MATEMÁTICA UNIVERSITÁRIA
n$^{\underline{os}}$ 38/39 – julho/dezembro 2005 – pp. 35–68

Introdução à probabilidade e aos processos estocásticos:

uma exposição para quem não sabe nada do assunto.

ARTUR O. LOPES

Vamos apresentar neste artigo algumas definições preliminares, exibir alguns exemplos de natureza simples e discutir também algumas (poucas) idéias básicas da teoria da probabilidade e dos processos estocásticos. Certas partes desta exposição têm um caráter levemente informal.

Nossa apresentação visa o leitor que não tem a menor idéia do assunto. O objetivo aqui é descrever o ponto de vista probabilístico ou estatístico (de se analisar um problema matemático) em um nível elementar.

O rigor matemático será apenas aquele necessário para que as idéias centrais fiquem claras ao leitor.

Considere fixado um conjunto K. Uma probabilidade P em K é uma lei que associa a cada subconjunto A de K um valor real não negativo $P(A)$ menor ou igual a um. O valor $P(K)$ se assume ser igual a um. Algumas propriedades mais (definição 2) serão requeridas para P.

Fixado o conjunto K, denotamos por $\mathcal{P}(K) = \{A | A \subset K\}$, o conjunto das partes de K.

Infelizmente, em muitos casos interessantes não se pode definir P (com as tais propriedades que desejamos) em todos os subconjuntos A de K, ou seja sobre a classe de conjuntos $\mathcal{P}(K)$.

Sendo assim, é necessário introduzir a seguinte definição:

Definição 1: Uma família $\mathcal{A}$ de subconjuntos de um conjunto K é chamado de σ-álgebra sobre K no caso em que:

a) o conjunto K pertence a $\mathcal{A}$,

b) se A pertence a $\mathcal{A}$, então o complemento $K - A$ também pertence a $\mathcal{A}$,

c) se $(A_n)_{n \in \mathbf{N}}$ é uma coleção enumerável de conjuntos em $\mathcal{A}$, então a união $\cup_{n \in \mathbf{N}} A_n$ também está em $\mathcal{A}$.

Note que $\mathcal{A} \subset \mathcal{P}(K)$. Note também que se $(A_n)_{n \in \mathbf{N}}$ é uma coleção enumerável de conjuntos em $\mathcal{A}$, então a interseção $\cap_{n \in \mathbf{N}} A_n = K - \cup_{n \in \mathbf{N}} (K - A_n)$ está em $\mathcal{A}$.

Os conjuntos A em $\mathcal{A}$ são chamados de *conjuntos $\mathcal{A}$-mensuráveis*, ou, simplemente *conjuntos mensuráveis*, se estiver claro qual a sigma-álgebra $\mathcal{A}$ a qual estamos nos referindo.

Quando K for um conjunto finito ou enumerável a única sigma-álgebra que iremos considerar sobre K será $\mathcal{A} = \mathcal{P}(K)$.

Observamos que quando K é finito, então $\mathcal{P}(K)$ é finito, mas quando K é enumerável infinito, então $\mathcal{P}(K)$ não é enumerável.

Definição 2: Uma probabilidade P sobre uma σ-álgebra $\mathcal{A}$ de subconjuntos de K é uma lei que associa a cada A em $\mathcal{A}$ um número real $P(A)$ tal que

a) $P(K) = 1$,

b) se E_n , $n \in \mathbf{N}$, é uma coleção enumerável de subconjuntos de $\mathcal{A}$ tal que $E_n \cap E_m = \emptyset$, para $m \neq n$, então $P(\cup_{n \in \mathbf{N}} E_n) = \sum_{n \in \mathbf{N}} P(E_n)$.

Chamamos o par $(K, \mathcal{A})$, como definido acima, de espaço mensurável.

Uma vez fixada a probabilidade P sobre a sigma-álgebra $\mathcal{A}$ chamamos a tripla $(K, \mathcal{A}, P)$ de espaço de probabilidade.

Um exemplo simples aparece quando jogamos um dado: existem seis faces e intuitivamente sabemos que cada uma tem probabilidade $1/6$ de sair. Podemos modelar tal problema com $K = \{1,2,3,4,5,6\}$, e $P(\{i\}) = 1/6$, onde $i \in K$. A sigma álgebra $\mathcal{A}$ é o conjunto das partes $\mathcal{P}(K)$. A probabilidade de sair 1 ou 2 é 2/6. Ou seja

$$P(\{1,2\}) = P(\{1\}) + P(\{2\}) = 1/6 + 1/6 = 2/6.$$

Este é um problema não determinístico. Não podemos afirmar qual face irá sair ao lançar o dado, podemos apenas falar na probabilidade de sair uma certa face.

A idéia que o conceito geral de probabilidade traduz é que um conjunto $A \in \mathcal{A}$ tal que $P(A) = 0$ é de natureza desprezível (probabilisticamente falando), ou seja, $w \in A$ não ocorre em termos probabilísticos (ou estatísticos). Por sua vez, um conjunto A tal que $P(A) = 1$ traduz o fato que $w \in A$ ocorre com certeza em termos probabilísticos. Estamos tratando aqui de eventos aleatórios, sendo assim só se pode falar da probabilidade de algo ocorrer. O conjunto A descreve um conjunto de elementos w com certas propriedades. Quanto maior for o valor $P(A)$, maior a probabilidade de $\omega \in A$ ocorrer.

No exemplo do dado considerado antes, apenas o conjunto vazio tem probabilidade zero. Neste caso, não aparece de maneira muito transparente o que estamos tentando destacar acima. O problema é que tomamos um exemplo muito simples. Outros mais complexos aparecerão em breve.

Quando K for finito ou enumerável, ou seja da forma $K = \{k_1, k_2, \ldots, k_n\}$ ou $K = \{k_1, k_2, \ldots, k_n, \ldots\}$, a probabilidade P sobre $\mathcal{A} = \mathcal{P}(K)$ fica especificada apenas pelos valores $p_i = P(\{k_i\})$.

É usual se denotar o espaço K onde P atua por Ω e denominá-lo de espaço amostral. Um elemento w em Ω será denominado de uma amostra de Ω ou também de um evento.

Definição 3: Considere K equipado com uma σ-álgebra $\mathcal{A}$ e V outro conjunto equipado com uma σ-álgebra $\mathcal{G}$. Uma função $\phi : K \to V$

é chamada de mensurável se $\phi^{-1}(A) \in \mathcal{A}$ para todo $A \in \mathcal{G}$.

Consideraremos aqui apenas funções mensuráveis $\phi : K \to V$, em que V é finito ou enumerável, $V = \{v_1, v_2, ..., v_i, ..\}$, e contido em $\mathbb{R}$ (ou ainda em $\mathbb{R}^m$ para algum m natural). O conjunto K não precisa necessariamente estar dentro de um $\mathbb{R}^k$.

Quando $\mathcal{A}$ for $\mathcal{P}(K)$, o conjunto das partes de K, então qualquer $\phi : K \to V$ é mensurável.

Em qualquer caso é usual a notação $X : (\Omega, \mathcal{A}) \to (V, \mathcal{G})$, ou seja se utiliza letras maiúsculas, para descrever a função $\phi = X$ com as propriedades acima descritas. A função mensurável X é usualmente denominada de variável aleatória. É comum denotar os elementos de Ω por w e os elementos onde X toma valores por x, logo x denota um elemento em V (a letra minúscula x correspondendo a maiúscula X). Assim se conisderarmos uma $Y : (\Omega, \mathcal{A}) \to (V, \mathcal{G})$ os elementos de V serão denominados de y.

Vamos dar um exemplo de um jogo que ilustra tal conceito: considere um dado com seis faces e que será jogado uma vez. O jogo é o seguinte: se sair a face 1 ou 2 ganhamos 1 real, caso contrário ganhamos 2 reais. Considere $X : \{1,2,3,4,5,6\} = \Omega \to \{1,2\}$, tal que

$$X(1) = 1,\ X(2) = 1\,, X(3) = 2\,, X(4) = 2\,, X(5) = 2,\ X(6) = 2.$$

A função X descreve o que vamos ganhar quando se joga o dado em função da face que sai. Neste caso é natural concluir que

$$P(\{w | X(w) = 1\}) = 1/3 \quad \text{e} \quad P(\{w | X(w) = 2\}) = 2/3.$$

Seguindo a notação descrita acima, a sigma-álgebra a ser considerada em Ω é $\mathcal{A} = \mathcal{P}(\{1,2,3,4,5,6\})$. Ainda $V = \{1,2\}$ e $\mathcal{G} = \mathcal{P}(\{1,2\})$.

É fácil ver que soma, produtos e compostas de funções mensuráveis determinam novas funções mensuráveis.

Como outro exemplo de probabilidade, considere uma cidade que possui população de $N = 10.000$ habitantes e que cada habitante utiliza um e apenas um de dois provedores de internet. Suponhamos que num dado dia 7.000 habitantes utilizam o provedor 1 e 3.000 usam o provedor 2.

Neste caso é natural tomar Ω como o conjunto dos 10.000 habitantes, $\mathcal{A}$ como a classe das partes $\mathcal{P}(\Omega)$ e ainda considerar a probabilidade P em $\mathcal{P}(\Omega)$ tal que

$$P(A) = \frac{\text{número de pessoas no conjunto } A \subset \Omega}{10.000}.$$

Seja $X : \Omega \to V = \{1,2\}$, que associa cada habitante a seu provedor no dia em questão. Como Ω e V são finitos (logo $\mathcal{A} = \mathcal{P}(\Omega)$) então X é mensurável.

