)$, a distribuição de probabilidade $\hat{P}$ de X é determinada sôbre $(R^{(1)}, \mathcal{B}_1)$ pela função de distribuição $F(x)$.

Reciprocamente, dada uma probabilidade $\hat{P}$ em $(R^{(1)}, \mathcal{B}_1)$ e uma função $X : \Omega \longrightarrow R^{(1)}$, queremos saber se é possível construir um espaço de probabilidade conveniente $(\Omega, \mathcal{F}, P)$, de modo que X seja variável aleatória e $\hat{P}$ sua distribuição de probabilidade:

$$(\Omega, \mathcal{J}, P) \longrightarrow X \longrightarrow (R^{(1)}, \mathcal{B}_1, \hat{P})$$

Definição 0.22 - Dada uma função $X : \Omega \to \Omega'$ e uma σ-álgebra $\mathcal{B}$ de suconjuntos do Ω', a σ-álgebra gerada por X em Ω, $\mathcal{B}(X)$, é a classe de todos os subconjuntos de Ω da forma $[X \in B]$, $B \in \mathcal{B}$.

As relações(0.16) asseguram que $\mathcal{B}(X)$ é, de fato, uma σ-álgebra ; $\mathcal{B}(X)$ é a menor σ-álgebra que torna $X : \Omega \to \Omega'$ mensurável e contém os eventos que só dependem de X.

Então, para responder ao problema acima descrito, tomemos $\mathcal{J}$ como a σ-álgebra gerada por X :

$X : \Omega \longrightarrow R^{(1)}$, $\mathcal{J} = \mathcal{B}_1(X)$, $\mathcal{B}_1$ = álgebra de Borel em $R^{(1)}$, e no espaço probabilizável $(\Omega, \mathcal{B}_1(X)$ vamos definir a probabilidade P pela igualdade

$$P[X \in B] = \hat{P}(B) \quad , \quad \forall B \in \mathcal{B}_1(X) \quad . \qquad (0.23)$$

mas isto só é possível se todo elemento de $\mathcal{B}_1(X)$ fôr representado de um único modo como a imagem inversa, por X, de um boreliano B ; caso contrário, se existirem borelianos distintos $B_1, B_2 \in \mathcal{B}_1$ tais que :

$$[X \in B_1] = [X \in B_2] \quad ,$$

então (0.23) tem sentido só se vale

$$\hat{P}(B_1) = \hat{P}(B_2) \quad .$$

Então:

Proposição 0.24 - Seja $X : \Omega \longrightarrow R^{(1)}$ tal que $X(\Omega) = F \subset R^{(1)}$. Se, para cada $B \in \mathcal{B}_1$ com $B \subset F^c$ tivermos $P(B) = 0$, então está definida univocamente em $(\Omega, \mathcal{B}_1(X))$ uma probabilidade P pela relação

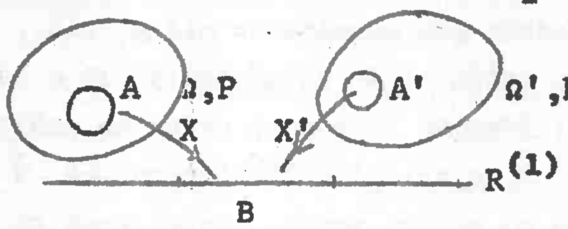
$$\hat{P}(X \in B) = \hat{P}(B) \quad , \quad \forall B \in \mathcal{B}_1 \quad ,$$

e $\hat{P}$ é a distribuição de probabilidade da variável aleatória X. (X definida em $(\Omega, \mathcal{B}_1(X), P)$).

Definição 0.25 - Duas variáveis aleatórias X e X' definidas respectivamente nos espaços de probabilidade $(\Omega, \mathcal{J}, P)$ e $(\Omega', \mathcal{J}', P')$ têm a mesma distribuição ou são igualmente distribuídas se as suas respectivas distribuições de probabilidade $\hat{P}$ e $\hat{P}'$ são iguais, isto é:

$$\hat{P}(B) = \hat{P}'(B) \quad , \quad \forall\, B \in \mathcal{B}_1 \quad \text{ou}$$

$$P[X \in B] = P'[X' \in B], \forall\, B \in \mathcal{B}_1$$



Exemplo: As variáveis aleatórias X e Y citadas nos itens a) e b) que seguem à definição 0.17 são igualmente distribuídas.

Em particular, duas variáveis aleatórias igualmente distribuídas têm a mesma função de distribuição.

Definição 0.26 - Duas variáveis aleatórias X e Y sôbre $(\Omega, \mathcal{J}, P)$ são iguais quase certamente ou $X = Y$ com probabilidade 1 se o conjunto dos pontos de Ω em que X e Y assumem valores distintos tem probabilidade nula. Indicamos:

$$X \overset{q.c.}{=} Y \quad \text{ou} \quad X = Y \ (q.c.) \quad \text{ou} \quad P[X = Y] = 1 \quad \text{ou} \quad P[X \neq Y] = 0$$

É claro que variáveis aleatórias iguais quase certamente têm a mesma distribuição. A relação $X \overset{q.c.}{=} Y$ é uma relação de equivalência e a cada classe de equivalência corresponde, univocamente, uma função de distribuição.

Reciprocamente, dada uma função $F : R^{(1)} \longrightarrow [0,1]$ satisfazendo às condições (0.21) que caracterizam uma função de distribuição, existe uma variável aleatória em espaço de probabilidade conveniente, cuja função de distribuição seja a função F dada; trata-se da função coordenada $\tilde{X}$ definida em $(R^{(1)}, \mathcal{B}_1, P)$ por:

$$\tilde{X} : R^{(1)} \longrightarrow R^{(1)} \quad ; \quad \tilde{X}(x) = x \quad , \quad \text{sendo } P = \hat{P} \text{ definida}$$

por (0.19).

De fato, como $\tilde{X}(R^{(1)}) = R^{(1)}$, não há conjuntos $B \in \mathcal{B}_1$ contidos no complementar da imagem por $\tilde{X}$ de $R^{(1)}$ logo, pela proposição 0.24, está definida em $(R^{(1)}, \mathcal{B}(\tilde{X})) = (R^{(1)}, \mathcal{B}_1)$ uma probabilidade P, tal que

$$P[B] = P[\tilde{X} \in B] = \hat{P}[B] \quad .$$

e $\hat{P}(B)$ é obtido a partir dos intervalos contidos em B, uma vez que, para cada intervalo $[x_1, x_2)$ temos:

$$\hat{P}[\ x_1 \leq x < x_2] = F(x_2) - F(x_1) .$$

Em suma, dada uma variável aleatória X em um espaço de probabilidade $(\Omega, \mathcal{F}, P)$, a ela associamos uma probabilidade $\hat{P}$ em $(R^{(1)}, \mathcal{B}_1)$, que é a distribuição de X; tôdas as variáveis aleatórias (mesmo as definidas em outro espaço de probabilidade) cuja distribuição coincida com $\hat{P}$ são consideradas iguais a X sob o ponto de vista da probabilidade; em particular, sempre existe no espaço $(R^{(1)}, \mathcal{B}_1)$ uma função coordenada $\tilde{X}$ que tem a mesma distribuição de X. Dada uma probabilidade $\hat{P}$ em $(R^{(1)}, \mathcal{B}_1)$, que pode ser dada por uma função de distribuição $F : R^{(1)} \longrightarrow [0,1]$ por (0.19), podemos sempre considerar a classe das variáveis aleatórias cuja distribuição é $\hat{P}$; um representante desta classe é a função coordenada. Notemos que as variáveis aleatórias de distribuição igual a $\hat{P}$ podem estar definidas em espaços de probabilidade diferentes.

Fixemos a atenção nas variáveis aleatórias igualmente distribuídas sôbre um mesmo espaço $(\Omega, \mathcal{F}, P)$. Sejam

$X : \Omega \longrightarrow R^{(1)}$ e $Y : \Omega \longrightarrow R^{(1)}$, igualmente distribuídas.

Então X e Y são funções mensuráveis de $(\Omega, \mathcal{F})$ em $(R^{(1)}, \mathcal{B}_1)$, isto é:

$$\mathcal{B}_1(X) \subset \mathcal{F}$$

$$\mathcal{B}_1(Y) \subset \mathcal{F}.$$

onde $\mathcal{B}_1(X)$ e $\mathcal{B}_1(Y)$ são sub- σ-álgebras de $\mathcal{F}$.

<u>Como X e Y são igualmente distribuídas, dado</u> $B \in \mathcal{B}_1$, <u>o evento</u> $A(B) \in \mathcal{F}$ <u>definido por</u>:

$$A(B) = X^{-1}(B) \Delta Y^{-1}(B) = [X \in B] \Delta [Y \in B] = ([X \in B] - [Y \in B]) +$$
$$+ ([Y \in B] - [X \in B]) ,$$

<u>é desprezível</u>:

$$P[\Delta(B)] = 0$$

Seja $\mathcal{Q}$ a classe de todos os subconjuntos de Ω que têm probabilidade P nula. Então todo evento que depende de X, isto é, evento da forma $[X \in B]$ para algum $B \in \mathcal{B}_1$, difere de um evento que depende de Y por um evento de $\mathcal{Q}$.

Em resumo, dadas duas variáveis aleatórias igualmente distribuídas X e Y, os eventos das σ-álgebras $\mathcal{B}_1(X)$ e $\mathcal{B}_1(Y)$ geradas por elas são obtidos tomando os eventos da outra σ-álgebra e reunindo elementos da classe $\mathcal{Q}$.

Definição 0.27 - Uma sub-σ-álgebra $\mathcal{F}_1 \subset \mathcal{F}$ de um espaço de probabilidade $(\Omega, \mathcal{F}, P)$ é completa (em relação a P) se $\mathcal{F}_1$ contém todos os subconjuntos de seus conjuntos de probabilidade P nula.

Proposição 0.28 - A σ-álgebra completa gerada por uma variável aleatória X, $\overline{\mathcal{B}_1(X)}$, coincide com a σ-álgebra completa gerada por qualquer variável aleatória de mesma distribuição que X definida no mesmo espaço, e é igual à menor σ-álgebra que contém os eventos de $\mathcal{B}_1(X)$ e os eventos de probabilidade P nula:

$$\overline{\mathcal{B}_1(X)} = \mathcal{F}(\mathcal{B}_1(X) \cup \mathcal{Q})$$

Em particular, se $X \overset{q.c.}{=} Y$ então $\overline{\mathcal{B}_1(X)} = \overline{\mathcal{B}_1(Y)}$.

No decorrer dêste trabalho, sempre que não houver possibilidade de confusão, diremos "variável aleatória X" em vez de "classe de equivalência da variável aleatória X" e "σ-álgebra $\mathcal{B}_1(X)$ dos eventos que dependem de X" em vez de σ-álgebra completa $\overline{\mathcal{B}_1(X)}$ dos eventos que dependem da classe de equivalência de X"

O que importa é ter presente a idéia de como uma função de distribuição F determina uma distribuição de probabilidade $\hat{P}$ em $(R^{(1)}, \mathcal{B}_1)$ e de como tal distribuição está associada a uma variável aleatória (classe de equivalência) definida em um espaço de probabilidade $(\Omega, \mathcal{F}, P)$ que por sua vez está associada à uma sub-σ álgebra (completa) de $\mathcal{F}$, e reciprocamente:

$$F \longleftrightarrow \hat{P} \longleftrightarrow \underset{\text{(classe de X)}}{X} \longleftrightarrow \overline{\mathcal{B}_1(X)}$$

Resta estudar a composição de uma função Borel-mensurável da re ta com uma variável aleatória:

Se $f : R^{(1)} \longrightarrow R^{(1)}$ é Borel-mensurável e $X : \Omega \longrightarrow R^{(1)}$ é uma variável aleatória, então $f \circ X = f(X)$ é também variável aleatória cuja distribuição está determinada pela distribuição de X :

$$\begin{tikzcd}(\Omega, \mathcal{J}, P) \arrow[r, "X"] \arrow[rr, bend right, "f \circ X"'] & (R^{(1)}, \mathcal{B}_1, \hat{P}) \arrow[r, "f"] & (R^{(1)}, \mathcal{B}_1, \hat{P}')\end{tikzcd}$$

De fato, basta notar que $\hat{P}_{f\circ X}[B] = P[f\circ X \in B] = P[X \in f^{-1}(B)] = \hat{P}_X(f^{-1}(B))$ pois, temos $[(f\circ X) \in B] = [X \in f^{-1}(B)]$,
e como para todo $B \in \mathcal{B}_1$ também $f^{-1}(B) \in \mathcal{B}_1$, todo evento de $\mathcal{B}_1(f\circ X)$ está em $\mathcal{B}_1(X)$, isto é, todo evento que depende só de $f\circ X$ depende só de X , ou ainda, $f\circ X$ é $\mathcal{B}_1(X)$-mensurável:

$$\mathcal{B}_1(f\circ X) \subset \mathcal{B}_1(X) .$$

Esta afirmação e mais a recíproca estão contidas na seguinte proposição:

<u>Proposição</u> 0.29 - <u>Dada uma variável aleatória</u> X <u>sôbre</u> $(\Omega, \mathcal{J}, P)$, <u>a classe de tôdas as variáveis aleatórias sôbre</u> Ω <u>que dependem só de</u> X ($\mathcal{B}_1(X)$-<u>mensuráveis</u>) <u>coincide com a classe das variáveis aleatórias da forma</u> $f\circ X$, <u>onde</u> f <u>é uma função real Borel-mensurável</u> (f é única, a menos de uma $\hat{P}_X$ equivalência).

Na demonstração, vamos usar o resultado seguinte, que contém apenas um procedimento básico da Teoria da Medida adaptado à linguagem da Probabilidade.

<u>Proposição</u> 0.30 - <u>Se uma classe</u> $\mathcal{L}$ <u>de funções reais</u> $\mathcal{J}$-<u>mensuráveis definidas em</u> Ω <u>satisfaz às condições</u>:

(i) $Y_1, Y_2 \in \mathcal{L}$, $\alpha, \beta \geq 0 \Rightarrow \alpha Y_1 + \beta Y_2 \in \mathcal{L}$;

(ii) $Y_n \in \mathcal{L}$, $Y_n \uparrow Y \Rightarrow Y \in \mathcal{L}$;

(iii) para todo $A \in \mathcal{J}$, $1_A \in \mathcal{L}$ (1_A = indicador de A é a função igual a 1 nos pontos de A e nula fora de A) ,

então $\mathcal{L}$ <u>contém tôdas as variáveis aleatórias não negativas sôbre</u> $(\Omega, \mathcal{J}, P)$.

Queremos mostrar que, para cada variável aleatória $\mathcal{B}_1(X)$-mensurável Z, existe uma função real mensurável f tal que $Z = f \circ X$. Mostremos inicialmente para as variáveis aleatórias não negativas; na proposição 0.30, seja $\mathcal{L}$ a classe das funções $\mathcal{B}_1(X)$-mensuráveis que podem ser escritas na forma $Z = f \circ X$, para algum f. Então, se $Y_1 = f_1 \circ X$, $Y_2 = f_2 \circ X$, $\alpha, \beta \geq 0$, segue $\alpha Y_1 + \beta Y_2 = (\alpha f_1 + \beta f_2) \circ X$, logo (i) está satisfeita por $\mathcal{L}$; do mesmo modo, se $Y_n = f_n \circ X$, $Y_n \uparrow Y$, então $Y = f \circ X$ onde $f = \lim \uparrow f_n$ é mensurável, logo (ii) está satisfeita por $\mathcal{L}$; se $A \in \mathcal{B}_1(X)$, então $A = [X \in B]$ para algum $B \in \mathcal{B}_1$, logo $1_A = 1_{[X \in B]} = 1_B \circ X$ e (iii) está satisfeita e portanto tôdas as variáveis aleatórias não negativas em $(\Omega, \mathcal{B}_1(X))$ são da forma $f \circ X$.

Para as variáveis aleatórias em geral, basta notar que elas são a diferença de duas variáveis aleatórias não negativas:

$$Z = Z^+ - Z^-$$

0.3 <u>Esperança matemática e densidade de probabilidade</u>

Seja $(\Omega, \mathcal{F}, P)$ um espaço de probabilidade e X uma variável aleatória. Se X é integrável com relação à medida P, então:

<u>Definição</u> 0.31 - <u>A esperança matemática ou valor esperado</u> EX <u>de</u> X <u>é a integral</u>:

$$EX = \int_\Omega X \, dP = \int_\Omega X(\omega) \, dP(\omega) = \int_\Omega X(\omega) \, P(d\omega) .$$

Notemos que dizer que X é integrável equivale a dizer $E|X| < \infty$. Se $\hat{P}_X$ é a distribuição de probabilidade de X, então, para cada $B \in \mathcal{B}_1$, $P[X \in B] = \hat{P}_X[B]$ e $\hat{P}_X$ determina a esperança de X, conforme a proposição seguinte:

<u>Proposição</u> 0.32 - <u>Sejam</u> X <u>e</u> X' <u>variáveis aleatórias sôbre</u> $(\Omega, \mathcal{F}, P)$ <u>e</u> $(\Omega', \mathcal{F}', P')$ <u>respectivamente, tendo a mesma distribuição de probabilidade</u> $\hat{P}$. <u>Então, se</u> ϕ <u>é</u> uma função real mensurável em $(R^{(1)}, \mathcal{B}_1)$,

$$\begin{array}{l}(\Omega, \mathcal{F}) \xrightarrow{X} \\ \qquad\qquad\qquad (R^{(1)}, \mathcal{B}_1) \xrightarrow{\Phi} (R^{(1)}, \mathcal{B}_1) \\ (\Omega', \mathcal{F}') \xrightarrow{X'}\end{array}$$

temos:

$$E\,\phi(X) = \int_{\Omega} \phi(X(\omega))\, P(d\omega) = \int_{R^{(1)}} \phi(x)\, d\,\hat{P}(x) = \int_{\Omega} \phi(X'(\omega'))P'(d\omega) =$$

$$= E\,\phi(X') \qquad (0.33)$$

<u>onde a existência de uma das integrais implica a existência das outras e as igualdades.</u>

A demonstração se faz novamente através da proposição 0.30.

Seja $\mathcal{E}$ a classe das funções reais $\mathcal{B}_1$-mensuráveis satisfazendo às igualdades (0.33); pela linearidade da integral e pelo lema de Fatou, vemos que as condições (i) e (ii) da proposição(0.30) estão satisfeitas. Por outro lado, seja $B \in \mathcal{B}_1$ e $\phi = 1_B$ = indicador de B. Como X e X' têm a mesma distribuição,

$$E(1_B o X) = \int_{\Omega} (1_B o X)(\omega)\, P(d\omega) = \int_{[X \in B]} 1.d\,P =$$

$$= P[X \in B] = \hat{P}(B) = P[X' \in B] = \int_{[X' \in B]} 1.\, d\,P' =$$

$$= \int_{\Omega'} (1_B o X')(\omega)\, P'(d\omega) = E(1_B o X')$$

logo 1_B está em $\mathcal{E}$, para $B \in \mathcal{B}_1$ e (iii) está satisfeita. Segue que tôdas as funções $\mathcal{B}_1$-mensuráveis não negativas satisfazem às igualdades (0.33); pela linearidade da integral e a igualdade:

$$\phi = \phi^{+} - \phi^{-}$$

segue que (0.33) é satisfeita por qualquer função real mensurável, c.q.d.

