Conceitos Básicos de Processos Estocásticos

Professor Gregorio Saravia Atuncar

Departamento de Estatística

Universidade Federal de Minas Gerais

2011

Introdução

Nos anos que tenho ministrado a disciplina de Processos Estocásticos no Curso de Graduao em Estatística da Universidade Federal de Minas Gerais tenho notado que não temos um texto adequado e acessível para os estudantes, principalmente em Português.

Pretendemos, com essas notas de aula, tentar superar esse problema e esperamos dessa maneira contribuir no aprendizado da disciplina.

Assumimos, para acompanhar essas notas, que o estudante tem familiaridade com Cáalculo Integral e Algebra Linear. Assumimos também que tem familiaridade com Conceitos de Probabilidade em nível inicial. Mesmo assim, fazemos, no primeiro capítulo um resumo dos conceitos básicos de Probabilidade.

As primeiras anotações que levaram a essa versão estão baseadas em anotações de aulas de estudantes que foram meus alunos na disciplina. Em particular agradeço a Ismênia B. de Magalhães e Fernanda N. de Assis que emprestaram suas notas de aula. Muitos estudantes contribuiram com observações feitas em versões preliminares. Citar todos eles sería impossível. Agradeço também a Clodio P. de Almeida pela sua assistência na preparação da versão final.

Belo Horizonte, 24 de maio de 2011.

Gregorio Saravia Atuncar

Contents

1 Capítulo 1. Introdução e Revisão de Probabilidade.

Faremos neste capítulo uma breve revisão dos conceitos mais importantes de Probabilidade. Assumimos que o leitor está familiarizado com o tópico. A quem não estiver, recomendamos uma leitura cuidadosa de Ross (2002), por exemplo.

Na segunda seção serão apresentados os conceitos básicos de Processos Estocásticos.

1.1 Probabilidade

Imaginemos um experimento aleatório Ξ com espaço amsotral S. Consideremos n repetições independentes de Ξ e seja $A \subset S$.

Definição 1.1.1 Define-se a probabilidade frequentista do evento A como

$$P(A) = lim_{n \to \infty} \frac{n_A}{n}.$$

Nessa definição, n_A é o numero de vezes em que o evento A ocorre.

O leitor pode facilmente verificar que :

(A1)$P(A) \geq 0$,

(A2) $P(S) = 1$

(A3) Se A e B são dois eventos disjuntos, então $P(A \cup B) = P(A) + P(B)$

(A1), (A2) e (A3) são conhecidos como axiomas de probabilidade. Diremos neste caso que P é aditiva.

Quando substituimos (A3) por

(A3') $P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$, se A_1, A_2,. . . é tal que $A_i \cap A_j = \emptyset$ para $i \neq j$,

diremos que P é σ - aditiva.

No que segue, adotaremos a definição axiomática de probabilidade. Isto é, ao falar de uma probabilidade P, assumimos que P é uma função que a cada evento $E \subset S$, associa sua probabilidade $P(A)$ satisfazendo (A1), (A2) e (A3) ou (A3').

A partir dos axiomas sobre uma probabilidade aditiva, pode-se provar que:

P1. Para A e B dois eventos quaisquer, $P(A \cup B) = P(A) + P(B) - P(A \cap B)$,

P2. $P(A^c) = 1 - P(A)$, onde A^c é o evento complementar de A,

P3. Se $A \subset B$, então $P(A) \leq P(B)$,

P4. $P(A) \leq 1$,

P5. Se A_1, A_2, . . ., A_n são eventos disjuntos, então $P(\bigcup_{i=1}^n A_i) = \sum_{i=1}^n P(A_i)$

Deixamos como exercício para o leitor provar essas propriedades a partir dos axiomas. A propriedade 1 pode ser estendida para mais de dois eventos. Por exemplo, se A, B e C são tres eventos, tem-se que

$$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C)$$

A partir de P1 temos também que $P(A \cap B) = P(A) + P(B) - P(A \cup B)$. Como $P(A \cup B) \leq 1$, temos então que

$$P(A \cap B) \geq P(A) + P(B) - 1$$

Essa ultima relação é conhecida como Desigualdade de Bonferroni. Tal resultado pode ser estendido para mais de dois eventos. Observe que

$$P(A \cap B \cap C) = P(A \cap (B \cap C)) \geq P(A) + P(B \cap C) - 1,$$

usando novamente a desigualdade de Bonferroni, teremos que $P(B \cap C) \geq P(B) + P(C) - 1$. Então finalmente obtemos que $P(A \cap B \cap C) \geq P(A) + P(B) + P(C) - 2$

Em geral, se A_1, A_2, . . ., A_n são n eventos quaisquer, pode-se provar usando o principio de indução matematica que

$$P(A_1 \cap A_2 \cap ... \cap A_n) \geq P(A_1) + P(A_2) + ...P(A_n) - (n-1)$$

Uma probabilidade P satisfazendo (A1), (A2) e (A3), satisfaz (A3') se e somente se satisfaz a propriedade chamada continuidade no vazio que estabelecemos, sem prova, a seguir.

Antes de estabelecer essa propriedade, precisamos da seguinte

Definição 1.1.2: Dizemos que uma sequência de eventos A_1, A_2, . . ., decresce monotonamente para o vazio se para cada n, $A_{n+1} \subset A_n$ e $\cap_{n=1}^{\infty} = \emptyset$. Adotaremos a notação $A_n \downarrow \emptyset$

Continuidade no vazio: Se $A_n \downarrow \emptyset$, então $P(A_n) \downarrow 0$.

Naturalmente essa propriedade pode ser estabelecida em forma mais geral. Dizemos que a sequência de eventos A_1, A_2, . . .; decresce monotonamente para A se para cada n, $A_{n+1} \subset A_n$ e $\cap_{n=1}^{\infty} A_n = A$. A notação $A_n \downarrow A$ será adotada. Se $A_n \downarrow A$, então $P(A_n) \downarrow P(A)$. Diremos que a sequência A_1, A_2, . . cresce monotonamente para o evento A se para cada n, $A_n \subset A_{n+1}$ e $\cup_{n=1}^{\infty} A_n = A$. Se $A_n \uparrow A$, então $P(A_n) \uparrow P(A)$.

Um conceito muito importante em Teoria de Probabilidade é o de Probabilidade Condicional. Seja um evento A tal que $P(A) > 0$. Para um outro evento B, define-se a probabilidade condiconal de B dado que o evento A ocorreu, como a razão $\frac{P(A \cap B)}{P(A)}$. Denota-se essa probabilidade por $P(B/A)$. Isto é,

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

.

O leitor pode verificar que $P(B/A)$ satisfaz os tres axiomas de probabilidade, mas agora com o espaço amostral restrito a A. Ou seja:

(A1C) $P(B/A) \geq 0$,

(A2C): $P(A/A) = 1$,

(A3C). Se B e C são dois eventos disjuntos, então $P((B \cup C)/A) = P(B/A) + P(C/A)$.

Deixamos como exercicio para o leitor checar essas propriedades. E a partir desses axiomas, formalizar e provar propriedades similares a P1-P5. Se P é σ - aditiva, a probabilidade condicional também é σ - aditiva e portanto contínua no vazio.

Exemplo 1.1.1 Consideremos o problema do lançamento de dois dados. Calculemos a probabilidade de que a soma dos resultados seja maior que 10 dado que a soma dos resultados é par.

Solução. O espaço amostral associado a esse experimento contem 36 elementos, todos equiprováveis. Defina os eventos:

A: Soma dos resultados é par, B: Soma maior que 10

$P(A) = \frac{18}{36} = 0,5$. Os únicos resultados em que a soma é par e a soma é maior que 10 é o resultado (6,6). Sendo assim, $P(B/A) = \frac{1/36}{0,5} = \frac{1}{18}$.

Recomendamos ao leitor que ao abordar um problema sobre calculo de probabilidade, defina claramente os eventos envolvidos no problema. Deve ser, sempre, esse o primeiro passo. Isso não somente facilita a soluçãoo do problema como facilita também a comunicação com eventuais interlocutores.

Exemplo 1.1.2. Consideremos uma linha de produção em grande escala que produz 4% de itens defeituosos. Tres itens dessa linha são escolhidos ao acaso para inspeção. Calculemos a probabilidade de que seja escolhido no maximo um item defeituoso.

Solução. O espaço amostral associado ao experimento de escolher tres itens é $S = \{DDD, DDN, DND, NDD, NND, NDN, DNN, NNN\}$. Nesse espaço, DND, por exemplo representa defeituosos na primeira e terceira retiradas e não defeituoso na segunda.

Defina o evento A: No máximo um item defeituoso é retirado. Esse evento pode ser representado por
$A = \{NND, NDN, DNN, NNN\}$ e
$P(A) = P(NND)+P(NDN)+P(DNN)+P(NNN) = 3(0,96)^2(0,04)+(0,96)^3 = 0,9953$.

Observe que estamos assumindo que os resultados das retiradas são independentes. Essa suposição, neste casso é razoável pois a linha de produção é em grande escala e portanto as retiradas estão sendo feitas de um numero grande de itens. Em casos como esses podemos fazer tal aproximação . A maneira de ilustração, suponha que durante um período de operação foram produziods 10.000 itens. Em média serão produzidos 400 itens defeituosos. A probabilidade de que o segundo item observado seja defeituoso, dado que o primeiro foi defeituoso é igual a $\frac{399}{9999} = 0,0399$ que é bem próximo de $0,04$ que é a probabilidade de que o segundo seja defeituoso. Essa aproximação já não é boa se temos um número pequeno de itens produzidos. Por exemplo, se são produzidos 100 itens as correspondentes probabilidades são 0,0303 e 0,04.

Calcular, nesse exemplo, a probabilidade de que nenhum defeituoso foi observado dado que no maximo um defeituoso foi retirado. Para abordar esse ultimo problema, defina o evento B: Nenhum defeituoso foi observado. $P(B/A) = \frac{P(A \cap B)}{P(A)} = \frac{(0,96)^3}{0,9953} = 0,8889$

Assim como definimos $P(B/A)$, podemos também definir $P(A/B) = \frac{P(A \cap B)}{P(B)}$ e a partir dessas definições temos que

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

Essa última relação é conhecida como **Regra de multiplicação**. Tal regra pode ser estendida a mais de dois eventos. isto é, por exemplo

$$P(A \cap B \cap C) = P(A \cap B)P(C/A \cap B) = P(A)P(B/A)P(C/(A \cap B)).$$

Em geral, sejam A_1, A_2, . . ., A_n eventos quaisquer, a probabilidade da intersecção desses eventos é dada por

$$P(A_1 \cap A_2 \cap ... \cap A_n) = P(A_1)P(A_2/A_1)...P(A_n/(A_1 \cap A_2... \cap A_{n-1}))$$

Observe que no exemplo 1.2, $P(B) = \frac{3}{36}$ que é diferente de $P(B/A) = \frac{1}{18}$. Quando $P(B/A) = P(A)$, diremos que os eventos A e B são independentes.

Uma definição equivalente: Os eventos A e B são independentes se $P(A \cap B) = P(A)P(B)$.

A prova dessa equivalência é simples: Suponha que $P(B/A) = P(B)$. Pela regra da multiplicação temos que $P(A \cap B) = P(A)P(B/A)$. Substituindo o segundo fator, temos $P(A \cap B) = P(A)P(B)$. Reciprocamente, suponha que $P(A \cap B) = P(A)P(B)$, então $P(B/A) = \frac{P(A)P(B)}{P(A)} = P(B.)$

Um resultado muito importante é o Teorema de Bayes que estabelecemos a seguir. Antes disso daremos a seguinte definição:

Diremos que os eventos E_1, E_2, . . ., E_k formam uma partição do espaço amostral S se :
(i) $E_i \cap E_j = \emptyset$ para $i \neq j$,
(ii) $\bigcup_{i=1}^{k} E_j = S$.

Teorema da Probabilidade Total. Consideremos a partição E_1, E_2, . . ., E_k do espaço amostral S e um evento F quaisquer. Então

$$P(F) = \sum_{i=1}^{k} P(E_i)P(F/E_i).$$

A prova é muito simples: O evento F pode ser representado por uma união de eventos disjuntos pois

$F = F \cap (\bigcup_{i=1}^{k} E_i) = \bigcup_{i=1}^{k}(F \cap E_i)$.

A segunda igualdade é verdadeira pela propriedade distributiva da interseção com respeito à união.

Desde que $E_i \cap E_j = \emptyset$, teremos que $(F \cap E_i) \cap (F \cap E_j) = \emptyset$ para $i \neq j$. Então, pela propriedade P5 temos que $P(F) = \sum_{i=1}^{k} P(E_i \cap F)$. Aplicando a regra do produto a cada termo da soma temos que $P(F) = \sum_{i=1}^{k} P(E_i)P(F/E_i)$.

Como consequência imediata do resultado anterior temos o Teorema de Bayes que estabelece que se E_1, E_2, . . ., E_k é uma partição do espaço amostral S, então para $j = 1, ..., k$,

$$P(E_j/F) = \frac{P(E_j)P(F/E_j)}{\sum_{i=1}^{k} P(E_i)P(F/E_i)}$$

1.2 Variáveis Aleatórias

Introduzimos nesta seção o conceito de variável aleatória. O leitor deve lembrar que no exemplo 2, os reultados do experimento aleatório não são representados por numeros. O objetivo de uma variável aleatória é justamente quantificar os resultados de um experimento aleatório. Nesse exemplo, contemos o numero de defeituosos retirados e chamemos ao resultado de X. Assim, $X(NNN) = 0$, $X(NND) = X(NDN) = X(DNN) = 1$, $X(DDN) = X(DND) = X(NDD) = 2$ e $X(DDD) = 3$. Fazendo essa associação, o nosso espaço amostral original transformou-se em $R_X = \{0, 1, 2, 3\}$.

Formalizando, diremos que uma variável aleatória real, X, é uma função que associa um numero real a cada resultado de um experimento aleatório. Assumiremos sempre que X é uma variável aleatória real. Desde que uma variável aleatória X é uma função, o conjunto de todos os possíveis valores que ela pode assumir será chamado de contradominio de X e será representado por R_X.

Uma variável aleatória pode ser discreta ou contínua. Diremos que é discreta se assume um número finito ou infinito enumerável de valores. Caso contrario, diremos que X é uma variável aleatória continua.

Se X é uma variável aleatória discreta definimos a função de probabilidade como a função que a cada valor de X associa a sua correspondente probabilidade. Isto é, $p_j = P(X = x_j)$.

No exemplo 1.1.2 temos

$p_0 = P(X = 0) = P(NNN) = (0.96)^3 = 0.884736$,

$p_1 = P(X = 1) = P(DNN) + P(NDN) + P(NND) = 3(0.04)(0.96)^2 = 0.110592$,

$p_2 = P(X = 2) = P(DDN) + P(DND) + P(NDD) = 3(0.04)^2(0.96) = 0.004608$ e

$p_3 = P(X = 3) = (0.04)^3 = 0.000064$.

Tais valores podem ser apresentados da forma seguinte:

Table 1: Função de Probabilidade

0	0.884736
1	0.110592
2	0.004608
3	0.000064

A partir da função de probabilidade de uma variável aleatória discreta, definimos a função de distribuição, que associa a cada valor real, x, a probabilidade da variável X ser menor ou igual que x. Formalmente a função de distribuição F availada em x, é definida por

$$F(x) = P(X \leq x)$$

.

A partir da definição temos que se a variável aleatória discreta assume os valores x_1, x_2, . . ., x_k, a função de distribuição está definida por

$$F(x) = P(X \leq x) = \sum_{j:x_j \leq x} p_j$$

Observe que se $x < y$, o evento $\{a \in \Re : a \leq x\} \subset \{a \in \Re : a \leq y\}$, então $F(x) \leq F(y)$. Isto é, a função de distribuição é não decrescente. O leitor pode provar também que F é contínua à direita e que $\lim_{x \rightarrow\ -\infty} F(x) = 0$ e $\lim_{x \rightarrow \infty} F(x) = 1$. Essas últimas propriedades decorrem da continuidade no vazio.

A função de distribuição tem sido definida a partir da função de probabilidade. Podemos também fazer o caminho inverso, isto é, definir a função de probabilidade a partir da função de distribuição. De fato, se $x_1 < x_2 < ... < x_k$ são os valores que X pode assumir, temos que $p(x_j) = F(x_j) - F(x_{j-1})$.

A esperança de uma variável aleatória discreta é definida por

$$E(X) \;=\; \sum_{j=1}^{k} x_j P(X = x_j).$$

Denotaremos por μ_X a esperança de X e quando não houver lugar a confussão, denotaremos apenas por μ. Se Y é uma função de X, isto é, se $Y = g(X)$, a esperança de Y pode ser calculada como

$$E(Y) = \sum_{j=1}^{k} g(x_j) P(X = x_j).$$

Um caso particular é quando $g(X) = (X - \mu_X)^2$. Temos neste caso,

$$\begin{aligned} E(Y) &= E(X - \mu_X)^2 \\ &= \sum_{j=1}^{k} (x_j - \mu_X)^2 P(X = x_j), \end{aligned}$$

e essa será chamada Variância de X e será denotada por σ_X^2. O leitor pode provar que $\sigma_X^2 = E(X^2) - (E(X))^2$. Pode provar também , como um outro caso particular, fazendo $g(X) = aX + b$, que $E(aX + b) = aE(X) + b$.

Um outro caso particular é quando $g(X) = X^n$, $E(g(X)) = E(X^n) = \sum_{j=1}^{k} x_k^n P(X = x_k)$ é chamado o momento de ordem n de X.

Quando X é uma variável aleatória continua, existe uma função f, não negativa tal que $\int_{-\infty}^{\infty} f(x)dx = 1$. Tal função é chamada função de densidade de probabilidade de X. Se $A \subset \Re$, definimos $P(A) = \int_A f(x)dx$.

Se $A = [a, b]$, pela definição de P temos que $P(A) = P(a \leq X \leq b) = \int_a^b f(x)dx$. Observe que $P(X = a) = P(a \leq X \leq a) = \int_a^a f(x)dx = 0$. Isto é, se X é variável aleatória contínua, a probabilidade de que ela assuma um valor específico é zero para quaisquer valor real. Como consequência disso temos que se X é contínua, então

$$P(a < X \leq b) = P(a \leq X \leq b) = P(a \leq X < b) = P(a < X < b)$$

Se X é variável aleatória contínua, a função de distribuição é definida por $F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)dt$. Feita essa definição, temos que, pelo Teorema Fundamental do Calculo que $f(x) = F^{'}(x)$ se F é derivável em x.

Em forma análoga ao caso discreto, definimos a esperança de X como

$$E(X) = \int xf(x)dx.$$

E se $Y = g(X)$, $E(Y) = \int g(x)f(x)dx$.

Consideremos agora duas variáveis aleatórias, X e Y e consideremos o vetor (X, Y). Chamaremos a (X, Y) um vetor aleatório. Consideremos primeiro o caso em que tanto X quanto Y são discretas. A função que a cada ponto (x_i, y_j) do contradominio do vetor (X, Y) associa sua probabilidade será chamada função de probabilidade conjunta do vetor (X, Y), isto é

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

A partir da função de probabilidade conjunta, podemos encontrar as funções de probabilidade de X e de Y. Cada uma delas é chamada função de probabilidade marginal de X e de Y respectivamente.

A obtenção de cada uma dessas funções é uma aplicação do Teorema da Probabilidade Total. Observemos que o vetor (X, Y) transforma um espaço amostral S no espaço amostral $R_X X R_Y$ (Produto cartesiano de R_X e R_Y). Nesse novo espaço podemos definir uma partição fazendo $E_i = \{(x_i, y_j) : y_j \in R_Y\}$. Claramente $E_i \cap E_k = \emptyset$ se $i \neq k$ e $\cup_{i:x_i \in R_X} E_i = R_X X R_Y$.