Seja $\phi : (K, \mathcal{A}, P) \to (V, \mathcal{G})$ mensurável, então fica naturalmente definido uma probabilidade $\tilde{P}$ sobre $(V, \mathcal{G})$, através de $\tilde{P}(B) = P(\phi^{-1}(B))$ para cada $B \in \mathcal{G}$. É fácil ver que $\tilde{P}$ é uma probabilidade sobre $(V, \mathcal{G})$. Algumas vezes denotaremos tal $\tilde{P}$ por P_ϕ, para enfatizar que foi obtida de P através de ϕ.

Dado $X : (\Omega, \mathcal{A}, P) \to (V, \mathcal{G})$, diremos que $\tilde{P} = P_X$ é a probabilidade induzida pela função mensurável (variável aleatória) X e pela probabilidade P.

Tal probabilidade P_X é denominada de *distribuição da variável aleatória* X. Quando X toma valores reais, ou seja $V = \mathbb{R}$, a probabilidade P_X estará definida para subconjuntos B da reta real.

Em breve veremos que a distribuição P_X de X é na verdade mais importante que a própria probabilidade P.

No exemplo que estamos considerando acima, onde definimos $X : \Omega \to V = \{1,2\}$, obtemos, a partir da probabilidade P inicialmente considerada, uma nova probabilidade $\tilde{P} = P_X$ definida acima como $\tilde{P}(\{1\}) = 0.7$ e $\tilde{P}(\{2\}) = 0.3$. Usando a propriedade aditiva da probabilidade, então é claro que disto segue que $\tilde{P}(\{1,2\}) = 1$. Ainda $P(\emptyset) = 0$. Deste modo ficaram explícitos os valores de $\tilde{P} = P_X$ sobre $\mathcal{P}(\{1,2\})$.

Este é um dos exemplos mais simples de probabilidade que conseguimos imaginar.

Podemos então dizer que utilizar o provedor 1 tem probabilidade 0.7 e utilizar o provedor 2 tem probabilidade 0.3, ou seja $P_X(\{1\}) = 0.7$ e $P_X(\{2\})=0.3$. É preferível a notação $P(X=1)=0.7$ e $P(X=2)=0.3$.

Acima $P(X = 1)$ significa $P(\{w \mid X(w) = 1\})$, etc...

Fixado X, para simplificar a notação, muitas vezes não se diferencia a probabilidade P (inicialmente considerada) da distribuição P_X, omitindo assim expressões com $\tilde{P}$ ou P_X (que age sobre subconjuntos de V) e usando apenas a letra P (que age sobre subconjuntos de Ω). Logo, quando esta claro de qual X falamos, não deve ser motivo de confusão falar de $P(B)$ para um conjunto $B \subset V$. Neste caso, $P(B) = P(\{\omega \in \Omega \,|\, X(\omega) \in B\})$.

Vamos agora definir o que é um processo estocástico.

Definição 4: Seja $(\Omega, \mathcal{A}, P)$ espaço de probabilidade, $(S, \mathcal{G})$ um espaço mensurável e ainda uma família de variáveis aleatórias X_t indexadas por um parâmetro $t \in T$, onde $T \subset \mathbb{R}$, e onde cada $X_t : (\Omega, \mathcal{A}, P) \to (S, \mathcal{G})$ é mensurável. Dizemos que tal $(X_t)_{t \in T}$ é um processo estocástico. No presente texto S **é sempre finito ou enumerável** e assim $\mathcal{G} = \mathcal{P}(S)$.

Espaço de parâmetros - O conjunto $\mathrm{T} \neq \emptyset$ contido em $\mathbb{R}$ é denominado espaço de parâmetros ou índices do processo. O conjunto T possui uma ordem e vamos pensar que para cada $t \in T$ a variável X_t descreve o que acontece com o processo no tempo t.

Dois casos importantes são:

Parâmetro discreto -$T = \mathbb{N}$, ou $\mathbb{Z}$, ou ainda $\{1, 2, ..., n\}$.

Parâmetro contínuo - $T = [a, b]$, ou $T = \{t \in \mathbb{R} \,|\, t \geq 0\} = \mathbb{R}^+$ ou ainda $T = \mathbb{R}$.

Definição 5: Espaço de Estados - é o conjunto S, ou seja, o elenco dos possíveis valores de cada variável aleatória X_t.

Quando S é finito, S será suposto da forma $\{1, 2, ..., m\}$ ou ainda $\{1, 2, ..., m\}^k$, onde m e k são números naturais.

O único caso que iremos tratar aqui é quando $T = \mathbb{N}$.

Para todo $w \in \Omega$ fixo, e t também fixo, $X_t(w)$ determina o valor do processo no tempo t avaliado em w e algumas vezes é denotado por w_t. Quando $w \in \Omega$ está fixo e t variável, os valores $X_t(w) = w_t$ descrevem a evolução temporal ao longo do tempo $t \in T$. Usamos a letra grega ω para denotar $\omega = \{w_t\}_{t \in T} \in S^T$ associado a um certo w. Observe que $w \in \Omega$ e $\omega \in S^T$. Por exemplo, se $S = \{1, 2, 3\}$ e $T = \mathbb{N}$, um ω

poderia ser a seqüência ordenada infinita $(2, 2, 3, 1, 2, 3, 1, 2, 3, 3, ...) = (w_1, w_2, w_3, ..., w_n, ...)$.

Note ainda que usaremos aqui a seguinte notação: se o conjunto A é definido por

$$A = \{w | X_{t_1}(w) = a_1, X_{t_2}(w) = a_2, ..., X_{t_n}(w) = a_n\} =$$

$$\{X_{t_1} = a_1, X_{t_2} = a_2, ..., X_{t_n} = a_n\}$$

e B por

$$B = \{w | X_{s_1} = b_1, X_{s_2} = b_2, ..., X_{s_m} = b_m\},$$

então

$$A \cap B = \{w | X_{t_1}(w) = a_1, X_{t_2}(w) = a_2, ..., X_{t_n}(w) = a_n,$$

$$X_{s_1}(w) = b_1, X_{s_2}(w) = b_2, ..., X_{s_m}(w) = b_m\},$$

ou seja, sem maior preocupação com a ordem dos tempos envolvidos.

Concretamente, se

$$A = \{X_2 = 1, X_6 = 3, X_8 = 4\},$$

e

$$B = \{X_1 = 4, X_2 = 1, X_8 = 3, X_9 = 2\},$$

podemos denotar $A \cap B$, indistintamente como

$$A \cap B = \{X_1 = 4, X_2 = 1, X_6 = 3, X_8 = 3, X_8 = 4, X_9 = 2\},$$

ou como

$$A \cap B = \{X_2 = 1, X_6 = 3, X_8 = 4, X_1 = 4, X_2 = 1, X_8 = 3, X_9 = 2\}.$$

Note que neste caso

$$\{X_1 = 4, X_2 = 1, X_6 = 3, X_8 = 3, X_8 = 4, X_9 = 2\} = \emptyset,$$

porque não se pode existir w tal que $X_8(w) = 4$ e $X_8(w) = 3$.

Não confunda este conjunto com

$$\emptyset \neq \{X_1 = 4, X_2 = 1, X_6 = 3, X_8 \in \{3,4\}, X_9 = 2\} =$$

$$\{X_1 = 4, X_2 = 1, X_6 = 3, X_8 = 3, X_9 = 2\} \cup$$

$$\{X_1 = 4, X_2 = 1, X_6 = 3, X_8 = 4, X_9 = 2\}.$$

Note ainda que

$$\{X_1 \in \{2,4\}, X_2 = 1, X_4 \in \{3,4\}\} \neq$$

$$\{X_1 = 4, X_2 = 1, X_4 = 3\} \cup$$

$$\{X_1 = 2, X_2 = 1, X_4 = 4\}.$$

Prosseguindo com o exemplo da cidade com $N = 10.000$ habitantes, vamos supor que a cada mês se faça uma enquete e cada pessoa informe qual internet está usando. Vamos estabelecer que se vai realizar enquetes em três oportunidades seguidas com intervalo de um mês. Fica assim determinado que $T = \{1,2,3\}$, $S = \{1,2\}$ e X_t descreve qual provedor uma determinada pessoa estava utilizando no dia da enquete t-ésima, $t \in \{1,2,3\}$.

Para simplificar assumimos que em cada mês toda pessoa w utiliza um e apenas um provedor. Neste caso, é natural tomar $\Omega = \{1, 2, ..., N\}$, e uma amostra w é um habitante da cidade. Neste caso, $X_t(w) = w_t \in S$, $t \in \{1,2,3\}$, descreve o provedor utilizado pelo indivíduo w na enquete t-ésima.

Um elemento ω poderia ser, por exemplo, $\omega = (1,2,1) \in S^T$. Este ω corresponde a indivíduos w tais que usavam a internet 1 no mês 1, a internet 2 no mês 2 e a internet 1 no mês 3.

As perguntas que estaremos preliminarmente interessados em analisar envolvem por exemplo: qual o valor de

$$P(X_1 = 2, X_2 = 2, X_3 = 1) =$$

$$P(\{w \in \Omega \text{ tais que } X_1(w) = 2, X_2(w) = 2, X_3(w) = 1\})?$$

Para efetuar este cálculo, contamos cada indivíduo que na primeira enquete usava o provedor 2, na segunda o mesmo provedor 2 e na terceira trocou para o provedor 1. A seguir dividimos o número obtido por $N = 10.000$.