Em particular, para ϕ =identidade, a proposição 0.32 diz que duas variáveis aleatórias igualmente distribuídas têm a mesma esperança; variaveis aleatoriasde uma mesma classe de equivalência em um espaço dê probabilidade têm a mesma esperança. Das igualdades (0.33), vamos destacar a seguinte, no caso particular ϕ =identidade:

$$E\,X = \int_{R^{(1)}} x\, d\,\hat{P}(x) \quad , \qquad (0.34)$$

a qual mostra claramente que a esperança é um conceito associado à distribuição de X, ou à classe de equivalência de X.

<u>Corolário</u> 0.35 - Se X é uma variável aleatória sôbre $(\Omega, \mathcal{F}, P)$ e $\Phi : R^{(1)} \longrightarrow R^{(1)}$ é mensurável com $E|\Phi(X)| < \infty$, então, sendo $F(x)$ a função de distribuição de X, resulta:

$$E\,\Phi(X) = \int_{\Omega} \Phi(X)\, dP = \int_{R^{(1)}} \Phi(x)\, d\hat{P}(x) = \int_{-\infty}^{+\infty} \Phi(x)\, dF(x) \quad . \qquad (0.36)$$

Em particular, se

$$E|X| < \infty \quad , \quad E\,X = \int_{-\infty}^{+\infty} x \; dF(x)$$

Segue da última fórmula que $E\,X$ é a abscissa sôbre $R^{(1)}$ do centro de gravidade da distribuição de massas determinada por $F(x)$, daí ser chamada também de <u>valor médio</u>.

Pela linearidade da integral, temos que a esperança de uma soma finita de variáveis aleatórias é a soma das esperanças das parcelas.

Uma variável aleatória com esperança nula $E\,X = 0$, é chamada <u>centrada</u>.

Uma propriedade com aplicações freqüentes é a <u>desigualdade de Jensen</u>: Se X é tal que $E|X| < \infty$ e $\Phi(x)$ é uma função convexa definida num intervalo que contém os valores assumidos por X, vale:

$$\Phi(E\,X) \leq E(\Phi(X)) \quad .$$

Em particular:

$$|E\,X| \leq E|X|$$

<u>Definição</u> 0.37 - Se a distribuição de probabilidade $\hat{P}$ de uma variável aleatória X sôbre $(\Omega, \mathcal{F}, P)$ admite uma densidade $\delta(x)$ com relação à medida de Lebesgue μ na reta, então dizemos que $\delta(x)$ é a <u>densidade de probabilidade</u> de X.

Se X admite densidade $\delta(x)$, valem:

$$F(x) = \int_{-\infty}^{x} \delta(t)\, d_{\mu}(t) \quad (F = \text{função de distribuição de } X)$$

$$\hat{P}(a,b) = \int_{a}^{b} \delta(t)\, d\mu(t) \qquad (0.38)$$

$$E\,X = \int_{\Omega} X\, d\,P = \int_{R^{(1)}} x\, d\,\hat{P}(x) = \int_{-\infty}^{+\infty} x\, \delta(x)\, d\,\mu(x) \quad .$$

As condições para a existência da densidade de probabilidade são dadas pelo Teorema de Radom-Nikodym; antes de enunciar o teorema, lembremos que:

- uma medida $\hat{P}$ sôbre $(R^{(1)}, \mathcal{B}_1)$ é absolutamente contínua com relação a μ se a condição $\mu(B) = 0$ implica $\hat{P}(B) = 0$ para todo $B \in \mathcal{B}_1$;
- $\delta(x)$ é a densidade de uma medida $\hat{P}$ com relação a μ se para todo $B \in \mathcal{B}_1$ tivermos:

$$\hat{P}(B) = \int \delta(x)\, d\mu\,(x) \qquad (0.39)$$

e usamos a notação:

$$\hat{P} = \delta . \mu$$

- tôda função $\mathcal{B}_1$-mensurável positiva é a densidade de uma medida P' em relação a μ, sendo P' definida por (0.39) e absolutamente contínua com relação a μ.
- uma medida ν sôbre $(\Omega, \mathcal{F})$ é σ-finita se Ω é a reunião enumerável de conjuntos de medida ν finita.

Exemplo: a medida de Lebesgue μ na reta.

<u>Teorema</u> 0.40 - <u>(de Radom-Nikodym)</u> - <u>Seja</u> ν <u>uma medida absolutamente contínua com relação a uma medida positiva</u> σ <u>-finita</u> μ, <u>ambas sôbre</u> $(\Omega, \mathcal{F}, P)$. <u>Então existe uma única classe</u> $\tilde{f}$ <u>de funções numéricas</u> $\mathcal{F}$ <u>-mensuráveis positivas</u> μ <u>-equivalentes tais que cada</u> $f \in \tilde{f}$ <u>é uma densidade de</u> ν <u>com relação a</u> μ; f <u>é também chamada derivada de Radom-Nikodym de</u> ν <u>com relação a</u> μ. Se ν é finita então a densidade de ν com relação a μ é integrável.

0.4 - Processos estocásticos

As situações mais freqüentes não se descrevem por uma variável aleatória apenas, e sim por uma seqüência (finita ou infinita) de variáveis aleatórias, correspondendo a observações sucessivas de um mesmo aspecto da experiência aleatória no decorrer do tempo ou de diferentes aspectos dela para cada etapa de observação.

Definição 0.41 - Uma seqüência $(X_n)_{n \geq 1}$ de variáveis aleatórias definidas sôbre um mesmo espaço de probabilidade $(\Omega, \mathcal{J}, P)$ é um processo aleatório ou processo estocástico ou simplesmente processo sôbre $(\Omega, \mathcal{J}, P)$.

De modo análogo às variáveis aleatórias, interessa-nos estudar a distribuição de probabilidade de um processo.

Para isto consideremos no espaço $R^{(\infty)}$ de tôdas as seqüências infinitas $x = (x_1, x_2, \ldots)$ de números reais os retângulos n-dimensionais, que são os conjuntos do tipo:

$$C = \{ x = (x_1, x_2, \ldots, x_n, \ldots) \in R^{(\infty)} ; \; x_1 \in I_1, x_2 \in I_2, \ldots, x_n \in I_n \} \qquad (0.42)$$

onde $I_1, I_2, \ldots, I_n$ são intervalos finitos ou infinitos.

Observemos que a álgebra $\mathcal{a}(C)$ sôbre $R^{(\infty)}$ gerada pela classe C de todos os retângulos de dimensão finita é formada tomando reuniões finitas disjuntas duas a duas dêsses retângulos, tendo em vista a observação 0.10

Seja $\mathcal{B}_\infty$ a σ-álgebra de Borel em $R^{(\infty)}$, que é a σ-álgebra $\mathcal{J}(C)$ gerada pela classe C de todos os retângulos de dimensão finita.

Então o processo $X = (X_1, X_2, \ldots)$ é uma aplicação mensurável de $(\Omega, \mathcal{J}, P)$ em $(R^{(\infty)}, \mathcal{B}_\infty)$, isto é:

$$X = (X_1, X_2, \ldots) : (\Omega, \mathcal{J}, P) \longrightarrow (R^{(\infty)}, \mathcal{B}_\infty), \qquad (0.43)$$

$$[X \in B] = [\omega \in \Omega ; \; (X_1(\omega), X_2(\omega), \ldots) \in B] \in \mathcal{J}, \; \forall \; B \in \mathcal{B}_\infty$$

Para melhor compreender a situação vamos examinar o caso em que temos apenas duas variáveis aleatórias X_1 e X_2. Sejam $\hat{P}_{X_1}$ e $\hat{P}_{X_2}$ as distribuições de probabilidade de X_1 e X_2. Vamos considerar a função (vetor aleatório).

$$X = (X_1, X_2) \; : \; \Omega \longrightarrow R^{(2)} \; ; \; X(\omega) = (X_1(\omega), X_2(\omega)) \in R^{(2)} .$$

Um retângulo em $R^{(2)}$, é um conjunto da forma

$$C = \{x = (x_1, x_2) \in R^{(2)} \; ; \; x_1 \in I_1, x_2 \in I_2\}, \; I_1, I_2 \text{ sendo}$$

intervalos finitos ou infinitos.

Um cilindro em $R^{(2)}$ é um conjunto da forma:

$$C = \{x = (x_1, x_2) \in R^{(2)} \; ; \; x_1 \in B \, , x_2 \in B'\} \; , \; B, B' \in \mathcal{B}_1$$

A classe $\mathcal{C}$ dos retângulos de $R^{(2)}$ é estável por intersecção e a σ-álgebra $\mathcal{B}_2$ gerada por $\mathcal{C}$ (σ-álgebra de Borel) coincide com a σ-álgebra gerada pela classe $\mathcal{D}$ dos cilindros de $R^{(2)}$. Como um retângulo $\mathcal{B}$ em $R^{(2)}$ pode ser identificado com o retângulo de dimensão 2 em $R^{(\infty)}$ obtido pelo produto cartesiano de B com $R^{(1)} \times R^{(1)} \times \ldots$, podemos considerar o espaço probabilizável $(R^{(2)}, \mathcal{B}_2)$ contido em $(R^{(\infty)}, \mathcal{B})$, assim como $(R^{(1)}, \mathcal{B}_1)$ contido em $(R^{(2)}, \mathcal{B}_2)$. Anàlogamente, para as σ-álgebras sôbre Ω geradas por X_1 e por (X_1, X_2) vale a inclusão:

$$\mathcal{B}_1(X_1) \subset \mathcal{B}_2(X_1, X_2) = \{ [(X_1, X_2) \in B], \; B \in \mathcal{B}_2 \}$$

Em $(R^{(2)}, \mathcal{B}_2)$ o vetor aleatório $X = (X_1, X_2)$ define a distribuição de probabilidade conjunta $\hat{P}_2$ por

$$\hat{P}_2(B \times B') = P[(X_1, X_2) \in B \times B'] =$$

$$= P[X_1 \in B, X_2 \in B'], \; \forall \; B, B' \in \mathcal{B}_1 \qquad (0.44)$$

Em geral, a distribuição de probabilidade conjunta $\hat{P}_2(B \times B')$ é diferente de $\hat{P}_{X_1}(B) \cdot \hat{P}_{X_2}(B')$. Quando se dá a igualdade para quaisquer $B, B' \in \mathcal{B}_1$ dizemos que X_1 e X_2 são independentes.

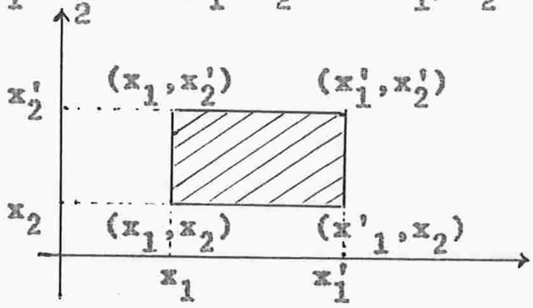
Se (Y_1, Y_2) é vetor aleatório com mesma distribuição conjunta que (X_1, X_2), então $\mathcal{B}_2(Y_1, Y_2)$ coincide com a σ-álgebra gerada por $\mathcal{B}_2(X_1, X_2) \cup \mathcal{Q}$, onde $\mathcal{Q}$ é a classe dos eventos de probabilidade P nula.

A função de distribuição do vetor aleatório (X_1, X_2), é a função $F(x_1, x_2)$ definida por:

$$F(x_1, x_2) = \hat{P}_2(X_1 < x_1, X_2 < x_2) \quad , \tag{0.46}$$

de modo que, para todo retângulo $C = [x_1, x_1') \times [x_2, x_2')$ teremos:

$$\hat{P}_2(C) = F(x_1', x_2') - F(x_1, x_2') - F(x_1', x_2) + F(x_1, x_2) \tag{0.47}$$



Para vetores aleatórios com mais de dois elementos, procedemos de modo análogo e os resultados são os seguintes:

Dado um processo estocástico sôbre $(\Omega, \mathcal{F}, P)$, para cada inteiro $n \geq 1$ o vetor aleatório $X = (X_1, X_2, \ldots, X_n)$ tem uma distribuição de probabilidade $\hat{P}_n$:

$$(\Omega, \mathcal{F}, P) \xrightarrow{(X_1, X_2, \ldots X_n)} (R^{(n)}, \mathcal{B}_n, \hat{P}_n) \, ,$$

e a família $(\hat{P}_n)_{n \geq 1}$ é consistente no seguinte sentido:

$$\hat{P}_n(B) = \hat{P}_{n+1}(B) \text{ para todo } B \in \mathcal{B}_n \qquad (\text{note que } \mathcal{B}_n \subset \mathcal{B}_{n+1}) \tag{0.48}$$

As sub-σ-álgebras $\mathcal{B}_n(X_1, \ldots, X_n) \subset \mathcal{F}$ dos eventos que dependem só das n primeiras variáveis aleatórias do processo satisfazem à condição:

$$\mathcal{B}_n(X_1, \ldots, X_n) \subset \mathcal{B}_{n+1}(X_1, \ldots, X_n, X_{n+1}) \subset \mathcal{F} \, , \quad n \geq 1$$

A classe

$$\mathcal{F}_0 = \bigcup_{n \geq 1} \mathcal{B}_n(X_1, \ldots, X_n)$$

de todos os eventos que dependem apenas de um número finito de variá-veis aleatórias é uma álgebra (mas não σ -álgebra) e a σ -álgebra $\mathcal{F}(\mathcal{F}_0)$ gerada por $\mathcal{F}_0$ coincide com a σ-álgebra $\mathcal{F}(X_1, X_2 \dots)$ gerada pelo processo $(X_1, X_2, \dots)$, isto é, sendo:

$$(X_1, X_2, \dots) : (\Omega, \mathcal{F}) \longrightarrow (R^{(\infty)}, \mathcal{B}_\infty)$$

$$\mathcal{F}(X_1, X_2, \dots) \overset{\text{def.}}{=} \mathcal{B}_\infty(X_1, X_2, \dots) = \{[(X_1, X_2, \dots) \in B], B \in \mathcal{B}_\infty\}$$

temos:

$$\mathcal{F}(X_1, X_2, \dots) = \mathcal{F}(\mathcal{F}_0) = \mathcal{F}\left(\bigcup_{n \geq 1} \mathcal{B}_n(X_1, \dots, X_n)\right) \subset \mathcal{F} \tag{0.49}$$

As funções de distribuição das probabilidades $\hat{P}_n$ são também chamadas funções de distribuição multivariadas.

Definimos a função de distribuição n-variada F_n por:

$$F_n(x_1, \dots, x_n) = \hat{P}_n(X_1 < x_1, \dots, X_n < x_n) \tag{0.50}$$

Propriedades das funções de distribuição multivariadas de um processo:

(0.51)

F_1) Dados os intervalos finitos $I_k = [a_k, b_k)$, $k = 1, \dots, n$, sendo $C = [a_1, b_1) \times \dots \times [a_n, b_n)$ e $\Delta_{I_k} G(x_1, \dots, x_k, \dots, x_n) = G(x_1, \dots, b_k, \dots, x_n) - G(x_1, \dots, a_k, \dots, x_n)$, $k = 1, \dots, n$, vale:

$$\hat{P}_n(C) = (\Delta_{I_1} \circ \dots \circ \Delta_{I_n})(F_n(x_1, \dots, x_n)) \geq 0 \;;$$

F_2) Continuidade pela esquerda: Se $x^{(k)} = (x_1^{(k)}, \dots, x_n^{(k)})$ e $x_j^{(k)} \uparrow x_j$; $j = 1, \dots, n$, então, sendo $x = (x_1, \dots, x_n)$,

$$F_n(x^{(k)}) \uparrow F_n(x)$$

F_3) Normalização: Todos os limites de F_n existem para os $x_j \uparrow \infty$ ou os $x_j \downarrow -\infty$. Se algum $x_j \downarrow -\infty$ então $F_n(x) \downarrow 0$.

Se todos os $x_j \uparrow \infty$, $j = 1, \dots, n$, então $F_n(x) \uparrow 1$.

F_4) Consistência: A família $(F_n)_{n \geq 1}$ das funções de distribuição multivariadas satisfaz à condição:

$$\lim_{x_n \uparrow +\infty} F_n(x_1, \ldots, x_n) = F_{n-1}(x_1, \ldots, x_{n-1}) \ , \ n \geq 1$$

A questão fundamental que se apresenta é a da existência de uma probabilidade $\hat{P}$ sôbre $(R^{(\infty)}, \mathcal{B}_\infty)$, que seja a "distribuição de probabilidade" do processo $(X_1, X_2, \ldots)$, isto é, satisfazendo à condição:

$$\hat{P}(B) = P\left[(X_1, X_2 \ldots) \in B \right] , \ \forall B \in \mathcal{B}_\infty . \qquad (0.52)$$

Teorema 0.53 - (da extensão de Kolmogorov). Dada uma família consistente $(\hat{P}_n)_{n \geq 1}$ de probabilidades sôbre $(R^{(n)}, \mathcal{B}_n)$, $n \geq 1$, existe uma única probabilidade $\hat{P}$ sôbre $(R^{(\infty)}, \mathcal{B}_\infty)$ cuja restrição a cada $\mathcal{B}_n$ coincide com $\hat{P}_n$.
(Demonstração em Breiman, página 24).

Definição 0.54 - Dado um processo $(X_1, X_2, \ldots)$ sôbre $(\Omega, \mathcal{F}, P)$ a probabilidade $\hat{P}$ satisfazendo à condição (0.52) e cuja existência é assegurada pelo Teorema de Kolmogorov é chamada distribuição de probabilidade do processo.

Definição 0.55 - Dois processos X e X' são equivalentes se êles têm a mesma distribuição de probabilidade $\hat{P}$.

Proposição 0.56 - Dois processos são equivalentes se, e sòmente se êles têm as mesmas funções de distribuição multivariadas.

Proposição 0.57 - Dada uma família $(F_n(x))_{n \geq 1}$ de funções definidas em $R^{(n)}$, $n = 1, 2, \ldots$, com valores em $[0, 1]$ e satisfazendo às quatro condições (0.51), existe um espaço de probabilidade conveniente e um processo sôbre êle, cujas funções de distribuição multivariadas são as funções dadas.

Basta tomar a distribuição de probabilidade $\hat{P}$ determinada sôbre $(R^{(\infty)}, \mathcal{B}_\infty)$ pelas funções $(F_n(x))_{n \geq 1}$ (Kolmogorov) e considerar em $(R^{(\infty)}, \mathcal{B}_\infty, \hat{P})$ o processo coordenado $\tilde{X} = (\tilde{X}_1, \tilde{X}_2, \ldots)$ onde

$$\tilde{X}_n(x_1, x_2, \ldots) = x_n \ , \ n \geq 1 \quad .$$

Se consideramos as sub-σ-álgebras de $\mathcal{F}$ em $(\Omega, \mathcal{F}, P)$, dos eventos que dependem sòmente das n primeiras variáveis aleatórias de um processo $(X_1, X_2, \ldots)$, obtemos a seqüência crescente

$$\mathcal{B}_1(X_1) \subset \mathcal{B}_2(X_1, X_2) \subset \ldots \quad \mathcal{B}_n(X_1, \ldots, X_n) \subset \ldots$$

À classe dos processos equivalentes ao processo $(X_1, X_2, \ldots)$ associemos a seqüência crescente de σ-álgebras completas $(\overline{\mathcal{B}_n(X_1, \ldots, X_n)})_{n \geq 1}$, geradas por $\mathcal{B}_n(X_1, \ldots, X_n) \cup \mathcal{O}$, onde $\mathcal{O}$ é a classe dos eventos de probabilidade P nula.