Então para $A = \{Y = y_j\}$, temos que

$$P(A) = P(Y = y_j) = \sum P(A \cap E_i) = \sum_{i:x_i \in R_X} P(X = x_i, Y = y_j).$$

Em forma análoga podemos obter a função de probabilidade marginal de Y.

Dado que $X = x_i$, define-se a função de probabilidade condicional de Y como

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

Quando para todo i e para j, tem-se que

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

dizemos que X e Y são independentes. Equivalentemente, diremos que X e Y são independentes se para todo i e para todo j , $P(X = x_i, Y = y_j) = P(X = x_i)P(Y = y_j)$.

Em forma analoga, define-se a função de probabilidade condicional de X dado que $Y = y_j$.

Os momentos condicionais de Y dado que que $X = x_j$ são simplesmente os momentos da função de probabilidade condicional, isto é, por exemplo, a esperança condicional de Y dado que $X = x_i$, é definida por

$$E(Y/X = x_j) = \sum_{j:y_j \in R_Y} y_j P(Y = y_j/X = x_i).$$

Se $W = g(Y)$, define-se a esperança condicional de W dado que $X = x_i$ como

$$E(W/X = x_i) = E(g(Y)/X = x_i) = \sum_{j:y_j \in R_Y} g(y_j) P(Y = y_j/X = x_i).$$

Em particular, $E(Y^2/X = x_i) = \sum_{j:y_j \in R_Y} y_j^2 P(Y = y_j/X = x_i)$.

Define-s e a variancia condicional de Y dado que $X = x_i$ por

$$Var(Y/X = x_i) = E(Y^2/X = x_i) - (E(Y/X = x_i))^2.$$

Quando os dois componentes do vetor (X, Y) são continuos, existe uma função não negativa f tal que para $A \subset \Re^2$, tem-se que $P((X, Y) \in A) = \int \int_A f(x, y) dx dy$. Tal função f é chamada função de densidade conjunta do vetor (X, Y) e satisfaz a condição $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y)dxdy = 1$.

Não provaremos aqui, mas pode-se obter as funções de densidade marginais tanto de X quanto de Y a partir da função de densidade conjunta de (X, Y), a saber:

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y)dy.$$

$$f_Y(y) = \int_{-\infty}^{\infty} f(x, y)dx.$$

Definem-se também as funções de densidade condicionais de X e de Y.

$$f(x/y) = \frac{f(x, y)}{f(y)},$$

$$f(y/x) = \frac{f(x, y)}{f(x)}.$$

Diremos que X e Y são independentes se

$$f(x/y) = f(x),$$

para todo x e todo y. Equivalentemente se para todo x e todo y, $f(x, y) = f_X(x)f_Y(y)$, diremos que X e Y são independentes.

Exemplo 1.2.1 Considere o vetor (X, Y) com função de densidade conjunta

$$f(x, y) = ae^{-(x+2y)} \qquad 0 < x < y < \infty$$

Encontrar as funções de densidade marginais de X e de Y e a função de denisdade condicional de Y dado que $X = 2$.

Solução: O valor de a é encontrado a partir da condição de que a integral sobre o plano da função de densidade conjunta é igual a 1, isto é $1 = \int_0^{\infty} \int_x^{\infty} ae^{-(x+2y)}dydx$. Resolvendo a integral obtem-se $a = 6$.

A função de densidade marginal de X é igual a

$$f_X(x) = \int_x^{\infty} 6e^{-(x+2y)}dy = 3e^{-x} \int_x^{\infty} 2e^{-2y}.$$

Resolvendo a integral, obtemos $f(x) = 3e^{-3x}$. O procedimento vale para $x > 0$. Então

$$f_X(x) = 3e^{-3x} \qquad x > 0.$$

Para $y > 0$,

$$f_Y(y) = \int_0^y 6e^{-(x+2y)}dx = 6e^{-2y}\int_0^y e^{-x}dx = 6e^{-2y}(1-e^{-y}).$$

Então

$$f_Y(y) = 6e^{-2y}(1-e^{-y}) \qquad y > 0.$$

Se $X = 4$,

$$f(y/4) = \frac{f(4,y)}{f_X(4)} = \frac{6e^{-(4+2y)}}{3e^{-3(4)}} = 2e^{-2(y-4)} \qquad y > 4$$

Em forma análoga ao caso discreto, definem-se os momentos condicionais de Y dado que $X = x$ como os momentos das funções de densidade condicionais.

Dado um vetor (X, Y), e $W = g(X, Y)$, define-se a esperança de W como

$E(W) = \sum_i \sum_j g(x_i, y_j)P((X, Y) = (x_i, y_j))$ no caso discreto e

$E(W) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x, y)f(x, y)dxdy$ no caso continuo.

A covariancia entre X e Y, denotada por σ_{XY} é definida por $E[(X-\mu_X)(Y-\mu_Y)]$, isto é

$$\sigma_{XY} = E[(X-\mu_X)(Y-\mu_Y)]$$

Deixamos como exercicio para o leitor, calcular σ_{XY} no exemplo anterior.

Um outro exercicio é provar que $\sigma_{XY} = E(XY) - E(X)E(Y)$.

O vetor $\underline{m} = (E(X), E(Y))$ é chamado vetor de médias do vetor (X, Y) e a matriz

$$\Sigma = \begin{pmatrix} \sigma_X^2 & \sigma_{XY} \\ \sigma_{XY} & \sigma_Y^2 \end{pmatrix}$$

é chamada matriz de variancias e covariancia do vetor (X, Y). A desigualdade ede Cauchy-Schwarz garante que o determinante deΣ é não negativa. O determinante é 0 apenas se Y ou X é uma função linear da outra.

Exemplo 1.2.2 Calcule a esperança condicional de Y dado que $X = 4$.

Solução. No exemplo anterior calculamos a função de densidade condicional de Y dado que $X = 4$, então

$$E(Y/X = 4) = \int_4^\infty 2ye^{-2(y-4)}dy = \frac{1}{2} + 4.$$

1.3 Definições Básicas de Processos Estocásticos

Antes de iniciar com as definições básicas, falemos de alguns exemplos simples que ajudarão entender os conceitos envolvidos na teoria de Processos Estocasticos. Voce leitor, esta familiarizado com Variáveis aleatórias e/ou vetores aleatórios. Nos anos que tenho ministrado a disciplina de Processos Estocasticos para o Curso de Graduação em Estatistica da UFMG, tenho percebido que o estudante tem certa dificuldade quando esses conceitos são introduzidos.

Nas minhas primeiras leituras sobre Processos Estocásticos encontrei o seguinte exemplo que pode não ser útil do ponto de vista técnico mas ajuda entender o que é um processo estocastico. Tal exemplo fala de uma pessoa que vamos chamar de João. João recebe um cavalo de presente, mas ele não entende nada sobre cavalos. Depois de um tempo queria cortar a cauda do cavalo. Ele perguntou para um entendido sobre a altura à qual deveria cortar a cauda. O entendido disse para ele que corte à altura que ele achar conveniente pois seja qual for a altura, terá alguem que ache muito curta a cauda e terá alguem que ache muito comprida. E acrescentou: Mesmo voce, depois de cortar, pode mudar de ideia no dia seguinte.

Esse exemplo permite visualizar que a altura da cauda depende da pessoa que está opinando(w) e também do instante de tempo em que a pessoa está opinando. Quer dizer, se X representa o comprimento da cauda do cavalo, X é uma função de w e n, sendo w a pessoa que emite a opinião e n o dia em que tal pessoa emite a opinião.

Podemos, então, definir $X = \{X(w, n) : w \in \Omega, n \in I\}$ como um Processo estocastico.
Para w fixo, $X(w) = \{X(w, n) : n \in I\}$ será uma realização do processo.

Pensemos num segundo exemplo: Considere uma linha de produção. Essa linha possui 100 máquinas. Todo dia, escolhe-se uma máquina e observa-se o número de peças defeituosas por ela produzidas. Neste caso, $X = \{X(w, n) : w = 1, 2, ..., 100; n = 1, 2, ...\}$ define o processo associado ao número de peças defeituosas produzidas por máquina.

Definição 1.3.1. Definimos formalmente um processo estocastico $X = \{X_t : t \in I\}$ como uma familia de variáveis aleatòrias . I será chamado espaço de parâmetros. Em nossa disciplina assumiremos $I = [0, \infty)$ ou $I = \{0, 1, 2, ...\}$. No primeiro caso diremos que o processo é com parametro de tempo continuo e no segundo, com parâmetro de tempo discreto.

Sabemos que uma variável aleatória real é uma função que a cada ponto w de um espaço amostral associa um número real. Assumiremos que no processo X, todas as variáveis estão definidas sobre um mesmo espaço amostral. O contradominio de todas as variáveis aleatórias será chamado espaço de estados do processo e denotaremos por E. Se E é discreto, diremos que o processo estocastico é discreto, se E é continuo, diremos que o processo é continuo.

Exemplo 1.3.1. Seja $X = \{X_t : t \geq 0\}$, onde X_t= Preço de uma ação no instante t. Nesse caso, $I = T = [0, \infty)$, $E = \{x : x > 0\}$.

Exemplo 1.3.2. $X = \{X_n : n \geq 0\}$ onde X_n = Número de itens defeituosos encontrados, no dia n, em uma linha de produção. Nesse caso, $I = \{0, 1, 2, ...\}$ e $E = \{0, 1, 2, ...\}$.

Desde que um processo é uma familia de variáveis aleatórias,podemos falar das distribuições delas. Para quaisquer conjunto finito $\{t_1, t_2, ..., t_k\} \subset I$, a distribuição do vetor $(X_{t_1}, X_{t_2}, ..., X_{t_k}\}$ é chamada a distribuição finito dimensional do processo. Em particular, para um valor de t fixo, a distribuição de X_t será a distribuição unidimensional de X_t. A totalidade das distribuições finito-dimensionais de um processo determinam , sobre condições gerais, a distribuição do processo. Esse resultado é dado pelo Teorema de Consistência de Kolmogorov. Infelizmente foge ao alcance da nossa disciplina. O leitor interessado pode ver, por exemplo, Kolmogorov (1956).

Um processo estocastico é chamado Gaussiano se todos os vetores finito-dimensionais possuim distribuição normal multivariada.

Associada a X_t temos sua média $\mu_t = E(X_t)$ e sua variância $\sigma_t^2 = Var(X_t)$. A função média do processo é definida por $\mu = \{\mu_t : t \in I\}$ e a função variancia por $\sigma^2 = \{\sigma_t^2 : t \in I\}$. A função covariancia de um processo é definida por $\{\sigma(s,t), s, t \in I\} = \{Cov(X_s, X_t), s, t \in I\}$.

Diremos que um processo estocastico é estacionário se $\mu_t = u$, $\sigma_t^2 = \sigma^2$ para todo t e

$Cov(X_s, X_t)$ é uma função que depende apenas de $|s - t|$. O processo será dito estritamente estacionário se para todo k, s e $0 < t_1 < t_2 < ... < t_k$ os vetores $(X_{t_1}, X_{t_2}, ..., X_{t_k})$ e $(X_{t_1+s}, X_{t_2+s}, ..., X_{t_k+s})$ possuim a mesma distribuição conjunta. Em particular, X_t e X_{t+s} são identicamente distribuidos para quaisquer valores de s e t. Alguns autores usam os conceitos de estacionario no sentido amplo e estacionario respectivamente.

Dado um processo estocastico, a diferença $X_{t+s} - X_s$ é definida como incremento do processo em um itervalo de comprimento t. Diremos que o processo possui incrementos independentes se os incrementos em intervalos disjuntos de tempo são independentes. Isto é, dados $0 < t_1 < t_2 < t_3 < t_4 < ... < t_k$, as variáveis $X_{t_2} - X_{t_1}$, $X_{t_3} - X_{t_2}$, . . . e $X_{t_k} - X_{t_{k-1}}$ são independentes.

Se a distribuição dos incrementos $X_{t+s} - X_t$ depende apenas do comprimento s do intervalo $(t, t + s]$, diremos que o processo possui incrementos estacionários.

De particular interesse em nossa disciplina serão os processos Markovianos. Diremos que um processo estocastico possui a propriedade de Markov se, dado o valor de X_t, as distribuições de X_s para $s > t$, não dependem dos valores de X_u para $u < t$. Quer dizer que o comportamento do processo no futuro não depende do passado quando o estado presente do processo é conhecido.

Definição 1.3.2. Formalmente, diremos que um processo estocastico possui a propriedade de Markov se para $0 < t_1 < t_2 < ... < t_n < t$,

$$P(X_t \in A/X_{t_1} = x_1, X_{t_2} = x_2, ..., X_{t_n} = x_n) = P(X_t \in A/X_{t_n} = x_n).$$

A função

$$P(x, s, t, A) = P(X_t \in A/X_s = x) \qquad t > s,$$

é chamada função de probabilidade de transição e é básica para o estudo dos processos de Markov.

A distribuição de X_0 é chamada distribuição inicial do processo. Se a distribuição inicial e a função de probabilidade de transição são conhecidas, todas as distrbuições finito dimensionais podem ser calculadas, como pode se ver a seguir: (Por simplicidade, assumir que o processo é discreto com espaço de estados finito $E = \{0, 1, 2, ..., M\}$)

a) Distribuições unidimensionais:

$$\begin{aligned} P(X_t = j) &= \sum_{i=0}^{M} P(X_0 = i, X_t = j) \\ &= \sum_{i=0}^{M} P(X_0 = i)P(X_t = j/X_0 = i) = \sum_{i=0}^{M} P(X_0 = i)P(i, 0, t, j) \end{aligned}$$

b) Distribuições finito dimensionais:

$$\begin{aligned} & P(X_{t_1} = i_1, X_{t_2} = i_2, ..., X_{t_k} = i_k) \\ = & \sum_{i=0}^{M} P(X_0 = i, X_{t_1} = i_1, X_{t_2} = i_2, ..., X_{t_k} = i_k) \\ = & \sum_{i=0}^{M} P(X_0 = i)P(X_{t_1} = i_1/X_0 = i)P(X_{t_2} = i_2/X_{t_1} = i_1)...P(X_{t_k} = i_k/X_{t_{k-1}} = i_{k-1}) \\ = & \sum_{i=0}^{M} P(X_0 = i)P(i, 0, t_1, i_1)P(i_1, t_1, t_2, i_2)...P(i_{t_{k-1}}, t_{k-1}, t_k, i_k). \end{aligned}$$

Daremos mais detalhes e exemplos no proximo capitulo em que abordarmos as cadeias de Markov.

1.4 Exercícios

1.4.1 Uma linha de produção, em grande escala, produz 2% de itens defeituosos. Todos os dias no inicio de operação, 5 itens são escolhidos para inspeção. Se um ou mais desses itens resultar defeituoso, as máquinas são calibradas.
a) Encontre a função de probabilidade do número de dias transcorridos até a primeira calibração,
b) Se nos primeiros 4 dias de operação não precisou calibrar as máquinas, qual é a probabilidade de que nos proximos 4 dias não precise calibração?,
c) Que hipotese voce está assumindo para abordar o problema?

1.4.2 A função de densidade de uma variável aleatória contínua é dada por $f(x) = \frac{1}{2}x$ para $0 < x < \alpha$ e 0 caso contrário.
a) Encontre a função de distribuição de X e calcule $P(X < 1/X < 1,5)$,

b) Encontre $Var(X)$.

1.4.3 A variável aleatória X tem função de distribuição dada por $F(a) = 0$ para $a \leq 0$ e $F(a) = 1 - e^{-0,05a}$ para $a > 0$.
a) Encontre a função de densidade de X,
b) Calcule $P(X > 20)$ e $P(X > 40/X > 20)$,
c) Encontre $E(X)$.

1.4.4 Considere X uma variável aleatória de Poisson com parámetro λ. Mostre que

$$E(X^n) = \lambda E\left\{(X+1)^{(n-1)}\right\}$$

Você sabe que $E(X) = \lambda$. Use o resultado anterior e calcule $Var(X)$ e $E(X^3)$

1.4.5 Considere a variável aleatória contínua X cuja função de densidade é dada por $f(x) = ax^2$ para $0 < x < 2$ e 0 caso contrário
a) Encontre a função de distribuição de X
b) Calcule $P(0.5 < X < 1.4|0.8 < X < 1.6)$

1.4.6 Considere o vetor (X, Y) distribuido uniformemente na região R definida por $R = \{(x, y) : 0 < y < 3, y - 3 < x < 3 - y\}$.
a) Encontre as funções de densidade marginal de X e de Y,
b) Calcule $P(-1 < X < 1/Y = 2)$ e $P(-1 < X < 1/1 < Y < 2)$,

1.4.7 a) Sejam X e Y variáveis aleatórias independentes continuas com funções de densidade f_X e f_Y respectivamente. Defina $Z = X + Y$. Prove que a função de distribuição de Z é dada por $F_Z(z) = \int_{-\infty}^{\infty} f_X(y)F_Y(z - y)dy$. A partir desse resultdo encontre a função de densidade de Z.
b) Considere X e Y i.i.d. $N(0, 1)$. Use o resultado de (a) e encontre a função de densidade da soma dessas variáveis aleatórias.

1.4.8 Considere X_1, X_2, . . ., X_n i.i.d. distribuidos de acordo a uma uniforme no intervalo $[a, b]$.
a) Defina $U_n = max\{X_1, X_2, ..., X_n\}$. Encontre a função de distribuição de U_n,
b) Considere $u < b$ e encontre $\lim_{n\to\infty} F_n(u)$,
c) Considere $u \geq b$ e encontre $\lim_{n\to\infty} F_n(u)$,

1.4.9 Sejam X e Y i.i.d. com distribuição geometrica com parametro p

a) Encontre a distribuição de $Z = X + Y$,
b) Se $Z = k$, encontre a distribuição de X.

1.4.10 Considere $X \sim exp(\lambda)$ e $Y \sim exp(\lambda)$. Suponha que X e Y são independentes, defina $U = X + Y$ e $V = \frac{X}{X+Y}$
a) Mostre que U e V são independentes,
b) Encontre a distribuição de V

1.4.11 a)Considere X e Y com distribuição binomial com parametros n, p e m,p respectivamente. Se X e Y são independentes, prove que $X + Y$ tem distribuição binomial. Quais são os parametros,
b) Qual a distribuição de X se $X + Y = r$?.

1.4.12. Resolva o Exercicio 1.4.11 se X e Y são independentes com distribuição de Poisson com parametros λ e β respectivamente.

1.4.13. Considere X e Y independentes com distribuição exponencial com parametros λ e β respectivamente. Defina $U = min\{X, Y\}$. Encontre a distribuição de U. **Obs.** Considere a representação $f(x) = \lambda e^{-\lambda x}$ para $x > 0$.

1.4.14. No Exercicio anterior, calcule $P(U = X)$.

1.4.15 Considere X_1, X_2, . . ., X_n indpendentes com distribuição exponencial com parametro λ_i respectivamente. Defina $U = min\{X_1, X_2, ..., X_n\}$. Encontre a distribuição de U e calcule $P(U = X_1)$.

1.5 Referências

Allen, A. O. Probability, Statistics and Queueing Theory with Computer Science Applications, 2d. ed. Academic Press, N. York. 1990.

Karlin, S. e Taylor, H. A First Course in Stochastic Processes, 2d. ed. Academic Press, N. York. 1975.

Kolmogorov, A. N. Foundations of the Theory of Probability. Second English ed. Chelsea Publishing Company, N. York. 1956.