Os possíveis $\omega = (w_1, w_2, w_3)$ tomariam valores em

$$\{1,2\}^3 = \{1,2\} \times \{1,2\} \times \{1,2\}.$$

Note que diferentes pessoas $w \in \Omega$ podem determinar o mesmo valor $\omega = (w_t)_{t \in T} \in \{1,2\}^3$.

Um conceito de fundamental importância em probabilidade é o de probabilidade condicional.

Definição 6: Fixado $(\Omega, \mathcal{A}, P)$, denotamos por

$$P(A \,|\, B) = \frac{P(A \cap B)}{P(B)},$$

a probabilidade de ocorrer A dado que ocorreu B. Isto só faz sentido, é claro, se $P(B) \neq 0$.

Por exemplo, para saber qual a probabilidade de um estudante do colégio B passar no exame vestibular da universidade A, considera-se o quociente

$$\frac{\text{número de estudantes do colégio } B \text{ que passaram na universidade } A}{\text{número de estudantes do colégio } B}.$$

Definição 7: Fixada uma probabilidade P, dizemos que o evento definido pelo conjunto A é independente do evento definido pelo conjunto B se

$$P(A \cap B) = P(A)\, P(B).$$

No caso em que $P(B) \neq 0$, temos que

$$P(A \,|\, B) = P(A).$$

Esta propriedade descreve o fato que para os conjuntos A, B em consideração, a ocorrência de A não tem nada a ver, em termos estatísticos, com a ocorrência de B.

Definição 8: Sejam (X, Ω, P) onde $X(w) \in S$ e (Y, Ω, P_1) onde $Y(w) \in S_1$. Diremos que X é independente de Y se para qualquer elementos $a \in S \subset \mathbb{R}$ e $b \in S_1 \in \mathbb{R}$ vale que

$$P(X = a \,|\, Y = b) = P(X = a),$$

ou seja, se

$$P(X = a, Y = b) = P(\{X = a\} \cap \{Y = b\}) = P(\{X = a\})\, P(\{Y = b\}).$$

Como exemplo, considere X a variável que descreve a face que sai quando se joga um dado a primeira vez e Y a variável que descreve a face que sai quando jogamos o dado uma segunda vez.

Sejam $a, b \in \{1, 2, 3, 4, 5, 6\}$ fixados, então

$$P(X = a, Y = b) = 1/36 = (1/6)^2.$$

Isto porque, temos ao todo 6^2 possibilidades de saidas de pares ordenados de faces $(x, y) \in \{1, 2, 3, 4, 5, 6\} \times \{1, 2, 3, 4, 5, 6\}$. O par (a, b) corresponde a apenas uma possibilidade. Cada par tem a mesma chance de sair, logo tem mesma probabilidade.

Sendo assim,

$$P(X = a, Y = b) = 1/36 = 1/6 \;\, 1/6 = P(X = a)\, P(Y = b).$$

Logo, a face que sai no primeiro lançamento do dado é independente do que sai no segundo lançamento. É usual colocar o tempo como parâmetro e assim denominar $X = X_1$ e $Y = X_2$. Se fossemos lançar a moeda uma terceira vez, o resultado seria X_3

Voltando ao modelo do uso da internet, poderíamos definir uma variável Y tal que $Y(w) = 4$ se a renda mensal do indivíduo w é abaixo de $4.000,00$ reais e $Y(w) = 5$ caso contrário. Neste caso, $S_1 = \{4, 5\}$, e se por acaso X é independente de Y, então existe uma clara indicação de que o uso da internet 1 ou 2 é independente da classe de renda do indivíduo w.

Podemos nos perguntar também: qual a probabilidade de uma pessoa utilizar a internet 1 na terceira enquete, dado que utilizou a internet 2 nas duas primeiras? Será que nesta questão específica existe independência?

Para responder tal pergunta devemos calcular

$$P(X_3 = 1 \mid X_1 = 2, X_2 = 2) = \frac{P(X_1 = 2, X_2 = 2, X_3 = 1)}{P(X_1 = 2, X_2 = 2)}.$$

Vai existir independência (do que acontece no tempo 3 em função do uso anterior no tempo 1 e 2), se, por acaso,

$$\frac{P(X_1 = 2, X_2 = 2, X_3 = 1)}{P(X_1 = 2, X_2 = 2)} = P(X_3 = 1).$$

Não nos parece natural que vá ocorrer independência, pois existe sempre uma certa dose de inércia nas índoles das pessoas: se um indivíduo usava a internet 2 no mês 2, então o valor da probabilidade que ele vá continuar usando a internet 2 no mês 3 é maior do que o valor da probabilidade que ele passe a usar a internet 1 no mês 3.

Para responder com certeza a pergunta acima seria necessário obter os dados exatos sobre os habitantes da tal cidade nestas três oportunidades e fazer a conta acima.

Consideramos finalmente o caso mais interessante em que $T = \mathbb{N}$ e $X_t : \Omega \to \{1, 2\}$ que vai descrever a evolução temporal ilimitada do uso do provedor de cada habitante w da cidade. Uma pergunta natural que podemos nos fazer neste caso é a seguinte: será que existem os limites

$$\lim_{t\to\infty} P(X_t = 1) = \pi_1,$$

e

$$\lim_{t\to\infty} P(X_t = 2) = \pi_2?$$

Outra questão: será que existe o limite

$$\lim_{t\to\infty} P(X_t = 1 \mid X_1 = 2)?$$

Um dos objetivos da teoria é atacar questões desta natureza.

Para cada amostra $w \in \Omega$ fixada, seja a seqüência $\omega = (w_t)_{t\in\mathbb{N}} = (X_t(w))_{t\in\mathbb{N}} \in \{\omega : \mathbb{N} \to S\} = S^{\mathbb{N}}$, que será denominada de caminho amostral. Na verdade os $\omega = (w_t)_{t\in\mathbb{N}}$ desempenham na teoria um papel mais fundamental do que os w.

Os exemplos do mundo real, no entanto, muitas vezes aparecem de maneira natural no domínio dos $w \in \Omega$.

Alertamos o leitor que, fixado o processo estocástico $(X_t)_{t\in\mathbb{N}}$, é usual não fazer muita distinção entre w e ω e também entre Ω e $S^{\mathbb{N}}$. Ou seja, podemos falar em $\omega \in \Omega$ ou $w \in S^{\mathbb{N}}$. Faremos esta distinção apenas quando isto for necessário para esclarecer algum ponto.

Vamos considerar a seguir uma classe importante de processos estocásticos.

Definição 9: Fixados $(\Omega, \mathcal{A}, \mathcal{P})$, dizemos que o processo X_t tomando valores em S (enumerável) e com parâmetro $t \in T = \mathbb{N}$ é independente se para cada n e cada seqüência $t_1 < t_2 < ... < t_n$, $t_i \in T = \mathbb{N}$, e para cada seqüência de valores $a_1, a_2, ..., a_n \in S$ vale que

$$P(X_{t_1} = a_1, X_{t_2} = a_2, ..., X_{t_n} = a_n) =$$

$$P(X_{t_1} = a_1)\, P(X_{t_2} = a_2) .....P(X_{t_n} = a_n).$$

Vamos voltar ao exemplo do jogo com um dado que mencionamos antes. Como vimos, neste jogo $P(X = 1) = 1/3$ e $P(X = 2) = 2/3$. Vamos agora jogar o dado sucessivamente e X_t vai descrever o que ganhamos na jogada $t \in T = \mathbb{N}$ em função da face que saiu. É natural assumir que para cada t fixo, a variável X_t é descrita também por X como acima.

Uma conta fácil (levando em conta o conjunto das possibilidades) mostra que

$$P(X_1 = 1, X_2 = 2, X_3 = 1) = \frac{2\,.4\,.2}{6^3} = \frac{16}{216} =$$

$$\frac{2}{6}.\frac{4}{6}.\frac{2}{6} = P(X_1 = 1)\, P(X_2 = 2)\, P(X_3 = 1).$$

Procedendo de maneira semelhante é fácil ver que

$$P(X_{t_1} = a_1, X_{t_2} = a_2, ..., X_{t_n} = a_n) =$$

$$P(X_{t_1} = a_1)\, P(X_{t_2} = a_2).....P(X_{t_n} = a_n),$$

para qualquer seqüência $t_1 < t_2 < ... < t_n$ e $a_i \in \{1, 2\}, i \in \{1, 2, ..., n\}$.

Mais explicitamente,

$$P(X_{t_1} = a_1,\, X_{t_2} = a_2, ...,\, X_{t_n} = a_n) = (\frac{1}{3})^k\, (\frac{2}{3})^{n-k},$$

onde k é o número de valores 1 entre os n valores $a_1, a_2, ..., a_n$.

Ou seja, o processo estocástico associado a jogar o dado sucessivas vezes (e ver obtemos $X = 1$ ou $X = 2$) é um processo independente.

Outra questão: vamos jogar o dado n vezes e denotar por $X_1, X_2, ..., X_n$ os resultados obtidos sucessivamente em ordem de aparecimento; qual a probabilidade de se obter k vezes $X_i = 1$ (ou seja, sair a face 1 ou 2 do dado), entre os $i \in \{1, 2, 3.., n\}$? Ora existem

$$\frac{n!}{(n-k)!\, k!}$$

possibilidades de isto ocorrer no universo de 2^n ocorrências de $X = 1$ ou $X = 2$, ou seja dos possíveis resultados X_i que se obtém ao jogar o dado n vezes.

Cada uma das ocorrências tem probabilidade $(\frac{1}{3})^k\, (\frac{2}{3})^{n-k}$. Aqui estamos usando a expressão acima que segue da independência do processo.