A condição F_1 de (0.51) dá a probabilidade $\hat{P}$ de um particular evento da forma $C = I_1 x \ldots x I_n$, a partir da função de distribuição n-variada F_n.

Em geral, se $B \in \mathcal{B}_n$ é um conjunto de Borel de $R^{(n)}$, temos:

$$\hat{P}(B) = \int_B \ldots \int d_{x_1, \ldots, x_n} F(x_1, \ldots, x_n)$$

(integral de Lebesgue-Stieltjes) .

Se existir uma função real positiva mensurável δ definida em $R^{(n)}$ e Lebesgue-integrável tal que:

$$F(x_1, \ldots, x_n) = \int_{-\infty}^{x_1} \ldots \int_{-\infty}^{x_n} \delta(\mu_1, \ldots, \mu_n)\, d\mu_1 \ldots d\mu_n$$

para todo $x_1, \ldots, x_n$ dizemos que δ é a densidade de $F(x_1, \ldots, x_n)$. Neste caso temos a probabilidade dada pela densidade δ :

$$\hat{P}(B) = \int_B \ldots \int \delta(\mu_1, \ldots, \mu_n)\, d\mu_1 \ldots d\mu_n .$$

CAPÍTULO I - ALGUNS TÓPICOS DA TEORIA DAS PROBABILIDADES

I.1 - Os espaços $L^p(\Omega, \mathcal{F}, P)$: convergências, projeção ortogonal, teorema da representação de Riesz, integrabilidade uniforme, subespaços.

Dado um espaço de probabilidade $(\Omega, \mathcal{F}, P)$, indicamos com $L(\Omega, \mathcal{F}, P)$ o espaço vetorial das (classes de equivalência das) variáveis aleatórias definidas sôbre $(\Omega, \mathcal{F}, P)$.

Para cada $p \geq 1$, $L^p(\Omega, \mathcal{F}, P)$ é o subespaço vetorial das variáveis aleatórias X com p-ésima potência $|X|^p$ integrável, isto é:

$$L^p = \{X \in L ; \int_\Omega |X|^p \, dP < \infty\}$$

Em $L^p(\Omega, \mathcal{F}, P)$, $p \geq 1$, está definida uma norma:

$$\|X\|_p = \left(\int_\Omega |X|^p \, dP\right)^{\frac{1}{p}}$$

Cada L^p com a norma acima é um espaço de Banach, isto é, é completo no sentido de que dada uma seqüência $(X_n)_{n \geq 1}$ de elementos de L^p com $\|X_n - X_m\|_p \to 0$ para $n, m, \to \infty$, existe $X \in L^p$ tal que $\|X_n - X\|_p \to 0$ para $n \to \infty$.

Os espaços L^p formam uma seqüência decrescente:

$$L \supset L^1 \supset L^2 \supset \dots \supset L^p \supset \dots$$

e se $p < q$, temos $\|X\|_p \leq \|X\|_q$.

Em L temos a noção de convergência quase certa: $X_n \to X$ quase certamente (q.c.) se $P(X_n \to X) = 1$ ou $P(X_n \not\to X) = 0$.

Teorema I.1 - (de Egoroff) - Dada uma seqüência de variáveis aleatórias $(X_n)_{n \geq 1}$ que converge quase certamente para uma variável aleatória X, existe, para todo $\varepsilon > 0$ um evento A com $P(A) < \varepsilon$ tal que $X_n \to X$ uniformemente em A^c.

A convergência em norma de L^p é chamada também convergência em média de ordem p. A relação entre a convergência quase certa e a convergência em norma do L^1 é dada num sentido pelo

Teorema I.2 - (da convergência dominada de Lebesgue) - Sejam $(X_n, X \in L(\Omega, \mathcal{F}, P)$. Se $X_n \xrightarrow{q.c.} X$ e se existir um elemento $Y \in L^1$ tal que $|X_n| \leq |Y|$ q.c., $n \geq 1$, então

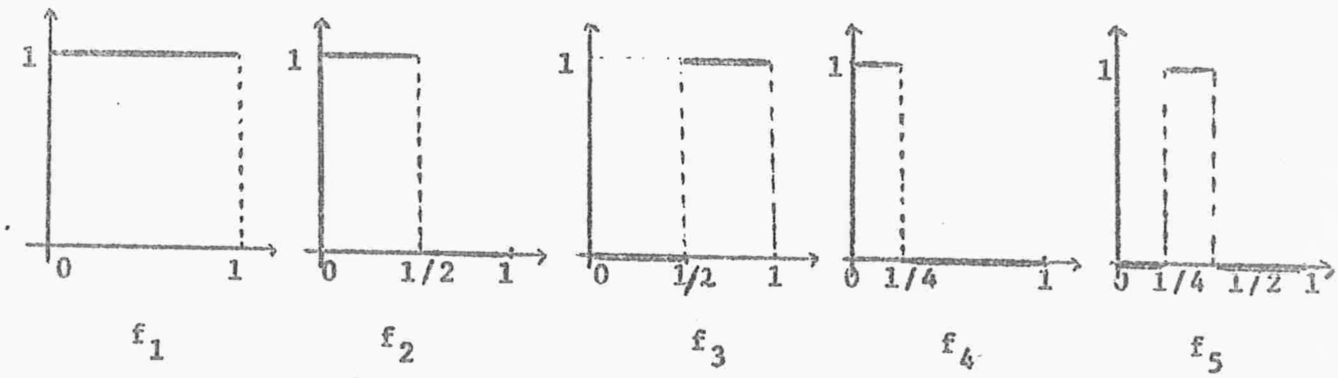
$X_n, X \in L^1$ e $||X_n - X||_1 \to 0$.

[Veremos adiante uma espécie de recíproca dêste teorema].

A convergência em média de ordem 1 não implica a convergência quase certa. Um contra exemplo é bem conhecido:

Para cada $h \geq 1$ seja a subdivisão de $[0,1]$ em 2^h intervalos iguais e façamos $0 \leq k < 2^h$ e

$$X_{2^h+k}(x) = 1_{\left[\frac{k}{2^h}, \frac{k+1}{2^h}\right]}(x) = \text{indicador do intervalo } \left[\frac{k}{2^h}, \frac{k+1}{2^h}\right]$$



As funções $X_1, X_2, \ldots$ assim obtidas convergem para 0 em L^1 mas o conjunto dos pontos em que $X_n \to 0$ tem medida de Lebesgue nula, isto é: $\mu\,[X_n \to 0] = 0$.

Mas a convergência quase certa e em L^1 seguem de uma condição mais simples de verificar:

Proposição I.3 - Para tôda seqüência $(X_n)_{n \geq 1}$ em L^1 tal que $\sum_{n \geq 1} ||X_n||_1 < \infty$, a série $\sum_{n \geq 1} X_n$ converge quase certamente e em L^1.

Ainda em L temos a noção da convergência estocástica ou em probabilidade P . Dizemos que uma seqüência $X_n \in L$ converge estocàsticamente ou em probabilidade para $X \in L$ se:

$$P\left[\,|X_n - X| > \varepsilon\,\right] \to 0 \quad \text{para } n \to \infty .$$

Indicamos esta convergência por $X_n \xrightarrow{P} X$

A convergência $X_n \xrightarrow{P} X$ equivale a: tôda subseqüência de

$(X_n)_{n \geq 1}$, tem uma subseqüência que converge q.c. para X.

Por outro lado a convergência quase certa implica convergência em probabilidade e a convergência em média de ordem p também implica a convergência em probabilidade.

O critério de Cauchy é válido também para as convergências q.c. e em probabilidade: tôda seqüência de Cauchy (q.c. ou em P) de L converge (q.c. ou em P) para um elemento de L.

Merece atenção especial o espaço $L^2(\Omega, \mathcal{F}, P)$ das variáveis aleatórias de quadrado integrável. L^2 é o único dos subespaços de L que é espaço de Hilbert: um espaço vetorial munido do produto interno, completo em relação à métrica por êle determinada. O produto interno em L^2 é definido por:

$$< X, Y > = \int_\Omega X\,Y\; dP \qquad (X, Y \in L^2) ,$$

e a norma a êle associada:

$$\|X\|_2 = \left(< X, X >\right)^{\frac{1}{2}} = \left[\int X^2\, dP\right]^{\frac{1}{2}}$$

Proposição I.4 - Num espaço de Hilbert uma série ortogonal $\sum_{n \geq 1} f_n$ é convergente se, e sòmente se $\sum_{n \geq 1} \|f_n\|^2 < \infty$.

Para espaços de Hilbert a noção de projeção ortogonal; os seguintes resultados são úteis no estudo do operador esperança condicional em L^2;

Teorema I.5 - (da projeção ortogonal) - Seja K um subespaço vetorial fechado do espaço de Hilbert H. Para todo $f \in H$, existe um único elemento de K chamado projeção ortogonal de f sôbre K, $pr_K f$, tal que $f - pr_K f$ seja ortogonal a K. Ainda, $pr_K f$ é o único elemento de K em que $\inf_{g \in K} \|g - f\|$ é atingido - $pr_K f$ é o elemento de K que mais se aproxima de f na distância dada pelo produto interno de H.

$$pr_K f \in K \;, \quad f - pr_K f \perp K \;.$$

Corolário I.6 - a) Se K é um subespaço vetorial fechado de um espaço de Hilbert H e se $K^\perp$ designa seu subespaço ortogonal, todo elemento $f \in H$ se escreve de um único modo como a soma de um elemento de K e de um elemento de $K^\perp$:

$$f = pr_K f + pr_{K^\perp} f$$

b) O projetor ortogonal pr_K sôbre um subespaço vetorial fechado K de um espaço de Hilbert H é um operador linear de norma 1 cuja imagem $pr_K(H)$ é o espaço K e o núcleo $\{ h ; pr_K(h) = 0\}$ é o espaço ortogonal $K^{\perp}$; (a norma de um operador F é definida por: $||F|| = \sup \{||F(x)||; ||X|| \leq 1\}$

c) O biortogonal $(K^{\perp})^{\perp}$ de todo subespaço vetorial fechado K de um espaço de Hilbert H coincide com K.

Um elemento $X \in L(\Omega, \mathcal{J}, P)$ é <u>quase certamente limitado</u> se $P[X > \alpha] = 0$ para algum α, a classe $L^{\infty}(\Omega, \mathcal{J}, P)$ das variáveis aleatórias q.c. limitadas ou essencialmente limitadas é um subespaço vetorial de L e

$$||X||_{\infty} = \inf \{ \alpha ; P[X > \alpha] = 0 \}$$

<u>é uma norma em L^{∞}</u> ; L^{∞} <u>com a norma</u> $||X||_{\infty}$ <u>é também espaço de Banach</u>, e a convergência em L^{∞} é a <u>convergência quase uniforme</u>, isto é, $X_n \xrightarrow{L^{\infty}} X$ se e só se $X_n \longrightarrow X$ uniformemente exceto para um conjunto de probabilidade P nula.

Seja E o espaço vetorial das variáveis aleatórias em escada de $L(\Omega, \mathcal{J}, P)$.

Então, vale a seguinte proposição citada por Neveu ([2] , Ex. II-6-3 pg. 56)

<u>Proposição</u> I.7 - Vale

$$E \subset L^{\infty} \subset L^q \subset L^p \subset L^1 \qquad (1 < p < q < \infty)$$

e E e L^r são densos em L^s $(1 \leq s \leq r \leq \infty)$. Para que $X_n \xrightarrow{L^p} X$ é suficiente que $X_n \xrightarrow{L^q} X$ se $q > p$.

Seguem dois resultados que serão utilizados no estudo do conceito de esperança condicional. As demonstrações se encontram em Neveu ([2], páginas 108 e 109) .

<u>Proposição</u> I.8 - (Teorema de representação de Riesz) - Seja $1 \leq p \leq \infty$ e seja p' dado por $\frac{1}{p} + \frac{1}{p'} = 1$. Tôda (classe de equivalência da) variável aleatória

$$X \in L^p(\Omega, \mathcal{J}, P)$$

define uma forma linear contínua F_X sôbre $L^{p'}$ pela fórmula:

$$F_X(\xi) = \int_\Omega X \xi \, dP \qquad (\xi \in L^{p'}) \quad . \tag{I.9}$$

Ainda, a norma de F_X é dada por: $||F_X|| = ||X||_p$.

Reciprocamente, quando $1 \leq p' < \infty$, tôdas as formas lineares contínuas sôbre $L^{p'}$ são dadas pela fórmula (I.9) .

<u>Corolário</u> I.10 - <u>Para que uma forma linear contínua</u> F <u>sôbre</u> $L^\infty(\Omega, \mathcal{F}, P)$ <u>seja da forma</u> $F(\xi) = \int X \xi$ <u>para algum</u> $X \in L^1$, <u>é necessário e suficiente que a função aditiva de conjuntos</u> $F(1_A)$ <u>seja</u> σ-<u>aditiva sôbre</u> $\mathcal{F}$.

<u>Definição</u> I.11 - Uma seqüência $(X_n)_{n \geq 1}$ de elementos de $L^1(\Omega, \mathcal{F}, P)$ é uniformemente integrável se

$$\sup_{n \geq 1} \int_{\{|X_n| > a\}} |X|_n \, dP \to 0$$

quando $a \to \infty$.

Exemplo: Tôda seqüência $(X_n)_{n \geq 1}$ de $L^1(\Omega, \mathcal{F}, P)$ majorada em valor absoluto por uma variável aleatória integrável X , seja $|X_n| \leq X \in L^1, n \geq 1$, é uniformemente integrável. Em particular tôda família finita de variáveis aleatórias integráveis é uniformemente integrável. Transcrevemos aqui de Neveu ([2], proposição II-5-4 página 50) a seguinte proposição que situa o conceito de integrabilidade uniforme :

<u>Proposição</u> I.12 - <u>Para tôda seqüência</u> $(X_n)_{n \geq 1}$ <u>de elementos de</u> $L^1(\Omega, \mathcal{F}, P)$ <u>e todo</u> $X \in L(\Omega, \mathcal{F}, P)$, <u>as seguintes condições são equivalentes.:</u>

(i) $(X_n)_{n \geq 1}$ é uniformemente integrável e $X_n \xrightarrow{P} X$ para $n \to \infty$;

(ii) X é integrável e $X_n \xrightarrow{L^1} X$ para $n \to \infty$.

Vimos na proposição 0.30 que, dada uma variável aleatória $X \in L(\Omega, \mathcal{F}, P)$ e a sub-σ-álgebra $\mathcal{B}_1(X) \subset \mathcal{F}$ dos eventos que dependem só de X , a classe de todos os elementos de $L(\Omega, \mathcal{F}, P)$ que são $\mathcal{B}_1(X)$-mensuráveis coincide com a classe dos elementos da forma $f \circ X$, onde f é Borel-mensurável. Dadas duas variáveis dêste tipo

$f_1 \circ X$ e $f_2 \circ X$, uma combinação linear destas com coeficientes reais ainda é do mesmo tipo:

$$\alpha\,(f_1 \circ X) + \beta(f_2 \circ X) = (\alpha\, f_1 + \beta\, f_2) \circ X$$

e é fácil ver que os elementos dêsse tipo formam um subespaço vetorial de $L(\Omega, \mathcal{F}, P)$, que é o subespaço vetorial das variáveis aleatórias definidas sôbre $(\Omega, \mathcal{B}_1(X), P)$, e que indicamos $L(\Omega, \mathcal{B}_1(X), P)$.

Se tomarmos a σ-álgebra completa $\overline{\mathcal{B}_1(X)}$, então $L(\Omega, \overline{\mathcal{B}_1(X)}, P)$ é o espaço vetorial das classes de equivalência das variáveis aleatórias sôbre $(\Omega, \overline{\mathcal{B}_1(X)}, P)$, sendo que cada classe de equivalência contém tôdas as funções da forma $f \circ X$, onde f é mensurável.

Para cada $p \geq 1$, seja o subespaço vetorial $L^p(\overline{\mathcal{B}_1(X)}) = L(\Omega, \overline{\mathcal{B}_1(X)}, P) \cap L^p(\Omega, \mathcal{F}, P)$ das variáveis aleatórias de potência p-ésima integrável definidas sôbre $(\Omega, \overline{\mathcal{B}_1(X)}, P)$. De modo análogo ao da demonstração da Proposição 0.30, demonstra-se que $L^p(\overline{\mathcal{B}_1(X)})$ coincide com o espaço vetorial das classes de equivalência das variáveis aleatórias da forma $f \circ X$, onde f é função real mensurável, com potência p-ésima integrável em relação à distribuição $\hat{P}_X$ de X .

O raciocínio é o mesmo quando substituimos a variável aleatória X pelo vetor aleatório $\mathbb{X} = (X_1, \ldots, X_n) : (\Omega, \mathcal{F}, P) \to (R^{(n)}, \mathcal{B}_n)$, sendo que, nêste caso, temos os subespaços vetoriais $L^P(\overline{\mathcal{B}_n(\mathbb{X})})$ e $\mathcal{F}$ e sôbre $R^{(n)}$ a distribuição de probabilidade conjunta $\hat{P}_{X_1,\ldots,X_n}$. Isto explica o seguinte enunciado:

<u>Proposição</u> I.13 - A correspondência $f \longmapsto f \circ X$ é um isomorfismo (em têrmos de classes de equivalência) do espaço de Banach $L^P(R^{(n)}, \hat{P}_{X_1,\ldots,X_n})$ das funções de potência p-ésima integráveis com relação a $\hat{P}_{X_1,\ldots,X_n}$, sôbre o subespaço $L^P(\mathcal{B}_n(\mathbb{X})) \subset L^P(\Omega, \mathcal{F}, P)$.

Entretanto, tomando-se qualquer sub-σ-álgebra $\overline{\mathcal{D}} \subset \mathcal{F}$ completa, sabemos que: $L(\Omega, \overline{\mathcal{D}}, P)$ é um subespaço vetorial de $L(\Omega, \mathcal{F}, P)$ (respectivamente $L^P(\Omega, \overline{\mathcal{D}}, P)$ de $L^P(\Omega, \mathcal{F}, P)$). É interessante então procurar caracterizar todos os subespaços do tipo $L^P(\overline{\mathcal{D}})$, onde $\mathcal{D}$ percorre tôdas as sub-σ-álgebras completas de $\mathcal{F}$.

Como o sup e inf de duas funções mensuráveis é mensurável e o limite monótono de funções mensuráveis é mensurável, o espaço vetorial $L(\Omega, \mathcal{B}, P)$ (respectivamente $L^p(\Omega, \mathcal{B}, P)$, $p \geq 1$) deve ser estável para as operações de sup, inf, $\lim \uparrow$ e $\lim \downarrow$. Por outro lado, como êle contém tôdas as funções características e sempre $\Omega \in \mathcal{B}$, êle contém a função constante 1, que é igual a 1_Ω. As condições acima citadas caracterizam os subespaços vetoriais de $L(\Omega, \mathcal{F}, P)$ determinados por sub-σ-álgebras completas de $\mathcal{F}$, conforme a seguinte proposição transcrita de Neveu ([1], proposição 7 página 16):

Proposição I.14 - Os subespaços $L(\mathcal{B})$ (respectivamente $L^p(\mathcal{B})$, $p \geq 1$) obtidos quando $\mathcal{B}$ percorre as sub-σ-álgebras completas de $\mathcal{F}$ são exatamente os subespaços vetoriais fechados de $L(\Omega, \mathcal{F}, P)$ $L^p(\Omega, \mathcal{F}, P)$, $p \geq 1$) que são estáveis por sup, inf, $\lim \uparrow$ e $\lim \downarrow$ e contêm a função constante 1.