Ross, S. A First Course in Probability, 6th ed. Prentice Hall, N. Jersey, 2002.

2 Capítulo 2. Cadeias de Markov

No capitulo anterior apresentamos algumas definições básicas sobre processos Markovianos. Neste capitulo concentraremos nossa atenção nos processos Markovianos com espaços de estados discretos. Neste caso diremos que o processo é uma cadeia de Markov. Concentraremos inicialmente no caso em que o parametro de tempo é discreto e o espaço de estados é finito. Exemplos comuns desse tipo de cadeias acontecem na área de controle de qualidade. Na inspeção de itens de uma linha de produção podem-se encontrar itens defeituosos e esses defeitos podem ser de varios tipos. É de interesse do inspetor saber se a sequencia de defeitos guarda uma certa estrutura. Um outro exemplo é o que é chamado de Caminho aleatório restrito que será descrito na seguinte seção.

Sem perda de generalidade assumiremos $E = \{0, 1, ..., M\}$. Apresentaremos um estudo detalhado apenas das cadeias de Markov de primeira ordem e daremos alguns detalhes sobre as cadeias de ordem superior. Neste capítulo, enquanto não digamos o contrário, estaremos falando das cadeias de Markov de primeira ordem

2.1 Conceitos Básicos

Definição 2.1.1 Diremos que o processo $X = \{X_n : n = 0, 1, ...\}$ é uma cadeia de Markov, se para todo n,

$$P(X_{n+1} = j/X_0 = i_0, X_1 = i_1, ..., X_n = i) \quad = \quad P(X_{n+1} = j/X_n = i). \tag{1}$$

De maneira informal, podemos ver que a relação anterior estabelece que a distribuição condicional de X_{n+1} dada a historia do processo depende apenas do estado presente do processo. $P(X_{n+1} = j/X_n = i)$ define a probabilidade de transição, em um passo, do estado i para o estado j no instante de tempo n. Em geral essa probabilidade depende de i, j e n. Quando essa probabilidade não depender de n, diremos que a cadeia é homogenea no tempo ou que possui probabilidades de transição estacionárias. Neste caso, definimos

$$P_{ij} = P(X_{n+1} = j/X_n = i) \qquad n = 0, 1, 2, ...$$

As probabilidades de transição podem ser arranjadas em uma matriz quadrada de ordem $M + 1$ e será chamada Matriz de probabilidades de transição e denotada por P. Isto é

$$P = \begin{pmatrix} P_{00} & P_{01} & . & . & . & P_{0M} \\ P_{10} & P_{11} & . & . & . & P_{1M} \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ P_{M0} & P_{M1} & . & . & . & P_{MM} \end{pmatrix}$$

Observemos que para cada i, $\{P_{ij}, j = 0, 1, ..., M\}$ define uma função de probabilidade. Isto é, para cada i, temos que

$$P_{ij} \geq 0 \qquad \text{para } j = 0, 1, ..., M$$

e

$$\sum_{j=0}^{M} P_{ij} = 1.$$

Uma matriz P cujas entradas satisfazem essas duas condições é chamada Matriz Estocastica.

Distribuições Unidimensionais. A distribuição de X_0 é chamada distribuição inicial da cadeia e será denotada por $\underline{u}_0 = (u_{00}, u_{01}, ...u_{0M})$, um vetor de dimensão $M + 1$. Por abuso de notação, representaremos dessa forma um vetor linha. Se $\underline{u}_0$ e P são conhecidas, podemos encontrar a distribuição de X_n para $n = 1, 2, ...$

Calculemos , por exemplo, a distribuição de probabilidade de X_1. Essa distribuição de probabilidade será representada por $\underline{u}_1 = (u_{10}, u_{11}, ...u_{1M})$

$$\begin{aligned} P(X_1 = j) &= \sum_{i=0}^{M} P(X_0 = i, X_1 = j) \\ &= \sum_{i=0}^{M} P(X_0 = i) P(X_1 = j / X_0 = i) \\ &= \sum_{i=0}^{M} u_{0i} P_{ij}. \end{aligned} \tag{2}$$

Observe que $P(X_1 = j)$ é o produto escalar de $\underline{u}_0$ pela coluna correspondente ao estado j da matriz de probabilidades de transição P. Então $\underline{u}_1 = \underline{u}_0 P$. Isto é, o vetor que representa a função de probabilidade de X_1 é obtido como o produto matricial do vetor $\underset{\sim}{u}_0$ pela matriz P.

Em forma análoga podemos encontrar a distribuição de X_2. De fato,

$$\begin{aligned} P(X_2 = j) &= \sum_{i=0}^{M} P(X_1 = i, X_2 = j) \\ &= \sum_{i=0}^{M} P(X_1 = i) P(X_2 = j / X_1 = i) \\ &= \sum_{i=0}^{M} u_{1i} P_{ij}. \end{aligned} \tag{3}$$

Ou seja $\underset{\sim}{u}_2 = \underset{\sim}{u}_1 P = (\underset{\sim}{u}_0 P) P = \underset{\sim}{u}_0 P^2$.

O leitor pode perceber que se o vetor $\underset{\sim}{u}_k$ representando a função de probabilidade de X_k for conhecido, o vetor $\underset{\sim}{u}_{k+1}$ pode ser obtido como o produto matricial do vetor $\underset{\sim}{u}_k$ pela matriz P. Isto é, $\underset{\sim}{u}_{k+1} = \underset{\sim}{u}_k P$. Iterativamente obtem-se $\underset{\sim}{u}_{k+1} = \underset{\sim}{u}_k P = \underset{\sim}{u}_{k-1} P^2 = ... = \underset{\sim}{u}_0 P^{k+1}$. Ou seja, a ditribuição de probabilidade de X_{k+1} pode ser obtida também como o produto matricial do vetor representando a distribuição inicial da cadeia pela matriz P^{k+1}. Veremos mais para frente que P^{k+1} é a matriz de probabilidades de transição de ordem $k+1$.

Quer dizer que se conhecemos a distribuição inicial e a matriz de probabilidades de transição, todas as distribuições unidimensionais podem ser calculadas.

Distribuições Multivariadas. Falemos agora das distribuições finito-dimensionais. Para isso, precisamos calcular as probabilidades de transição de ordem superior. Temos definido P_{ij} como a probabilidade de transição do estado i para o estado j em um passo. Definamos $P_{ij}^{(2)}$ como a probabilidade de transição do estado i para o estado j em dois passos. Isto é

$$P_{ij}^{(2)} = P(X_{n+2} = j / X_n = i).$$

Essa probabilidade condicional pode ser calculada como a função de probabilidade marginal de uma função de probabilidade condicional conjunta, ou seja

$$\begin{aligned} P_{ij}^{(2)} &= P(X_{n+2} = j / X_n = i) \\ &= \sum_{k=0}^{M} P(X_{n+1} = k, X_{n+2} = j / X_n = i) \end{aligned}$$

$$= \sum_{k=0}^{M} P(X_{n+1} = k/X_n = i)P(X_{n+2} = j/X_{n+1} = k)$$
$$= \sum_{k=0}^{M} P_{ik}P_{kj}.$$

Esse resultado estabelece que a probabilidade de transição, de segunda ordem, do estado i para o estado j é calculada somando as probabilidades de todas as trajetorias que vão do estado i para um estado intermediário k em um passo e do estado k para o estado j em mais um passo.

O leitor pode observar que $P_{ij}^{(2)}$ é o produto escalar da linha correspondente ao estado i pela coluna correspondente ao estado j da matriz P. Isto é, se definirmos $P^{(2)}$ como a matriz de probabilidades de transição em dois passos ou matriz de transição de segunda ordem, temos que $P^{(2)} = PP = P^2$.

Deixamos como exercicio para o leitor provar que a matriz de probabilidades de transição em tres passos ou de terceira ordem, $P^{(3)}$ pode ser calculada como $P^{(3)} = PP^{(2)} = PP^2 = P^3$. As entradas da matriz P^3 representam as probabilidades de transição em tres passos. Isto é,

$$P_{ij}^{(3)} = P(X_{n+3} = j/X_n = i).$$

Em geral, as matrizes de probabilidades de transição em m passos ou de ordem m, $P^{(m)}$ podem ser calculadas como as correspondentes potencias da matriz P.

Observação Como consequência da relação 1, pode-se provar que

$$P(X_{n+m} = j/X_0 = i_0, ..., X_n = i) = P(X_{n+m} = j/X_n = i). \tag{4}$$

Provemos, por exemplo, essa relação para $m = 2$:

$$P(X_{n+2} = j/X_0 = i_0, ..., X_n = i)$$
$$= \sum_{k=0}^{M} P(X_{n+1} = k, X_{n+2} = j/X_0 = i_0, ..., X_n = i)$$
$$= \sum_{k=0}^{M} P(X_{n+1} = k/X_0 = i_0, ..., X_n = i)P(X_{n+2} = j/X_0 = i_0, ..., X_n = i, X_{n+1} = k)$$
$$= \sum_{k=0}^{M} P(X_{n+1} = k/X_n = i)P(X_{n+2} = j/X_{n+1} = k)$$

$$\begin{aligned} &= \sum_{k=0}^{M} P(X_{n+1}=k, X_{n+2}=j/X_n=i) \\ &= P(X_{n+2}=j/X_n=i). \end{aligned}$$

Deixamos como exercicio para o leitor fazer a prova para $m = 3$.
Pode-se provar também que para n_1, n_2, . . ., n_r tais que $n < n_1 < n_2 < ...n_r$,

$$\begin{aligned} & P(X_{n_1}=i_1, X_{n_2}=i_2, ..., X_{n_r}=i_r/X_0=i_0, X_1=i_1, ..., X_n=i) \\ = \; & P(X_{n_1}=i_1, X_{n_2}=i_2, ..., X_{n_r}=i_r/X_n=i). \end{aligned}$$

Equações de Chapman- Kolmogorov. Essas equações são de muita utilidade no estudo das cadeias de Markov. Elas estabelecem que a probabilidade de transição de ordem $n + m$ de um estado i para um estado j pode ser calculada somando todos os produtos das probabilidades de transição do estado i para um estado intermediario k em m passos pelas correspondentes probabilidades de transição do estado k para o estado j em n passos. Isto é

$$P_{ij}^{(m+n)} = \sum_{k=0}^{M} P_{ik}^{(m)} P_{kj}^{(n)}.$$

Exemplo 2.1.1 Calcule a distribuição de probabilidade de X_1 na cadeia de Markov com espaço de estados $E = \{0, 1, 2, 3\}$, distribuição inicial

$$\underline{u}_0 = \left\{ \begin{array}{cccc} (0,2 & 0,4 & 0,3 & 0,1) \end{array} \right\}$$

e matriz de probabilidade de transição definida por

$$P = \begin{pmatrix} 0,4 & 0,1 & 0,3 & 0,2 \\ 0,1 & 0,3 & 0,4 & 0,2 \\ 0,2 & 0,1 & 0,3 & 0,4 \\ 0,3 & 0,1 & 0,3 & 0,3 \end{pmatrix}$$

Solução. O vetor $\underline{u}_1$ é obtido multiplicando o vetor $\underline{u}_0$ pela matriz P e obtem-se

$$\underline{u}_1 = (u_0, u_1, u_2, u_3) = \left(\begin{array}{cccc} 0,2 & 0,4 & 0,3 & 0,1) \end{array} \right) \begin{pmatrix} 0,4 & 0,1 & 0,3 & 0,2 \\ 0,1 & 0,3 & 0,4 & 0,2 \\ 0,2 & 0,1 & 0,3 & 0,4 \\ 0,3 & 0,1 & 0,3 & 0,3 \end{pmatrix}$$

Fazendo o produto, obtem-se

$$\underline{u}_1 = \begin{pmatrix} 0,21 & 0,18 & 0,34 & 0,27 \end{pmatrix}$$

Exemplo 2.1.2. No exemplo anterior, calcular a matriz de probabilidades de transição de ordem 2.

Solução.

$$P^2 = \begin{pmatrix} 0,4 & 0,1 & 0,3 & 0,2 \\ 0,1 & 0,3 & 0,4 & 0,2 \\ 0,2 & 0,1 & 0,3 & 0,4 \\ 0,3 & 0,1 & 0,3 & 0,3 \end{pmatrix} \begin{pmatrix} 0,4 & 0,1 & 0,3 & 0,2 \\ 0,1 & 0,3 & 0,4 & 0,2 \\ 0,2 & 0,1 & 0,3 & 0,4 \\ 0,3 & 0,1 & 0,3 & 0,3 \end{pmatrix}$$

Fazendo o produto, obtem-se

$$P^2 = \begin{pmatrix} 0,29 & 0,12 & 0,31 & 0,28 \\ 0,21 & 0,16 & 0,33 & 0,30 \\ 0,27 & 0,12 & 0,31 & 0,30 \\ 0,28 & 0,12 & 0,31 & 0,29 \end{pmatrix}$$

O leitor pode usar recursos computacionais e calcular as matrizes de probabilidades de transição de ordem 3 e 4. Obterá respectivamente

$$P^3 = \begin{pmatrix} 0,274 & 0,124 & 0,312 & 0,290 \\ 0,256 & 0,132 & 0,316 & 0,296 \\ 0,272 & 0,124 & 0,312 & 0,292 \\ 0,273 & 0,124 & 0,312 & 0,291 \end{pmatrix}$$

e

$$P^4 = \begin{pmatrix} 0,2714 & 0,1248 & 0,3124 & 0,2914 \\ 0,2676 & 0,1264 & 0,3132 & 0,2928 \\ 0,2712 & 0,1248 & 0,3124 & 0,2916 \\ 0,2713 & 0,1248 & 0,3124 & 0,2915 \end{pmatrix}$$

Na matriz P^3, o valor $0,256$, por exemplo, representa $P(X_{n+3} = 0/X_n = 1)$.

Tendo calculado as probabilidades de transição de ordem superior, temos condição de calcular as distribuições finito dimensionais. Isto é, podemos calcular a distribuição do vetor $(X_{n_1}, X_{n_2}, ..., X_{n_k})$ para quaisquer valor de k e $0 \leq n_1 \leq n_2 \leq ...n_k$.

Exemplo 2.1.3. Com a informação do Exemplo 2.1, calculemos, a maneira de ilustração, $P(X_3 = 2, X_5 = 0, X_6 = 1, X_{10} = 2)$.

Solução. Calculemos primeiro $P(X_3 = 2)$. Usando os conceitos anteriores, calculamos $\underline{u}_3 = \underline{u}_0 P^3$ e obtemos

$$\underline{u}_3 = \left(\begin{array}{cccc} 0,2661 & 0,1272 & 0,3136 & 0,2931 \end{array} \right)$$

Dai obtem-se $P(X_3 = 2) = 0,3136$ e

$$\begin{aligned} & P(X_3 = 2, X_5 = 0, X_6 = 1, X_{10} = 2) \\ = \;& P(X_3 = 2)P(X_5 = 0/X_3 = 2)P(X_6 = 1/X_5 = 0)P(X_{10} = 2/X_6 = 1) \\ = \;& P(X_3 = 2)P^{(2)}_{20}P^{(1)}_{01}P^{(4)}_{12} \\ = \;& 0,3136(0,272)((0,1)(0,3132) \\ = \;& 0,00267157. \end{aligned}$$

2.2 Alguns Exemplos Importantes

Apresentaremos a seguir alguns exemplos interessantes do ponto de vista prático. No primeiro deles, conhecido como Modelo de Estoque tem-se como objetivo dimensionar o estoque de um produto visando lucro máximo a longo prazo. O segundo exemplo aborda o problema de Mobilidade Social e o terceiro é um exemplo trazido da Genética.

Exemplo 2.2.1. Modelo de Estoque. Um determinado item é estocado para ser vendido em períodos de operação. Assume- se que o tempo de renovação do estoque é nulo ou desprezível. A demanda do item em um período de operação é uma variável aleatória inteira D com função de probabilidade $P(D = k) = p_k$. Por simplicidade assumamos que existe inteiro positivo e finito, V, tal $0 \leq D \leq V$. O nível do estoque é observado no inicio de cada período e a política de estoque é determinada especificando dois números inteiros positivos, a e A com $a < A$. Se o nível é menor ou igual que a, um pedido é feito de tal maneira que o período seja iniciado com A itens. Se o nível é maior que a, inicia-se o período com o número de itens disponíveis. Defina X_n: nível do estoque no inicio do dia n. O espaço de estados desta cadeia é $E = \{0, 1, ..., A\}$. Calculemos a matriz de probabilidades de transição.

Defina $q_l = \sum_{k=l}^{V} p_k$

Se $i \leq a$, as probabilidades de transição P_{ij} são iguais a P_{Aj} pois nesse caso, o estoque é renovado até atingir o nível A. Isto é:
$P_{i0} = P_{A0} = P(D \geq A) = \sum_{k=A}^{V} p_k = q_A$,
$P_{i1} = P_{A1} = P(D = A - 1) = p_{A-1}$,
$P_{i2} = P_{A2} = P(D = A - 2) = p_{A-2}$,

.
.
.
$P_{iA} = P_{AA} = P(D = 0) = p_0$.

Se $i > a$,
$P_{i0} = P(D \geq i) = q_i$,
$P_{i1} = p_{i-1}$,
$P_{i2} = p_{i-2}$,
.
.
.
$P_{ii} = p_0$.

Observação $P_{A0} = P(D \geq A)$ pois pode ter demanda não satisfeita. Isto é, um freguês pode chegar na loja e o vendedor pode não ter o produto para vender.

A maneira de ilustração, suponha $a = 3$ e $A = 6$. A matriz de transição nesse caso será

$$P = \begin{pmatrix} q_6 & p_5 & p_4 & p_3 & p_2 & p_1 & p_0 \\ q_6 & p_5 & p_4 & p_3 & p_2 & p_1 & p_0 \\ q_6 & p_5 & p_4 & p_3 & p_2 & p_1 & p_0 \\ q_6 & p_5 & p_4 & p_3 & p_2 & p_1 & p_0 \\ q_4 & p_3 & p_2 & p_1 & p_0 & 0 & 0 \\ q_5 & p_4 & p_3 & p_2 & p_1 & p_0 & 0 \\ q_6 & p_5 & p_4 & p_3 & p_2 & p_1 & p_0 \end{pmatrix}$$

Continuação do exemplo. A função de probabilidade da demanda é definida a seguir

$$\begin{pmatrix} 0 & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\ 0,017 & 0,090 & 0,210 & 0,278 & 0,232 & 0,123 & 0,041 & 0,008 & 0,001 \end{pmatrix}$$

Calculando os valores de q_i e p_i, encontramos

$$P = \begin{pmatrix} 0,050 & 0,123 & 0,232 & 0,278 & 0,210 & 0,090 & 0,017 \\ 0,050 & 0,123 & 0,232 & 0,278 & 0,210 & 0,090 & 0,017 \\ 0,050 & 0,123 & 0,232 & 0,278 & 0,210 & 0,090 & 0,017 \\ 0,050 & 0,123 & 0,232 & 0,278 & 0,210 & 0,090 & 0,017 \\ 0,405 & 0,278 & 0,210 & 0,090 & 0,017 & 0,000 & 0,000 \\ 0,173 & 0,232 & 0,278 & 0,210 & 0,090 & 0,017 & 0,000 \\ 0,050 & 0,123 & 0,232 & 0,278 & 0,210 & 0,090 & 0,017 \end{pmatrix}$$

De interesse nesse tipo de problema é, a partir da distribuição da demanda, definir apropriadamente os valores de a e A de tal forma a maximizar os lucros e com alta probabilidade garantir a satisfação da demanda. Isto é, o problema a ser abordado é o dimensionamento do estoque. Tal problema será tratado posteriormente.