Logo a probabilidade que buscamos vale

$$\frac{n!}{(n-k)!\, k!}\, (\frac{1}{3})^k\, (\frac{2}{3})^{n-k}.$$

Mais geralmente, se ocorrer no jogo uma probabilidade p de sair $X = 1$ e uma probabilidade $1 - p$ de ocorrer $X = 2$ em uma jogada, a probabilidade de ocorrer um total de k vezes o $X = 1$ em n jogadas é igual a

$$\frac{n!}{(n-k)!\, k!}\, p^k\, (1-p)^{n-k}.$$

Estamos supondo, é claro, que existe independência entre as sucessivas jogadas.

Esta distribuição é denominada de Binomial (n, p). Para cada k temos um valor e a soma destes valores para $k = 0, 1, 2, .., n$ é igual a 1.

Para checar que a soma destes valores é exatamente igual a 1, podemos usar o Binômio de Newton:

$$1 = (p + (1-p))^n = \sum_{k=0}^{n} \frac{n!}{(n-k)!\,k!}\, p^k\,(1-p)^{n-k},$$

Vamos agora dar um exemplo prático do uso da Teoria das Probabilidades. Uma companhia aérea possui um avião com s lugares. Ela sabe que em geral ocorre que um certo número de pessoas compram a passagem mas não aparecem na hora do vôo. Então ela vende v lugares para o vôo e $v > s$. Através da experiência passada, a companhia sabe que, em termos estatísticos, existe uma probabilidade p de comparecimento. Ou seja, cada indivíduo, entre as v pessoas que compram passagem, comparece ao vôo com probabildade p. Qual o risco de que o número de pessoas que comparecem ao vôo supere o número de assentos s? Ora, primeiro note que se pode supor independência na análise da questão. Qual probabilidade $r(j)$ de aparecerem j dos v passageiros que compraram a passagem? Resposta:

$$r(j) = \frac{v!}{(v-j)!\,j!}\, p^j\,(1-p)^{n-j}.$$

Logo, a probabilidade de que o número de pessoas que comparecem a um certo vôo supere o número de assentos s é

$$\sum_{j=s+1}^{v} r(j).$$

Definição 10: Fixados $(\Omega, P, \mathcal{A})$, $(V, \mathcal{V}) = (V, \mathcal{P}(V))$, onde $V \subset \mathbb{R}$ é finito ou enumerável, e uma função mensurável $X : (\Omega, P, \mathcal{A}) \to (V, \mathcal{P}(V))$, chamamamos de integral de X em relação a P, o valor real

$$\sum_{x \in V} x\, P(\{w \text{ tal que } X(w) = x\}),$$

e que será denotado por $\int X dP = \int X(w) dP(w)$

É usual chamar $\int X dP$ de *esperança da variável aleatória* X.

Note que dadas as variáveis aleatórias $X : \Omega \to V \subset \mathbb{R}$ e $Y : \Omega \to V \subset \mathbb{R}$, então $\int (X+Y)dP = \int XdP + \int YdP$.

Vamos calcular agora $E(X_1)$ no caso do exemplo do processo independente em que jogamos sucessivamente um dado.

Segue da definição acima que

$$\begin{aligned} E(X_1) &= \int X_1 dP \\ &= 1.P(X_1 = 1) + 2.P(X_1 = 2) = 1\,.\,1/3\; + 2\,.\,2/3\; = 5/3. \end{aligned}$$

Este valor corresponde ao lucro médio esperado quando se joga a moeda uma vez.

Note que neste caso também vale para qualquer tempo i que $E(X_i) = 5/3$.

Para saber a riqueza acumulada até a terceira jogada deveríamos considerar a variável aleatória $S_3 = X_1 + X_2 + X_3$.

A variável S_3 está definida para elementos

$$(w_1, w_2, w_3) \in \{1,2,3,4,5,6\}^3.$$

Para simplificar a notação é usual considerar que embora inicialmente $X_1 : \{1,2,3,4,5,6\} \to \{1,2\}$, podemos usar a mesma expressão X_1 para denotar $X_1 : \{1,2,3,4,5,6\}^3 \to \{1,2\}$, ou seja, $X_1(w_1,w_2,w_3) = X_1(w_1)$.

O mesmo vale analogamente para X_2 e X_3. Este procedimento é completamente geral e será utilizado no texto em outros casos similares sem nenhuma menção a esta pequena sutileza. Logo X_1, X_2, X_3 e S_3 podem ser consideradas todas sobre o mesmo domínio $\{1,2,3,4,5,6\}^3$.

Como a integral é aditiva na funcão obtemos então que

$$E(S_3) = E(X_1) + E(X_2) + E(X_3) = 3\,.\,5/3 = 5.$$

S_3 descreve o ganho em três jogadas e é uma variável aleatória tomando valores em $\mathbb{R}$. Ainda, $E(S_4) = E(X_1) + E(X_2) + E(X_3) + E(X_4) = 20/3$.

Se tivermos que estimar o que se ganha na média neste jogo em três jogadas, a resposta será $E(S_3) = 5$.

Podemos considerar outros jogos em que dependendo da face que sai uma moeda quando esta é lançada quatro vezes seguidas se ganha uma certa quantia, ou então um jogo de roleta, etc...

Da mesma forma como antes, se pode calcular a esperança do ganho auferido em cada um destes diversos jogos. Se tivermos que tomar uma decisão sobre qual jogo seria mais lucrativo jogar, a escolha mais sábia seria participar daquele que tem o maior valor esperado do ganho.

Vamos apresentar mais um exemplo da importância do cálculo do valor esperado. No caso da venda de passagem da companhia aérea discutido previamente, assuma que para cada passageiro que execeda os s assentos disponíveis será necessário pagar uma diária de hotel (até que saia o próximo vôo no dia seguinte) de 200,00 reais. Qual será a despesa média esperada $L(v)$ oriunda do procedimento de vender v passagens com $v > s$ para um avião de s lugares?

A resposta é

$$L(v) = \sum_{j=s+1}^{v} r(j) \times (j-s) \times 200,00.$$

Agora devemos calcular, em função do valor do preço de cada passagem, o lucro obtido com a venda de v passagens e comparar com o prejuízo $L(v)$ oriundo do eventual comparecimento de mais de s passageiros com a venda de v passagens.

Para simplificar nossa análise não estamos levando em conta a influência no número de assentos necessários no avião no dia seguinte (na eventualidade de comparecimento de mais de s passageiros num certo dia).

A questão relevante para a companhia aérea seria então encontrar para qual v ocorre o valor máximo de lucro esperado. Não vamos discutir aqui as eventuais questões éticas envolvidas no problema.

Após estes exemplos vamos voltar ao nosso tópico principal, ou seja os Processos Estocásticos.

Voltemos agora ao caso geral de processo estocástico $(X_t)_{t\in\mathbb{N}}$, sobre um espaço de probabilidade $(\Omega, P, \mathcal{A})$ fixado, onde $X_t : \Omega \to S$ para um certo conjunto finito ou enumerável S. O caso que realmente nos interessa é quando $t \in \mathbb{N}$, ou seja, t ilimitado. Neste caso, podemos, por exemplo, analisar o comportamento limite de um caminho amostral $(w_t)_{t\in\mathbb{N}}$ (com probabilidade 1) obtido do processo, etc...

Por exemplo, no caso em que $S = \{1,2\}$, podemos nos perguntar: fixado $\omega = (w_1, w_2, w_3, .., w_n, ...)$ será que existe o limite

$$\lim_{n\to\infty} \frac{\#\,\text{vezes que aparece o 1 entre os valores } w_1, w_2, w_3, ..., w_{n-1}}{n}.$$

Como se pode calcular tal valor? Este tipo de resultado ocorre, por exemplo, para processos independentes e é contemplado pela Lei dos Grandes Números, um dos teoremas mais importantes da Probabilidade (ver [G] ou [N] para enunciado preciso).

Existem $\omega = (w_t)_{t\in\mathbb{N}}$ para os quais tal limite não existe. O que desejamos saber é se existe um conjunto $K \subset \Omega = \{1,2\}^{\mathbb{N}}$ tal que $P(K) = 1$ e para $\omega \in K$, vale que existe o limite acima. Esta é a essência da visão probabilística (ou estatística) de analisar o problema.

Concretamente, no caso em que se joga sucessivamente uma moeda, e associamos 1 à cara e 2 à coroa, se $X_n \in \{1,2\}$ determina a face que saiu na jogada n-ésima, sabemos intuitivamente que quando a moeda é jogada um número grande de vezes, na média, na metade delas sai cara. Este fato segue da Lei dos Grandes Números e do fato que o processo X_n, neste caso, é independente.

Note que é **possível** que saia infinitas vezes cara, o que corresponderia ao evento $\omega = (1,1,1,1,1,1,1,1,..)$. O fato é que elementos desta forma estão fora de um conjunto $K \subset \Omega = \{1,2\}^{\mathbb{N}}$ de probabilidade 1 onde vale que

$$\lim_{n\to\infty} \frac{\#\,\text{vezes que aparece o 1 entre os valores } w_1, w_2, ..., w_{n-1}}{n} = \frac{1}{2}.$$

Voltemos ao exemplo em que uma companhia vende v lugares para um avião de s lugares, onde $v > s$. Vamos supor que a cada dia é oferecido um vôo seguindo esta política, onde v e s estão fixos. Suponha que X_n descreva o número de passageiros que aparecem no dia n. O processo estocástico X_n toma valores em $S = \{0, 1, 2, 3, .., v\}$. É razoável supor que X_n é um processo independente. Conforme calculado antes, $L(v)$ é o valor esperado de gasto com hotel oriundo desta política. Uma das conseqüências da Lei dos Grandes Números é que, neste caso, se calcularmos o valor médio durante, digamos, 100 dias, iremos obter aproximadamente o valor $L(v)$. Mais precisamente, seja L_n o gasto no dia n com hotel para passageiros que não conseguem lugar no vôo (quando $X_n > s$), então da LGN segue que

$$\lim_{n\to\infty} \frac{L_1 + L_2 + ... + L_n}{n} = L(v).$$

Sendo assim, o gasto estimado em 100 dias seria aproximadamente $100\, L(v)$.