Corolário I.15 - Se $(\mathcal{B}_n)_{n \geq 1}$ é uma seqüência crescente de sub-σ-álgebras completas de $\mathcal{F}$ em $(\Omega, \mathcal{F}, P)$ e se $\mathcal{B}_\infty$ é a σ-álgebra completa gerada pela classe $\bigcup_{n \geq 1} \mathcal{B}_n$, então, $\bigcup_{n \geq 1} L^p(\mathcal{B}_n)$ é um subespaço vetorial denso de $L^p(\mathcal{B}_\infty)$. (Demonstração em Neveu, [1], página 18).

I.2 Independência estocástica

Definição I.16 - Dado um espaço de probabilidade $(\Omega, \mathcal{F}, P)$ as sub-σ-álgebras $\mathcal{A}_1, \ldots, \mathcal{A}_n$ de $\mathcal{F}$ são independentes se para quaisquer eventos $A_i \in \mathcal{A}_i$, $i = 1, \ldots, n$ se tem

$$P\left(\bigcap_{i=1}^{n} A_i\right) = \prod_{i=1}^{n} P(A_i) .$$

Para uma seqüência infinita de σ-álgebras temos a

Definição I.17 - Uma seqüência $(\mathcal{A}_i)_{i \geq 1}$ de sub-σ-álgebras de $\mathcal{F}$ é independente se qualquer subseqüência finita extraída da seqüência dada fôr independente.

A definição equivale a dizer que $\mathcal{A}_1, \ldots, \mathcal{A}$ são independentes para todo $n \geq 2$.

Definição I.18 - Os eventos $A_1, \ldots, A_n$ de $(\Omega, \mathcal{F}, P)$ são independentes se as σ-álgebras $\{A_i, A_i^c, \Omega, \emptyset\}$, $1 \leq i \leq n$ forem independentes.

Proposição I.19 - Dadas as variáveis aleatórias $X_1,\dots,X_n$ as σ-álgebras $\mathcal{B}_1(X_i) \subset \mathcal{F}$, $1 \leq i \leq n$ geradas por elas são independentes se, e sòmente se a sua distribuição de probabilidade conjunta $\hat{P}_{X_1,\dots,X_n}$ fôr igual ao produto das distribuições de probabilidade das variáveis aleatórias dadas:

$$\hat{P}_{X_1,\dots,X_n} = \hat{P}_{X_1} \dots \hat{P}_{X_n} \quad \text{sôbre} \quad (R^{(n)}, \mathcal{B}_n) \tag{I.20}$$

Temos:

$$X_i : (\Omega, \mathcal{F}, P) \to (R^{(1)}, \mathcal{B}_1), \quad 1 \leq i \leq n .$$

$$(X_1, \dots, X_n) : (\Omega, \mathcal{F}, P) \longrightarrow (R^{(n)}, \mathcal{B}_n)$$

e para mostrar (I.20) é suficiente mostrar que ela vale para os retângulos $B_1 \times \dots \times B_n$, $B_i \in \mathcal{B}_1$ $1 \leq i \leq n$ (proposição 0.13). Por outro lado, pela definição da distribuição,

$$\hat{P}_{X_1 \dots X_n}(B_1 \times \dots \times B_n) = P\left[(X_1, \dots, X_n) \in B_1 \times \dots \times B_n\right] =$$

$$= P\left[X_1 \in B_1, \dots, X_n \in B_n\right] = P\left\{\bigcap_{i=1}^{n} \left[X_i \ni B_i\right]\right\} \tag{I.21}$$

e como para cada $\left[X_i \in B_i\right] \in \mathcal{B}_1(X_i)$, $1 \leq i \leq n$, a independência das σ-álgebras $\mathcal{B}_1(X_1), \dots, \mathcal{B}_1(X_n)$ implica (I.20).

Reciprocamente, se vale (I.20) então, em particular para os retângulos se tem

$$\hat{P}_{X_1 \dots X_n}(B_1 \times \dots \times B_n) = \hat{P}_{X_1}(B_1) \dots \hat{P}_{X_n}(B_n)$$

e por (I.21) temos

$$\hat{P}\left\{\bigcap_{i=1}^{n} \left[X_i \in B_i\right]\right\} = \hat{P}_{X_1 \dots X_n}(B_1 \times \dots \times B_n) = \hat{P}_{X_1}(B_1) \dots \hat{P}_{X_n}(B_n)$$

para quaisquer escôlhas de $B_i \in \mathcal{B}_1(X_i)$, logo as σ-álgebras $\mathcal{B}_1(X_1), \dots, \mathcal{B}_n(X_n)$ são independentes.

Definição I.22 - Nas condições da proposição (I.19) dizemos que as variáveis aleatórias $X_1, \dots, X_n$ são independentes.

A generalização para um número infinito de variáveis aleatórias é óbvia:

Definição I.23 - Uma seqüência $X_1, X_2, \dots$ de variáveis aleatórias é independente se, para quaisquer k-uplas $X_{n_1}, \dots, X_{n_k}$ extraídas da seqüência tivermos ,

$$\hat{P}_{X_{n_1} \dots X_{n_k}} = \hat{P}_{X_{n_1}} \dots \hat{P}_{X_{n_k}}$$

para todo $k \geq 2$.

Na realidade bastaria ter tomado

$$\hat{P}_{X_1 \dots X_n} = \hat{P}_{X_1} \dots \hat{P}_{X_n} \quad \text{para todo } n \geq 2 .$$

A noção de independência não depende do representante tomado na classe de equivalência da variável aleatória, como se vê pela proposição (I.19) que dá a independência em têrmos de distribuição.

Exprimindo a condição (I.20) por funções de distribuição, temos:

Proposição I.24 - Uma condição necessária e suficiente para que a seqüência $X_1, X_2, \dots$ seja independente é que para todo $n \geq 2$ e tôda n-upla $(t_1, \dots, t_n)$ se tenha

$$F_{X_1 \dots X_n}(t_1, \dots, t_n) = F_{X_1}(t_1) \dots F_{X_n}(t_n) \qquad (I.25)$$

Proposição I.26 - Se $Y_1, \dots, Y_n$ pertencem, respectivamente aos espaços $L^1(\mathcal{A}_1), \dots, L^1(\mathcal{A}_n)$ e se $\mathcal{A}_1, \dots, \mathcal{A}_n$ são independentes então a variável aleatória produto $Y_1 \dots Y_n$ está em $L^1(\Omega, \mathcal{F}, P)$ e temos :

$$\int_\Omega Y_1 \dots Y_n \, dP = \prod_{i=1}^{n} \int_\Omega Y_i \, dP \qquad (I.27)$$

A demonstração se apoia na proposição 0.31; a classe das variáveis aleatórias para as quais vale a proposição contém os indicadores

$Y_i = 1_{B_i}$; $B_i \in \mathcal{Q}_i$, $i = 1, \ldots, n$ (a igualdade se reduz a $1 = 1$); pela linearidade e continuidade da integral, as condições da proposição 0.31 estão satisfeitas e portanto a demonstração está feita para as variáveis aleatórias não negativas e para completá-la basta considerar a relação $Y = Y^+ - Y^-$.

<u>Corolário</u> I.28 - Se $X_1, \ldots, X_n$ são independentes e $f_1, \ldots, f_n$ são integraveis, então $\prod_{i=1}^{n} f_i(X_i)$, é integrável e

$$\int \prod_{i=1}^{n} f_i(X_i) \, dP = \prod_{i=1}^{n} \int_{\Omega} f_i(X_i) \, dP = \prod_{i=1}^{n} \int_{R^{(1)}} f_i \, d\hat{P}_{X_i}$$

Ver a proposição 0.30 .

Notemos que a fórmula acima equivale a:

$$E(f_1(X_1) \ldots f_n(X_n)) = E f_1(X_1) \ldots E f_n(X_n) \qquad (I.29)$$

se $X_1, \ldots, X_n$ são independentes e $E| f_i(X_i) | < \infty$, $i=1, \ldots, n$.

<u>Proposição</u> I.30 - Se $\mathcal{Q}_1, \mathcal{Q}_2, \ldots$ são independentes e se I_j, $j = 1, 2, \ldots$ são partes duas a duas disjuntas do conjunto dos inteiros ≥ 1, a seqüência

$$\left(\bigcup_{i \in I_j} \mathcal{Q}_i \right)_{j \geq 1}$$

é ainda independente .

<u>Definição</u> I.31 - Dada uma seqüência $X_1, X_2, \ldots$ de variáveis aleatórias sôbre $(\Omega, \mathcal{F}, P)$, um <u>evento</u> $A \in \mathcal{F}$ é chamado <u>assintótico</u> se $A \in \mathcal{F}(X_n, X_{n+1}, \ldots)$ (= σ-álgebra gerada pelas v.a. $X_n, X_{n+1}, \ldots$), para todo $n \geq 1$. A classe dos eventos assintóticos é a σ-álgebra assintótica:

$$\bigcap_{n=1}^{\infty} \mathcal{F}(X_n, X_{n+1}, \ldots)$$

Exemplos de eventos assintóticos:

1. Sejam $B_1, B_2, \ldots$ borelianos em $\mathcal{B}_1$; o evento:
$[X_n \in B_n \text{ i.v.}] = [X_n \in B_n \text{ infinitas vêzes}] =$
$= \{\omega; X_n(\omega) \in B_n$ vale para um número infinito de índices $n\} = \lim_{n \uparrow \infty} \bigcup_{n=m}^{\infty} [X_n \in B_n] = \bigcap_{m=1}^{\infty} \bigcup_{n=m}^{\infty} [X_n \in B_n] =$
$= \lim_n \sup [X_n \in B_n]$ é assintótico.

2. No lançamento da moeda perfeita, seja $X_n = \pm 1$, $n \geq 1$ conforme o resultado da n-ésima jogada seja cara ou corôa. Se c_n é qualquer seqüência de números reais, o evento $\{\omega; \sum_{n \geq 1} c_n X_n \text{ converge}\}$ é assintótico.

3. Para a mesma seqüência $(X_n)_{n \geq 1}$ do exemplo acima, o evento
$$\left\{ \omega; \frac{X_1 + \ldots + X_n}{n} \not\to \frac{1}{2} \right\}$$
é assintótico.

Dois resultados envolvendo independência e eventos assintóticos são de uso freqüente:

<u>Proposição</u> I.32 - (<u>lei</u> 0-1 <u>de Kolmogorov</u>) - <u>Se um processo</u> $X_1, X_2, \ldots$ <u>tem as variáveis aleatórias independentes, todo evento assintótico é trivial</u> - <u>tem probabilidade</u> 0 <u>ou</u> 1. (Demonstração em Breiman, página 40).

<u>Lema I.33</u> (<u>de Borel-Cantelli</u>) I. Se $(A_n)_{n \geq 1}$ é uma seqüência de eventos de $(\Omega, \mathcal{F}, P)$ tal que $\sum_{n=1}^{\infty} P(A_n) < \infty$, então $P[A_n \text{ i.v.}] = P(\lim_n \sup A_n) = 0$.

II. Se $(A_n)_{n \geq 1}$ é uma seqüência de eventos independentes de $(\Omega, \mathcal{F}, P)$ tal que $\sum_{n=1}^{\infty} P(A_n) = \infty$ então $P[A_n \text{ i.v.}] = 1$. (Demonstração em Breiman, página 41).

I.3 Probabilidade e Esperança Condicional

Seja $\mathcal{D}$ uma sub-σ-álgebra de $\mathcal{F}$ em $(\Omega, \mathcal{F}, P)$ e consideremos a medida P restrita a $\mathcal{D}$, isto é, o espaço de probabilidade

$(\Omega, \mathcal{D}, P)$. Dada uma variável aleatória X integrável sôbre $(\Omega, \mathcal{F}, P)$ seja a função:

$$\gamma(D) = \int_D X \, dP \quad , \quad D \in \mathcal{D} \tag{I.34}$$

Então $\gamma(D)$ é uma medida finita que é absolutamente contí - nua em relação à restrição de P a $(\Omega, \mathcal{D}, P)$. Pelo teorema de Radom-Nikodym, existe uma função real $\mathcal{D}$-mensurável $E^{\mathcal{D}}(X)$ tal que :

$$\gamma(D) = \int_D X \, dP = \int_D E^{\mathcal{D}}(X) \, dP \quad , \quad \forall D \in \mathcal{D} \tag{I.35}$$

e $E^{\mathcal{D}}(X)$ está definida a menos de uma P-equivalência.

<u>Definição</u> I.36 - <u>A esperança condicional</u> $E^{\mathcal{D}}(X)$ <u>de uma variável aleatória</u> X <u>em relação a uma sub-</u> σ<u>-álgebra</u> $\mathcal{D} \subset \mathcal{F}$ <u>é a derivada de Radom-Nikodym de</u> γ <u>em relação a</u> P, <u>caracterizada por</u> (I.35). Indicamos também a esperança condicional pelo símbolo $E(X \mid \mathcal{D})$

<u>Definição</u> I.37 - <u>Dado um evento</u> $A \in \mathcal{F}$, <u>a probabilidade condicional de</u> A <u>dada a sub-</u> σ<u>-álgebra</u> $\mathcal{D} \subset \mathcal{F}$ é a esperança condicional do indicador 1_A em relação a $\mathcal{D}$:

$$P(A \mid \mathcal{D}) = E(1_A \mid \mathcal{D})$$

$$P[A \cap D] = \int_{A \cap D} 1 \, dP = \int_D 1_A \, dP = \int_D P(A \mid \mathcal{D}) \, dP$$

logo $P(A \mid \mathcal{D})$ é uma função $\mathcal{D}$-mensurável que satisfaz a igualdade acima $\forall D \in \mathcal{D}$.

Em particular, as expressões esperança condicional dado um evento B, $E(X \mid B)$, e probabilidade condicional dado um evento B, $P(A \mid B)$, se referem ao caso em que a σ-álgebra $\mathcal{D}$ é gerada por $\{B\}$, isto é, $\mathcal{D} = \{B, B^c, \phi, \Omega\}$

<u>Definição</u> I.38 - <u>Quando a sub- σ -álgebra $\mathcal{D}$ em relação à qual tomamos esperança condicional é a σ -álgebra completa gerada por</u> n <u>variáveis aleatórias</u> $X_1, \ldots, X_n$, $\mathcal{B}_n(X_1, \ldots, X_n)$, <u>dizemos esperança condicional de</u> X <u>da</u>das as variáveis $X_1, \ldots, X_n$ e indicamos:

$$E^{\mathcal{B}_n(X_1,\ldots,X_n)} X = E(X|X_1,\ldots,X_n)$$

Como $E(X|X_1, \ldots, X_n)$ é uma função $\mathcal{B}_n(X_1, \ldots, X_n)$-mensurável, pela proposição 0.30 generalizada ao $R^{(n)}$ existe uma função mensurável f definida em $R^{(n)}$ e a valores reais tal que

$$E(X|X_1, \ldots, X_n) = f(X_1, \ldots, X_n) .$$

$$\begin{array}{c}(\Omega, \mathcal{F}, P) \xrightarrow{(X_1,\ldots,X_n)} (R^{(n)}, \mathcal{B}_n) \xrightarrow{f} (R^{(1)}, \mathcal{B}_1) \\ \underbrace{\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad}_{E(X|X_1,\ldots,X_n) = f(X_1,\ldots,X_n)} \\ \mathcal{B}_1(E(X|X_1,\ldots,X_n)) \subset \mathcal{D} = \mathcal{B}_n(X_1,\ldots,X_n)\end{array}$$

Em particular, se a σ -álgebra $\mathcal{B}$ é gerada por uma partição enumerável $(B_j)_{j\geq 1}$ de Ω, para tôda variável aleatória X integrável vale:

$$E^{\mathcal{B}} X = \sum_{j\geq 1} E(X|B_j) \cdot 1_{B_j}$$

logo,

$$E^{\mathcal{B}} X = \sum_{j\geq 1} \left[\frac{1}{P(B_j)} \int_{B_j} X \, dP\right] 1_{B_j} \qquad (I.39)$$

pois, integrando-se cada parcela do segundo membro em B_j temos:

$$\int_{B_j} \left(\frac{1}{P(B_j)} \int_{B_j} X \, dP\right) 1_{B_j} \, dP = \frac{P(B_j)}{P(B_j)} \int_{B_j} X \, dP = \int_{B_j} E(X|B_j) \, dP .$$

Então, neste caso particular, a esperança condicional em cada evento B_j representa a média dos valores de X em B_j. Por outro lado, em cada evento B_j a probabilidade condicional de um evento dado B_j é dada por:

$$P(A|B_j) = \frac{P(A\cap B_j)}{P(B_j)} , \qquad (j \geq 1) \qquad (I.39')$$

Propriedades da Esperança Condicional: (I.40)

1. $E(X|\mathcal{D})$ é linear em X para $\mathcal{D}$ fixado.
2. Se $X \geq 0$ q.c. então $E(X|\mathcal{D}) \geq 0$ q.c.
3. Se $\mathcal{D} \subset \mathcal{E}$ então $E^{\mathcal{D}}(E^{\mathcal{E}} X) = E^{\mathcal{D}} X$; em particular, $E^{\mathcal{D}}(E^{\mathcal{D}} X) = E^{\mathcal{D}} X$.
4. Se as σ-álgebras $\mathcal{B}_1(X)$ e $\mathcal{D}$ são independentes, então $E^{\mathcal{D}} X = E(X)$ (proposição I.28) .
5. Se $\psi(x)$ é uma função convexa definida num intervalo contendo a imagem de X, vale:

   $\psi(E^{\mathcal{D}} X) \leq E^{\mathcal{D}} [\psi(X)]$; em particular $|E^{\mathcal{D}} X| \leq E^{\mathcal{D}} |X|$
   (a desigualdade contrária vale se ψ é ~~convexa~~ côncava) .
6. Se $\lim_{n\to\infty} X_n \overset{q.c.}{=} X$ e se $E[\sup_n |X_n|] < \infty$ então

$$\lim_{n\to\infty} E^{\mathcal{D}} X_n \overset{q.c.}{=} E^{\mathcal{D}} X \ .$$

O estudo da esperança condicional como operador em $L^2(\Omega, \mathcal{J}, P)$ nos fornece uma interpretação geométrica dêsse conceito e por outro lado, a extensão para $L^1(\Omega, \mathcal{J}, P)$ permite formar uma idéia global da situação na qual tomamos as esperanças condicionais de uma função integrável como elementos dos diferentes subespaços $L^p(\Omega, \mathcal{D}, P) \subset L^1(\Omega, \mathcal{J}, P)$.

Se $\mathcal{D} \subset \mathcal{J}$ é uma sub-σ-álgebra completa, pela proposição I.14 $L^2(\mathcal{D}) = L^2(\Omega, \mathcal{D}, P)$ é um subespaço vetorial fechado do espaço de Hilbert $L^2(\Omega, \mathcal{J}, P)$ e a projeção ortogonal:

$$pr_{\mathcal{D}} : L^2(\Omega, \mathcal{J}, P) \longrightarrow L^2(\mathcal{D})$$

satisfaz à condição:

$$\int_\Omega [X - pr_{\mathcal{D}} X] Z \, dP = 0 \quad \forall Z \in L^2(\mathcal{D})$$

ou

$$\int_\Omega X Z \, dP = \int_\Omega (pr_{\mathcal{D}} X) Z \, dP \quad \forall Z \in L^2(\mathcal{D}) \qquad (I.41)$$

Notando que as condições (I,41) e

$$\int_D X \, dP = \int_D (pr_{\mathcal{D}} X) \, dP \quad \forall D \in \mathcal{D} \qquad (I.41')$$

são equivalentes e comparando com (I.35) concluimos:

<u>No espaço de Hilbert</u> $L^2(\Omega, \mathcal{F}, P)$ <u>a esperança condicional em relação a uma sub-</u> σ <u>-álgebra completa</u> $\mathcal{D} \subset \mathcal{F}$ <u>é a projeção ortogonal sôbre o subespaço vetorial fechado</u> $L^2(\mathcal{D})$.