Neste caso estamos assumindo que tanto o número de itens vendidos quanto o número de itens procurados é uma variável aleatória inteira. Portanto o nível de estoque é também uma variável aleatória inteira. Quando essas variáveis são contínuas, a abordagem é diferente e náo será feita nesta disciplina. O leitor interessado nesse assunto pode iniciar estudos nessa direção e uma boa referênica é Oksendal (1998).

Exemplo 2.2.2. Modelo de Mobilidade Social. Esse exemplo foi apresentado por Prais (1955). Ele usou os dados obtidos por Glass e Hall (1954). Seguindo tais autores, classifiquemos uma população economicamente ativa nos seguintes grupos:

1: Operários não qualificados,
2: Operários semi- qualificados,
3: Operários qualificados,
4: Funções administrativas de nível médio,
5: Funções administrativas de nível superior,
6: Cargos em nível gerencial,
7: Cargos em nível executivo.

O filho de uma pessoa no nível i será do nível j com probabilidade P_{ij}. A partir dos dados fornecidos por Glas e Hall, Prais obteve o estimador da matriz de probabilidades de transição. No próximo capítulo definiremos o estimador da matriz de probabilidades de transição de uma cadeia de Markov.

$$P = \begin{pmatrix} 0,388 & 0,146 & 0,202 & 0,062 & 0,140 & 0,047 & 0,015 \\ 0,107 & 0,267 & 0,227 & 0,120 & 0,206 & 0,053 & 0,020 \\ 0,035 & 0,101 & 0,188 & 0,191 & 0,357 & 0,067 & 0,061 \\ 0,021 & 0,039 & 0,112 & 0,212 & 0,430 & 0,124 & 0,062 \\ 0,009 & 0,024 & 0,075 & 0,123 & 0,473 & 0,171 & 0,125 \\ 0,000 & 0,013 & 0,041 & 0,088 & 0,391 & 0,312 & 0,155 \\ 0,000 & 0,008 & 0,036 & 0,083 & 0,364 & 0,235 & 0,274 \end{pmatrix}$$

Exemplo 2.2.3. Caminho Aleatório Finito. Um caminho aleatório finito é uma cadeia de Markov com espaço de estados $E = \{a, a+1, a+2, ..., a+M\}$. Sem perda de generalidade, assuma $a = 0$. Se a cadeia está no estado i, pode continuar em i ou fazer uma transição para um dos estados vizinhos. Isto é, a matriz de probabilidades de transição é dada por

$$P = \begin{pmatrix} P_{00} & P_{01} & 0 & 0 & . & . & . & P_{0M} \\ P_{10} & P_{11} & P_{12} & 0 & . & . & . & P_{1M} \\ 0 & P_{21} & P_{22} & P_{23} & 0 & . & . & 0 \\ . & . & . & . & . & . & . & . \\ . & . & . & . & . & . & . & . \\ . & . & . & . & . & . & . & . \\ 0 & . & . & . & . & P_{(M-1)(M-2)} & P_{M-1)(M-1)} & P_{(M-1)M} \\ P_{M0} & 0 & 0 & 0 & . & . & P_{M(M-1)} & P_{MM} \end{pmatrix}$$

De grande interesse na analise de uma cadeia de Markov são as probabilidades de transição e a distribuição inicial da cadeia pois como temos dito anteriormente, a partir delas poderemos calcular todas as distribuições da cadeia. Mais importante ainda é o comportamento da cadeia quando ela esteja em equilibrio. Isto é, por exemplo no Modelo de Estoque, é de interesse prático, como é que a cadeia vai se comportar depois que o dono da loja tenha estado operando por um periodo longo de tempo. No Modelo de Mobilidade Social é de interesse a estrutura da população a longo prazo. Ou seja, por exemplo para definir políticas públicas, é de interesse saber qual a percentagem da população economicamente ativa em cada grupo. Nesse sentido, estaremos interessados na distribuição $\underline{u} = \lim_{n\to\infty} \underline{u}_n$. A distribuição $\underline{u}$ será chamada distribuição invariante da cadeia.

Para que exista distribuição invariante, algumas propriedades a cadeia precisa satisfazer. Apresentamos a seguir algumas definições visando estabelecer as condições que uma cadeia de Markov precisa satisfazer para que possua distribuição invariante.

2.3 Irredutibilidade

Definição 2.3.1: Diremos que o estado j é acessível desde o estado i se existe uma trajetória indo do estado i para o estado j. Isto é, se existe inteiro não negativo n_{ij}, finito tal que $P_{ij}^{(n_{ij})} > 0$. Usaremos a notação $i \to j$ para indicar que o estado j é acessível desde o estado i.

Definição 2.3.2: Diremos que dois estados, i e j comunicam se j é acessível desde i e i é acessível desde j. Isto é, se existem números n_{ij} e n_{ji}, finitos, tais que $P_{ij}^{(n_{ij})} > 0$ e $P_{ji}^{(n_{ji})} > 0$. Usaremos a notação $i \leftrightarrow j$ para indicar que os estados i e j comunicam entre eles.

A relação de comunicação define uma relação de equivalência. Isto é:

(a) $i \leftrightarrow i$,
(b)Se $i \leftrightarrow j$, então $j \leftrightarrow i$,
(c) Se $i \leftrightarrow j$ e $j \leftrightarrow k$, então $i \leftrightarrow k$,

A prova das duas primeiras condições é trivial. Vejamos a prova de (c).

Desde que $i \leftrightarrow j$ e $j \leftrightarrow k$, existem números inteiros n_{ij} e n_{jk}, finitos tais que $P_{ij}^{(n_{ij})} > 0$ e $P_{jk}^{(n_{jk})} > 0$. Então $P^{(n_{ij}+n_{jk})} = \sum_{l=0}^{M} P_{il}^{(n_{ij})} P_{lk}^{(n_{jk})} \geq P_{ij}^{(n_{ij})} P_{jk}^{(n_{jk})} > 0$. A igualdade é válida pois é a equação de Chapman-Kolmogorov.

Quer dizer que se $i \to j$ e $j \to k$, então $i \to k$. Deixamos para o leitor provar que $k \to i$.

O espaço de estados $E = \{0, 1, ..., M\}$ pode ser particionado como $E = C_1 \cup C_2 ... \cup C_r \cup T$. Nessa decomposição do espaço de estados, para i=1, 2, . . ., r todos os estados em C_i comunicam entre eles mas se $i \neq j$, nenhum estado de C_i é acessível desde quaisquer estado de C_j e viceversa. O conjunto T é formado por estados a partir dos quais existe trajetórias para algum estado em alguma das classes C_1, . . ., C_r mas estando em alguma dessas classes não é possível mais voltar ao conjunto T. Os conjuntos C_1, C_2, . . ., C_r são chamados classes de equivalencia. Os estados do conjunto T serão chamados **Estados Transitórios.**

Definição 2.3.3: Diremos que uma cadeia é irredutível se a relação de equivalência determina uma única classe de equivalência. Isto é, diremos que a cadeia é irredutível se todos os estados comunicam entre eles. No Modelo de estoque e no Modelo de Mobilidade Social, todos os estados comunicam. Portanto em ambos os casos a cadeia é irredutível.

Exemplo 2.3.1 Considere a cadeia de Markov com espaço de estados $E = \{0, 1, 2, ..., 8\}$ e matriz de transição

$$P = \begin{pmatrix} 0,3 & 0,7 & 0,0 & 0,0 & 0,0 & 0,0 & 0 & 0,0 & 0,0 \\ 0,4 & 0,6 & 0,0 & 0,0 & 0,0 & 0,0 & 0 & 0,0 & 0,0 \\ 0,0 & 0,0 & 0,2 & 0,3 & 0,5 & 0,0 & 0 & 0,0 & 0,0 \\ 0,0 & 0,0 & 0,1 & 0,6 & 0,3 & 0,0 & 0 & 0,0 & 0,0 \\ 0,0 & 0,0 & 0,6 & 0,0 & 0,4 & 0,0 & 0 & 0,0 & 0,0 \\ 0,1 & 0,2 & 0,0 & 0,0 & 0,0 & 0,0 & 0 & 0,5 & 0,2 \\ 0,0 & 0,0 & 0,3 & 0,4 & 0,0 & 0,0 & 0 & 0,3 & 0,0 \\ 0,0 & 0,0 & 0,2 & 0,0 & 0,5 & 0,1 & 0 & 0,0 & 0,2 \\ 0,3 & 0,1 & 0,2 & 0,1 & 0,0 & 0,0 & 0 & 0,2 & 0,1 \end{pmatrix}$$

Nesse caso podemos definir $C_1 = \{0, 1\}$, $C_2 = \{2, 3, 4\}$ e $T = \{5, 6, 7, 8\}$. Os estados em C_1 comunicam entre eles , os estados em C_2 também comunicam entre eles, mas desde quaisquer estado em C_1 não podemos ir para um estado em C_2 e viceversa. Além do mais desde um estado em T podemos ir para algum estado na classe C_1 ou na classe C_2, mas não poderemos retornar ao conjunto T.

O leitor pode calcular, neste exemplo, as matrizes de probabilidades de transição de ordem 2, 4, 8, 16. Apresentamos a seguir as matrizes de ordem 2 e 16. Poderá observar a partir das matrizes que nenhum estado na classe C_1 comunica com os estados da classe C_2 e viceversa. Pode observar também, a partir da matriz P^{16} que desde quaisquer estado, existe probabilidade zero de ir para algum estado na classe T. Voltaremos a esse exemplo posteriormente.

$$P^2 = \begin{pmatrix} 0,37 & 0,63 & 0,00 & 0,00 & 0,00 & 0,00 & 0 & 0,00 & 0,00 \\ 0,36 & 0,64 & 0,00 & 0,00 & 0,00 & 0,00 & 0 & 0,00 & 0,00 \\ 0,00 & 0,00 & 0,37 & 0,24 & 0,39 & 0,00 & 0 & 0,00 & 0,00 \\ 0,00 & 0,00 & 0,26 & 0,39 & 0,35 & 0,00 & 0 & 0,00 & 0,00 \\ 0,00 & 0,00 & 0,36 & 0,18 & 0,46 & 0,00 & 0 & 0,00 & 0,00 \\ 0,17 & 0,21 & 0,14 & 0,02 & 0,25 & 0,05 & 0 & 0,04 & 0,12 \\ 0,00 & 0,00 & 0,16 & 0,33 & 0,42 & 0,03 & 0 & 0,00 & 0,06 \\ 0,07 & 0,04 & 0,38 & 0,08 & 0,30 & 0,00 & 0 & 0,09 & 0,04 \\ 0,16 & 0,28 & 0,11 & 0,13 & 0,23 & 0,02 & 0 & 0,02 & 0,05 \end{pmatrix}$$

$$P^{16} = \begin{pmatrix} 0,36 & 0,64 & 0,00 & 0,00 & 0,00 & 0,00 & 0,00 & 0,00 & 0,00 \\ 0,36 & 0,64 & 0,00 & 0,00 & 0,00 & 0,00 & 0,00 & 0,00 & 0,00 \\ 0,00 & 0,00 & 0,34 & 0,25 & 0,41 & 0,00 & 0,00 & 0,00 & 0,00 \\ 0,00 & 0,00 & 0,34 & 0,25 & 0,41 & 0,00 & 0,00 & 0,00 & 0,00 \\ 0,00 & 0,00 & 0,34 & 0,25 & 0,41 & 0,00 & 0,00 & 0,00 & 0,00 \\ 0,17 & 0,30 & 0,18 & 0,13 & 0,22 & 0,00 & 0,00 & 0,00 & 0,00 \\ 0,02 & 0,03 & 0,32 & 0,24 & 0,39 & 0,00 & 0,00 & 0,00 & 0,00 \\ 0,05 & 0,09 & 0,29 & 0,22 & 0,35 & 0,00 & 0,00 & 0,00 & 0,00 \\ 0,17 & 0,30 & 0,18 & 0,13 & 0,22 & 0,00 & 0,00 & 0,00 & 0,00 \end{pmatrix}$$

Neste exemplo foi fácil identificar as classes de equivalência e o conjunto T. Em geral, para decompor a cadeia, se for possível decompor, escolhemos um estado quaisquer e observamos os estados com os quais ele comunica.

Exemplo 2.3.2 A partir das seguintes matrizes de transição, definir as classes de equivalência

$$P_1 = \begin{pmatrix} 0 & 0,2 & 0,3 & 0 & 0,5 & 0 & 0 & 0 & 0 & 0 \\ 0,4 & 0 & 0 & 0 & 0 & 0 & 0,6 & 0 & 0 & 0 \\ 0 & 0,1 & 0 & 0 & 0,9 & 0 & 0 & 0 & 0 & 0 \\ 0,7 & 0 & 0 & 0 & 0 & 0 & 0,3 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0,4 & 0 & 0 & 0,6 & 0 & 0 \\ 0 & 0,2 & 0,1 & 0,3 & 0,2 & 0 & 0 & 0,2 & 0 & 0 \\ 0 & 0,3 & 0,2 & 0 & 0,1 & 0 & 0 & 0 & 0 & 0,4 \\ 0 & 0,2 & 0,6 & 0 & 0,1 & 0 & 0 & 0 & 0,1 & 0 \\ 0 & 0,3 & 0,2 & 0 & 0,5 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0,4 & 0,1 & 0 & 0,2 & 0,3 & 0 & 0 & 0 & 0 \end{pmatrix}$$

$$P_2 = \begin{pmatrix} 0,0 & 0 & 0 & 0 & 0,7 & 0 & 0 & 0,3 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0,3 & 0,2 & 0 & 0,1 & 0,4 \\ 0 & 0 & 0 & 0,2 & 0,4 & 0 & 0 & 0,4 & 0 & 0 \\ 0,2 & 0 & 0,5 & 0 & 0 & 0 & 0 & 0,3 & 0 & 0 \\ 0,6 & 0 & 0 & 0,3 & 0 & 0 & 0 & 0,1 & 0 & 0 \\ 0 & 0,2 & 0 & 0 & 0 & 0,1 & 0 & 0 & 0,4 & 0,3 \\ 0 & 0,5 & 0 & 0 & 0 & 0,1 & 0,1 & 0 & 0,1 & 0,2 \\ 0 & 0 & 0,3 & 0,7 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0,2 & 0 & 0 & 0 & 0,3 & 0,1 & 0 & 0,1 & 0,3 \\ 0 & 0,1 & 0 & 0 & 0 & 0,4 & 0 & 0 & 0,5 & 0 \end{pmatrix}$$

No primeiro caso, escolhamos o estado 3. A partir da matriz P_1, podemos definir a trajetória

$$3 \to 0 \to 2 \to 4 \to 7 \to 8 \to 1 \to 6 \to 9 \to 5 \to 3.$$

Essa trajetória tem probabilidade positiva pois

$$\begin{aligned} & P(3 \to 0 \to 2 \to 4 \to 7 \to 8 \to 1 \to 6 \to 9 \to 5 \to 3) \\ \geq \; & P_{30}P_{02}...P_{53}. \end{aligned}$$

Temos encontrado, então, uma trajetória que iniciando no estado 3, visita todos os estados da cadeia e volta ao estado 3. Concluimos então que a cadeia é irredutível.

No segundo caso, escolhemos o estado 3. A partir da matriz P_2, podemos ver que os estados 3, 0, 7, 2 e 4 formam uma classe de equivalência e os estados 1, 5 ,6, 8 e 9 definem uma segunda classe de equivalência. Neste caso, então, a cadeia é redutível e a decomposição da cadeia, $E = C_1 \cup C_2 = \{0, 2, 3, 7, 4\} \cup \{1, 5, 6, 8, 9\}$.

2.4 Periodicidade

Definição 2.4.1. Definimos o período de um estado i, representado por $d(i)$, como o máximo divisor comum de todos os inteiros n tais que $P_{ii}^{(n)} > 0$. Isto é,

$$d(i) \quad = \quad m.d.c\{n : P_{ii}^{(n)} > 0\}.$$

Observemos que se $P_{ii} > 0$, o estado i tem período 1 pois $d(1)$ precisa dividir 1. No exemplo 2.3.2, a partir da matriz P_1 podemos ver que $P_{11}^{(2)} \geq P_{10}P_{01} > 0$ e $P_{11}^{(3)} \geq P_{16}P_{62}P_{21} > 0$. Neste caso, $d(1)$ é igual a 1 pois $d(1)$ precisa dividir 2 e 3. Se um estado i tem período igual a 1, diremos que o estado é aperiódico.

Provaremos a seguir que todos os estados de uma classe de equivalência têm o mesmo período. Isto é, o período é uma propriedade de classe.

Proposição 2.4.1. Se $i \leftrightarrow j$, então $d(i) = d(j)$.

Prova. Desde que $i \leftrightarrow j$, existem inteiros não negativos, n_{ij} e n_{ji} tais que $P_{ij}^{(n_{ij})} > 0$ e $P_{ji}^{(n_{ji})} > 0$. Sejam m e n tais que $P_{ii}^{(n)} > 0$ e $P_{jj}^{(m)} > 0$.

$P_{jj}^{(n_{ji}+n+n_{ij})} \geq P_{ji}^{(n_{ji})}P_{ii}^{(n)}P_{ij}^{(n_{ij})} > 0$. Quer dizer então que $d(j)$ é um divisor de $n_{ji}+n+n_{ij}$. Temos também que $P_{jj}^{(n_{ji}+n_{ij})} \geq P_{ji}^{(n_{ji})}P_{ij}^{(n_{ij})} > 0$. Isso mostra que $d(j)$ é também um divisor de $n_{ji}+n_{ij}$ e portanto é um divisor da diferença $n = (n_{ji}+n+n_{ij})-n_{ji}+n_{ij}$. Isso prova que $d(j) \leq d(i)$ pois $d(i)$ é o maior divisor de n. Deixamos como exercicio para o leitor provar que $d(i) \leq d(j)$ e concluir que $d(i) = d(j)$.

Exemplo 2.4.1. Considere a cadeia de Markov com espaço de estados $E = \{0, 1, 2, 3, 4, 5\}$

e matriz de transição

$$P = \begin{pmatrix} 0,0 & 0,3 & 0,0 & 0,0 & 0,0 & 0,7 \\ 0,6 & 0,0 & 0,4 & 0,0 & 0,0 & 0,0 \\ 0,0 & 0,5 & 0,0 & 0,5 & 0,0 & 0,0 \\ 0,0 & 0,0 & 0,9 & 0,0 & 0,1 & 0,0 \\ 0,0 & 0,0 & 0,0 & 0,4 & 0,0 & 0,6 \\ 0,3 & 0,0 & 0,0 & 0,0 & 0,7 & 0,0 \end{pmatrix}$$

Deixamos como exercicio provar que a cadeia é irredutível.

Vejamos a periodicidade do estado 0. Observe que $P_{00} = 0$, $P_{00}^{(2)} \geq P_{01}P_{10} > 0$. Suponha que $P_{00}^{(2k)} > 0$. A partir dessa suposição, provemos que $P_{00}^{(2(k+1))} > 0$. De fato, $P_{00}^{(2(k+1))} \geq P_{00}^{(2k)} P_{01}P_{10} > 0$. Então, pelo Principio de Indução Matemática, temos que $P_{00}^{(2k)} > 0$ para todo inteiro positivo.

Provemos, usando o Principio de Indução Matematica, que $P_{00}^{(2k+1)} = P_{02}^{(2k+1)} = P_{04}^{(2k+1)} = 0$ para todo inteiro k.

Claramente $P_{00}^{(3)} = P_{02}^{(3)} = P_{04}^{(3)} = 0$.