Desejamos esclarecer um ponto importante sobre como são em geral introduzidos os modelos na teoria.

Considere $T = \mathbb{N}$, as variáveis $X_t : \Omega \to S$, $t \in \mathbb{N}$, a probabilidade P, etc... Como sempre S é finito, ou, se infinito, então enumerável. Considere agora para n fixo, uma seqüência também fixada de tempos ordenados, $t_1 < t_2 < ... < t_n \in T$.

Considere uma seqüência $A_1, A_2, ..., A_n \subset S$ fixada. Considere

$$P(\{\, w \in \Omega \ \text{ tais que } \ X_{t_1}(w) \in A_1,\, X_{t_2}(w) \in A_2, ...,\, X_{t_n}(w) \in A_n\}) =$$

$$P(X_{t_1} \in A_1,\, X_{t_2} \in A_2, ...,\, X_{t_n} \in A_n).$$

Dizemos que $P(\, X_{t_1} \in A_1,\, X_{t_2} = A_2, ...,\, X_{t_n} = A_n\,)$ determina a distribuição conjunta das variáveis $X_{t_1}, X_{t_2}, ..., X_{t_n}$ quando exaurimos os possíveis

$$A_1, A_2, ..., A_n \subset S$$

A informação acima descrita e que é fornecida pelas distintas possibilidades de todos os n, $t_1 \leq t_2 \leq .. \leq t_n$, e $A_1, A_2, ..., A_n \subset S$ são denominadas das distribuições finito dimensionais. Ela é obtida a partir do conhecimento explícito de P, de Ω das X_t, onde $t \in T = \mathbb{N}$ etc...

As distintas classes de Processos Estocásticos que são analisadas em diversos livros, em geral, não são apresentadas do modo descrito inicialmente. Se assume em cada modelo, na maioria das vezes, que o processo satisfaz certas propriedades baseadas na distribuições finito dimensionais. Isto é natural quando se busca encontrar modelos e não se parte de um exemplo concreto. Após obtermos uma série de distintos modelos teóricos é que fará sentido analisar um determinado problema do mundo real. Então, nos perguntaremos: em qual dos diversos modelos teóricos anteriormente estudados melhor se encaixa o fenômeno natural em análise? Quanto maior for a riqueza e diversidade de modelos que tivermos a nossa disposição, com mais precisão será descrito o fenômeno natural e melhor poderão ser as nossas previsões futuras a seu respeito. Como se diz, "se na sua caixa de ferramentas, a única que está disponível é um martelo, todo problema vai lhe parecer com um prego".

É necessário que as informações das distribuições finito dimensionais satisfaçam certas regras de compatibilidade.

Mais exatamente, são necessárias apenas duas condições de compatibilidade: a primeira, denominada aditiva, é que para qualquer n, $j \in \{1, 2, ..., n\}$, $t_1 \leq t_2 \leq ... \leq t_n \in T$, $A_1, A_2, ..., A_{j-1}, A_{j+1}, ..., A_n \subset S$ e $A_j, B_j \subset S$ fixados, vale que

$$P(X_{t_1} = A_1, .., X_{t_{j-1}} = A_{j-1}, X_{t_j} \in A_j, X_{t_{j+1}} = A_{j+1}, .., X_{t_n} = A_n) +$$

$$P(X_{t_1} = A_1, .., X_{t_{j-1}} = A_{j-1}, X_{t_j} \in B_j, X_{t_{j+1}} = A_{j+1}, .., X_{t_n} = A_n) =$$

$$P(X_{t_1} = A_1, ..., X_{t_{j-1}} = A_{j-1}, X_{t_j} \in A_j \cup B_j, X_{t_{j+1}} = A_{j+1}, ..., X_{t_n} = A_n).$$

A segunda condição de compatibilidade afirma que para qualquer n, $j \in \{1, 2, ..., n\}$,

$$t_1 < t_2 < ... < t_{j-1} < t_j < t_{j+1} < ... < t_n \in T,$$

$$a_1, a_2, ..., a_{j-1}, a_{j+1}, ..., a_n \in S$$

fixados, vale que

$$\sum_{r \in S} P(X_{t_1} = a_1, .., X_{t_{j-1}} = a_{j-1}, X_{t_j} = r, X_{t_{j+1}} = a_{j+1}, .., X_{t_n} = a_n) =$$

$$P(X_{t_1} = a_1, ..., X_{t_{j-1}} = a_{j-1}, X_{t_{j+1}} = a_{j+1}, ..., X_{t_n} = a_n).$$

A condição acima descreve apenas o seguinte: fixado um tempo t_j qualquer, somar as probabilidades sobre todas as possibilidades r em S, é a mesma coisa que não fazer restrição alguma no tempo t_j

Repetindo, em geral não se parte do conhecimento explícito da P, de Ω, das $X_t : \Omega \to S$, etc..., mas sim supõe-se que P satisfaça algo (em termos das distribuições finito dimensionais) além, é claro, das condições de compatibilidade.

Um teorema de Kolmogorov afirma que se as informações fornecidas pelas distribuições finito dimensionais satisfazem as tais regras de compatibilidade, então existem P, Ω e $\mathcal{F}$, tais que a partir destes se recupera o que foi dado de maneira implícita.

Sendo assim, nos referiremos no futuro a primeira forma de introduzir um processo estocástico como explícita e a esta última forma (a partir das distribuições finito-dimensionais) como implícita.

No teorema de Kolmogorov, no caso em que $T = \mathbb{N}$, o conjunto Ω pode ser escolhido como $S^{\mathbb{N}}$, e as variáveis X_t, $t \in \mathbb{N}$, de tal forma que se $w = (a_1, a_2, a_3, ..., a_t, ...) \in S^{\mathbb{N}}$ então $X_t(w) = a_t$.

Desta forma, por exemplo,

$$\{w \in \Omega | X_1(w) = a_1, X_2 = a_2, X_4 = a_4\} =$$

$$\{a_1\} \times \{a_2\} \times S \times \{a_4\} \times S^{\mathbb{N}} \subset S^{\mathbb{N}}.$$

Para a formalização completa dos tópicos descritos acima referimos o leitor a alguns bons textos de Probabilidade e Processos Estocásticos (ver por exemplo [B], [D], [N], [G], [KT1], [KT2]) .

Observamos que P, Ω e $\mathcal{F}$ não são únicos. Ou seja, podemos obter distintos P, Ω etc... a partir da mesma informação obtida das distribuições finito-dimensionais. Dito isto, a partir de agora, um processo estocástico $\{X_n\}_{n \in \mathbb{N}}$ tomando valores em S será, nada mais nada menos que uma probabilidade P sobre o conjunto $\Omega = S^{\mathbb{N}}$. Ainda, se $\omega = (w_0, w_1, w_2, ..., w_n, ..) \in \Omega = S^{\mathbb{N}}$, então $X_n(w) = w_n$. Sobre a sigma-álgebra $\mathcal{F} \subset \mathcal{P}(\Omega)$ definida no conjunto $\Omega = S^{\mathbb{N}}$, diremos no momento apenas que ela contém todos os conjuntos da forma

$$\{w \in \Omega | X_{t_1}(w) = a_1, X_{t_2} = a_2, ..., X_{t_n} = a_n\} \subset S^{\mathbb{N}},$$

onde $n \in \mathbb{N}$, $t_1 < t_2 < ... < t_n$ e $a_1, a_2, ..., a_n \in S$.

Vamos dar um exemplo interessante e que é descrito de maneira implícita.

Uma matriz $\mathcal{P} = (P_{ij})_{i,j=1,2,..,n}$ da forma n por n é dita estocástica se os elementos P_{ij} da matriz são não negativos e a soma de cada linha é igual a 1.

Para simplificar a exposição considere $S = \{1, 2\}$ e $T = \mathbb{N}$.

Neste caso, um exemplo de matriz estocástica seria

$$\mathcal{P} = \begin{pmatrix} 1/3 & 2/3 \\ 4/7 & 3/7 \end{pmatrix}.$$

Um vetor $\pi = (\pi_1, \pi_2)$ é denominado de vetor de probabilidade sobre $S = \{1, 2\}$ se $\pi \in \mathbb{R}^2$ é tal que $\pi_1, \pi_2 \geq 0$ e $\pi_1 + \pi_2 = 1$.

Por exemplo $\pi = (2/5, 3/5)$.

Fixado $\mathcal{P}$ e π vamos definir primeiro a probabilidade P sobre $\Omega = \{1, 2\}^{\mathbb{N}}$ para certos tipos de conjunto.