Observação: da condição (I.41) podemos deduzir a propriedade: se Z é variável aleatória em L^2 $\mathcal{D}$ -mensurável então para todo $X \in L^2$ vale $E^{\mathcal{D}}(XZ) = Z\,E^{\mathcal{D}}(X)$.

Nem todo projetor ortogonal de $L^2(\Omega, \mathcal{F}, P)$ é uma esperança condicional. A seguinte proposição é conseqüência da proposição I.14.

<u>Proposição</u> I.42 - Para que um projetor ortogonal de $L^2(\Omega, \mathcal{F}, P)$ seja uma esperança condicional é necessário e suficiente que êle seja positivo e mantenha a função constante 1 invariante.

Para obter uma interpretação da esperança condicional em $L^1(\Omega, \mathcal{F}, P)$ vamos usar o corolário I.10 do Teorema da Representação de Riesz .

Seja $X \in L^1(\Omega, \mathcal{F}, P)$ fixado e $\mathcal{D} \subset \mathcal{F}$ uma sub- σ -álgebra completa; vamos indicar com F_X a restrição ao subespaço de Banach $L^\infty(\Omega, \mathcal{D}, P) \subset L^\infty(\Omega, \mathcal{F}, P)$ do funcional linear contínuo determinado por X :

$$F_X : L^\infty(\Omega, \mathcal{D}, P) \longrightarrow R^{(1)}$$

$$Z \longrightarrow F_X(Z) = \int_\Omega X Z \, dP$$

A função de conjunto $F_X(1_D)$, $D \in \mathcal{D}$ é σ -aditiva, pois

$$F_X(1_D) = \int X 1_D \, dP = \int_D X \, dP$$

e a integral de uma variável aleatória integrável é σ -aditiva como função do conjunto de integração. Pelo corolário I.10 existe uma variável aleatória $\bar{X} \in L^1(\mathcal{D})$ tal que

$$F_X(Z) = \int_\Omega X Z \, dP = \int_\Omega \bar{X} Z \, dP \qquad , \quad \forall Z \in L^\infty(\Omega, \mathcal{D}, P) \tag{I.43}$$

Então, comparando (I.43) com (I.35) concluimos que a esperança condicional de X dado $\mathcal{D}$ é a variável aleatória $\bar{X}$ cuja existência é assegurada pelo corolário I.10.

De modo análogo o teorema I.8 da representação de Riesz dá a e esperança condicional em cada um dos subespaços de Banach $L^p(\Omega, \mathcal{F}, P)$ $L^1(\Omega, \mathcal{F}, P)$ em relação a como sendo a variável aleatória que define funcionais lineares contínuos nesses subespaços.

I.4 Martingais

Definição I.44 - Dada uma seqüência crescente $(\mathcal{A}_n)_{n\geq 1}$ de sub- σ -álgebras em $(\Omega, \mathcal{F}, P)$, chama-se martingal (MG) adaptado à seqüência $(\mathcal{A}_n)_{n\geq 1}$ qualquer seqüência de variáveis aleatórias integráveis $(X_n)_{n\geq 1}$ tais que:

(MG 1) $X_n \in L^1(\mathcal{A}_n)$ $n \geq 1$

(MG 2) $E^{\mathcal{A}_n}(X_{n+1}) = X_n$, $n \geq 1$.

Em particular uma seqüência $(X_n)_{n\geq 1}$ de variáveis aleatórias integráveis é chamada martingal se ela fôr um martingal adaptado à seqüência das sub- σ -álgebras $(\mathcal{B}_n(X_1, \ldots, X_n))_{n\geq 1}$, isto é:

(MG) $E(X_{n+1} | X_1, \ldots, X_n) = X_n$ $n \geq 1$

A condição (MG) equivale a

(MG') $\int_A X_m \, dP = \int_A X_n \, dP$ $\forall A \in \mathcal{B}_n(X_1, \ldots, X_n)$ e $m \geq n$.

ou

(MG") $\int_A X_{n+1} \, dP = \int_A X_n \, dP$ $\forall A \in \mathcal{B}_n(X_1, \ldots, X_n)$ $n \geq 1$.

Proposição I.45 - Se $(Y_n)_{n\geq 1}$ é uma seqüência de variáveis aleatórias independentes centradas ($E\,Y_n = 0$, $n \geq 1$) de $L^1(\Omega, \mathcal{F}, P)$, então a seqüência $(X_n)_{n\geq 1}$, onde $X_n = Y_1 + \ldots + Y_n$ é um MG.

Inicialmente notemos que a função $\gamma(x_1, \ldots, x_n) = x_1 + \ldots + x_n$ é função mensurável de $(R^{(n)}, \mathcal{B}_n)$ em $(R^{(1)}, \mathcal{B}_1)$; pela proposição 0.30 generalizada ao caso do vetor aleatório $X = (X_1, \ldots X_n)$ e da relação

$$X_n = \gamma(Y_1, \ldots Y_n) ,$$

segue

$$\mathcal{B}_n(X_1, \ldots, X_n) = \mathcal{B}_n(Y_1, \ldots, Y_n)$$

e então, pela propriedade (I.40) nº 4 e a linearidade da esperança condicional,

$$E(X_{n+1} | X_1, \ldots, X_n) = E(Y_1 + \ldots + Y_n + Y_{n+1} | Y_1, \ldots, Y_n) =$$

$$= E(Y_1 + \ldots + Y_n | Y_1, \ldots, Y_n) + E(Y_{n+1} | Y_1, \ldots, Y_n) =$$

$$= Y_1 + \ldots + Y_n + E\,Y_{n+1} = X_n + 0 = X_n \;.$$

Os teoremas fundamentais da teoria dos martingais baseiam-se na idéia de <u>amostragem opcional</u>. Dado um processo $X_1, X_2, \ldots$ sôbre $(\Omega, \mathcal{F}, P)$, o processo transformado por amostragem opcional associa, a cada ponto $\omega \in \Omega$, uma subseqüência da seqüência $X_1(\omega), X_2(\omega), \ldots$ sendo que a observação das n primeiras variáveis aleatórias $X_1, \ldots, X_n$ no ponto ω determina se $X_n(\omega)$ pertence ou não ao processo transformado nêsse ponto amostral.

<u>Definição</u> I.46 - Seja $X_1, X_2, \ldots$ um processo sôbre $(\Omega, \mathcal{F}, P)$: Uma seqüência $m_1, m_2, \ldots$ de variáveis aleatórias com valores inteiros é uma <u>seqüência de variáveis amostrais</u> se :

a) $1 \leq m_1 \leq m_2 \leq \ldots$

b) $[m_k = j] \in \mathcal{B}_j(X_1, \ldots, X_j)$ para todo $k, j \geq 1$

Então a seqüência $(\tilde{X}_n)_{n \geq 1}$ definida por

$$\tilde{X}_n = X_{m_n} \qquad n \geq 1 \tag{I.47}$$

é chamada processo derivado por amostragem opcional a partir do processo $X_1, X_2, \ldots$ dado.

Observação: Para casa $k \geq 1$, m_k determina uma partição de em eventos de $\mathcal{B}_j$, $j \geq 1$ pela relação:

$$\Omega = \bigcup_{j \geq 1} [m_k = j]$$

e a amostragem opcional permite analisar o processo dado separadamente nos eventos de tal partição.

<u>Teorema</u> I.48 - <u>Seja</u> $X_1, X_2, \ldots$ <u>um</u> MG , $m_1, m_2, \ldots$ <u>variáveis amostrais</u>, $\tilde{X}_1, \tilde{X}_2, \ldots$ <u>o processo derivado por amostragem opcional</u>. <u>Se as seguintes condições estiverem satisfeitas</u>:

a) $E\,|\tilde{X}_k| < \infty \quad , \quad k \geq 1$

b) $\lim\limits_{N} \int\limits_{[m_n > N]} |X_N| \, dP = 0 \qquad n \geq 1$

<u>então o processo</u> $\tilde{X}_1, \tilde{X}_2, \ldots$ <u>é um</u> MG .

A demonstração se encontra em Breiman, página 85 .

<u>Corolário</u> I.49 - <u>Sob as condições do teorema</u> I.48 <u>e se além</u> disso

$$\overline{\lim_{n \to \infty}}\; E\,|X_n| < \infty$$

tem-se

1) $E\,X_1 = E\,\tilde{X}_n \leq \overline{\lim}\; E\,X_n$

2) $E\,|\tilde{X}_n| \leq 2\,\overline{\lim}\; E\,|X_n| - E\,X_1$

A demonstração se encontra em Breiman, página 87.

No caso particular em que as variáveis aleatórias são de quadrado integrável, podemos dar uma interpretação geométrica do conceito de martingal.

<u>Proposição</u> I.50 - <u>Se</u> $(X_n)_{n \geq 1}$ <u>é um</u> MG <u>em</u> $L^2(\Omega, \mathcal{F}, P)$ <u>a seqüência</u> $Y_1 = X_1$, $Y_n = X_n - X_{n-1}$ $(n \geq 2)$ <u>das diferenças sucessivas das variáveis aleatórias</u> X_n <u>é uma seqüência ortogonal</u>. <u>Então, pelo teorema de Pitágoras</u>:

$$E(X_n^2) = \sum_{m=1}^{n} E\,(Y_m^2) \qquad n \geq 1$$

Usamos a expressão seqüência ortogonal no sentido de que dois elementos distintos quaisquer são ortogonais. Vamos mostrar que para todo $m \leq n$ temos:

$$\int Y_m Y_{n+1} \, dP = 0 .$$

Pela observação que segue a fórmula (I.41'), como $Y_m \in L^2(X_1, \dots, X_n)$ para $m \leq n$, temos:

$$E(Y_m Y_{n+1} \mid X_1, \dots, X_n) = Y_m E(Y_{n+1} \mid X_1, \dots, X_n) =$$

$$= Y_m E(X_{n+1} - X_n \mid X_1, \dots, X_n) = Y_m X_n - Y_m X_n = 0$$

Logo, usando a definição de $E(Y_m Y_{n+1} \mid X_1, \dots, X_n)$:

$$\int Y_m Y_{n+1} \, dP = \int E(Y_m Y_{n+1} \mid X_1, \dots, X_n) \, dP = 0$$

para $m \leq n$, c.q.d.

Encontra-se em Neveu ([1], capítulo 3) um estudo detalhado dos martingais de quadrado integrável, que dispensamos, porque vamos aplicar apenas o estudo dos martingais em $L^1(\Omega, \mathcal{F}, P)$.

__Teorema__ I.51 - __Se__ $(X_n)_{n \geq 1}$ __é um__ MG __em__ $L^1(\Omega, \mathcal{F}, P)$ __então para tôdo__ $a > 0$ __e__ $k \geq 1$ __vale:__

1) $$a P\left[\max_{1 \leq j \leq k} X_j \geq a\right] \leq \int_{\left[\max_{1 \leq j \leq k} X_j \geq a\right]} X_k \, dP \leq E|X_k|$$

2) $$a P\left[\min_{1 \leq j \leq k} X_j < -a\right] \leq (E|X_k| - E X_1)$$

Seja

$$A_1 = [X_1 \geq a], \quad A_i = [X_j < a, \ 1 \leq j < i \ ; \ X_i \geq a], \ i > 1$$

Então os A_i são determinados pelas i primeiras variáveis aleatórias $X_1, \dots, X_i$ do MG e

$$A = \left[\max_{1 \leq j \leq k} X_j \geq a\right] = \sum_{j=1}^{k} A_j$$

Então

$$\int_A X_k \, dP = \sum_{j=1}^{k} \int_{A_j} X_k \, dP \overset{(MG')}{=\!=} \sum_{j=1}^{k} \int_{A_j} X_j \, dP \geq a \sum_{j=1}^{k} P(A_j) =$$

$$= a\, P(A)$$

e a primeira desigualdade de 1) está demonstrada; a segunda desigualdade é óbvia.

Para a desigualdade 2), tomemos

$$m_1 = 1$$

$$m_2 = \begin{cases} k & \text{se } X_j \geq -a \text{ para } j = 1, \ldots, k \\ \inf \{ j ; X_j < -a \} & \text{caso contrário} \end{cases}$$

$$m_n = k \quad n \geq 3$$

É imediato que os m_n são variáveis amostrais; tomando

$$\tilde{X}_n = X_{m_n}$$

as condições (a) e (b) do Teorema I.48 são fàcilmente verificadas, e o nôvo processo $(\tilde{X}_n)_{n \geq 1}$ é um MG . Então:

$$E X_{m_2} = \int_{[X_{m_2} \geq -a]} X_{m_2} \, dP + \int_{[X_{m_2} < -a]} X_{m_2} \, dP \leq E|X_k| - a\, P[X_{m_2} < -a]$$

Mas, de

$$P\left[\min_{1 \leq j \leq k} X_j < -a \right] = P\left[X_{m_2} < -a \right] \qquad \text{segue}$$

$$P\left[\min_{1 \leq j \leq k} X_j < -a \right] \leq \frac{1}{a} \left(E|X_k| - E X_{m_2} \right)$$

e como

$$E X_{m_2} \overset{(MG')}{=\!=} E X_{m_1} = E X_1 ,$$

vem

$$P\left[\min_{1 \leq j \leq k} X_j < -a \right] \leq \frac{1}{a} \left(E|X_k| - E X_1 \right) \qquad \text{c.q.d.}$$

Vamos citar um resultado mais preciso, devido a Doob [Cap.VII

§ 3) ; a importância dos teoremas precedente e seguinte está no fato de que o número k de variáveis aleatórias envolvidas não importa, podendo ser inclusive infinito, contanto que haja o último elemento do martingal.

<u>Teorema</u> I.52 - <u>Se</u> $(X_j)_{1 \leq j \leq k}$ <u>é um</u> MG <u>de variáveis aleatórias não negativas, então</u>

$$E\left[\max_{1 \leq j \leq k} X_j^{\alpha}\right] \leq \frac{e}{e-1} + \frac{e}{e-1} E\left[X_k \log^+ X_k\right], \quad \text{se} \quad \alpha = 1$$

$$\leq \frac{\alpha}{\alpha - 1} E\left[X_k^{\alpha}\right], \quad \text{se} \quad \alpha > 1 .$$

onde: e = base do logaritmo neperiano

$\log^+ Z = \text{máx}\,(0, \log Z)$

Note que o coeficiente na segunda desigualdade é função decrescente de α, que tende ao número $\underline{e}$ para $\alpha \longrightarrow \infty$.

Para estudar a convergência dos martingais em L^1 vamos introduzir o conceito de <u>processo truncado</u>.

Sendo $m_1, m_2, \ldots$ uma seqüência de variáveis amostrais associada a um processo dado $(X_n)_{n \geq 1}$, podemos considerar o processo truncado em um dado índice M ($M \geq 1$ fixado) e obtido por amostragem opcional, tomando:

$$m_{n,M} = \min\,(M, m_n)$$
$$\tilde{X}_{n,M} = X_{m_{n,M}} \tag{I.53}$$

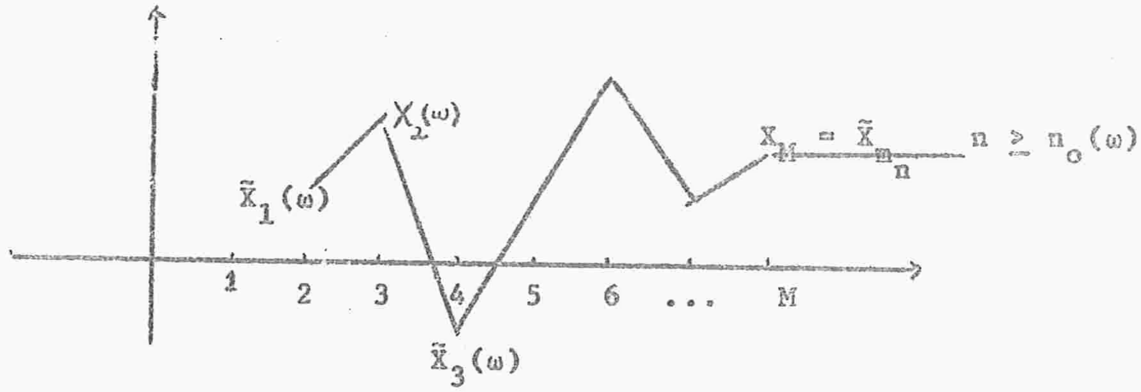
Se

$$\Omega_{n,M} = \left(\omega \in \Omega \,;\, m_n(\omega) \geq M\right)$$

então

$$\tilde{X}_{n,M} = X_M \quad \text{em} \quad \Omega_{n,M}$$

isto é, $\tilde{X}_{n,M}$ é um processo que pára em X_M ; para cada ponto amostral $\omega \in \Omega$ fixado, $\tilde{X}_n(\omega)$ como função de n é <u>constante</u> para $n \geq M$:



Vamos tomar variáveis amostrais e o MG truncado em M e obtido por amostragem opcional, fazer $M \to \infty$ e analisar a convergência.

As variáveis amostrais são:

Dados $a < b$ reais fixados:

$$m_1 = \min \{ n \,;\, X_n \leq a \} \text{ se tal existe, } \infty \text{ caso contrário}$$

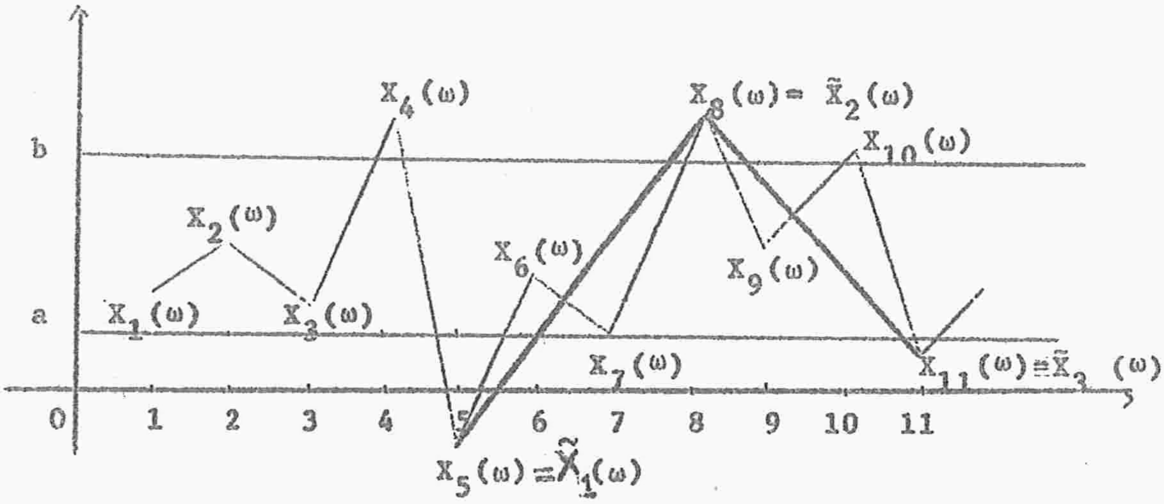
$$m_2 = \min \{ n \geq m_1 \,;\, X_n \geq b \} \text{ se tal existe, } \infty \text{ caso contrário} \tag{I.54}$$

Em geral :

$$m_{2n+1} = \min \{ n \geq m_{2n} \,;\, X_n \leq a \} \text{ se tal existe, } \infty \text{ caso contrário}$$

$$m_{2n+2} = \min \{ n \geq m_{2n+2} \,;\, X_n \geq b \} \text{ se tal existe, } \infty \text{ caso contrário}$$

Para cada $\omega \in \Omega$, os $m_n(\omega)$ são os índices sucessivos dos vértices da poligonal $X_1(\omega)\, X_2(\omega) \ldots X_n(\omega) \ldots$, que não pertencem ao interior da faixa determinada por a e b, sendo que, para n ímpar, m_n indica vértice abaixo da reta de ordenada $\underline{a}$ e, para n par, m_n indica vértice acima da reta de ordenada $\underline{b}$.