Suponha que $P_{00}^{(2k+1)} = P_{02}^{(2k+1)} = P_{04}^{(2k+1)} = 0$
$P_{00}^{(2(k+1)+1)} = P_{00}^{(2k+3)} = \sum_{j=0}^{4} P_{0j}^{(2k+1)} P_{j0}^{(2)}$. Nesse somatório, os valores de j tal que $P_{j0}^{(2)}$ é positivo são os correpondentes a $j = 0, 2, 4$, mas pela hipotese de indução, os correspondentes fatores dessas probabilidade no somatório são iguais a zero. Isso prova que $P_{00}^{(2k+1)} = 0$ para todo inteiro k. A prova de que $P_{02}^{(2k+1)} = 0$ e $P_{04}^{(2k+1)} = 0$ é similar e deixamos como exercicio para o leitor.

Portanto, o estado 0 tem período $d(0) = 2$. Desde que a cadeia é irredutível, a cadeia tem período 2.

Definição 2.4.2. Se uma cadeia é irredutível e um dos estados tem período 1 (portanto todos os estados têm período 1) , diremos que a cadeia é aperiódica. Em nossa disciplina concentraremos os estudos em cadeias aperiódicas.

2.5 Recorrência

O conceito de recorrência será mais importante no caso em que a cadeia possui espaço de estados infinito. No caso em que a cadeia é finita, existe pelo menos um estado recorrente e em particular, se a cadeia é irredutível, todos os estados serão recorrentes. Abordaremos esse conceito posteriormente ao abordar problemas com espaço de estados infinito.

2.6 Simulação de uma cadeia de Markov

Considere uma cadeia de Markov com distribuição inicial $\underline{u}_0$ e matriz de probabilidades de transição P. O seguinte algoritmo simula uma realização de tamanho $N+1$ dessa cadeia.

Passo 1. Gere uma observação da variável aleatória com distribuição de probabilidade definida por $\underline{u}_0$. Chame x_0 essa observação,

Passo 2. Gere uma observação de variável aleatória com distribuição de probabilidade definida pela linha correspondente às probabilidades de transição a partir de x_0, chame essa observação de x_1,

.
.
.

Passo N. Gere uma observação de variável aleatória com distribuição de probabilidade definida pela linha correspondente às probabilidades de transição a partir de x_{N-1}. da matriz P, chame essa observação de x_N,

x_0, x_1, x_2, . . . , x_N é uma realização de tamanho N da cadeia em questão.

2.7 Cadeias de Markov com espaço de estados infinito.

Sem perda de generalidade assumiremos que o espaço de estados é $E = \{0, 1, 2, ...\}$ ou $E = \{..., -2, -1, 0, 1, 2, ...\}$.

No primeiro caso a matriz de probabilidades de transição é dada por

$$P = \begin{pmatrix} P_{00} & P_{01} & P_{02} & . & . & . \\ P_{10} & P_{11} & P_{12} & . & . & . \\ P_{20} & P_{21} & P_{22} & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \end{pmatrix}$$

Os conceitos de irredutibilidade e periodicidade são so mesmos que no caso de espaço de estados finito. Vejamos o conceito de **Recorrência**.

Definição 2.7.1. Diremos que o estado i é recorrente se $\sum_{n=0}^{\infty} P_{ii}^{(n)} = \infty$.

Proposição 2.7.1. Recorência é uma propriedade de classe. Isto é, se $i \leftrightarrow j$ e i é recorrente, então j também é recorrente.

Prova. Precisamos provar que $\sum_{m=0}^{\infty} P_{jj}^{(m)} = \infty$. Todo m tal que $m \geq n_{ji} + n_{ij}$ pode ser escrito como $m = n_{ji} + n + n_{ij}$ para algum $n \geq 0$. Pela equação de Chapman- Kolmogorov temos que

$$P_{jj}^{(m)} \geq P_{ji}^{(n_{ji})} P_{ii}^{(n)} P_{ij}^{(n_{ij})}.$$

Então

$$\begin{aligned} \sum_{m=n_{ji}+n_{ij}}^{\infty} P_{jj}^{(m)} &\geq \sum_{n=0}^{\infty} P_{ji}^{(n_{ji})} P_{ii}^{(n)} P_{ij}^{(n_{ij})} \\ &= P_{ji}^{(n_{ji})} P_{ij}^{(n_{ij})} \sum_{n=0}^{\infty} P_{ii}^{(n)} = \infty. \end{aligned}$$

A proposição anterior permite dizer que se uma cadeia é irredutível e possui um estado recorrente, então todos os estados são recorrentes.

Exemplo 2.7.1 Exemplo de cadeia recorrente. Considere o caminho aleatório irrestrito (com espaço de estados $E = \{..., -2, -1, 0, 1, 2, ...\}$), $P_{i(i+1)} = p$ e $P_{i(i-1)} = q$.

Consideremos o estado 0. Intuitivamente podemos provar que $P_{00}^{(2n+1)} = 0$ pois é impossível que a cadeia, iniciando em 0, depois de um número impar de passos esteja novamente em 0. Para calcular $P_{00}^{2n)}$, raciocinamos em forma analoga: A cadeia , depois de 2n passos, estará no estado 0 se a mitade do número de transições, n, for para a direita e a mitade para a esquerda. A probabilidade disso acontecer é dada pela distribuição binomial. Isto é $P_{00}^{(2n)} = \frac{(2n)!}{n!n!} p^n q^n$.

Usando a aproximação de Stirling, para n grande, essa probabilidade pode ser aproximada por $\frac{(4pq)^n}{\sqrt{\pi n}}$. Então para m suficientemente grande,

$$\sum_{n=m}^{\infty} P_{00}^{(2n)} \approx \sum_{n=m}^{\infty} \frac{(4pq)^n}{\sqrt{\pi n}}.$$

Sabe-se que se $p + q = 1$, então $pq \leq \frac{1}{4}$ e que a igualdade ocorre se e somente se $p = q = \frac{1}{2}$. Portanto a série diverge para infinito se e somente se $p = q = \frac{1}{2}$. Quer dizer, então que o estado 0 é recorrente se e somente se $p = q = \frac{1}{2}$. Como a cadeia é irredutível, todos os estados são recorrentes nesse caso.

Definição 2.7.2. Seja i um estado recorrente. Dizemos que i é recorrente positivo se $\lim_{n\to\infty} P_{ii}^{(n)} = u_i > 0$. Se $\lim_{n\to\infty} P_{ii}^{(n)} = u_i = 0$, diremos que o estado é recorrente nulo.

Estabelecemos, sem prova, que uma cadeia finita possui pelo menos um estado recorrente positivo. Portanto, se a cadeia é finita e irredutível, todos os estados são recorrentes positivos.

Definição 2.7.3. Se uma cadeia é irredutível, aperiódica e recorrente, ela é dita ser **Ergodica**. Se é irredutível, aperiódica e recorrente positiva, ela é dita **Fortmente Ergodica**.

2.8 Distribuição Invariante.

Definição 2.8.1 Tanto no caso de espaço de estados finito quanto no caso de espaço de estados infinito, se uma cadeia é fortemente ergodica, para cada estado i existe $\lim_{n\to\infty} P_{ii}^{(n)} = u_i > 0$.. Além do mais, esse limite não depende do estado inicial da cadeia, isto é, $\lim_{n\to\infty} P_{ji}^{(n)} = u_i$ para todo $j \in E$. Nesses casos define-se a distribuição invariante da cadeia e é representada pelo vetor $\underline{u} = (, ..., u_{-2}, u_{-1}, u_0, u_1, u_2, ...)$.

Se a cadeia é finita, aperiodica e irredutível, existe a distribuição invariante e pode ser facilmente calculada. Calcula-se $\lim_{n\to\infty} P^n$. Essa matriz existe e tem a forma

$$P = \begin{pmatrix} u_0 & u_1 & . & . & . & u_M \\ u_0 & u_1 & . & . & . & u_M \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ u_0 & u_1 & . & . & . & u_M \end{pmatrix}$$

Exemplo 2.8.1 Consideremos o Modelo de estoque (Exemplo 2.2.1). Com a distribuição de probabilidade para a demanda e os valores de $a = 3$ e $A = 6$ considerados naquela ocasião, a matriz de probabilidades de transição foi encontrada,

$$P = \begin{pmatrix} 0,050 & 0,123 & 0,232 & 0,278 & 0,210 & 0,090 & 0,017 \\ 0,050 & 0,123 & 0,232 & 0,278 & 0,210 & 0,090 & 0,017 \\ 0,050 & 0,123 & 0,232 & 0,278 & 0,210 & 0,090 & 0,017 \\ 0,050 & 0,123 & 0,232 & 0,278 & 0,210 & 0,090 & 0,017 \\ 0,405 & 0,278 & 0,210 & 0,090 & 0,017 & 0,000 & 0,000 \\ 0,173 & 0,232 & 0,278 & 0,210 & 0,090 & 0,017 & 0,000 \\ 0,050 & 0,123 & 0,232 & 0,278 & 0,210 & 0,090 & 0,017 \end{pmatrix}$$

Nesse exemplo calculando as potências da matriz P encontramos

$$P^{16} = \begin{pmatrix} 0,118574 & 0,156795 & 0,231488 & 0,241485 & 0,169016 & 0,0697004 & 0,0129418 \\ 0,118574 & 0,156795 & 0,231488 & 0,241485 & 0,169016 & 0,0697004 & 0,0129418 \\ 0,118574 & 0,156795 & 0,231488 & 0,241485 & 0,169016 & 0,0697004 & 0,0129418 \\ 0,118574 & 0,156795 & 0,231488 & 0,241485 & 0,169016 & 0,0697004 & 0,0129418 \\ 0,118574 & 0,156795 & 0,231488 & 0,241485 & 0,169016 & 0,0697004 & 0,0129418 \\ 0,118574 & 0,156795 & 0,231488 & 0,241485 & 0,169016 & 0,0697004 & 0,0129418 \\ 0,118574 & 0,156795 & 0,231488 & 0,241485 & 0,169016 & 0,0697004 & 0,0129418 \end{pmatrix}$$

Observe que todas as entradas em cada coluna são iguais. Encontramos, neste caso, que o vetor representando a distribuição invariante é

$$\underline{u} = (0,118574, 0,156795, 0,231488, 0,241485, 0,169016, 0,0697004, 0,0129418)$$

Qual a a interpretação da distribuição invariante de cadeia?. Por definição, $u_i = \lim_{n\to\infty} P_{ji}^{(n)}$ e esse limite não depende de j. Quer dizer, então que u_i representa a probabilidade de que a cadeia se encontre no estado i, após ela entrar em equilibrio. No exemplo do modelo de estoque, 0,231488, por exemplo, representa a probabilidade de que existam 2 unidades no estoque, 0,118574 representa a probabilidade de que o estoque esteja no nível zero. Isto é, que não tenhamos item para vender. Cabe ao gerente da loja determinar a política de estoque de acordo aos valores dessas probabilidades.

Exemplo 2.8.2 No Modelo de Mobilidade social, calculando as potências da matriz P encontramos

$$P^{32} = \begin{pmatrix} 0,0226346 & 0,0413963 & 0,0875287 & 0,127222 & 0,409756 & 0,182188 & 0,129274 \\ 0,0226346 & 0,0413963 & 0,0875287 & 0,127222 & 0,409756 & 0,182188 & 0,129274 \\ 0,0226346 & 0,0413963 & 0,0875287 & 0,127222 & 0,409756 & 0,182188 & 0,129274 \\ 0,0226346 & 0,0413963 & 0,0875287 & 0,127222 & 0,409756 & 0,182188 & 0,129274 \\ 0,0226346 & 0,0413963 & 0,0875287 & 0,127222 & 0,409756 & 0,182188 & 0,129274 \\ 0,0226346 & 0,0413963 & 0,0875287 & 0,127222 & 0,409756 & 0,182188 & 0,129274 \\ 0,0226346 & 0,0413963 & 0,0875287 & 0,127222 & 0,409756 & 0,182188 & 0,129274 \end{pmatrix}$$

De acordo a esses resultados, por exemplo, $12,7222\%$ da população economicamente ativa, estarão ocupando cargos administrativos em nível médio. Essas probabilidades permitirão fazer um planejamento e definir políticas públicas.

Observação. Temos encontrado que P^{16} e P^{32} definem as distribuições invariantes nos exemplos 2.4.1 e 2.4.2 respectivamente. A rigor falta provar que se para algum inteiro positivo m, a matriz P^m é tal que as entradas da primeira coluna são iguais,. . ., as entradas da última coluna são iguais, então as matrizes P^n para $n \geq m$ são da mesma forma. Veja exercicio 2.9.14.

Temos definido a distribuição invariante de uma cadeia, quando ela existe, a partir dos limites $\lim_{n\to\infty} P_{ji}^{(n)}$. Uma outra forma de encontrar a distribuição invariante é a seguinte:

Lembremos que $\underline{u}_{k+1} = \underline{u}_k P$. Se existir distribuição invariante teremos que

$$lim_{k\to\infty}\underline{u}_{k+1} = \lim_{k\to\infty} \underline{u}_k = \underline{u}.$$

Portanto, teremos que

$$\underline{u} = \underline{u}P.$$

Podemos, então encontrar a distribuição invariante resolvendo esse sistema linear de equações.

Vale lembrar que tal sistema de equações envolve um número de condições igual ao número de incognitas mais 1 pois além dos $\underline{u}_k$ satisfazer o sistema, eles precisam satisfazer a condição

$$u_0 + u_1 + ... + u_n = 1.$$

Sabemos que um sistema de equações desse tipo pode não ter solução ou pode ter solução única ou ter multiplas soluções. **A condição de Ergodicidade Forte garante que a solução existe e é única**

2.9 Exercícios

2.9.1.Considere a cadeia de Markov com $E = \{0, 1, 2, 3, 4\}$. Se $P(X_0 = 2) = 1$ e

$$P = \begin{pmatrix} 0,2 & 0,3 & 0,0 & 0,1 & 0,4 \\ 0,1 & 0,3 & 0,3 & 0,0 & 0,3 \\ 0,0 & 0,3 & 0,7 & 0,0 & 0,0 \\ 0,2 & 0,4 & 0,1 & 0,0 & 0,3 \\ 0,0 & 0,2 & 0,3 & 0,3 & 0,2 \end{pmatrix}.$$

a) Encontre a distribuição de probabilidade de X_4,
b) Encontre $P(X_4 = 3, X_6 = 2, X_{10} = 1, X_{15} = 2)$.

2.9.2. Considere uma cadeia de Markov com matriz de probabilidades de transição P. Temos estabelecido que para todo i, $\sum_{j=0}^{M} P_{ij} = 1$ e que $P_{ij} \geq 0$.
a) Prove que esses resultados valem para as probabilidades de transição de segunda ordem. Isto é, prove que para todo i, $\sum_{j=0}^{M} P_{ij}^{(2)} = 1$ e $P_{ij} \geq 0$.
b) Prove que esses resultados valem também para quaisquer valor de n. Isto é, prove que para todo $i \in E$, $P_{ij}^{(n)} \geq 0$ e $\sum_{j=0}^{M} P_{ij}^{(n)} = 1$.

2.9.3 Considere uma cadeia de Markov com espaço de estados E. Para cada i, j. k em E, prove que:
a) $P(X_1 = i/X_0 = j, X_2 = k) = \frac{P_{ji}P_{ik}}{P_{jk}^{(2)}}$,
b) Para k_1, k_2 e k_3 tais que $1 < k_1 < k_2 < k_3$, $P(X_{k_2} = i/X_{k_1} = j, X_{k_3} = k) = \frac{P_{ji}^{(k_2-k_1)}P_{ik}^{(k_3-k_2)}}{P_{jk}^{(k_3-k_1)}}$.

2.9.4 Prove que para toda cadeia de Markov,

$$P(X_0 = i_0/X_1 = i_1, ..., X_n = i_n) = P(X_0 = i_0/X_1 = i_1).$$

2.9.5. Considere o Modelo de Estoque. Suponha que a demanda, D, tem distribuição binomial com parametros $n = 20$ e $p = 0,4$. Defina $a = 6$ e $A = 15$. Encontre a matriz de probabilidades de transição. A cadeia é irredutível?. Se a resposta for afirmativa, encontre a distribuição invariante.

2.9.6. No exercicio anterior, voce acha conveniente definir a e A iguais aos valores considerados?. Porque não definir $A = 20$?

2.9.7. No Modelo de Mobilidade Social (Exemplo 2.2.2), use os dados do exemplo apresentado e
a) calcule a probabilidade de que o neto de uma pessoa na classe de operário qualificado venha a pertencer à clase Funções administrativas de nível superior,
b) Calcule a percentagem de tal população, quando o processo estiver em equilibrio, na classe de operários qualificados.

2.9.8. João e Maria jogam a lançar uma moeda honesta. Se a moeda resultar cara, Jo ao ganha um real da Maria e perde um real se a moeda resultar coroa. Antes do primeiro jogo, Jo ao tinha 8 reais e Maria 10. Defina X_n: Capital de João depois do jogo n. Encontre a matriz de probabilidade de transição da cadeia associada ao jogo e classifique os estados.

2.9.9. Considere o jogo anterior e assuma que os jogadores fazem o seguinte acordo: Se um deles ficar sem dinheiro, recebe do outro jogador, com probabilidade $\frac{1}{3}$, um real. Encontre nesse caso a matriz de probabilidade de transição, prove que a cadeia é fortemente ergodica e encontre a distribuição invariante.

2.9.10. Considere a cadeia de Markov com matriz de probabilidade de transição

$$P_1 = \begin{pmatrix} 0,2 & 0,5 & 0,3 & 0,0 & 0,0 & 0,0 & 0,0 & 0,0 \\ 0,1 & 0,0 & 0,9 & 0,0 & 0,0 & 0,0 & 0,0 & 0,0 \\ 0,0 & 0,6 & 0,4 & 0,0 & 0,0 & 0,0 & 0,0 & 0,0 \\ 0,0 & 0,0 & 0,0 & 0,3 & 0,7 & 0,0 & 0,0 & 0,0 \\ 0,0 & 0,0 & 0,0 & 0,6 & 0,4 & 0,0 & 0,0 & 0,0 \\ 0,1 & 0,3 & 0,0 & 0,2 & 0,1 & 0,2 & 0,0 & 0,1 \\ 0,0 & 0,0 & 0,2 & 0,3 & 0,1 & 0,3 & 0,1 & 0,0 \\ 0,2 & 0,1 & 0,0 & 0,0 & 0,2 & 0,1 & 0,1 & 0,3 \end{pmatrix}.$$

Classifique os estados e encontre as classes de equivalência.

2.9.11. Considere agora a cadeia com matriz de probabilidades de transição

$$P_2 = \begin{pmatrix} 0,2 & 0,4 & 0,3 & 0,0 & 0,1 & 0,0 & 0,0 & 0,0 \\ 0,1 & 0,0 & 0,9 & 0,0 & 0,0 & 0,0 & 0,0 & 0,0 \\ 0,0 & 0,6 & 0,4 & 0,0 & 0,0 & 0,0 & 0,0 & 0,0 \\ 0,0 & 0,0 & 0,0 & 0,3 & 0,7 & 0,0 & 0,0 & 0,0 \\ 0,0 & 0,0 & 0,0 & 0,6 & 0,4 & 0,0 & 0,0 & 0,0 \\ 0,1 & 0,3 & 0,0 & 0,2 & 0,1 & 0,2 & 0,0 & 0,1 \\ 0,0 & 0,0 & 0,2 & 0,3 & 0,1 & 0,3 & 0,1 & 0,0 \\ 0,2 & 0,1 & 0,0 & 0,0 & 0,2 & 0,1 & 0,1 & 0,3 \end{pmatrix}.$$

A única diferença com a matriz anterior é que neste caso, $P_{04} > 0$. Classifique os estados e encontre as classes de equivalência.