Por definição,

$$P(X_0 = a_0, X_{t_1} = a_1, X_{t_2} = a_2, X_{t_3} = a_3, ..., X_{t_n} = a_n) =$$

$$P(\{w \in \Omega \,|\, X_0 = a_0, X_{t_1} = a_1, X_{t_2} = a_2, X_{t_3} = a_3, ..., X_{t_n} = a_n\}) =$$

$$\pi_{a_0} \mathcal{P}^{t_1}_{a_0\,a_1} \mathcal{P}^{t_2-t_1}_{a_1\,a_2} \mathcal{P}^{t_3-t_2}_{a_2\,a_3} .... \mathcal{P}^{t_n-t_{n-1}}_{a_{n-1}\,a_n},$$

onde $n \in \mathbb{N}$, $a_i \in \{1, 2\}$ e $0 = t_0 < t_1 < t_2 < ... < t_n$.

Por exemplo, no caso da matriz dois por dois e do vetor de probabilidade descrito acima, obtemos que

$$\frac{12}{105} = \frac{3}{5}\,\frac{4}{7}\,\frac{1}{3} = \pi_2\, \mathcal{P}_{2\,1}\, \mathcal{P}_{1\,1} =$$

$$P(\{2\} \times \{1\} \times \{1\} \times \{1, 2\}^{\mathbb{N}}) = P(X_0 = 2, X_1 = 1, X_2 = 1).$$

Afirmamos que, neste caso, as regras de compatibilidade estão satisfeitas para tal P e assim pelo Teorema de Kolmogorov esta P pode ser considerada sobre uma certa sigma-álgebra $\mathcal{F} \subset \mathcal{P}(\Omega)$. Todos os conjuntos da forma

$$\{w \in \Omega \,|\, X_0 = a_0, X_{t_1} = a_1, X_{t_2} = a_2, X_{t_3} = a_3, ..., X_{t_n} = a_n\},$$

estão em $\mathcal{F}$.

Este processo é nosso primeiro exemplo de Processo de Markov. Em breve vamos falar um pouco mais detalhadamente sobre tais processos.

Os resultados acima são descritos detalhadamente no que se denomina de Teoria da Medida ([B],[C],[F]).

Vamos apresentar agora uma ilustração concreta da maneira implícita de apresentar um processo estocástico. Para isto será necessário descrever, antes de mais nada, uma propriedade muito importante.

Regra de Bayes: Seja $(P, \Omega, \mathcal{A})$ um espaço de probabilidade. Considere duas variáveis aleatórias, uma X tomando valores em S_1 e outra Y tomando valores em S_2.

Então, para $s_1 \in S_1$ fixo

$$P(X = s_1) = \sum_{s_2 \in S_2} P(X = s_1 | Y = s_2)\ P(Y = s_2).$$

Esta propriedade segue trivialmente de

$$\begin{aligned} P(X = s_1) &= \sum_{s_2 \in S_2} (X = s_1 \,\mathrm{e}\, Y = s_2) \\ &= \sum_{s_2 \in S_2} P(X = s_1) | Y = s_2)\ P(Y = s_2), \end{aligned}$$

que por sua vez segue da propriedade b) da definição 2.

Uma versão um pouco mais geral desta propriedade afirma:

Regra de Bayes: Seja $(P, \Omega, \mathcal{A})$ um espaço de probabilidade. Considere uma variável aleatória X tomando valores em S e um conjunto $A \in \mathcal{A}$.

Então,

$$P(w \in A) = \sum_{s \in S} P(w \in A | X = s)\ P(X = s).$$

A demonstração deste fato é a mesma do caso anterior.

De maneira heurística podemos dizer que a regra de Bayes desempenha em probabilidade um papel semelhante ao do teorema fundamental no Cálculo. Mais exatamente, um informação global (uma integral no Cálculo) $P(w \in A)$ é obtida através de uma totalização de informações localizadas (a derivada no Cálculo) $\sum_{s \in S} P(w \in A | X = s)\ P(X = s)$.

A análise das diversas propriedades de um processo estocástico será tão mais complexa quanto mais intensa forem as relações de dependência entre as variáveis. Nesta hierarquia de dificuldade, os mais simples são os processos independentes, depois seguem os Processos de Markov que serão analisados brevemente a seguir.

Considere uma pessoa que possui um capital inicial $c > 0$, onde c é um número natural, e que vai participar de um jogo em que uma moeda (com probabilidade $1/2$ de sair cara e $1/2$ de sair coroa) é lançada sucessivamente.

O jogador ganha um real se sai cara e perde um real se sai coroa.

Ele gostaria de atingir um capital fixado $C > c > 0$. O jogo termina quando a fortuna do jogador atinge o valor C ou o valor 0. É natural supor que se o capital inicial c é grande e próximo de C, então existe maior probabilidade de o jogador atingir seu objetivo de alcançar a fortuna C do que quando o capital inicial c for próximo de zero. Como quantificar tais probabilidades? Vamos denotar por $p(c)$ a probabilidade de o jogador entrar em bancarrota, ou seja, atingir ao longo do jogo a fortuna 0 (antes de atingir o valor C).

Em princípio, apenas sabemos que $p(0) = 1$ e $p(C) = 0$.

Vamos modelar este problema através de um processo estocástico. Considere c fixado. Neste caso, tome o espaço de estados $S=\{0, 1, 2, .., C\}$, $T = \mathbb{N}$, X_t é o valor da fortuna no tempo t. Considere $\Omega = \{1, 2, ..., C\}^{\mathbb{N}}$ e a probabilidade $P = P_c$ será descrita a seguir. O ponto fundamental é que vamos evitar dizer quem é P explicitamente.

Note que $P(\{w | X_0 = c\}) = 1$, $P(\{w | X_0 \neq c\}) = 0$. Ainda, $P(X_1 = c) = 0$.

É mais natural descrever P através de condicionais, mais exatamente,

$$P(X_{t+1} = d + 1\ | X_t = d) = 1/2,$$

$$P(X_{t+1} = d - 1\ | X_t = d) = 1/2,$$

para qualquer $t \geq 1$ e $1 \leq d \leq C-1$.

É claro que segue do estabelecido acima que fixado d, para qualquer $b < d-1$, ou $b > d+1$, ou mesmo $b = d$, vale que

$$P(X_{t+1} = b \,|X_t = d) = 0.$$

Ainda, é natural assumir que

$$P(X_{t+1} = 0 \,|\, X_t = 0) = 1,$$

$$P(X_{t+1} = C \,|\, X_t = C) = 1.$$

Isto é o jogo termina ao ser alcançado um dos valores 0 ou C.

Note que estas informações implícitas e ainda a informação $P(X_0 = c) = 1$ são suficientes para calcular o valor de $P(A)$ para um conjunto qualquer $A \subset \Omega = \{1, 2, ..., C\}^{\mathbb{N}}$ que depende apenas de finitas restrições temporais.

De fato, por exemplo, considere c fixado,

$$P(X_0 = c\,,\, X_1 = c+1) = P(X_1 = c+1 \,|\, X_0 = c)\, p(X_0 = c) = \frac{1}{2}\,.\,1 = \frac{1}{2}.$$

Note que é natural no presente exemplo que, para $i \geq 0$, e d_i, d_{i+1}, $d_{i+2} \in S$ valha

$$P(X_{i+2} = d_{i+2} \,|\, X_i = d_i,\, X_{i+1} = d_{i+1}) = P(X_{i+2} = d_{i+2} \,|\, X_{i+1} = d_{i+1}),$$

pois o valor de X_i não influencia na probabilidade de $X_{i+2} = d_{i+2}$. É claro que a probabilidade de $X_{i+2} = d_{i+2}$ é influenciada pelo valor $X_{i+1} = d_{i+1}$. Por exemplo,

$$P(X_{i+2} = 5 \,|\, X_{i+1} = 3) = 0,$$

mas

$$P(X_{i+2} = 5 \,|\, X_{i+1} = 4) = \frac{1}{2}.$$

A propriedade

$$P(X_{i+2} = d_{i+2} \,|\, X_i = d_i,\, X_{i+1} = d_{i+1}) = P(X_{i+2} = d_{i+2} \,|\, X_{i+1} = d_{i+1}),$$

será denominada de markoviana, e ela afirma em sentido literário que a probabilidade do que vai acontecer num certo tempo depende apenas do valor da variável aleatória no tempo imediatamente anterior. Os processos de Markov serão aqueles que possuem tal propriedade.

Seja $d_2 \in S$, sob as hipóteses acima, podemos calcular via condicionais o valor $P(X_0 = c,\, X_1 = c+1,\, X_2 = d_2)$.

De fato,

$$
\begin{aligned}
&P(X_0 = c,\, X_1 = c+1,\, X_2 = d_2) = \\
&P(X_2 = d_2 \mid X_0 = c,\, X_1 = c+1)\, P(X_0 = c\,,\, X_1 = c+1) = \\
&P(X_2 = d_2 \,|X_1 = c+1)\, P(X_0 = c\,,\, X_1 = c+1).
\end{aligned}
$$

Observe agora que a informação $P(X_2 = d_2\,|X_1 = c+1)$ nos foi fornecida e o valor $P(X_0 = c\,,\, X_1 = c+1)$ já foi calculado antes. Logo, podemos obter o desejado.

Sendo assim, por exemplo, no caso em que $c = 1$,

$$P(\{w|X_0(w) = 1, X_1(w) = 2, X_2(w) = 1\}) = (\frac{1}{2})^2.$$

É usual denotar a probabilidade P quando assumimos que $X_0 = c$, ou seja, quando o capital inicial for c, por P_c. Este ponto de vista será importante a seguir. Vamos analisar a função $p(c)$ como função de c.

Este procedimento que nos permitiu calcular probabilidades de conjuntos A que dependem de três coordenadas pode ser também aplicado para conjuntos A que dependem de m coordenadas temporais através de um procedimento indutivo semelhante ao descrito acima.