Se m_{2P} é finito então p indica o número de vêzes que a poligonal $X_1 X_2 \ldots X_p$ cruza efetivamente no sentido ascendente a faixa determinada por a e b. Como $\{m_n = j\}$ só depende das variáveis aleatórias $X_1, \ldots, X_j$, temos uma seqüência de variáveis amostrais.

<u>Teorema</u> I.55 - <u>Se</u> $(X_n)_{n \geq 1}$ <u>é um</u> MG <u>tal que</u> $\overline{\lim_n}\, E|X_n| < \infty$ <u>então o conjunto dos pontos</u> ω <u>tais que</u> $(X_n(\omega))_{n \geq 1}$ <u>é uma seqüência oscilante tem probabilidade nula.</u>

Seja

$$\tilde{X}_{n,M} = X_{m_{n,M}}$$

o processo truncado obtido por amostragem opcional conforme (I.47) sendo as variáveis amostrais dadas por (I.54).

Para todo $n \geq 1$ vale:

$$E\,|\tilde{X}_{n,M}| = \int_{\{m_{n,M} < M\}} |\tilde{X}_{n,M}|\, dP + \int_{\{m_{n,M} \leq M\}} |\tilde{X}_{n,M}|\, dP \leq \sum_{j=1}^{M-1} E|X_j| +$$

$$+ E\,|X_M| = \sum_{j=1}^{M} E\,|X_j| < \infty$$

e como

$$\int_{\{m_{n,M} > N\}} |X_N|\, dP = \int_{\{m_{n,M} > N\}} |X_M|\, dP$$

concluimos que:

$$\lim_{N} \int_{\{m_{n,M} > N\}} |X_M| \, dP = \lim_{N} \int_{\{m_{n,M} > N\}} |X_M| \, dP = 0$$

e pelo Teorema I.48 $(\tilde{X}_{n,M})_{n \geq 1}$ é um MG para todo $M > 0$.

Seja:

$$Z_M = (\tilde{X}_{3,M} - \tilde{X}_{2,M}) + (\tilde{X}_{5,M} - \tilde{X}_{4,M}) + \ldots$$

Na realidade Z_M é uma soma finita pois, para $m_{2n} \geq M$ (o que se dá certamente se $2n \geq M$), se tem:

$$\tilde{X}_{2n+1,M} - \tilde{X}_{2n,M} = X_M - X_M = 0$$

Seja

$\beta_M = \max \{n ; m_{2n} < M\} =$ número de cruzamentos ascendentes da poligonal $X_1 \ldots X_M$

Então se k é inteiro não negativo, no conjunto

$$\{ \beta_M = k \}$$

temos

$$Z_M = (\tilde{X}_{3,M} - \tilde{X}_{2,M}) + \ldots + (\tilde{X}_{2k+1,M} - \tilde{X}_{2k,M}) = (\tilde{X}_{3,M} - \tilde{X}_{2,M}) + \\ + \ldots + (a - \tilde{X}_{2k,M}) + (\tilde{X}_{2k+1,M} - a)$$

O último têrmo acima só é positivo se

$$m_{2k+1} \geq M$$

e neste caso é igual a $(X_M - a)$

Em qualquer caso, sôbre $\{ \beta_M = k \}$ temos:

$$Z_M \leq - k(b - a) + (X_M - a)^+$$

ou, em geral:

$$Z_M \leq - \beta_M(b - a) + (X_M - a)^+ .$$

Tomando esperanças:

$$E\, Z_M \leq - (b - a)\, E\, \beta_M + E(X_M - a)^+ \qquad (I.56)$$

Como $(\tilde{X}_{n,M})_{n \geq 1}$ é um MG, temos

$$\tilde{X}_{n,M} = E\, (\tilde{X}_{n+1,M} \mid \tilde{X}_{1,M}, \ldots, \tilde{X}_{n,M})$$

logo,

$$E\,\tilde{X}_{n,M} = E\,[\,E(X_{n+1,M} \mid X_{1,M}, \ldots, X_{n,M})\,] = E\,\tilde{X}_{n+1,M}$$

para todo $n \geq 2$, e então $E\,Z_M = 0$ e da condição (I.56) resulta:

$$E\,\beta_M \leq \frac{E.(X_M - a)^+}{b - a} \qquad (I.57)$$

O número de cruzamentos ascendentes da poligonal cujo último elemento é X_M é uma variável aleatória β_M, que satisfaz à condição (I.57).

Vejamos o que acontece quando $M \longrightarrow \infty$

É claro que

$$\lim_{M\uparrow\infty} \uparrow m_{n,M} = m_n \;.$$

Então, se $\tilde{X}_n = X_{m_n}$, segue $\tilde{X}_n = \lim_{M\longrightarrow\infty} \tilde{X}_{n,M}$ e

vamos mostrar que $(\tilde{X}_n)_{n\geq 1}$ é ainda um MG : para todo $n \geq 1$:

$$E\,|\tilde{X}_n| = \int \lim_{M\to\infty} |\tilde{X}_{n,M}|\, d\,P \underset{(\text{Fatou})}{\leq} \underline{\lim}_{M\to\infty} E\,|\tilde{X}_{n,M}| \leq 2 \underline{\lim}_{N\to\infty} \overline{\lim}_{n\to\infty} E|X_n| - E\,X_1 \underset{(\text{hipótese})}{<} \infty$$

Como para todo $M \geq 1$ temos

$$m_{n,M} \leq m_n$$

vale

$$[\,m_n > N\,] \subset [\,m_{n,M} > N\,]$$

logo

$$\int_{[m_n > N]} |X_N|\; dP \leq \int_{[m_{n,M} > N]} |X_N|\; d\,P$$

e por conseguinte

$$\lim_{N\longrightarrow\infty} \int_{[m_n > N]} |X_N|\; dP = 0 \qquad \text{para todo } n \geq 1 \;,$$

e o Teorema I.48 assegura que $\tilde{X}_n$ é um MG .

Então o número total de cruzamentos ascendentes da poligonal $X_1 X_2 \dots$ é igual a

$$\beta_\infty = \lim_{M \to \infty} \beta_M ,$$

para o qual ainda vale (I.57), onde se faz $M \to \infty$.

Seja

$$S_{a,b} = \bigcap_{n=1}^{\infty} [m_n < \infty] = [\omega \in \Omega ; \text{ a poligonal } X_1(\omega)X_2(\omega)\dots$$
não está contida no interior da faixa determinada por a e b]

Por um lado $\beta_\infty = \infty$ sôbre $S_{a,b}$ e por outro lado,

$$E\beta_\infty \le \frac{1}{b-a} \overline{\lim_M} E(X_M - a)^+ \le \frac{1}{b-a} \left[\overline{\lim_M} E |X_M| + a \right] < \infty$$

o que só é possível se $P(S_{a,b}) = 0$.

Como podemos repetir o raciocínio para cada par $a < b$ de racionais, temos

$$P[\omega ; (X_n(\omega))_{n \ge 1} \text{ é uma seqüência oscilante}] = P\left(\bigcup_{\substack{a,b \in Q \\ a < b}} S_{a,b} \right) = 0 \qquad \text{(I.58)}$$

c.q.d.

Teorema I.59 - (<u>da convergência dos martingais em</u> L^1)

<u>Se</u> $(X_n)_{n \ge 1}$ <u>é um</u> MG <u>uniformemente integrável</u> (vide definição I.11), <u>existe uma variável aleatória integrável</u> X <u>tal que</u>

$$X_n \xrightarrow[L^1]{q.c.} X$$

<u>Ainda, vale</u>

$$X_n = E(X | X_1, \dots, X_n) \qquad (n \ge 1)$$

Como

$$E|X_n| = \int_{[|X_n| \le x]} |X_n| \, dP + \int_{[|X_n| > x]} |X_n| \, dP \le x + \int_{[|X_n| > x]} |X_n| \, dP ,$$

temos:

$$\overline{\lim_{n\to\infty}}\ E|X_n| \leq x + \overline{\lim_{n\to\infty}} \int_{[|X_n|>x]} |X_n|\ d\,P < \infty \text{ (integrabilidade uniforme)} \tag{I.60}$$

Pelo teorema I.55, deve ocorrer uma das alternativas:

A) existe uma variável aleatória X tal que $X_n \to X$ q.c.

B) $|X_n| \to \infty$ com probabilidade positiva.

O caso B) está excluído pelo lema de Fatou, pois, teríamos:

$$\infty = \int \underline{\lim_n} |X_n|\ dP \leq \underline{\lim_n} \int |X_n|\ d\,P \leq \overline{\lim_n}\ E|X_n|$$

em contradição com (I.60). Ainda:

$$\int |X|\ dP \leq \underline{\lim_n} \int |X_n|\ d\,P < \infty$$

logo X é integrável. Note que a condição $\overline{\lim_{n\to\infty}}\ E|X_n| < \infty$ mais fraca que a integrabilidade uniforme é suficiente para a convergência q.c.

Quanto à convergência em L^1: pelo teorema I.1 de Egoroff, dado $\delta > 0$ existe $A \in \mathcal{J}$ tal que $P(A) < \delta$ e

$$V_n = |X_n - X| \longrightarrow 0 \quad \text{uniformemente sôbre } A^c,$$

e

$$\overline{\lim_n}\ E\,V_n = \overline{\lim_n} \int_A |X_n - X|\ d\,P \leq \overline{\lim_n} \int_A |X_n|\ d\,p + \int_A |X|\ d\,p.$$

Substituindo

$$\int_A |X_n|\ d\,p$$

por

$$\int_A |X_n|\ d\,p = \int_{A\cap[|X_n|\leq x]} |X_n|\ d\,p + \int_{A\cap[|X_n|>x]} |X_n|\ d\,p \leq$$

$$\leq x\,\delta + \int_{[|X_n|>x]} |X_n|\ d\,p,$$

e também $\int_A |X|\ d\,P$ por uma expressão análoga, e fazendo $\delta \to 0$, obtemos

$$\overline{\lim_n}\ E\,V_n \leq \overline{\lim} \int_{[|X_n|>x]} |X_n|\ d\,P + \int_{[|X|>x]} |X|\ d\,P$$

Tomando limite para $x \uparrow \infty$, a integrabilidade uniforme de $(X_n)_{n \geq 1}$ e a integrabilidade de X nos dão

$$\overline{\lim}\, E\,|X_n - X| = 0, \quad \text{ou} \quad X_n \xrightarrow{L^1} X$$

Quanto à última parte, temos por hipótese

$$(\text{MG}') \qquad \int_A X_n \, dP = \int_A X_m \, dP \qquad \forall\, m > n\,, \ \forall\, A \in \mathcal{B}_n(X_1,\ldots,X_n)$$

Pelo já visto, $X_n \xrightarrow{L^1} X$, logo

$$\int_A X_n \, dP = \int_A X \, dP$$

para todo

$$A \in \mathcal{B}_n\,(X_1, \ldots, X_n) \quad ,$$

logo

$$X_n = E\,(X|X_1, \ldots, X_n) \qquad , \quad \text{c.q.d.}$$

Podemos então dizer que <u>todo</u> MG <u>uniformemente integrável é uma seqüência de esperanças condicionais</u>.

<u>Corolário</u> I.61 - <u>Seja</u> $Y_1, Y_2, \ldots$ <u>um processo sôbre</u> $(\Omega, \mathcal{F}, P)$ <u>e</u> Z <u>uma variável aleatória integrável</u>. <u>Então</u> :

$$E\,(Z|Y_1, \ldots, Y_n) \xrightarrow[L^1]{\text{q.c.}} E\,(Z|Y_2, Y_2, \ldots)$$

Seja
$$X_n = E\,(Z|Y_1, \ldots, Y_n) \ .$$

Então, como

$$|X_n| = |\,E(Z\,|\,Y_1,\ldots,Y_n)| \underset{(\text{I.40, n95})}{\leq} E\,(|Z|\,|Y_1,\ldots,Y_n) \ ,$$

vale

$$E\,|X_n| \leq E\,|Z| < \infty \qquad (\forall n \geq 1) \ .$$

De

$$\mathcal{B}_n(X_1, \ldots, X_n) \subset \mathcal{B}_n(Y_1, \ldots, Y_n)$$

segue

$$E(X_{n+1} | X_1,\dots,X_n) = E\{E[E(Z|Y_1,\dots,Y_n, Y_{n+1}) | Y_1,\dots,Y_n] | X_1,\dots,X_n\} =$$

$$= E[E(Z|Y_1,\dots,Y_n) \mid X_1,\dots,X_n] = E(X_n|X_1,\dots,X_n) = X_n$$

donde concluimos que $(X_n)_{n\geq 1}$ é um MG tal que $\overline{\lim_n} E|X_n| < \infty$. Pe- demonstração da primeira parte do teorema I.59 concluímos que

$$X_n \xrightarrow{q.c.} X$$

e

$$E|X| \leq \overline{\lim_n}\ E|X_n| \leq E|Z|$$

Ainda,

$$\int_{\{|X_n| > x\}} |X_n|\, dP \leq \int_{\{|X_n| > x\}} E(|Z| \,|Y_1,\dots,Y_n)\, dP =$$

$$= \int_{\{|X_n| > x\}} |Z|\, dP \leq \int_{\{\sup_n |X_n| > x\}} |Z|\, dP$$

Seja agora

$$U = \sup |X_n| < \infty \quad q.c.$$

Como

$$\lim_{x\uparrow\infty} P\{U > x\} = 0 \quad o \quad MG \quad (X_n)_{n\geq 1}$$

é uniformemente integrável, logo pelo Teorema I.59 $X_n \xrightarrow{L^1} X$

Então para todo $A \in \mathcal{B}_N(Y_1,\dots,Y_N)$ temos

$$\lim_{n\to\infty} \int_A X_n\, dP = \int_A X\, dp$$

Mas para $n \geq N$,

$$\int_A X_n\, dP = \int_A Z\, dP ,$$

logo

$$Z = X \qquad q.c.$$

c.q.d.

CAPÍTULO II - APLICAÇÕES

II.1 - Uma recíproca do Teorema da Convergência Dominada de Lebesgue.

Seja dado um processo $(X_n)_{n \geq 1}$ sôbre $(\Omega, \mathcal{F}, P)$ de variáveis aleatórias não negativas integráveis e suponhamos que $X_n \xrightarrow{q.c.} X \in L^1(\Omega, \mathcal{F}, P)$; seja $U = \sup_n X_n$. O teorema I.2 da convergência dominada de Lebesgue assegura que, se $U \in L^1(\Omega, \mathcal{F}, P)$ então $X_n \xrightarrow{L^1} X$, isto é $E|X_n - X| \to 0$, ou, como

$$|E X_n - E X| \leq E|X - X_n| ,$$

$$E X_n \longrightarrow E X .$$

Mais geralmente, se $\mathcal{D}_o$ é qualquer sub-σ-álgebra de $\mathcal{F}$, então pela propriedade (I.40) nº 6 temos

$$E(X_n | \mathcal{D}_o) \longrightarrow E(X | \mathcal{D}_o) \quad \text{q.c.} \qquad (II.1)$$

O seguinte teorema mostra que a condição $U \in L^1$ não só é suficiente, mas também necessária, num sentido algo mais amplo, para a ocorrência de (II.1).

<u>Teorema</u> II.2 - <u>Se</u> $X_n \geq 0$, $X_n \longrightarrow X$ q.c., $X_n, X \in L^1$ <u>mas</u> $U = \sup_n X_n \notin L^1$, <u>então existem, em um espaço de probabilidade conveniente, variáveis aleatórias</u> $X^*, X_1^*, X_2^*, \ldots$ <u>com as mesmas distribuições de dimensões finitas que</u> $X, X_1, X_2, \ldots$ <u>e uma</u> σ <u>-álgebra</u> $\mathcal{C}$ <u>tais que</u>

$$P\left[E(X_n^* | \mathcal{C}) \longrightarrow E(X^* | \mathcal{C}) \right] = 0 \qquad (II.3)$$

Vamos reduzir o teorema ao caso especial em que cada X_n assume sòmente dois valores, 0 e $v_n > 0$ e em cada ponto amostral exatamente um X_n é positivo. Vamos então demonstrar o teorema sob as seguintes condições:

Sendo $p_n = P[X_n = v_n]$, $v_n > 0$ $(n \geq 1)$,

$$0 < p_n < 1 , \quad \sum_{n \geq 1} p_n = 1 , \; X \equiv 0 , \quad E U = E \sup_n X_n = \sum_{n \geq 1} p_n v_n = \infty$$

Para chegarmos a esta simplificação, vamos decompôr:

$$X_n - X = (X_n - X)^+ - (X_n - X)^- \ .$$

Notando que

$$\sup_n (X_n - X)^+ \geq U - X \notin L^1 \ , \quad (X_n - X)^+ \longrightarrow 0 \ \text{q.c.}$$

$$\sup (X_n - X)^- \leq X \in L^1 \ , \quad (X_n - X)^- \longrightarrow 0 \ \text{q.c.}$$

temos, para qualquer σ-álgebra $\mathcal{C}$, $E((X_n - X)^- \mid \mathcal{C}) \to 0$ q.c., logo

$$P\left[E(X_n \mid \mathcal{C}) \to E(X \mid \mathcal{C})\right] = P\left[E(X_n - X \mid \mathcal{C}) \to 0\right] =$$

$$= P\{E\left[(X_n - X)^+ - (X_n - X)^- \mid \mathcal{C}\right] \to 0\} = P\{E\left[(X_n - X)^+ \mid \mathcal{C}\right] \to 0\}$$

e o teorema estárá provado se mostrarmos que

$$P\{E\left[(X_n - X)^+ \mid \mathcal{C}\right] \to 0\} = 0$$

para alguma σ-álgebra $\mathcal{C}$ em um espaço de probabilidade conveniente, e assim já reduzimos o problema ao caso da variável aleatória limite $X \equiv 0$.