2.9.12. Finalmente, considere a cadeia com matriz de probabilidades de transição

$$P_3 = \begin{pmatrix} 0,2 & 0,2 & 0,3 & 0,0 & 0,1 & 0,0 & 0,2 & 0,0 \\ 0,1 & 0,0 & 0,9 & 0,0 & 0,0 & 0,0 & 0,0 & 0,0 \\ 0,0 & 0,4 & 0,4 & 0,0 & 0,0 & 0,2 & 0,0 & 0,0 \\ 0,0 & 0,2 & 0,0 & 0,3 & 0,5 & 0,0 & 0,0 & 0,0 \\ 0,0 & 0,0 & 0,0 & 0,6 & 0,4 & 0,0 & 0,0 & 0,0 \\ 0,1 & 0,3 & 0,0 & 0,2 & 0,1 & 0,2 & 0,0 & 0,1 \\ 0,2 & 0,0 & 0,2 & 0,1 & 0,1 & 0,3 & 0,1 & 0,0 \\ 0,2 & 0,1 & 0,0 & 0,0 & 0,2 & 0,1 & 0,1 & 0,3 \end{pmatrix}.$$

Classifique os estados e encontre as classes de equivalência.

2.9.13 Considere os tres exercicios anteriores. Em quais casos, a cadeia é Fortemente ergodica?. Nos casos em que a cadeia é fortemente ergodica, encontre a distribuição invariante.

2.9.14. Suponha que para um inteiro positivo, m, a matriz de probabilidade de transição de ordem m, P^m é tal que os elementos de cada coluna são iguais. Prove que $P^n = P^m$ para todo $n \geq m$.

2.9.15. Uma matriz estocastica é dita **Duplamente estocastica** se para todo $j \in E$, $\sum_{i=0}^{M} P_{ij} = 1$. Isto é, se as somas dos elementos de cada coluna são iguais a 1. Suponha que uma cadeia de Markov fortemente ergodica com espaço de estados $E = \{0, 1, ..., M\}$ tem matriz de transição duplamente estocastica e prove que a distribuição invariante atribui igual probabilidade a cada estado.

2.9.16. Um dado é lançado sucessivamente. Defina $Y_0 = 0$ e $Y_n = \sum_{i=1}^{n} X_i$, onde X_i é o resultado do i-esimo lançamento. Defina $Z_n = r_n$ em que r_n é o resto ao dividir Y_n por 13. Encontre $\lim_{n\to\infty} P(Z_n = 0)$.

2.9.17. Suponha que a distribuição inicial de uma cadeia de Markov é igual à distribuição invariante:
a) Prove que X_1, X_2, . . . sãoidenticamente distribuidas,
b) Prove que (X_0, X_1), (X_1, X_2), . . . são identicamente distribuidas.

2.9.18. Seja $\underline{u}$ a distribuição invariante de uma cadeia. Suponha que j e k são dois estados tais que para todo $i \in E$, $P_{ij} = cP_{ik}$ para alguma constante c positiva. Prove que $u_j = cu_k$.

2.9.19. Considere uma cadeia de Markov sobre os inteiros não negativos com matriz de probabilidades de transição definida por $P_{i(i+1)} = p$ e $P_{i0} = 1 - p$. Mostre que essa cadeia possui distribuição invariante e ela é única. Encontre tal distribuição.

2.9.20. Estabeleça as equações que definem a distribuição invariante no Exercicio 9.1. Tal sistema tem solução?. Se a resposta é afirmativa, resolva as equações.

2.9.21. Um sistema tem duas máquinas de serviço, dos quais apenas uma está em operação em quaisquer instante de tempo. A unidade em funcionamento pode quebrar em quaisquer dia com probabilidade p. Existe uma oficina que leva dois dias para consertar a máquina. A oficina é tal que pode trabalhar no conserto em apenas uma máquina. Forme uma cadeia de Markov tomando como estados o par (x, y) onde x é o número de máquinas em condições de operar no final de um dia e y é 1 se um dia de trabalho no conserto tem transcorrido sem ter reparado a máquina e 0 caso contrário. Defina os estados $0 : (2, 0)$, $1 : (1, 0)$, $2 : (1, 1)$, $3 : (0, 1)$. Mostre que, com $p + q = 1$, a matriz de probabilidades de transição é dada por

$$P_3 = \begin{pmatrix} q & p & 0 & 0 \\ 0 & 0 & q & p \\ q & p & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}.$$

Tal cadeia é fortemente ergodica?. Em caso afirmativo, encontre a distribuição invariante.

2.9.22 Suponha que as condições do tempo em um determinado dia, (dia com sol (S) ou dia chuvoso(C)) dependa dos dois dias enteriores. Suponha que se teve sol ontem e hoje, a probabilidade de amanhã ter sol é 0,8. se ontem foi chuvoso e hoje tem sol, a probabilidade de amanhã ter sol é0,6. Se ontem teve sol e hoje é chuvoso, a probabilidade de ter sol amanhã é0,4. Se ontem foi chuvoso e hoje também, a probabilidade de amanhã ter sol é 0,1. Defina $0 : (S, S)$, $1 : (S, C)$, $1 : (C, S)$ e 3:(C,C). Mostre que podemos modelar esse problema como uma cadeia de Markov. Encontre a matriz de probabilidades de transição e encontre a distribuição invariante se existir.

2.9.23. Considere um poligono regular de $2r+1$ vertices V_1, V_2, . . . , V_{2r+1}. Suponha que em cada ponto V_k é colocada uma massa $w_k^{1)}$ com $\sum_{k=0}^{2} r+1 w_k^{(1)} = 1$. Obtenha novas massas $w_k^{(2)}$ substituindo a antiga pela média das duas massas dos vertices vizinhos. Repete-se esse procediemento sucessivamente. Encontre $\lim_{n\to\infty} w_k^{(n)}$ para $k = 1, 2, ..., 2r + 1$.

2.10 Referências

Bhattacharya, R. N. ; Waymire, E. C. Stochastic Processes with applications. J. Wiley. N. York. 1990.

Bartholomew,D. J. Stochastic Models for Social Processes, 3rd Edition. J. Wiley. N. York. 1982

Glass, D. V. Social Mobility in Britain Routledge and Kegan Paul. 1954. London.

Karlin, S. ; Taylor, H. M. A first Course in Stochastic Processes. Academic Press. N. York. 1975.

Kemeny, J. G. ; Snell, J. L. Finite Markov Chains. Springer. N. York. 1976.

Oksendal, B. Stochastic Differenctial Equations, An Introduction with Applications, fifth Edition. Springer. 1998.

Prais, S. J. Measuring social mobility. J. R. Statistical Society. A118. Pgs 56-66. 1955.

3 Capítulo 3. Introdução a Inferência Estocástica

Apresentamos neste capitulo, as idéias básicas de Inferência Estatística para Processos Estocasticos. Na literatura este tópico é conhecido como Inferência Estocastica. Concentraremos nossa atenção apenas na estimação das probabilidades de transição e a distribuição invariante de uma cadeia de Markov. Assumiremos que a cadeia é de primeira ordem e fortemente ergodica.

As probabilidades de transição serão estimadas pelo Método de Máxima Verossimilhança. Para estimar a distribuiçao invariante decompomos a realização de uma cadeia em ciclos regenerativos (que serão definidos oportunamente) e aplicaremos A Lei Forte dos Grandes Números.

O leitor interessado apenas na definição e aplicações dos correspondentes estimadores, pode ir diretamente à fórmula (7) da seção 3.1 e ao Corolário 3.2.2 da seção 3.2. O leitor interessado em mais detalhes técnicos pode consultar Bhattacharya e Waymire (1990), Asmussen (1987) ou Derman (1956).

3.1 Estimação das Probabilidades de Transição

Revisemos os conceitos de **Função de verossimilhança**.

Considere X_1, X_2, . . ., X_n uma amostra aleatória de uma variável aleatória com função de densidade f que depende de um parâmetro θ. θ pode ser unidimensional ou p-dimensional. A função de verossimilhança, dado que $X_1 = x_1$, $X_2 = x_2$, . . ., $X_n = x_n$ denotada por $L(\theta/x_1, x_2, ..., x_n)$ é definida por

$$L(\theta/x_1, x_2, ..., x_n) = f(x_1; \theta)f(x_2; \theta)...f(x_n; \theta).$$

Observação. É conveniente e oportuno fazer uma comparação entre a função de verossimilhança e a função de densidade conjunta das observações independentes X_1, X_2,. . ., X_n avaliada em x_1, x_2, . . ., x_n; isto é $f(x_1; \theta)f(x_2; \theta)...f(x_n; \theta)$. Notemos que a função é a mesma. A diferença é que enquanto a função de densidade conjunta é uma função de x_1, x_2, . . ., x_n dado o valor de θ; a função de versoimilhança é uma função de θ dados x_1, x_2, . . . , x_n.

O estimador obtido pelo Método de Máxima Verossimilhança é aquele valor, $\hat{\theta}$, que maximiza a função de verossimilhança. Observemos que a função de verossimilhança é um produto. Se aplicarmos logaritmo à função, será mais facil obter o estimador por este método. É isso que normalmente se faz. A fundamentação técnica está na seguinte proposição da teoria de funções.

Proposição 3.1.1. Seja g uma função de θ. Se h é uma função estritamente crescente, então

$$argmax\,[g(\theta)] = argmax\,[h(g(\theta)]\,.$$

Isto é, o valor $\hat{\theta}$ que maximiza $g(\theta)$ é o mesmo que maximiza $h(g(\theta))$.

Essa propopsição permite, por exemplo, tomar o logaritmo da função de verosimilhança que é o que normalmente se usa na literatura de Inferência Estatística.

Exemplo 3.1.1 Seja $X_1,\ X_2,\ .\ .\ .,\ X_n$ uma amostra aleatória de uma distribuição exponencial (Usemos a representação $f(x) = \frac{1}{\lambda}exp(-\frac{1}{\lambda}x)$ para $x > 0$). A função de verssimilhança dada a amostra $x_1,\ x_2,\ .\ .\ .,\ x_n$ é

$$\begin{aligned} L(\lambda/x_1, x_2, ..., x_n) &= f(x_1;\lambda)...f(x_n;\lambda) \\ &= \frac{1}{\lambda}exp(-\frac{1}{\lambda}x_1)...\frac{1}{\lambda}exp(-\frac{1}{\lambda}x_n) \\ &= \frac{1}{\lambda^n}e^{-\frac{1}{\lambda}\sum_{i=1}^{n}x_i}. \end{aligned} \tag{5}$$

Tomando logaritmos, temos

$$\begin{aligned} l(\lambda/x_1, x_2, ..., x_n) &= lnL(\lambda/x_1, x_2, ..., x_n) \\ &= -nln(\lambda) - \frac{1}{\lambda}\sum_{i=1}^{n}x_i. \end{aligned}$$

Derivando com respeito a λ, temos que

$$\frac{\partial l(\lambda/x_1, x_2, ..., x_n)}{\partial\lambda} = -\frac{n}{\lambda} + \frac{1}{\lambda^2}\sum_{i=1}^{n}x_i.$$

Igualando a zero a derivada, obtemos

$$\hat{\lambda} = \frac{1}{n}\sum_{i=1}^{n}x_i = \overline{x}.$$

Deixamos como exercicio para o leitor, tomar a segunda derivada de $l(\lambda/x_1, x_2, ..., x_n)$ e concluir que de fato, $\hat{\lambda}$ corresponde a um ponto de máximo. Mais detalhes sobre o Método de Máxima Verossimilhança podem ser vistos em Hogg e Craig (1995).

Observação: Se a representação $f(x) = \lambda exp(-\lambda x)$ para $x > 0$, é usada para a distribuição exponencial, o estimador de máxima verossimilhança para λ será $\hat{\lambda} = \frac{n}{\sum_{i=1}^{n} x_i} = \frac{1}{\bar{x}}$. veja Exercicio 3.3.1.

No caso de termos uma realização, i_0, i_1, . . ., i_n, de tamanho $n+1$ de uma cadeia de Markov, usaremos o Método de Máxima Verossimilhança para encontrar os estimadores das probabilidades de transição P_{ij} para $i, j = 0, 1, ..., M$.

A função de verossimilhança neste caso é

$$\begin{aligned}
& L(P_{00}, P_{01}, ..., P_{MM}/i_0, i_1, ..., i_n) \\
= \ & P(X_0 = 0)P(X_1 = i_1/X_0 = i_0)P(X_2 = i_2/X_1 = i_1)...P(X_n = i_n/X_{n-1} = i_{n-1}) \\
= \ & u_{0i_0} P(X_1 = i_1/X_0 = i_0)P(X_2 = i_2/X_1 = i_1)...P(X_n = i_n/X_{n-1} = i_{n-1}) \\
= \ & u_{0i_0} \Pi_{j=0}^{n-1} P(X_{j+1} = i_{j+1}/X_j = i_j) \\
= \ & u_{0i_0} \Pi_{i,j \in E} P_{ij}^{n_{ij}}.
\end{aligned}$$

$u_{0i_0} = P(X_0 = i_0)$ e na última expressão, n_{ij} representa o número de transições do estado i para o estado j (em um passo).

Tomando logaritmos, temos

$$l(P_{00}, P_{01}, ..., P_{MM}/i_0, i_1, ..., i_N) = ln(u_{0i_0}) + \sum_{i=0}^{M} \sum_{j=0}^{M} n_{ij} ln(P_{ij}).$$

Temos, neste caso, um problema de maximização com restrições. Lembremos que para cada i, $\sum_{j=0}^{M} P_{ij} = 1$.

Usando o Método de Lagrange, definamos, para cada i, a função auxiliar

$$\begin{aligned}
& l^*(P_{00}, P_{01}, ..., P_{MM}/i_0, i_1, ..., i_N) \\
= \ & ln(u_{i_0}) + \sum_{i=0}^{M} \sum_{j=0}^{M} n_{ij} ln(P_{ij}) + \lambda(\sum_{j=0}^{M} P_{ij} - 1).
\end{aligned}$$

Derivando com respeito a P_{ij}, temos

$$\frac{\partial l^*(P_{00}, P_{01}, ..., P_{MM}/i_0, i_1, ..., i_N)}{\partial P_{ij}} = \frac{n_{ij}}{P_{ij}} + \lambda.$$

Dai obtem-se

$$\lambda P_{ij} \;=\; -n_{ij} \tag{6}$$

Somando em j, obtemos $\lambda = -\sum_{j=0}^{M} n_{ij} = -n_i$.

Nessa relação, n_i representa o numero de vezes que a cadeia visitou o estado i na realização sendo considerada.

Finalmnete, substituindo o valor de λ em (5)

$$\hat{P}_{ij} = \frac{n_{ij}}{n_i}. \tag{7}$$

Observemos que o estimador da probabilidade de transição P_{ij} é obtido como a razão entre o número de transições, na realização da cadeia, do estado i para o estado j e o número de vezes em que a cadeia esteve no estado i .

Exemplo 3.1.2 Somente para efeito de ilustração considere a realização de tamanho $n + 1 = 23$ de uma cadeia: 3 2 1 0 1 2 3 0 1 2 3 1 0 2 1 2 3 0 3 2 1 0 2.

Naturalmente, antes de encontrarmos os estimadores das probabilidades de transição, precisamos "estimar " o espaço de estados. Neste caso, apenas os estados 0, 1, 2, 3 aparecem na realização da cadeia. Podemos pensar que o espaço de estados é $E = \{0, 1, 2, 3\}$. A cadeia visitou o estado 3, 5 vezes. Dessas 5 vezes, a cadeia fez duas transições para o estado 0, duas para o estado 2 e uma transição para o estado 1. Então

$\hat{P}_{30} = \frac{2}{5}$, $\quad\hat{P}_{31} = \frac{1}{5}$, $\quad\hat{P}_{32} = \frac{2}{5}$, $\quad\hat{P}_{33} = 0$.

Observação 1. Os estimadores de P_{ij} quando i é o último estado visitado na realização da cadeia precisam de uma modificação, pois nesse caso existem n_i visitas ao estado i, mas apenas $n_i - 1$ transições. O denominador na definição de $\hat{P}_{ij}$ precisa ser, então $n_i - 1$. No exemplo anterior, o estado 2 é o último estado visitado e a cadeia visitou $n_2 = 7$ vezes o esatdo 2. Se usarmos $n_2 = 7$ no denominador, teriamos
$\hat{P}_{20} = \frac{0}{7}$, $\quad\hat{P}_{21} = \frac{3}{7}$, $\quad\hat{P}_{22} = \frac{0}{7}$, $\quad\hat{P}_{23} = \frac{3}{7}$.

Neste caso, a soma dessas probabilidades sería $\frac{6}{7}$. Com a modificção a ser implementada para o último estado visitado teremos:

$\hat{P}_{20} = \frac{0}{6}$, $\hat{P}_{21} = \frac{3}{6}$, $\hat{P}_{22} = \frac{0}{6}$, $\hat{P}_{23} = \frac{3}{6}$..

Naturalmente, quando n é "grande", maiormente não há diferença se usarmos denominador n_i ou $n_i - 1$.

Observação 2. No procedimento para encontrar os estimadores de máxima verossimilhanç para P_{ij}, observe que estamos assumindo que $P_{ij} > 0$.

As propriedades assintoticas dos estimadores de P_{ij} assim definidos foram estudadas por Derman (1956).

Observação 3. A abordagem feita para estimar as probabilidades de transição assume a realização de tamanho $n + 1$ da cadeia. Uma outra forma de abordar o problema é considerar várias realizações de tamanho m da cadeia. No Modelo de Mobilidade social descrito no capitulo 2, uma amostra de 3.500 pais foram classificados nas correspondentes classes (Glass e Hall (1954)). O estimador de P_{ij} foi calculado como $\hat{P}_{ij} = \frac{n_{ij}}{n_i}$, sendo n_i o número de pais na classe i e n_{ij}, o número de filhos na classe j.

Na abordagem inicial, precisariamos observar uma familia por 3500 gerações para encontrar os correspondentes estimadores.

3.2 Estimação da Distribuição Invariante

A existência da distribuição invariante de uma cadeia de Markov está estritamente ligada ao número de retornos aos estados recorrentes. No capitulo 2 vimos que se a cadeia é finita, ela possui pelo menos um estado recorrente positivo. Portanto se é irredutível, todos os estados sáo recorrente positivos. Sendo assim, se a cadeia é aperiodica, existe a distribuição invariante e ela é unica.

No caso em que o espaço de estados é infinito precisamos de mais algumas condições para garantir a existência e unicidade da distribuição invariante.

Intuitivamente, um estado é recorrente se podemos voltar a ele um número infinito de vezes e nesse caso, um estimador da probabilidade de que uma cadeia, em equilibrio, esteja no estado j será a fração $\frac{n_j}{n+1}$, sendo $n + 1$ o tamanho da realização da cadeia e n_j, o número de vezes em que a cadeia esteve no estado j. Daremos a seguir parte da fundamentação técnica desse estimador definido intuitivamente .

Seja f uma função de valores reais sobre o espaço de estados e defina

$$S_n \quad = \quad \sum_{m=0}^{n} f(X_m) \qquad \text{n=0, 1, 2, . . .} \tag{8}$$

Como um caso particular, podemos definir $f(j) = 1$ e $f(k) = 0$ para $k \neq j$. Então $\frac{S_n}{n+1}$ é exatamente a fração $\frac{n_j}{n+1}$.