Em resumo, para cada c fixado, as condicionais (e mais a informação $P(\{X_0 = c\}) = 1$) nos permitiram calcular completamente as

$$P_c(\, X_{t_1} = a_1, .., X_{t_{j-1}} = a_{j-1}, X_{t_j} = a_j, X_{t_{j+1}} = a_{j+1}, .., X_{t_n} = a_n),$$

e, neste caso, é mais natural a introdução do processo estocástico de maneira implícita do que por via explícita.

Pode-se mostrar que as condições de compatibilidade estão satisfeitas neste caso.

Para cada c, em função das condições de compatibilidade, considere pelo teorema de Kolmogorov o processo $X_t : \Omega \to S$, $t \in \mathbb{N}$, P_c, a probabilidade sobre $\Omega = \{0,1,2,...,C\}^{\mathbb{N}}$ e a sigma-álgebra $\mathcal{F}$ (de subconjuntos de Ω). Destacamos o fato que esta sigma-álgebra $\mathcal{F}$ **não é** o conjunto $\mathcal{P}(\Omega)$. A sigma-álgebra $\mathcal{F}$ não depende de c.

Lembre que as famílias de variáveis aleatórias X_t que consideramos aqui sempre satisfazem a propriedade: se $w = (w_0, w_1, w_2, .., w_n, ...)$ então $X_t(w) = w_t$.

Nosso objetivo é resolver o problema: quem é $p(c)$? Ou seja, calcular a probabilidade $P_c(B)$ do conjunto

$$B_c = B = \{w = (w_0, w_1, w_2, w_3, ...w_t, ..) \text{ tal que } w_0 = c,\ w_t = 0$$

$$\text{para algum } t > 0, \text{ e ainda } w_s \neq C \text{ para todo } r \text{ tal que } 0 < r < t\ \} =$$

$$\{w \,|\, X_0(w) = c,\ X_t(w) = 0 \text{ para algum } t \text{ e } X_r \neq C \text{ para } 0 < r < t\}.$$

Note que o conjunto B_c acima depende de infinitas informações.

Este conjunto B_c, pode-se mostrar (não faremos isto aqui neste momento), está na sigma-álgebra $\mathcal{F}$ (que existe em função das condições de compatibilidade). Fica claro, deste modo, a necessidade de se ter de considerar na teoria uma sigma-álgebra aonde fica bem definido o conceito de probabilidade. Apenas as distribuições finito-dimensionais não bastam e conjuntos mais complexos, que dependem de infintas coordenadas, aparecem de maneira natural em problemas concretos.

Existem eventos $w \in \Omega$ que nunca atingem qualquer dos valores C ou 0. Por exemplo, no caso c=2 e C=5, o evento w=$(w_0, w_1, w_2, w_3, ... w_t, ...) = (2,3,2,3,2,3,2,3,2,3...)$ possui tal propriedade.

Considere o conjunto

$$D = D_c = \{w = (w_0, w_1, w_2, w_3, ...w_t, ..) \in \Omega = \{0,1,2,...,C\}^{\mathbb{N}},$$

$$\text{tal que } w_0 = c,\ w_t \neq C \text{ e } w_t \neq 0 \text{ para todo } t \in \mathbb{N}\}.$$

O conjunto D está também em $\mathcal{F}$ e pode-se mostrar que $P(D) = P_c(D) = 0$. Este resultado é descrito em [N] [G] quando é analisado o assim chamado passeio aleatório.

Para ilustrar tal afirmação ao leitor vamos analisar um caso particular. Considere $C = 3$ e fixemos $c = 1$, então o conjunto D acima tem apenas um elemento

$$D = \{(1, 2, 1, 2, 1, 2, 1, 2, 1, 2, ...)\}.$$

Destacamos aqui o fato que um conjunto com apenas um elemento pode ter, eventualmente, probabilidade positiva.

Note que

$$D = \cap_{n=1}^{\infty} \{w | X_0 = 1, X_1 = 2, X_2 = 1, X_3 = 2, ..., X_{2n-1} = 2, X_{2n} = 1\},$$

e portanto D é um conjunto mensurável em $\mathcal{F}$ (obtido por interseção de conjuntos de $\mathcal{F}$ indexados por um conjunto enumerável $n \in N$) e assim faz sentido perguntar pelo valor $P(D)$.

Como vale a propriedade que se $V \subset U$, então $P(V) \leq P(U)$ (pois, $P(V) \leq (P(V) + P(U - V)) = P(U)$), então, dado $n \in \mathbb{N}$,

$$P(D) \leq P(\{w | X_0 = 1, X_1 = 2, X_2 = 1, X_3 = 2, ...,$$

$$X_{2n-1} = 2, X_{2n} = 1\}) = (\frac{1}{2})^{2n-1},$$

para todo n. Note que tomando n grande o valor $(\frac{1}{2})^{2n-1}$ se torna arbitrariamente pequeno. Logo $P(D) = 0$.

Esclarecidos estes pontos de grande importância, vamos voltar ao nosso problema.

Considere fixado o número natural c. Segue da regra de Bayes que para c tal que $C > c > 0$, vale que

$$p(c) = P_c(B_c) = P_c(\{w \in B_c\}) =$$

$$P_c(\{w \in B_c \,|\, X_1 = c + 1\}) \, P_c(\{X_1 = c + 1\}) +$$

$$P_c(\{w \in B_c \,|\, X_1 = c - 1\}) \, P_c(\{X_1 = c - 1\}) =$$

$$P_c(\{w \in B_c \,|\, X_1 = c + 1\}) \, \frac{1}{2} + P_c(\{w \in B_c \,|\, X_1 = c - 1\}) \, \frac{1}{2}.$$

O ponto fundamental agora é que para c tal que $0 < c < C$, vale que $p(c + 1) = P_c(\{w \in B_c \,|\, X_1 = c + 1\})$ e $p(c - 1) = P_c(\{w \in B_c \,|\, X_1 = c - 1\})$.

A afirmação acima, que é absolutamente intuitiva, requer uma prova formal a qual não será apresentada aqui.

Chegamos assim a equação de diferenças

$$P(\{w \mid X_0(w)=c,\ X_t(w)=0 \text{ para algum } t\})=p(c)=\frac{1}{2}(p(c-1)+p(c+1)),$$

com a condição inicial e final respectivamente $p(0) = 1$ e $p(C) = 0$. A solução desta equação é

$$p(c) = 1 - \frac{c}{C},$$

como pode ser confirmado por substituição na equação.

Esta fórmula dá o valor exato da dependência da probabilidade $p(c)$ em função da proximidade de c a C.

Em conclusão: utilizando a Regra de Bayes, ou seja condicionando, obtivemos uma equação de diferenças e a partir daí obtivemos a solução do problema que buscávamos. Não foi necessário calcular a probabilidade de nenhum conjunto específico! Variações destas idéias nos permitem obter soluções explícitas do valor de certas probabilidades que se deseja encontrar em inúmeros casos.

Uma das Regras de Ouro da Probabilidade: na dúvida, condicione!

Vamos dar a seguir um outro exemplo da maneira de descrever uma família de processos estocásticos através de uma maneira implícita.

Definição 11: Dizemos que o processo X_t tomando valores em S (enumerável) com parâmetro $t \in T = \mathbb{N}$ é estacionário se para cada n e cada seqüência $t_1 < t_2 < ... < t_n$, onde $t_i \in T$, $t > 0$ e para cada seqüência $a_1, a_2, ..., a_n$, onde $a_i \in S$ vale que

$$P(X_{t_1} = a_1,\ X_{t_2} = a_2, ...,\ X_{t_n} = a_n) =$$

$$P(X_{t_1+t} = a_1,\ X_{t_2+t} = a_2, ...,X_{t_n+t} = a_n).$$

Os processos estacionários são aqueles em que um deslocamento uniforme de um valor t, em todos os tempos envolvidos na distribuição conjunta, não altera esta.

Neste caso, por exemplo, para s fixo qualquer em S

$$P(X_1 = s) = P(X_{1+1} = s) = P(X_2 = s).$$

Consideramos acima $t = 1$, $n = 1$, $a_1 = s$ e $t_1 = 1$.

Ainda,

$$P(X_3 = s) = P(X_{2+1} = s) = P(X_2 = s) = P(X_1 = s),$$

e assim por diante...

Estas propriedades podem não ser verdadeiras se o processo X_t não é estacionário. Por exemplo, no caso do jogo da moeda descrito anteriormente em que o capital inicial era c, temos que $P(X_0 = c) = 1$, mas $P(X_1 = c) = 0$. Logo, neste caso, o processo não é estacionário.

Muitos processos importantes são estacionários.

Vamos finalizar este artigo com um exemplo de processo estacionário. Seja $S = \{1, 2, 3\}$ e $T = \mathbb{N}$.

Seja uma matriz $\mathcal{P} = (P_{ij})_{i,j=1,2,3}$ da forma três por três tal que os elementos P_{ij} da matriz são positivos e a soma de cada linha é igual a 1.

Tal matriz é um caso particular de matriz estocástica.

Por exemplo,

$$\mathcal{P} = \begin{pmatrix} 1/3 & 1/3 & 1/3 \\ 4/7 & 2/7 & 1/7 \\ 2/5 & 2/5 & 1/5 \end{pmatrix}$$

é uma matriz que satisfaz as condições acima estipuladas.