Usando ainda a mesma notação do enunciado do teorema, com $X \equiv 0$, seja a variável aleatória com valores inteiros:

$$\nu = \min\{n \ ; \ X_n \geq U - 1\}$$

Então, se

$$A_k = [\nu = k] \ , \qquad k \geq 1 \ \text{inteiro}$$

vale:

$$\Omega = \sum_{k \geq 1} A_k \ , \qquad \sum_{k \geq 1} P(A_k) = 1$$

Para cada k, tomemos uma função simples s_k nula fora de A_k, tal que

$$0 \leq s_k \leq X_k$$

e

$$E(s_k) \geq \int_{A_k} X_k \, dP - \frac{1}{2^k}$$

Então

$$\sup_{k \geq 1} s_k = \sum_{k \geq 1} s_k ,$$

logo

$$E \sup_{k \geq 1} s_k = \sum_{k \geq 1} E s_k \geq \sum_{k \geq 1} \int_{A_k} X_k \, dP - 1 \geq$$

$$\geq \sum_{k \geq 1} \int_{A_k} U \, dP - 1 - 1 = E U - 2 = \infty$$

Como $s_k \leq X_k$, para qualquer σ -álgebra C a condição

$$P \{ E(s_k | C) \to 0 \} = 0$$

implica

$$P \{ E(X_k | C) \longrightarrow 0 \} = 0 .$$

E assim o teorema se reduz ao caso em que $X \equiv 0$, $X_k = s_k$.= função simples, $k \geq 1$ e como $\sum_{k \geq 1} A_k = \Omega$, em cada ponto amostral no máximo um X_k é positivo.
Podemos ordenar novamente estas funções, excluindo as que são q.c. nulas, e se o conjunto A_0 em que tôdas essas funções se anulam tiver probabilidade positiva, podemos acrescentar o indicador I_{A_0} à seqüência e assim reduzimos o teorema ao caso de uma seqüência de variáveis aleatórias de L^1, indicadas ainda com a mesma notação $X_1, X_2, \ldots$, tal que $X_n \to 0$ q.c., cada variável aleatória X_n assume sòmente dois valores 0 e $v_n > 0$, e em cada ponto amostral sòmente uma delas é positiva. Ainda, se $p_n = P\{X_n = v_n\}$, temos $0 < p_n < 1$ e

$$\sum_{n \geq 1} p_n = 1 , \quad e \quad E U = E \sup_{n \, 1} X_n = \sum_{n \geq 1} p_n v_n = \infty \qquad \text{(II.4)}$$

Pelo exposto acima, o teorema estará demonstrado se o tivermos demonstrado com o seguinte enunciado:

Dada uma seqüência de variáveis aleatórias $(X_n)_{n \geq 1}$ que assumem sòmente dois valores 0 e $v_n > 0$, sendo que em cada ponto amostral sòmente um X_n é positivo, e satisfazendo à condição $\sum_{n \geq 1} p_n v_n = \infty$ onde $p_n = P\{X_n = v_n\}$ $(n \geq 1)$, existem, em um espaço de probabilidade conveniente, variáveis aleatórias X_n^*, $n \geq 1$ com as mesmas distribuições conjuntas das variáveis aleatórias dadas, e uma σ-álgebra C tal que:

$$P\left[E(X_n^* \mid C) \to 0\right] = 0$$

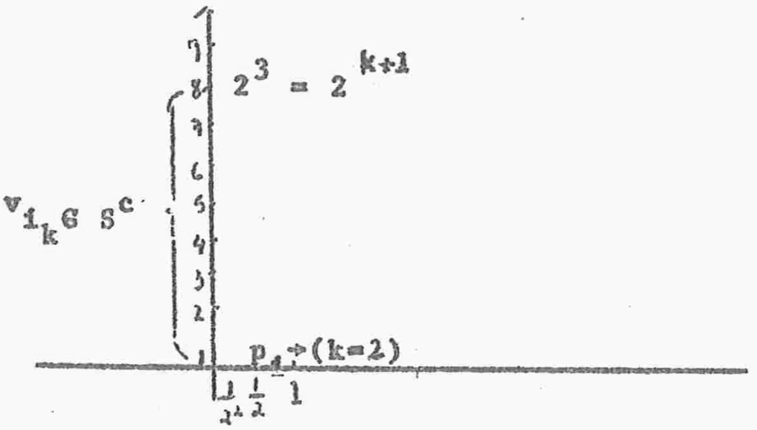
<u>Lema</u> II.4 - Dada a seqüência $(p_n)_{n \geq 1}$ de números reais com $0 < p_n < 1$ $(n \geq 1)$ $\sum_{n \geq 1} p_n = 1$ e a seqüência de números positivos $(v_n)_{n \geq 1}$, existe uma variável aleatória Z com valores inteiros tal que

$$P\left[\, Z = n \,\right] = p_n \;, \qquad n \geq 1 \;.$$

Como $0 < p_1 < 1$ existe o menor inteiro k tal que $\dfrac{1}{2^k} < p_1 \leq \dfrac{1}{2^{k-1}}$, ou $1 < 2^k p_1 \leq 2$.

Para cada $n \geq 1$ vamos indicar por S_n o conjunto dos índices $i \geq 2$, tais que

$$2^{k+n} \leq v_i < 2^{k+n+1} \qquad \text{(II.5)}$$



Seja

$$r_n = \sum_{i \in S_n} p_i \;, \quad t_n = r_n + 2^{-(n+k)} \;, \quad t = \sum_{n \geq 1} t_n = r + 2^{-k} \qquad \text{(II.6)}$$

$$r = \sum_{n \geq 1} r_n = \sum_{i \in S} p_i \;, \quad \text{onde } S = \bigcup S_n = \{\, i \;;\; i \geq 2 \text{ e } v_i \geq 2^{k+1} \}$$

Vamos definir as variáveis aleatórias independentes com valores inteiros ν, Z_0, Z_1, ... pelas suas distribuições, estendendo o espaço de probabilidade se necessário, como segue:

$$P\left[\nu = 0\right] = 1 - t \;;\; P\left[\nu = n\right] = t_n \;(n \geq 1) \qquad \text{(II.7)}$$

$$P\left\{Z_o = 1\right\} = \frac{p_1 - 2^{-k}}{1 - t} \quad ; \quad P\left\{Z_o = i\right\} = \begin{cases} \dfrac{p_i}{1 - t} & \text{se } i \geq 2 \text{ , } i \notin S \\ 0 & \text{caso contrário} \end{cases}$$

para $n \geq 1$:

$$P\left\{Z_n = 1\right\} = \frac{2^{-(k+n)}}{t_n} \quad ; \quad P\left\{Z_n = i\right\} = \begin{cases} \dfrac{p_i}{t_n} & \text{se } i \in S_n \\ 0 & \text{caso contrário} \end{cases} \tag{II.7}$$

As fórmulas (II.7) definem distribuições de probabilidade pois, notando que S^c, S_1, S_2, ... formam uma partição do conjunto dos índices $\mathbb{N}$ e considerando as fórmulas (II.6), temos:

$$P\left(\nu = 0\right) + \sum_{n \geq 1} P\left(\nu = n\right) = 1 - t + \sum_{n \geq 1} t_n = 1$$

$$P\left(Z_o = 1\right) + \sum_{i \geq 2} P\left(Z_o = i\right) = \frac{p_1 - 2^{-k}}{1 - t} + \sum_{\substack{i \geq 2 \\ i \notin S}} \frac{p_i}{1 - t} =$$

$$= \frac{1}{1 - t}\left\{p_1 - 2^{-k} + \sum_{\substack{i \geq 2 \\ i \notin S}} p_i\right\} = \frac{1}{1 - t}\left[1 - \sum_{i \in S} p_i - 2^{-k}\right] =$$

$$= \frac{1 - r - 2^{-k}}{1 - t} = \frac{1 - t}{1 - t} = 1 \tag{II.8}$$

$$P\left(Z_n = 1\right) + \sum_{i \geq 2} P\left(Z_n = i\right) = \frac{2^{-(k+n)}}{t_n} + \sum_{\substack{i \geq 2 \\ i \in S_n}} \frac{p_i}{t_n} =$$

$$= \frac{1}{t_n}\left[2^{-(k+n)} + \sum_{\substack{i \geq 2 \\ i \in S_n}} p_i\right] = \frac{1}{t_n}\left[2^{-(k+n)} + r_n\right] = \frac{t_n}{t_n} = 1$$

Então por construção as distribuições de probabilidade conjuntas das variáveis aleatórias ν, Z_o, Z_1, são dadas pela condição de independência:

$$P\left[\nu = k,\ Z_j = i_j\ ,\ 0 \leq j \leq n\right] = P\left[\nu = k\right] \prod_{0 \leq j \leq n} P\left[Z_j = i_j\right]$$

para quaisquer inteiros k , i_j , $0 < j \leq n$, para todo $n \geq 0$.

Vamos definir a variável aleatória Z pela técnica da amostragem opcional, em que a variável amostral é ν e a seqüência dada é $Z_o, Z_1, \ldots$. Se $\nu(\omega) = k$ então $Z = Z_\nu$ é tal que

$Z(\omega) = Z_{\nu(\omega)}(\omega) = Z_k(\omega)$. Temos:

$$P[Z_\nu = 1] = \sum_{n=0}^{\infty} P[\nu = n, Z_n = 1] \ \underline{\underline{(II.8)}} \ p_1 - 2^{-k} + \sum_{n\geq 1} 2^{-(k+n)} = p_1$$

Para $i > 1$, $i \notin S$,

$$P\{Z_\nu = i\} = P\{\nu = 0, Z_0 = i\} \ \underline{\underline{(II.8)}} \ p_i$$

Para $i \in S_n$,

$$P\{Z_\nu = i\} = P\{\nu = n, Z_n = i\} \ \underline{\underline{(II.8)}} \ p_i \qquad \text{c.q.d.}$$

Demonstração do teorema: Vamos definir X_n^*, $n \geq 1$ por:

$$X_n^* = \begin{cases} v_n & \text{sôbre } [Z_\nu = n] \\ 0 & \text{fora de } [Z_\nu = n] \end{cases}$$

e em cada ponto amostral sòmente uma X_n^* é positiva.

Então $X_1^*, X_2^*, \ldots$ tem as mesmas distribuições de dimensões finitas que $X_1, X_2 \ldots$ e para qualquer σ-álgebra $\mathcal{C}$ temos:

$$E\left[(X_n^* \mid \mathcal{C}\right] = v_n P[Z_\nu = n \mid \mathcal{C}]$$

Basta então construir uma σ-álgebra $\mathcal{C}$ tal que o evento

$$B = \{v_i P[Z_\nu = i \mid \mathcal{C}] \geq 1 \ \text{ infinitas vêzes}$$

tenha probabilidade 1.

Seja a σ-álgebra $\mathcal{B} = \mathcal{B}_\infty(Z_0, Z_1 \ldots)$ gerada pelas variáveis aleatórias $Z_0, Z_1, \ldots$ Os eventos $[Z_n = k]$, $n \geq 0$, $k \geq 1$ formam uma partição enumerável de Ω que gera a σ-álgebra $\mathcal{C}$.

Então para cada $n \geq 1$ fixado e sendo $i \in S_n$ temos; pelas fórmulas (II.7) e tendo em vista (I.39''):

$$P(Z_\nu = i \mid \mathcal{C}) = \begin{cases} 0 & \text{sôbre } [Z_n \neq i] \\ \dfrac{p_i}{\left(\dfrac{p_i}{t_n}\right)} = t_n & \text{sôbre } [Z_n = i] \end{cases}$$

Como $t_n \geq 2^{-(n+k)}$ e como para $i \in S_n$ temos $v_i \geq 2^{n+k}$, concluimos que para todo $i \in S_n$ e $n \geq 1$ vale

$$v_i \, P\left[Z \quad = i \,|\, C \right] = v_i \, t_n \geq 1 \qquad \text{sôbre } \left[Z_n = i \right] .$$

Para $n \geq 1$, sempre que ocorre $A_n = \left[Z_n \neq 1 \right]$ também ocorre

$$B_n = \left(v_i \, P \,|\, Z_{\,} = i \,|\, C \right) \geq 1 \ , \text{ para algum } i \in S_n \Big] , \quad n \geq 1 .$$

Os eventos A_n são independentes,

$$P(A_n) = 1 - \frac{2^{-(k+n)}}{t_n} = \frac{t_n - 2^{-(k+n)}}{t_n} = \frac{r_n}{t_n}$$

e

$$\sum_{n \geq 1} P(A_n) = \sum_{n \geq 1} \frac{r_n}{t_n}$$

Para mostrar que $B = \left[B_n \text{ i.v.} \right]$ tem probabilidade 1 basta mostrar que $A = \left[A_n \text{ i.v.} \right]$ tem probabilidade 1, e por Borel Cantelli (I.33), para isto é suficiente mostrar que a série $\sum_{n \geq 1} r_n/t_n$ diverge.

De (II.5) e (II.6), vem:

$$\text{Se} \quad r_n < 2^{-(n+k)} \ , \quad t_n < 2^{-(n+k-1)} \ , \ \text{e}$$

$$\frac{r_n}{t_n} \geq 2^{n+k-1} \, r_n = \sum_{i \in S_n} 2^{n+k+1} \, \frac{p_i}{4} \geq \sum_{i \in S_n} \frac{p_i \, v_i}{4}$$

Logo, se $r_n \geq 2^{-(n+k)}$ para infinitos índices n, então $\frac{r_n}{t_n} \geq \frac{1}{2}$ para infinitos n, e a série $\sum_{n \geq 1} \frac{r_n}{t_n}$ diverge.

Se $r_n < 2^{-(n+k)}$ para $n \geq n_o$,

$$\sum_n \frac{r_n}{t_n} \quad \sum_{n \geq n_o} \ \sum_{i \in S_n} \frac{p_i \, v_i}{4} = \sum_{i \in T} \frac{p_i \, v_i}{4}$$

onde

$$T = \{ i \geq 2, \ v_i \geq 2^{\,n_o + k} \} . \quad \text{Como}$$

$$\sum_{i\geq 1} p_i v_i = \infty \qquad e \qquad \sum_{i \notin T} p_i v_i \leq 2^{n_o+k} \sum_{i \notin T} p_i \leq 2^{n_o+k} \sum_{i\geq 1} p_i = 2^{n_o+k}$$

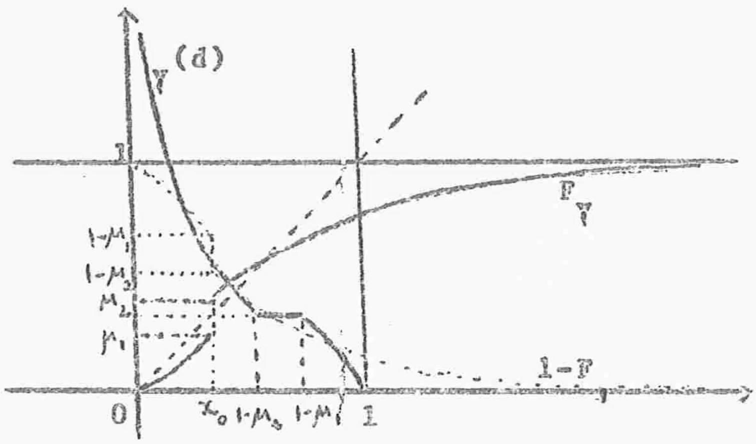
segue que

$$\sum_{i \in T} p_i v_i = \infty \quad , \text{ logo } \sum_{n \geq 1} \frac{r_n}{t_n} = \infty \qquad \text{c.q.d.}$$

II.2 - Uma aplicação do conceito de rearranjo decrescente de uma variável aleatória.

Encontramos em um artigo de G.H.Hardy e J.E. Littlewood [A maximal theorem with function - theoretic applications, Acta Math., vol. 54 (1930), pg. 81-116] as propriedades do rearranjo decrescente de uma variável aleatória não negativa limitada.

Seja Y uma variável aleatória não negativa limitada sôbre $(\Omega, \mathcal{F}, P)$ e $F_Y(x) = P\{Y < x\}$ a sua função de distribuição; os pontos de descontinuidade (saltos) de $F_Y(x)$ são enumeráveis.



Então (vide figura acima)

$$1 - F_Y(x) = P\{Y \geq x\}$$

é uma função cujo conjunto de pontos de descontinuidade é enumerável, que decresce de 1 a 0 .

A função inversa $Y^{(d)}(x) = \left[1 - F_Y(x) \right]^{-1}$ definida sôbre $[0,1]$ por

$$Y^{(d)} \left[1 - F_Y(x) \right] = x$$

é uma função decrescente para zero podendo ir ao infinito na origem e tem a seguinte propriedade: Se μ é a medida de Lebesgue em $[0,1]$ a distribuição de $Y^{(d)}$ com relação a μ é igual a distribuição $F_Y(y)$ de Y:

$$\mu \left[Y^{(d)} < x \right] = P \left[Y < x \right] = F_Y(x)$$

Os conjuntos tais que os valores de Y e $Y^{(d)}$ caem em um dado intervalo têm mesmas medidas, e se Y é integrável então $Y^{(d)}$ é também integrável; se ψ é qualquer função positiva.

$$\int_0^a \psi\,(Y(x))\,d\,x = \int_0^a \psi\,(Y^{(d)}(x))\,d\,x \qquad (II.9)$$

sempre que uma das integrais existe.

Na definição de $Y^{(d)}$, se em x_0 Y tem um salto de μ_1 a μ_2, então $Y^{(d)}$ é constante no intervalo $(1 - \mu_2, 1 - \mu_1)$.

Para que $Y^{(d)}$ esteja perfeitamente definida, basta tomar $Y^{(d)}$ contínua à esquerda em $(0,1]$, $Y^{(d)}$ é a variável aleatória decrescente equivalente a Y e chamamos <u>rearranjo decrescente</u> de Y.

A seguinte propriedade do rearranjo decrescente da variável aleatória Y foi deduzida por Hardy e Littlewood e pode ser encontrada no artigo acima citado:

<u>Proposição</u> II.10 - Para todo $a \geq 0$, uma condição necessária e suficiente para que seja

$$\int_0^a Y(x)\,\log^+ Y(x)\,d\,x < \infty$$

é que seja

$$\int_0^a \left[\frac{1}{x} \int_0^x Y^{(d)}(\mu)\,d\,\mu \right] d\,x < \infty$$

<u>Teorema</u> II.11 - <u>Se</u> Y <u>é uma variável aleatória não negativa integrável em</u> $(\Omega, \mathcal{F}, P)$ e

$$g(x) = \frac{1}{x} \int_0^x Y^{(d)}(\mu)\,d\,\mu \qquad x \in (0,1] \,,$$

então: a) para qualquer seqüência crescente $(A_n)_{n\geq 1}$ de σ-álgebras em $\mathcal{F}$ e qualquer $\lambda > 0$, temos

$$P\left[U > \lambda\right] \geq \mu\left[g > \lambda\right],$$

onde $U = \sup_{n\geq 1} E(Y \mid A_n)$

b) Para todo $\varepsilon > 0$ existe uma seqüência crescente $(C_n)_{n\geq 1}$ de σ-álgebras no intervalo $(0,1]$ tal que

$$\mu\left[V \geq k\,\varepsilon\right] = \mu\left[g \geq k\,\varepsilon\right], \qquad k = 0, 1, 2, \ldots$$

e $V \geq g - \varepsilon$, onde $V = \sup_{n\geq 1} E(Y^{(d)} \mid C_n)$.

c) Para todo $\varepsilon > 0$ e para qualquer seqüência de números reais $(Q_n)_{n\geq 1}$ em $(0,1)$, decrescente para zero, existem, em um espaço de probabilidade (Ω^*, P^*) conveniente, uma variável aleatória Y^* com a mesma distribuição de $Y^{(d)}$ e uma seqüência decrescente $(\mathcal{D}_n)_{n\geq 1}$ de σ-álgebras de Borel tais que, para todo j inteiro positivo:

$$P^*\left[W \geq j\,\varepsilon\right] \geq Q_k\,\mu\left[g \geq j\,\varepsilon\right],$$

onde $W = \sup_{n\geq 1} E(Y^* \mid \mathcal{D}_n)$

Demonstração: parte (a) A seqüência $Y_n = E(Y \mid \mathcal{A}_n)$ é um MG e pelo corolário (I.67)

$$E(Y \mid \mathcal{A}_n) \xrightarrow[L^1]{q.c.} E(Y \mid \mathcal{A}_\infty), \text{ onde } \mathcal{A}_\infty = \mathcal{F}\Big(\bigcup_{n\geq 1} \mathcal{A}_n\Big) \qquad (II.12)$$

Seja

$$U_k = \sup_{1\leq n\leq k} P(Y \mid \mathcal{A}_n),$$

então

$$U_k \uparrow U \qquad \text{quando } k \to \infty \qquad (II.13)$$

Pelo teorema I.5_ temos, para todo $a > 0$ e $k \geq 1$:

$$a\,P\left[U_k \geq a\right] \int_{\left[U_k \geq a\right]} E(Y \mid \mathcal{A}_k)\,dP. \qquad (II.14)$$

Se $a \downarrow \lambda$ os conjuntos $\left[U_k \geq a\right]$ crescem para $\left[U_k > \lambda\right]$ e pela continuidade da integral e propriedade monótona sequencial de P, a desigualdade (II.14) permite escrever para todo $k \geq 1$:

$$\lambda' P\left[U_k > \lambda'\right] \leq \int_{\left[U_k > \lambda'\right]} E(Y|\mathcal{a}_k)\, dP$$

Por (II.13), os conjuntos $\left[U_k \geq \lambda'\right]$ crescem para $\left[U > \lambda'\right]$ quando $k \to \infty$, e tendo em vista (II.12) temos

$$P\left[U > \lambda'\right] \leq \frac{1}{\lambda'} \int_{\left[U > \lambda'\right]} E(Y|\mathcal{a}_\infty)\, dP = \frac{1}{\lambda'} \int_{\left[U > \lambda'\right]} Y\, dP,$$

pois $\left[U > \lambda'\right] \in \mathcal{a}_\infty$.