Seja N_n o número de visitas ao estado j pela realização x_0, x_1, . . ., x_n da cadeia.

Se o estado j é recorrente, vimos no capitulo 2 que $\sum_{n=1}^{\infty} P_{jj}^{(n)} = \infty$.

Como consequência disso, $E(N_n/X_0 = j) = \infty$ quando $n \to \infty$, como pode ser visto a seguir:

Defina $1(X_k) = 1$ se $X_k = j$ e 0 caso contrario.

Então $N_n = \sum_{k=0}^{n} 1(X_k))$ e

$$\begin{aligned}
E(N_n/X_0 = j) &= E(\sum_{k=1}^{n} X_k/X_0 = j) \\
&= \sum_{k=1}^{n} E(X_k/X_0 = j) \\
&= \sum_{k=1}^{n} P(X_k = j/X_0 = j) \\
&= \sum_{k=1}^{n} P_{jj}^{(n)}.
\end{aligned}$$

Defina

$$T_j^{(1)} \quad = \quad min\{n > 0 : X_n = j\}, \tag{9}$$

$T_j^{(1)}$ é chamado tempo da primeira visita ao estado j. Se $X_0 = j$, $T_j^{(1)}$ é chamado tempo do primeiro retorno ao estado j. Na literatura de Processos Estocasticos, $T_j^{(1)}$ é chamado um tempo de parada ou tempo opcional. Não abordaremos na disciplina a teoria referente a tempos de parada.

Para $r > 1$, defina os tempos de retorno sucessivos ao estado j

$$T_j^{(r)} \;\; = \;\; min\{n > T_j^{(r-1)} : X_n = j\},$$

convencionalmente defina $T_j^{(r)} = \infty$ se não existir $n > T_j^{(r-1)}$ tal que $X_n = j$.

Dada a realização X_0, X_1, . . ., X_n, o conjunto de observações

$$X_{T_j^{(r)}+1}, X_{T_j^{(r)}+2}, ..., X_{T_j^{(r+1)}}$$

será chamado de bloco ou ciclo **regenerativo** e a contribuição na soma S_n de cada bloco será denotada por Z_r. Isto é, definamos

$$Z_r \;\; = \sum_{m=T_j^{(r)}+1}^{T_j^{(r+1)}} f(X_m).$$

Admitiremos, sem prova, que Z_1, Z_2, . . ., Z_{N_n} são independentes e identicamente distribuidas. A prova desse resultado baseia-se na chamada Propriedade Forte de Markov. O leitor interessado em ver os detalhes da prova pode consultar Bhattacharya e Waymire (1990) ou Asmussen (1987).

A idéia de decompor a cadeia em ciclos regenerativos foi proposta originalmente, em forma independente, por Athreya e Ney (1978) e por Nummelin (1978). Alguns detalhes podem ser vistos em Atuncar (1994).

Como consequencia desse resultado, aplicando a Lei Forte dos Grandes Números, temos que se $E(|Z_1|) < \infty$, então

$$\lim_{r\to\infty} \frac{1}{r} \sum_{i=1}^{r} Z_i = E(Z_1).$$

No que segue faremos uma suposição mais forte. Assumiremos que

$$E(\sum_{m=T_j^{(1)}+1}^{T_j^{(2)}} |f(X_m)|) < \infty.$$

Lembremos que N_n é o número de visitas , na realização da cadeia ao estado j.

S_n definido anteriormente pode ser escrito como

$$S_n = \sum_{m=0}^{T_j^{(1)}} f(X_m) + \sum_{r=1}^{N_n} Z_r - \sum_{m=n+1}^{T_j^{(N_n+1)}} f(X_m) \tag{10}$$

e

$$\begin{aligned}
\frac{S_n}{n+1} &= \frac{1}{n+1}\sum_{r=1}^{N_n} Z_r + \frac{1}{n+1}\left[\sum_{m=0}^{T_j^{(1)}} f(X_m) - \sum_{m=n+1}^{N_n+1} f(X_m)\right] \\
&= \frac{1}{n+1}\sum_{r=1}^{N_n} Z_r + R_n \\
&= \frac{N_n}{n+1}\frac{1}{N_n}\sum_{r=1}^{N_n} Z_r + R_n.
\end{aligned}$$

Ja provamos que se o estado j é recorrente, $E(N_n) \to \infty$ quando $n \to \infty$. Pode-se provar também que $N_n \to \infty$ e $R_n \to 0$ com probabilidade 1, quando $n \to \infty$.

Então, pela Lei Forte, temos que

$$\lim_{n\to\infty} \frac{1}{N_n}\sum_{r=1}^{N_n} Z_r = E(Z_1).$$

Se $f = 1$, temos que $Z_r = \sum_{m=T_j^{(r)}+1}^{T_j^{(r+1)}} 1 = T_j^{(r+1)} - T_j^{(r)}$. Então

$$\begin{aligned}
\lim_{n\to\infty} \frac{1}{N_n}\sum_{r=1}^{N_n} Z_r &= \lim_{n\to\infty} \frac{T_j^{(N_n+1)} - T_j^{(1)}}{N_n} \\
&= E(T_j^{(2)} - T_j^{(1)}).
\end{aligned}$$

E como consequencia disso, fazendo $f(X_m) = 1$ para todo m, temos:

$$\lim_{n\to\infty} \frac{n+1}{N_n} = E(T_j^{(2)} - T_j^{(1)}).$$

$E(T_j^{(2)} - T_j^{(1)})$ é o tempo médio de recorrência do estado j e $\frac{N_n}{n+1}$ é a proporção do tempo que a cadeia passa no estado j. Intuitivamente é claro que se o tempo médio de retorno ao estado j é $E(T_j^{(2)} - T_j^{(1)})$, quando $n \to \infty$, a cadeia permanece no estado j com probabilidade $\frac{1}{E(T_j^{(2)} - T_j^{(1)})}$.

Definição 3.2.1 Um estado j é recorrente positivo se $E(T_j^{(1)}/X_0 = j) < \infty$.

Estabelecemos, sem prova, a seguinte proposição:
Proposição 3.2.1. Suponha que o estado j é recorrente positivo e que f é uma função de valores reais tal que

$$E(|f(X_1)| + ... + |f(X_{T_j^{(1)}}|/X_0 = j) \quad < \quad \infty$$

Então

$$\lim_{n\to\infty} \frac{1}{n+1} \sum_{m=0}^{n} f(X_m) \quad = \quad \frac{E(f(X_1) + ... + f(X_{T_j^{(1)}})/X_0 = j)}{E(T_j^{(1)}/X_0 = j)},$$

(b) Se a cadeia é fortemente ergodica, o limite anterior não depende da distribuição inicial da cadeia.

Corolário 3.2.1. Se a cadeia é fortemente ergodica, então para todo $i \in E$ e $j \in E$,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{m=1}^{n} P_{ij}^{(m)} \quad = \quad \frac{1}{E(T_j^{(1)}/X_0 = j)}.$$

Para provar o corolário, defina $f(X_m) = 1$ se $X_m = j$ e 0 caso contrario. Como a cadeia visita uma vez o estado j em cada ciclo, nesse caso $Z_r = 1$ para todo r.

Antes de estabelecer o proximo corolário, dado um conjunto A, defina $\sharp A$ igual ao número de elementos de A. Em Matematica, $\sharp A$ é conhecido como cardinal de A.

Corolário 3.2.2. Como mais um caso particular da Proposição 3.2.1, trocando j por i, temos

$$\begin{aligned}\lim_{n\to\infty} \frac{\sharp\{m \leq n : X_m = i\}}{n+1} &= \frac{1}{E(T_i^{(1)}/X_0 = i)} \\ &= u_i. \end{aligned} \tag{11}$$

Para provar este corolário, basta definir $f(i) = 1$ e $f(k) = 0$ se $k \neq i$.

Baseados neste corolário, podemos definir o estimador de u_i; isto é

$$\hat{u}_i = \frac{\sharp\{m \leq n : X_m = i\}}{n+1}. \tag{12}$$

Observe que u_i coincide com a definição frequentista de probabilidade. Isto é, se X_0, X_1, . . ., X_n for uma amostra aleatória de uma variável aleatória X, $P(X = x_i)$ e seu correspondente estimador seriam definidos em forma similar a (11) e (12) respectivamente.

Provaremos a seguir que se $\underline{u} = (u_0, u_1, ...)$ com u_i definido em (11), $\underline{u}$ é a única distribuição invariante.

Para $r = 1, 2, 3, \ldots$; defina $N_i^{(r)} = \sharp\{m \in (T_j^{(r)}, T_j^{(r+1)}] : X_m = i\}$.

$N_i^{(r)}$ assim definido é igual ao número de vezes que a cadeia visita o estado i no ciclo r, ou seja $N_i^{(r)} = \sum_{m=T_j^{(r)}+1}^{T_j^{(r+1}} 1_i(X_m)$.

Com essa representação é fácil ver que $N_i^{1)}$, $N_i^{(2)}$, . . . é uma sequência de variáveis aleatórias independentes e identicamente distribuidas pois $N_i^{(r)}$ e $N_i^{(s)}$ para $r \neq s$ são funções de variáveis aleatórias em ciclos diferentes.

Por definição, $u_i \geq 0$. Provemos que $\sum_{i \in E} u_i = 1$:

Defina $\theta_j(i) = E(N_i^{(r)}/X_0 = j)$. Com essa definição temos que

$$\begin{aligned}
\sum_{i \in E} \theta_j(i) &= \sum_{i \in E} E(N_i^{(r)}/X_0 = j) \\
&= E(\sum_{i \in E} N_i^{(r)}/X_0 = j)) \\
&= E(T_j^{(2)} - T_j^{(1)}) \\
&= \frac{1}{u_j}.
\end{aligned}$$

A terceira igualdade vale porque o número de observações no ciclo regenerativo r é $T_j^{(r+1)} - T_j^{(r)}$ e essas variáveis sáo independentes e identicamente distribuidas.
Observemos que

$$\frac{1}{n+1} \sum_{r=1}^{N_n} N_i^{(r)} = \frac{1}{n+1} \sum_{i=1}^{n} N_i^{(r)} - \frac{1}{n+1} \sum_{i=N_n+1}^{n} N_i^{(r)}.$$

Então

$$\lim_{n\to\infty}\frac{1}{n+1}\sum_{r=1}^{N_n}N_i^{(r)} = u_i.$$

Também

$$\begin{aligned}\lim_{n\to\infty}\frac{1}{n+1}\sum_{r=1}^{N_n}N_i^{(r)} &= \lim_{n\to\infty}\frac{N_n}{n+1}\lim_{n\to\infty}\frac{1}{N_n}\sum_{r=1}^{N_n}N_i^{(r)}\\ &= u_j\theta_j(i).\end{aligned}$$

Portanto $u_i = u_j\theta_j(i)$ e

$$\begin{aligned}\sum_{i\in E}u_i &= \sum_{i\in E}u_j\theta_j(i)\\ &= u_j\sum_{i\in E}\theta_j(i)\\ &= \frac{u_j}{u_j}.\end{aligned}$$

Resta provar que $\sum_{i\in E}u_iP_{ij} = u_j$. A prova desse resultado e a unicidade foge ao alcance da disciplina e vamos omitir. Veja Bhattacharya e Waymire (1990).

3.3 Exercícios

3.3.1. Considere a distribuição exponencial com a representação $f(x) = \lambda exp(-\lambda x)$ para $x > 0$. Se X_1, X_2, . . ., X_n é uma amostra aleatória de X com essa função de densidade, encontre o estimador de máxima verossimilhana̧ de λ.

3.3.2. Seja X_1, X_2, . . ., X_n uma amostra aleatória. Encontre os estimadores de máxima verossimilhança dos correspondentes parámetros se:
a) $X \sim N(\mu,\sigma^2)$, b) $X \sim P(\lambda)$
c) $X \sim G(p)$ d) $X \sim b(20,p)$.

3.3.3. Considere a cadeia de Markov com matriz de probabilidades de transição dada no Exercicio 2.9.1. Simule uma realização de tamanho 65 da cadeia. Encontre os estimadores das probabilidades de transição e da distribuição invariante.

3.3.4. Repetir o Exercicio 3.3.3 com realizações da cadeia de tamanho 100, 200, 300, 500 e 2000. Em cada caso, defina

$$D_n = \sum_{i=0}^{4}\sum_{j=1}^{4} |\hat{P}_{ij} - P_{ij}|$$

$$A_n = \sum_{i=0}^{4} |\hat{u}_i - u_i|$$

a)Compare D_{65}, D_{100}, D_{200}, D_{300}, D_{500}, D_{2000}. Qual a sua conclusão?,
b) Compare A_{65}, A_{100}, A_{200}, A_{300}, A_{500}, A_{2000}. Qual a sua conclusão?.

3.3.5 Considerando os Exercicios 3.3.3 e 3.3.4, para cada tamanho da realização da cadeia observe os pares (i, j) tais que $n(i, j) = 0$.

3.3.6 Repita os Exercicios 3.3.3, 3.3.4 e 3.3.5 considerando a matriz

$$P = \begin{pmatrix} 0,3 & 0,0 & 0,5 & 0,2 \\ 0,1 & 0,6 & 0,0 & 0,3 \\ 0,0 & 0,7 & 0,3 & 0,0 \\ 0,2 & 0,5 & 0,0 & 0,3 \end{pmatrix}.$$

3.3.7. Para n=65 do exercicio 3.3.3 e para cada valor de n no exercicio 3.3.4, considere o ultimo estado visitado pela correspondente realização. Defina $\hat{P}_{ij} = \frac{n_{ij}}{n_i - 1}$ e $\tilde{P}_{ij} = \frac{n_{ij}}{n_i}$. Compare esses estimadores.

3.3.8. A seguinte, é uma porção de uma sequência DNA:

A C T G C A C G T G T G T C T G C A C G T A C T G C A T G C T C G T A C
T G T G T C A C T G T C G T C A C T G C C T G C A G T C A G T A C T C G
A C T C A C T G A C T C A C G T C G C T C A T G C A C G T G T C G T C A
C T A C T G C A C T A C T G A C T G A T G C A T A G T C A T C G T C A T
C G T C T C A G T A C T G C A T A T G C C A C T G C A T C G A A T G A C
T G C A G T C A C T T G C C A G T C A G T C T A A C T T G A A C A G T A
G C T A T G C A T G C T G C A A G T C A C T C G T G C A C T G C A C T G
C A A C T G C T G C G C A T G C A G T C A G T C A T G G T C A C T A C A
G T C G T C A T C A C T G A C T G C T C A C G T C C T A G T C A C T G C.

Assuma que essa porção ajusta uma cadeia de Markov de primeira ordem e estime as probabilidades de transição.

3.4 Referencias

Anderson, T. W e Goodman, L. A. Statistical Inference about Markov Chains. The Annals of Mathematical Statistics. Pgs 89-110. 1957.

Asmussen, S. Applied Probability and Queues. J. Wiley. N. York. 1987.

Athreya, K. B. e Fuh, C. D. Bootstrapping Markov chains: countable case. Journal of Statistical Planning and Inference. Pgs 311-331. 1992

Athreya, K. B. e Ney, P. A new approach to the limit theory of recuurent Markov chains. Trans. Amer. Math. Soc. 245 (1978) Pgs 493-501.

Atuncar, G. S. Statistical Inference for Real-valued Markov chains and some applications. Tese de doutorado. Department of Statistics. Iowa State University. 1994.

Bhattacharya, R. N. ; Waymire, E. C. Stochastic Processes with applications. J. Wiley. N. York. 1990.

Derman, C. Some asymptotic distribution theory for Markov chains with a denumerable number of states. Biometrika. Pgs 285-294. 1956

Hogg, R. e Craig, A. Introduction to Mathematical Statistics. Prentice Hall. N. York. 1995.

Karlin, S. ; Taylor, H. M. A first Course in Stocahstic Processes. Academic Press. N. York. 1975.

Nummelin, E. A splitting technique for Harris recurrent Markov chains. Zeitschrift fur Wahr. und verwandte Gebiete 43 (1978) pgs 309-318.

4 Capítulo 4. Processos Estocasticos com Parametro de Tempo Contínuo

Abordaremos neste capítulo os processos estocasticos com parametro de tempo contínuo. Vamos considerar o caso em que o espaço de etados é discreto (finito ou infinito enumerável) e iniciaremos com o modelo mais simples, conhecido como processo de Poisson homogeneo (no tempo). Seguidamente apresentaremos os processos de nascimento puro e na seção 3 os processos de nascimento e morte. Na seção 4 apresentamos alguns exemplos interessantes do ponto de vista prático.

4.1 Processos de Poisson

Definição 4.1.1(Processos de Contagem). Diremos que o processo $X = \{X_t : t \geq 0\}$ assumindo valores no espaço de estados $E = \{0, 1, 2, ...\}$ é um processo de contagem se:
(i) $X_0 = 0$,
(ii) X_t cresce somente em saltos,
(iii) Se $s < t$, então $X_s \leq X_t$,
(iv) É contínua à direita.

Exemplos tradicionais dos processos de contagem são aqueles que registram o número de ocorrências de um evento. Os processos que registram o número de acidentes que ocorrem em um trecho de estrada, o número de itens defeituosos produzidos por uma linha de produção, o número de chegadas a uma estação de serviço são exemplos destes processos.

Definição 4.1.2 (Processos de Poisson) Diremos que o processo de contagem $\{X_t : t \geq 0\}$ é um processo de Poisson homogeneo se possui incrementos independentes e estacionários e satisfaz as seguintes condições adicionais:

(i) $P(X_{t+h} - X_t = 1/X_t = i) = \lambda h + o(h)$ quando $h \downarrow 0$,

(ii) $P(X_{t+h} - X_t = 0/X_t = i) = 1 - \lambda h + o(h)$ quando $h \downarrow 0$,

(iii) $X_0 = 0$.

(i) quer dizer que aquela probabilidade é igual a λh mais uma função $g(h)$ tal que $\lim_{h \downarrow 0} \frac{g(h)}{h} = 0$.

Uma função satisfazendo essa condição é dita ser $o(h)$ quando $h \downarrow 0$. Por exemplo, se $g_1(h) = h^2$, $g_1(h) = o(h)$, mas $g_2(h) = \sqrt{h}$ não é de ordem $o(h)$ pois $\lim_{h\downarrow 0} \frac{\sqrt{h}}{h} = \lim_{h\downarrow 0} \frac{1}{\sqrt{h}} = \infty$. A função $g(h) = sen(h)$ também não é $o(h)$ pois $\lim_{h\downarrow 0} \frac{sen(h)}{h} = 1$.

Um exemplo importante que será usado é o seguinte: Se X é uma variável aleatória com distribuição exponencial com parametro λ, entáo $P(X \leq h) = \lambda h + o(h)$ quando $h \downarrow 0$. Vejamos a prova:

$$\begin{aligned}
P(X \leq h) &= 1 - e^{-\lambda h} \\
&= 1 - \left[1 - \lambda h + \sum_{n=2}^{\infty} \frac{(-\lambda h)^n}{n!} \right] \\
&= \lambda h + \sum_{n=2}^{\infty} \frac{(-\lambda h)^n}{n!}.
\end{aligned}$$

Resta provar que $\sum_{n=2}^{\infty} \frac{(-\lambda h)^n}{n!} = o(h)$. Para $0 < h < \frac{1}{\lambda}$,

$$\begin{aligned}
\sum_{n=2}^{\infty} \frac{(-\lambda h)^n}{n!} &= (-\lambda h)^2 \sum_{n=2}^{\infty} \frac{(\lambda h)^{n-2}}{n!} \\
&\leq (-\lambda h)^2 \sum_{n=2}^{\infty} \frac{(-\lambda h)^{n-2}}{(n-2)!} \\
&\leq (\lambda h)^2 \\
&= o(h).
\end{aligned}$$

É conveniente mencionar algumas propriedades das funções que são $o(h)$. Entre elas temos a que estabelece que soma de funções que são $o(h)$, são também $o(h)$. Isto é, se $g_1(h)$ e $g_2(h)$ são $o(h)$, então a soma $(g_1 + g_2)(h)$ é também $o(h)$ e se c é uma constatnte, então $cg_1(h)$ é também $o(h)$.