Seja um vetor $\pi = (\pi_1, \pi_2, \pi_3)$ de probabilidade sobre $S = \{1, 2, 3\}$, isto é, se $\pi \in \mathbb{R}^3$ é tal que $\pi_1, \pi_2, \pi_3 \geq 0$ e $\pi_1 + \pi_2 + \pi_3 = 1$.

Fixado $\mathcal{P}$ e π vamos definir primeiro a probabilidade P sobre $\Omega = \{1, 2, 3\}^{\mathbb{N}}$ para certos tipos de conjunto.

Por definição,

$$P(\, X_0 = a_0, X_{t_1} = a_1, X_{t_2} = a_2, X_{t_3} = a_3, ..., X_{t_n} = a_n\,) =$$

$$\pi_{a_0}\, \mathcal{P}^{t_1}_{a_0\, a_1}\, \mathcal{P}^{t_2 - t_1}_{a_1\, a_2}\, \mathcal{P}^{t_3 - t_2}_{a_2\, a_3} .... \mathcal{P}^{t_n - t_{n-1}}_{a_{n-1}\, a_n},$$

onde $n \in \mathbb{N}$, $a_i \in \{1, 2, 3\}$ e $0 = t_0 < t_1 < t_2 < ... < t_n$.

Pode-se mostrar que estão satisfeitas as condições de compatibilidade de Kolmogorov no presente caso.

Uma pergunta natural é; que propriedade deve possuir π para que o processo acima definido seja estacionário?

Ora

$$P(X_1 = a_1) = \sum_{a_0=1}^{3} P(X_0 = a_0, X_1 = a_1) = \sum_{a_0=1}^{3} \pi_{a_0}\, \mathcal{P}_{a_0\, a_1}.$$

Neste caso, fazendo a_1 variar em $\{1, 2, 3\}$ obtemos que em forma matricial, a expressão acima significa

$$\begin{pmatrix} \pi_1 & \pi_2 & \pi_3 \end{pmatrix} \begin{pmatrix} P_{11} & P_{12} & P_{13} \\ P_{21} & P_{22} & P_{23} \\ P_{31} & P_{32} & P_{33} \end{pmatrix} =$$

$$\begin{pmatrix} P(X_1 = 1) & P(X_1 = 2) & P(X_1 = 3) \end{pmatrix},$$

onde P_{ij} são as entradas da matriz $\mathcal{P}$.

Sendo assim, se desejamos que para todo $a_1 \in \{1, 2, 3\}$ valha que

$$P(X_1 = a_1) = P(X_0 = a_1) = \pi_{a_1},$$

então é necessário que

$$\pi\, \mathcal{P} = \pi.$$

Acima estamos fazendo o produto de um vetor π (uma matriz 3 por 1) por uma matriz $\mathcal{P}$ (três por três) e obtendo o vetor π.

Ou seja, buscamos um autovetor associado ao autovalor 1 (estamos acima considerando a matriz $\mathcal{P}$ agindo em vetores a esquerda).

Sendo assim π ser autovetor de $\mathcal{P}$ associado ao autovalor 1 é condição necessária para que o processo seja estacionário. Pode-se mostrar também que esta condição é suficiente (ver [N], [G]).

Outra pergunta natural é se tal autovetor π é único. A resposta no caso da matriz $\mathcal{P}$ em análise é que tal π é único.

Isto é sempre verdade para qualquer matriz estocástica $\mathcal{P}$ (da forma n por n) que tem todas as entradas estritamente positivas. Vamos mostrar este fato no caso de matrizes n por n.

Definição 12: Fixado S com cardinalidade n denotamos por $\Sigma = \Sigma_S$ o seguinte conjunto:

$$\Sigma = \{(p_1, p_2, ..., p_n) \mid \text{tal que } p_i \geq 0,\ p_1 + p_2 + ... + p_n = 1\}.$$

Proposição 1: *Seja $\mathcal{P}$ matriz estocástica da forma n por n. A função $T : \Sigma \to \Sigma$, tal que $u = T((p_1, p_2, ..., p_n)) = p\,\mathcal{P}$, onde $p = (p_1, p_2, ..., p_n)$, está bem definida.*

Demonstração: Para ver que T está bem definida analisemos primeiro que o caso em que $p = e_s \in \Sigma$ (o vetor que é nulo a menos da posição $s \in S$) para um s fixo. Denote então $v^s = e_s\mathcal{P} = T(e_s)$. As ordenadas de v^s são dadas por $(v_j^s)_{j\in S} = v^s = e_s\,\mathcal{P} = (P(s,j))_{j\in S}$.

Por exemplo, no caso de matriz estocástica 3 por 3, se $e_2 = (0, 1, 0)$ então

$$\begin{pmatrix} 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} P_{11} & P_{12} & P_{13} \\ P_{21} & P_{22} & P_{23} \\ P_{31} & P_{32} & P_{33} \end{pmatrix} = \begin{pmatrix} P_{21} & P_{22} & P_{23} \end{pmatrix} = v^2,$$

Note que $\sum_{j\in s} v_j^s = \sum_{j\in s} P(s,j) = 1$ (pois a matriz é estocástica). Logo $v^s \in \Sigma$.

Vamos agora ao caso geral: ora, um $p \in \Sigma$ qualquer é dado por $p = \sum_{s\in S} p_s e_s$, onde $\sum_{s\in S} p_s = 1$; o resultado então segue por linearidade: seja $(u_j)_{j\in S} = u = p\,\mathcal{P} = T(p)$, ora,

$$T(\sum_{s\in S} p_s e_s) = \sum_{s\in S} p_s T(e_s) = \sum_{s\in S} p_s v^s = u = (u_j)_{j\in S}.$$

Agora

$$\sum_{j\in S} u_j = \sum_{j\in S}\sum_{s\in S} p_s\, v_j^s = \sum_{s\in S}\sum_{j\in S} p_s\, v_j^s = \sum_{s\in S} p_s \sum_{j\in S} v_j^s = \sum_{s\in S} p_s\, 1 = 1.$$

Logo, $u = T(p) \in \Sigma$.

$\square$

No resultado acima não se exige que $\mathcal{P}$ tenha todas as entradas estritamente positivas.

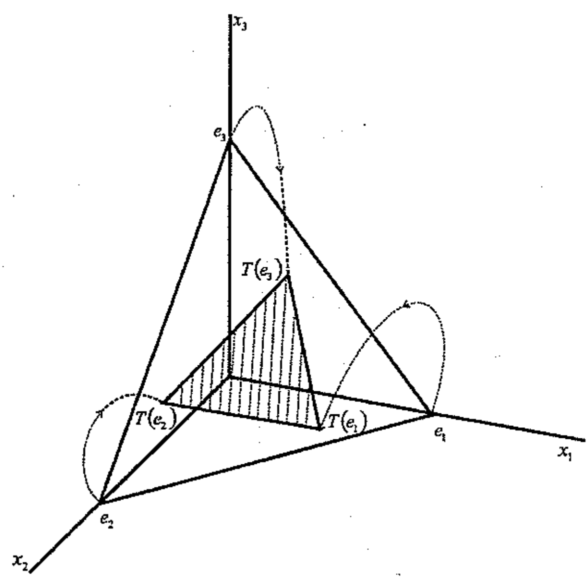

Figura 1

O Teorema do ponto fixo de Brouder [L] afirma que qualquer função contínua de um disco de $\mathbb{R}^{n-1}$ em si mesmo tem um ponto fixo. Segue da proposição acima (e do fato que Σ é homeomorfo a um disco de $\mathbb{R}^{n-1}$) que sempre se pode encontrar um autovetor π associado ao autovalor 1 para $\mathcal{P}$ estocástica (veja figura 1 que exibe o caso em que $n = 3$). De fato, considere $T : \Sigma \to \Sigma$ (que é contínua) e aplique o teorema. Logo, dada a matriz $\mathcal{P}$ podemos escolher π que nos determina pelo procedimento acima um processo estocástico estacionário.

Se a matriz estocástica $\mathcal{P}$ (da forma n por n) tiver todas as entradas positivas, a transformação T acima agindo em Σ será uma contração (ver definição em [L]) e assim o ponto fixo π será único. Deixamos a demostração deste fato a cargo do leitor. Logo, neste caso, o autovetor π associado ao autovalor 1 para $\mathcal{P}$ é único.

Para concluir gostaríamos de dizer que a Teoria dos Processos Estocásticos é atualmente uma área de intenso trabalho de pesquisa em Matemática. Problemas de grande complexidade e sofisticação matemática aparecem de maneira natural. Além disto, os modelos que são analisados nesta teoria são de grande utilidade em aplicações à diversas áreas do conhecimento: Engenharia, Biologia, Medicina, Economia, Geologia, Química, Metereologia, etc...

Referências

[B] P. Billingsley, Probability and Measure, John Wiley

[D] R. Durret, Probability: theory and examples, Duxbury Press

[G] G. Grimmett and D. Stirzaker, Probability and Random Processes, Oxford Press

[J] B. James, Probabilidade: um curso de nível intermediário, Projeto Euclides, IMPA

[C] A. Castro Junior, Um curso de Teoria da Medida, Projeto Euclides, IMPA

[F] P. Fernandez, Teoria da Medida, Projeto Euclides, IMPA

[KT1] S. Karlin and H. Taylor, A First Course in Stochastic Processes, Academic Press

[KT2] S. Karlin and H. Taylor, A Second Course in Stochastic Processes Academic Press

[L] E. L. Lima, Espaços Métricos, Projeto Euclides, IMPA

[N] J. Norris, Markov Chains, Cambridge Press

Artur O. Lopes
Instituto de Matemática - UFRGS
alopes@mat.ufrgs.br