Se $\lambda' \uparrow \lambda$ os conjuntos $\left[U > \lambda'\right]$ decrescem para $\left[U \geq \lambda\right]$, donde

$$P\left[U \geq \lambda\right] \leq \frac{1}{\lambda} \int_{\left[U \geq \lambda\right]} Y\, dP, \quad \text{para todo } \lambda > 0 . \qquad \text{(II.15)}$$

Se λ_0 é tal que $P\left[Y > \lambda_0\right] = 0$ então $P\left[U > \lambda_0\right] = 0$ e também $\mu\left[Y^{(d)} \geq \lambda_0\right] = 0$ e $\mu\left[g > \lambda_0\right] = 0$, donde, para todo λ_0 com $P\left[Y > \lambda_0\right] = 0$ a parte (a) é trivial.

Basta então mostrar (a) para os números λ tais que $P\left[Y \geq \lambda\right] > 0$.

Seja λ_1 tal que $P\left[Y \geq \lambda_1\right] > 0$, portanto $\mu\left[g \geq \lambda_1\right]$ 0 uma vez que $g \geq Y^{(d)}$ em $(0,1]$ e $Y^{(d)}$ tem a mesma distribuição que Y. Como g é monótona decrescente existe o maior número t de $(0,1]$ tal que $g(t) \geq \lambda_1$. Dado um evento A satisfazendo à condição

$$\frac{1}{P(A)} \int_A Y\, dP \geq \lambda_1 , \qquad \text{(II.16)}$$

Vale:

$$g(P(A)) = \frac{1}{P(A)} \int_0^{P(A)} Y^{(d)} dP \geq \frac{1}{P(A)} \int_A Y\, dP \geq \lambda_1$$

e como t é o maior número tal que $g(t) \geq \lambda_1$, temos

$$P(A) \geq t = \mu\left[g \geq \lambda_1\right] \qquad \text{(II.17)}$$

Por (II.15) o evento $\left[U \geq \lambda_1\right]$ satisfaz à condição (II.16) e portanto a (II.17), logo

$$P\left[U \geq \lambda_1\right] \leq \mu\left[g \geq \lambda_1\right]$$

para todo $\lambda_1 > 0$ tal que $P\left[U \geq \lambda_1\right] > 0$. Como os conjuntos $\left[U \geq \lambda_1\right]$ crescem para $\left[U \geq \lambda\right]$ quando $\lambda_1 \downarrow \lambda$, a última desigualdade implica a condição (a), c.q.d.

parte b) - Sôbre $(o,1]$, sejam os eventos dois a dois disjuntos:

$$C_n = \left[(n-1)\,\delta \leq g < n\,\delta\right] \qquad (n \geq 1) \qquad \text{(II.18)}$$

e seja $\mathcal{C}_n$ a σ-álgebra de Borel gerada por $C_1, \ldots, C_{n-1}$.

$$g(x) = \frac{1}{x} \int_0^x Y^{(d)}(x)\, dx .$$

$$C_n = \left[(n-1) \leq g < n\,\delta\right] \quad (n \geq 1)$$

$C_1 = \phi$ na figura.

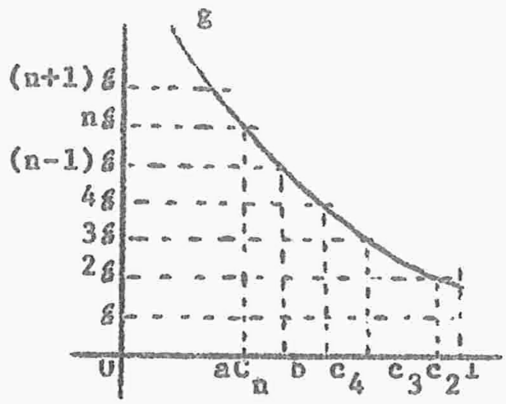


Figura (II.19)

Se C_n não é vazio, êle é um intervalo $a < \mu \leq b$. Por (I.39), e considerando que o menor evento de $\mathcal{C}_n$ que contém C_n é o intervalo $(0,b]$, temos:

$$E(Y^{(d)} \mid \mathcal{C}_n) = \frac{1}{b} \int_0^b Y^{(d)}(\mu)\, d\mu = g(b) \geq (n-1)\,\delta \quad \text{sôbre } C_n .$$

Assim, sôbre C_n vale

$$E(Y^{(d)} \mid \mathcal{C}_n) \geq g - \delta \qquad \text{para todo } n \text{ tal que } C_n \neq \phi,$$

donde

$$V = \sup_n E(Y^{(d)} \mid \mathcal{C}_n) \geq g - \delta \quad \text{em tôda parte.} \qquad \text{(II.20)}$$

Pela forma como definimos a função V, ela é mensurável em relação à σ-álgebra $\mathcal{F}\left(\bigcup_{n \geq 1} \mathcal{C}_n\right)$, e pela definição de C_n os eventos $\left[V \geq k\,\delta\right]$, são mensuráveis e para mostrar a relação

$$\mu\left[V \geq k\,\delta\right] = \mu\left[g \geq k\,\delta\right] ,$$

notemos que, aplicando a parte (a) à função $Y^{(d)}$ resulta

$$\mu\left[V \geq k\delta\right] \leq \left[g \geq k\,\delta\right]$$

e a desigualdade no sentido contrário é conseqüência de (II.20), pois, sendo $V > g$ em tôda parte temos:

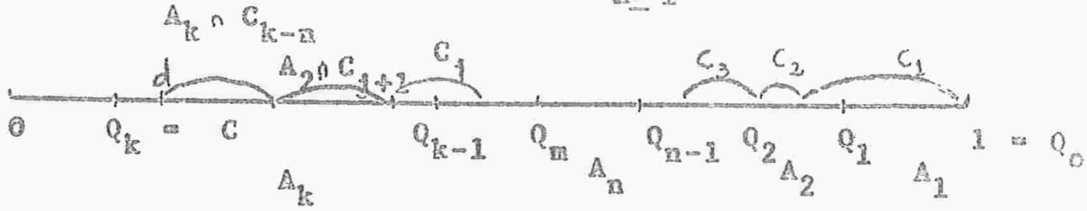
$$\left[g \geq k\delta\right] \subset \left[V \geq k\,\delta\right] \; .$$

e portanto

$$\mu\left[g \geq k\,\delta\right] \leq \mu\left[V \geq k\delta\right] , \qquad \text{c.q.d.}$$

parte c) - Seja $Q_o = 1$; $p_n = Q_{n-1} - Q_n$ $(n \geq 1)$. Então

$$p_n \geq 0 \quad \text{e} \quad \sum_{n \geq 1} p_n = 1$$



Vamos definir a variável aleatória Z sôbre $(0,1]$ com valores inteiros independente de g, $Y^{(d)}$ por meio da distribuição (extendendo, se necessário, o espaço de probabilidade $((0,1],\ \mathcal{B}_1, \mu)$ ao espaço $(\Omega^*, \mathcal{B}^*, P^*)$

$$P^*\left[Z = n\right] = p_n \qquad (n \geq 1)$$

Seja a variável aleatória Y definida sôbre $(\Omega^*, \mathcal{B}^*, P)$ pela distribuição de $Y^{(d)}$. Vamos construir a seqüência decrescente de σ-álgebras $(\mathcal{D}_n)_{n \geq 1}$ a partir dos eventos:

$$A_k = \left[Z = k\right] \qquad (k \geq 1)$$

e

$$C_i = \left[(i-1)\delta \leq g < i\,\delta\right] \quad (i \geq 1) \; .$$

definidas na demonstração da parte (b); para cada $n \geq 1$ seja:

$$\mathcal{D}_n = \mathcal{S}\,\{C_i \cap A_k\ (k > n,\ 1 \leq i \leq k-n),\ A_k \quad (k \geq 1)\} \qquad (II.21)$$

Como o menor evento de $\mathcal{D}_n$ que contém $C_i \cap A_k$ com $i > k - n > 0$ é $A_k - \sum_{j=1}^{k-n} C_j$, que é um intervalo $(c, d]$ quando não é o vazio, resulta, da igualdade (I.39) :

$$E(Y^* \mid \mathcal{D}_n) = \frac{1}{(d-c)} \int_c^d Y^{(d)} \, d\, P_{\geq} \, \mathcal{E}(d) \qquad \text{(II.22)}$$

$(Y^{(d)}$ decrescente)

sôbre $C_i \cap A_k$, para todo $i > k - n > 0$;

Então, como

$$\mathcal{E} \geq (i-1)\,\mathcal{E} > (k-n-1)\,\mathcal{E} \geq (k-n)\,\mathcal{E} \, ,$$

sôbre $C_i \cap A_k$, com $i > k - n > 0$, resulta

$$E(Y^* \mid \mathcal{D}_n) \geq (k-n)\,\mathcal{E} \quad \text{sôbre } C_i \cap A_k, \ i > k-n > 0$$

Para $k \leq i$, escolhamos $n = 1$, para $k > i$ escolhamos $n = k - i + 1$; sôbre $C_i \cap A_k$ vale:

$$W = \sup_n E(Y^* \mid \mathcal{D}_n) \geq \begin{cases} (k-1)\,\mathcal{E} & \text{se } k \leq i \\ (i-1)\,\mathcal{E} & \text{se } k > i \end{cases}$$

Então, sôbre o evento $\bigcup_{i,k>j} C_i \cap A_k = \bigcup_{i>j} C_i \cap \bigcup_{k>j} A_k$

vale

$$W = \sup_n E(Y^* \mid \mathcal{D}_n) \geq j\,\mathcal{E} \qquad \text{(II.23)}$$

Pela forma como definimos W ela é mensurável segundo $\mathfrak{F}(\bigcup_{n \geq 1} \mathcal{D}_n)$ e os conjuntos $[W \geq j\,\mathcal{E}]$, $j \geq 1$ são eventos de $\mathfrak{F}(\bigcup_{n \geq 1} \mathcal{D}_n)$. Como:

$$[\mathcal{E} \geq j\,\mathcal{E}] = \bigcup_{i>j} C_i \supset \bigcup_{i,k>j} C_i \cap A_k \, ,$$

resulta, por (II.23) e pela independência das variáveis aleatórias Z e $\mathcal{E}$:

$$P^*[W \geq j\,\mathcal{E}] \geq P^*[\mathcal{E} \geq j\,\mathcal{E}] \geq P^*\{\bigcup_{i,k>j} C_i \cap \bigcup_{k>j} A_k\} =$$

$$= P^*[\mathcal{E} \geq j\,\mathcal{E}]\, P^*[Z > j] = Q_k \, \mu[q \geq j\,\mathcal{E}] \qquad \text{c.q.d.}$$

Observação: Definimos no capítulo I os MG relativos a seqüências crescentes de σ-álgebras; <u>podemos definir também MG relativos a seqüências decrescentes de</u>

σ-álgebras e, tendo demonstrado uma relação equivalente a (II.14) para êste caso, o mesmo procedimento da demonstração da parte (a) permite extender (a) ao caso de σ-álgebras decrescentes.

Dizemos que uma distribuição μ sôbre a reta real domina uma distribuição γ se $\mu(x, \infty) \geq \gamma(x, \infty)$ para todo x.

Pela parte (a), para tôda Y não negativa de L^1 a distribuição P_g de g domina a de $\sup_n E(Y|\mathcal{A}_n)$, para qualquer seqüência monótona crescente ou decrescente de σ-álgebra $(\mathcal{A}_n)_{n\geq 1}$. A parte (b) afirma que P_g é a melhor possível para os $\mathcal{A}_n$ crescentes, no sentido de que se $\overline{P}$ é outra distribuição que domina a de $\sup E(Y|\mathcal{A}_n)$, então P é dominada pela distribuição P_g. A parte (c) afirma que P_g é ainda a melhor possível para os $\mathcal{A}_n$ decrescentes, embora em um sentido algo mais fraco.

Pelo teorema recíproco do teorema da convergência dominada de Lebesgue, vemos que é interessante determinar condições para a integrabilidade de $\sup_n X_n$.

Passaremos a estudar o caso especial em que as funções X_n são as esperanças condicionais de uma variável aleatória integrável não negativa sôbre $(\Omega, \mathcal{F}, P)$ com relação a uma seqüência monótona $(\mathcal{A}_n)_{n\geq 1}$ de sub-σ-álgebras de $\mathcal{F}$.

Se $0 \neq X \in L^1(\Omega, \mathcal{F}, P)$ o corolário I.61 nos permite afirmar que o MG $(E(X|\mathcal{A}_n))_{n\ 1}$ converge quase certamente e em L^1 para $E(X|\mathcal{A}_\infty)$, onde $\mathcal{A}_\infty = \mathcal{F}(\bigcup_{n\geq 1} \mathcal{A}_n)$. Pelo teorema I.52 temos:

$$E\,|\sup_n E(X|\mathcal{A}_n)| \leq \frac{e}{e-1} + \frac{e}{e-1} E\left[E(X|\mathcal{A}_\infty) \log^+ E(X|\mathcal{A}_\infty)\right]$$

e então basta procurar condições para a integrabilidade de $E(X|\mathcal{A}_\infty) \log^+ E(X|\mathcal{A}_\infty)$, e para isto é suficiente assegurar a integrabilidade de $E(X|\mathcal{A}_\infty) \log E(X|\mathcal{A}_\infty)$.

Como $\psi(x) = x \log x$ é uma função convexa, a propriedade (I.40) nº 5 permite escrever:

$$E(X|\mathcal{A}_\infty) \log E(X|\mathcal{A}_\infty) \leq E(X \log X \mid \mathcal{A}_\infty),$$

e concluimos que a condição $X \log X \in L^1$ é suficiente para que $\sup_n E(X|\mathcal{A}_n) \in L^1$.

O teorema seguinte mostra que a condição $X \log X \in L^1$ é, num certo sentido, também necessária para a integrabilidade do $\sup_n E(X|\mathcal{Q}_n)$:

Teorema II.24 - Se Y é uma variável aleatória não negativa integrável tal que $Y \log Y$ não é integrável, existe uma variável aleatória Y^* de mesma distribuição que Y em um espaço de probabilidade conveniente $(\Omega^*, \mathcal{B}^*, P^*)$, e uma seqüência monótona $(\mathcal{Q}^*_n)_{n \geq 1}$ de σ-álgebras que pode ser escolhida crescente ou decrescente, tal que

$$\sup_n E(Y^*|\mathcal{Q}^*_n) \notin L^1 \tag{II.25}$$

Demonstração: Podemos supôr que Y é uma função não crescente em $(o,1]$ com probabilidade de Lebesgue . Pelo resultado de Hardy e Littlewood [7] a condição $Y \log Y \notin L^1$ implica que

$$g(x) = \frac{1}{x} \int_0^x Y(\mu)\, d\mu \notin L^1 \quad .$$

Para escolher os $\mathcal{Q}^*_n$ crescentes satisfazendo (II.25) basta tomar para $\mathcal{Q}^*_n$ as σ-álgebras $\mathcal{C}_n$ da parte (b) do teorema II.11.

Para escolher os $\mathcal{Q}^*_n$ decrescentes notemos que, sendo $g \notin L^1$, temos

$$\sum_{k \geq 1} \mu\left[g \geq k\delta\right] = \infty$$

Então existe uma seqüência $(Q_k)_{k \geq 1}$ de números em $(o,1)$ que converge monòtonamente a zero [se $\sum_{k \geq 1} a_k = \infty$ com $a_k > 0$ $(k \geq 1)$, então $\sum_{k \geq 1} a_k b_k = \infty$ quando tomarmos $b_k = (\sum_{i=1}^{k} a_i)^{-1}$

satisfazendo:

$$\sum_{k \geq 1} Q_k \ \mu\left[g \geq k\delta\right] = \infty . \tag{II.26}$$

Então, para esta escolha de $(Q_n)_{k \geq 1}$ escolhendo os

$\mathcal{Q}_n^*$ como sendo os $\mathcal{D}_n$ da parte (c) do teorema II.11, temos

$$P^*\left[\sup_n E(Y^* \mid \mathcal{Q}_n^*) \geq k\,\delta\right] \geq Q_k\, \mu\left[\delta \geq k\,\delta\right] \qquad (\forall k \geq 1),$$

e por (II.26) resulta,

$$\sum_{k \geq 1} P^*\left[\sup_n E(Y^* \mid \mathcal{Q}_n^*) \geq k\,\delta\right] = \infty,$$

e portanto $\sup E(Y^* \mid \mathcal{Q}_n^*) \notin L^1$, o que completa a demonstração.

BIBLIOGRAFIA

1 - J. Neveu, MARTINGALES - Publicação do IME, 1970

2 - J. Neveu, Bases Mathématiques du Calcul des Probabilités - Ed. Masson et Cie., 1967

3 - Breiman, PROBABILITY - Addison Wesley Publishing Company, 1968

4 - M. Loève, PROBABILITY THEORY - D. Van Nostrand Company,Inc. 1960.

5 - David Blackwell and Lester E. Dubins, A CONVERSE TO THE DOMINATED CONVERGENCE THEOREM - Illinois Journal of Mathematics - Vol. 7 nº 3 (1963) páginas 508-504

6 - Doob, STOCHASTIC PROCESSES - John Wiley & Sons, Inc., N.Y. 1953.

7 - G.H.Hardy and J.E.Littlewood, A MAXIMAL THEOREM WITH FUNCTION THEORETIC APPLICATIONS - Acta Math.vol.54 (1930) paginas 81-116.