Usando essas propriedades, pode-se provar que se g_1, g_2, . . ., g_k são $o(h)$ e c_1, c_2, . . ., c_k são constantes, então $g = \sum_{i=1}^{k} c_i g_i$ é também $o(h)$.

O parametro λ envolvido em (i) e (ii) da Definição 4.1.2, é chamado taxa de ocorrência ou de nascimento do processo. Observe que (i) e (ii) implicam que $P(X_{t+h} - X_t \geq 2 / X_t = i) = o(h)$.

Esse último resultado estabelece que a probabilidade de acontecer mais de um evento em um pequeno intervalo de tempo é desprezível quando dividido pelo comprimento do intervalo. A interpretação disso é que dois ou mais eventos não ocorrem simultaneamente. Ou seja os saltos do processo são unitários.

De interesse no processo de Poisson, é calcular a função de probabilidade de X_t, isto é, queremos $P(X_t = j)$ para $t > 0$ e para $j = 0, 1, 2...,$

Para $t > 0$ e $j = 0, 1, 2, ...,$ defina $P_j(t) = P(X_t = j)$. Encontraremos essas probabilidades estabelecendo equações diferenciais para elas.

Existem duas maneiras da abordar o problema. Em ambas, é feita a decomposição do intervalo $[0, t+h]$ em dois intervalos. Na primeira considera-se $[0, t+h] = [0, h] \cup (h, t+h]$ e na segunda considera-se $[0, t+h] = [0, t] \cup (t, t+h]$. Dada a natureza da decomposição do intervalo, a primeira abordagem leva a estabelecer as equações diferenciais retrospectivas e a segunda leva a estabelecer as equações prospectivas. Na literatura em inglés tais equações são conhecidas como "backward differential equations" e "forward differential equations" respectivamente.

Abordaremos o problema usando a segunda decomposição.

Consideremos primeiramente $j = 0$:

$$\begin{aligned}
P_0(t+h) &= P(X_{t+h} = 0) \\
&= P(X_t = 0)P(X_{t+h} - X_t = 0/X_t = 0) \\
&= P_0(t)P(X_{t+h} - X_t = 0/X_t = 0) \\
&= P_0(t)[1 - \lambda h + o(h)].
\end{aligned}$$

A relação acima mostra que para que o processo esteja no estado 0 no instante $t+h$, precisa que o processo esteja no estado 0 no instante t e no intervalo $(t, t+h]$ não aconteça evento.

Voltando à relação acima, temos que

$$P_0(t+h) = P_0(t) - \lambda h P_0(t) + o(h)$$

Observe que estamos admitindo que $o(h)P_0(t) = o(h)$. Deixamos para o leitor a tarefa de dar uma explicação deste detalhe.

Dividindo por h e fazendo $h \downarrow 0$, obtemos

$$P_0'(t) + \lambda P_0(t) \ = \ 0$$

Multiplicando por $e^{\lambda t}$ essa equação, obtemos

$$\frac{d}{dt}[e^{\lambda t}P_0(t)] \ = \ 0$$

Desde que $X_0 = 0$, temos que $P_0(0) = 1$. A partir dali temos que para $t > 0$,

$$P_0(t) \ = \ e^{-\lambda t}$$

Considere agora $j = 1$:

$$\begin{aligned}
P_1(t+h) &= P(X_{t+h} = 1) \\
&= P_0(t)P(X_{t+h} - X_t = 1/X_t = 0) + P_1(t)P(X_{t+h} - X_t = 0/X_t = 1) \\
&= P_0(t)(\lambda h + o(h) + P_1(t)(1 - \lambda h + o(h))
\end{aligned}$$

Procedendo em forma analoga ao caso em que $j = 0$, e lembrando que $P_0(t) = e^{-\lambda t}$, temos

$$P_1(t+h) - P_1(t) + \lambda h P_1(t) \ = \ \lambda h e^{-\lambda t} + o(h)$$

Dividindo por h e fazemdo $h \downarrow 0$, temos

$$P_1'(t) + \lambda P_1(t) \ = \ \lambda e^{-\lambda t}$$

Resolvendo a equação diferencial (lembrando que $P_1(0) = 0$) , temos que para $t > 0$,

$$P_1(t) \ = \ \lambda t e^{-\lambda t}$$

Procedendo dessa maneira podemos encontrar $P_2(t)$, $P_3(t)$, . . . mas podemos mostrar, usando o Principio de Indução Matematica que para $t > 0$,

$$P_j(t) = \frac{e^{-\lambda t}(\lambda t)^j}{j!} \qquad \text{para } j = 1, 2, ...$$

Ja mostramos que a formula vale para $j = 1$. Suponha que a formula vale para j e mostremos que vale para $j + 1$.

$$\begin{aligned}
P_{j+1}(t+h) &= P(X_{t+h} = j+1) \\
&= P_j(t)P(X_{t+h} - X_t = 1/X_t = j) + P_{j+1}(t)P(X_{t+h} - X_t = 0/X_t = j+1) \\
&= P_j(t)(\lambda h + o(h)) + P_{j+1}(t)(1 - \lambda h + o(h)
\end{aligned}$$

A partir dali, rearrangando os termos, dividindo por h e fazendo $h \downarrow 0$, temos

$$P'_{j+1}(t) + \lambda P_{j+1}(t) = \frac{\lambda e^{-\lambda t}(\lambda t)^j}{j!}$$

Resolvendo a equação diferencial e usando a condição inicial $P_{j+1}(0) = 0$, obtemos

$$P_{j+1}(t) = \frac{e^{-\lambda t}(\lambda t)^{j+1}}{(j+1)!}$$

A partir desses resultados podemos encontrar as distribuições multidimensionais. Por exemplo para $0 \leq t_1 < t_2 < \infty$,

$$\begin{aligned} P(X_{t_1} = k_1, X_{t_2} = k_2) &= P(X_{t_1} = k_1)P(X_{t_2} = k_2 / X_{t_1} = k_1) \\ &= P(X_{t_1} = k_1)P(X_{t_2} - X_{t_1} = k_2 - X_{t_1} / X_{t_1} = k_1) \\ &= P(X_{t_1} = k_1)P(X_{t_2} - X_{t_1} = k_2 - k_1 / X_{t_1} = k_1) \end{aligned}$$

Desde que o processo de Poisson possui incrementos independentes, essa ultima expressão é igual a

$$P(X_{t_1} = k_1)P(X_{t_2} - X_{t_1} = k_2 - k_1) = \frac{e^{-\lambda t_1}(\lambda t_1)^{k_1}}{k_1!} \frac{e^{-\lambda(t_2 - t_1)}(\lambda(t_2 - t_1))^{k_2 - k_1}}{(k_2 - k_1)!}$$

Analogamente,

$$\begin{aligned} P(X_{t_1} = k_1, X_{t_2} = k_2, ..., X_{t_n} = k_n) &= \frac{e^{-\lambda t_1}(\lambda t_1)^{k_1}}{k_1!} \frac{e^{-\lambda(t_2 - t_1)}(\lambda(t_2 - t_1))^{k_2 - k_1}}{(k_2 - k_1)!} \\ &\quad ... \frac{e^{-\lambda(t_n - t_{n-1})}(\lambda(t_n - t_{n-1}))^{k_n - k_{n-1}}}{(k_n - k_{n-1})!} \end{aligned}$$

De interesse no processo de Poisson é a distribuição do tempo até a primeira ocorrência. Seja T_0 esse tempo. Observe que o evento $[T_0 > t]$ é equivalente ao evento $[X_t = 0]$. Então

$$P(T_0 > t) = P(X_t = 0) = e^{-\lambda t} \qquad \text{para } t > 0$$

Ou seja, T_0 tem distribuição exponencial com parámetro λ. Pode-se provar que se T_i é o tempo entre a i-esima e a $(i+1)$-esima ocorrência, temos que $T_1, T_2, \ . \ . \ .$ é uma sequência de variáveis aleatórias independentes e identicamente distribuidas de acordo a uma exponencail de parâmetro λ e como consequência, $W_n = T_1 + T_2 + ... + T_n$, tempo da n-esima ocorrência, tem distribuição gamma com parâmetros n e λ.

Observe que $E(T_i) = \frac{1}{\lambda}$, o recíproco da taxa de ocorrência. Essa relação é bastante intuitiva. Assuma por exemplo que $\lambda = 10$ ocorrências por hora, então $E(T_i) = \frac{1}{10}$ horas.

Problema Interessante Um problema que aparentemente não está relacionado com o processo de Poisson mas cuja solução pode ser encontrada estabelecendo uma equação diferencial é o seguinte:

O tempo de funcionamento, T_0, de uma máquina tem distribuição exponencial com parametro λ. Ao quebrar a maquina, é enviada para a oficina e o tempo de conserto, T_1, tem distribuição exponencial com parametro μ. Assuma que o tempo de transferência da maquina para a oficina é desprezível. Encontrar a probabilidade de que a maquina esteja funcionando no instante t. Assumir que no instante $t_0 = 0$, a maquina esteja funcionando.

Solução. Defina $X_t = 0$ se a maquina esta funcionando no instante t e $X_t = 1$, caso contrário e defina $P_0(t) = P(X_t = 0)$. Com estas definições, temos que

$$\begin{aligned} P_0(t+h) &= P(X_{t+h} = 0) \\ &= P_0(t)P(X_{t+h} - X_t = 0/X_t = 0) + P_1(t)P(X_{t+h} - X_t = 0/X_t = 1) \\ &= P_0(t)P(T_0 > h/X_t = 0) + P_1(t)P(T_1 \leq h/X_t = 1) \\ &= P_0(t)e^{-\lambda h} + P_1(t)[1 - e^{-\mu h}] \\ &= P_0(t)[1 - \lambda h + o(h)] + P_1(t)[\mu h + o(h)] \end{aligned}$$

Desde que $P_0(t) + P_1(t) = 1$, temos

$$P_0(t+h) - P_0(t) = -\lambda h P_0(t) + [1 - P_0(t)][\mu h + o(h)]$$

Dividindo por h e fazendo $h \downarrow 0$, obtemos

$$P_0^{'}(t) + (\lambda + \mu)P_0(t) = \mu$$

Multiplicando essa última equação por $e^{(\lambda+\mu)t}$, temos que

$$e^{(\lambda+\mu)t}[P'(t) + (\lambda + \mu)P_0(t)] = \mu e^{(\lambda+\mu)t}$$

A partir dali e levando em conta que $P_0(0) = 1$, obtemos

$$P_0(t) = \frac{\mu}{\lambda + \mu} + \frac{\lambda}{\lambda + \mu} e^{-(\lambda+\mu)t}$$

Observe que quando $t \to \infty$, $P_0(t) \to \frac{\mu}{\lambda+\mu}$. Qual a interpretação desse limite?. Voltaremos a este problema e suas generalizações posteriormente quando falarmos de processos de nascimento e morte.

Antes de proseguir com a leitura desta seção, sugerimos que o leitor aborde alguns problemas no final do capítulo (Exercicios 4.6.1 - 4.6.10)

4.2 Processos de Nascimento

Uma primeira generalização do processo de Poisson descrito anteriormente, é permitir que o parametro, λ, dependa do estado em que o processo se encontra. Isto é, substituir (i) e (ii) por

(i') $P(X_{t+h} - X_t = 1/X_t = j) = \lambda_j h + o(h)$ e

(i'') $P(X_{t+h} - X_t = 0/X_t = j) = 1 - \lambda_j h + o(h)$.

Os parametros λ_j são chamadas taxas de ocorrência quando o processo está no estado j. Procedendo em forma analoga ao caso em que λ não depende do estado, temos para $j = 0$,

$$\begin{aligned} P_0(t+h) &= P_0(t)P(X_{t+h} - X_t = 0/X_t = 0) + o(h) \\ &= P_0(t)[1 - \lambda_0 h + o(h)] \end{aligned}$$

A partir dai temos que $P_0(t+h) - P_0(t) + \lambda_0 h P_0(t) = 0$.

Lembrando que $P_0(0) = 1$, obtem-se

$$P_0(t) = e^{-\lambda_0 t}$$

Para $j = 1$, temos

$$\begin{aligned} P_1(t+h) &= P_1(t)P(X_{t+h} - X_t = 0/X_t = 1) + P_0(t)P((X_{t+h} - X_t = 1/X_t = 0) + o(h) \\ &= P_1(t)[1 - \lambda_1 h + o(h)] + P_0(t)[\lambda_0 h + o(h)] + o(h) \end{aligned}$$

Dai, obtem-se que

$$P_1(t+h) - P_1(t) = -\lambda_1 h P_1(t) + \lambda_0 h P_0(t) + o(h)$$

Dividindo por h e fazendo $h \downarrow 0$, temos

$$P_1'(t) + \lambda_1 P_1(t) = \lambda_0 e^{-\lambda_0 t}$$

Desde que $P_1(0) = 0$, obtemos

$$P_1(t) = e^{-\lambda_1 t} \int_0^t \lambda_0 e^{(\lambda_1 - \lambda_0)s} ds$$

Suponha que temos $P_j(t)$. Para obter $P_{j+1}(t)$ procedemos da seguinte maneira:

$$\begin{aligned} P_{j+1}(t+h) &= P(X_{t+h} = j+1) \\ &= P_{j+1}(t)P(X_{t+h} - X_t = 0/X_t = j+1) \\ &\quad +P_j(t)P(X_{t+h} - X_t = 1/X_t = j) + o(h) \\ &= P_{j+1}(t)[1 - \lambda_{j+1}h + o(h)] + P_j(t)[\lambda_j h + o(h)] + o(h) \end{aligned}$$

Dai obtemos

$$P_{j+1}(t+h) - P_{j+1}(t) + \lambda_{j+1} h P_{j+1}(t) = \lambda_j P_j(t) + o(h)$$

A partir dessa equação obtemos

$$\begin{aligned} P'_{j+1}(t) + \lambda_{j+1}P_{j+1}(t) &= \lambda_j P_j(t) \\ e^{\lambda_{j+1}t}[P'_{j+1}(t) + \lambda_{j+1}P_{j+1}(t)] &= \lambda_j e^{\lambda_{j+1}t} P_j(t) \end{aligned}$$

dali obtemos finalmente

$$P_{j+1}(t) = e^{-\lambda_{j+1}t} \int_0^t \lambda_j e^{\lambda_{j+1}s} P_j(s) ds$$

Observemos que um processo de nascimento é não decrescente. Neste caso, então não existe distribuição invariante.

Exemplo 4.2.1 Processo de Yule. Esse processo é muito comum na área biológica. Suponha que cada membro de uma população tem uma probabilidade $\lambda h + o(h)$ de gerar um novo membro em um intervalo de tempo de comprimento h. Assuma que no instante $t = 0$ o tamanho da população é $X_0 = N$. Assumindo independência e não interação entre os membros da população, temos que

$$\begin{aligned} P(X_{t+h} - X_t = 1/X_t = n) &= \frac{n!}{1!(n-1)!} [\lambda h + o(h)] [1 - \lambda h + o(h)]^{n-1} \\ &= n\lambda h + o(h) \end{aligned}$$

Para simplificar as contas, suponha $N = 1$. Ou seja $P_1(0) = 1$.

Da exposição precedente, pode-se mostrar que $P_0(t) = 0$ para todo $t > 0$.

Com isso, temos que $P'_1(t) + \lambda P_1(t) = 0$. resolvendo a equação e lembrando que $P_1(0) = 1$, temos que $P_1(t) = e^{-\lambda t}$.

Para encontrar $P_2(t)$, temos que

$$\begin{aligned} P_2'(t) + 2\lambda P_2(t\quad) &= \lambda P_1(t) \\ &= \lambda e^{-\lambda t} \end{aligned}$$

Resolvendo a equação diferencial, lembrando que $P_2(0) = 0$, encontra-se que

$$P_2(t) \;=\; e^{-\lambda t}(1 - e^{-\lambda t})$$

Deixamos como exercicio, provar por Indução Matemática que

$$P_n(t) \;=\; e^{-\lambda t}(1 - e^{-\lambda t})^{n-1}$$

4.3 Processos de Nascimento e Morte

Os processos contemplados até aqui são chamados processos de nascimento puro ou processos de chegada. Apresentamos a seguir os processos em que contempla-se mortes de elementos ou em aplicações reais, saidas de usuários de um sistema de serviços. Tais processos são chamados processos de nascimento e morte ou processos de chegada e saída. Para estes processos os postulados (i) e (ii) agora são

(i'') $P(X_{t+h} - X_t = 1/X_t = j) = \lambda_j h + o(h)$

$(ii''a)$ $P(X_{t+h} - X_t = -1/X_t = j) = \mu_j h + o(h)$

$(i''b)$ $P(X_{t+h} - X_t = 0/X_t = j) = 1 - (\lambda_j + \mu_j)h + o(h)$

Os parametros λ_j e μ_j são chamados taxas de nascimento e morte respectivamente quando o processo está no estado j. Do ponto de vista prático é muito natural definir $\mu_0 = 0$ pois se tem 0 elementos no sistema , ninguem pode sair ou "morrer".

Neste caso não vamos resolver as equações diferenciais. Vamos estabelece-las e a partir delas, assumindo que existe distribuição invariante do processo, encontraremos tal distribuição. No procedimento para encontrar a distribuição invariante, poderemos ver sobre qué condições (sobre os parametros) existe tal distribuição.

Do ponto de vista prático, a existência da distribuição invariante é relevante. No caso de um sistema de serviços, por exemplo, tal distribuição permitirá dimensionar o sistema visando custos mínimos de operação e satisfação dos clientes com alta probabilidade. Os critérios de dimensionamento podem ser estabelecidos pela gerência do sistema.

Para $j = 0$, temos

$$\begin{aligned} P_0(t+h) &= P_0(t)P(X_{t+h} - X_t = 0/X_t = 0) + P_1(t)P(X_{t+h} - X_t = -1/X_t = 1) + o(h) \\ &= P_0(t)[1 - \lambda_0 h + o(h)] + P_1(t)[\mu_1 h + o(h)] + o(h) \end{aligned}$$

A partir dali, obtemos

$$P_0^{'}(t) + \lambda_0 P_0(t) \quad = \mu_1 P_1(t) \tag{13}$$

Para $j \geq 1$, temos que

$$\begin{aligned} P_j(t+h) &= P_j(t)P(X_{t+h} - X - t = 0/X_t = j) + P_{j-1}(t)P(X_{t+h} - X_t = 1/X_t = j-1) \\ &\quad + P_{j+1}(t)P(X_{t+h} - X_t = -1/X_t = j+1) + o(h) \\ &= P_j(t)[1 - (\lambda_j + \mu_j)h + o(h)] + P_{j-1}(t)[\lambda_{j-1}h + o(h)] \\ &\quad + P_{j+1}(t)[\mu_{j+1}h + o(h)] + o(h) \end{aligned}$$

A partir dali, procedendo como nos casos anteriores, obtemos

$$P_j^{'}(t) \quad = \quad \lambda_{j-1}P_{j-1}(t) - (\lambda_j + \mu_j)P_j(t) + \mu_{j+1}P_{j+1}(t) \tag{14}$$

Se existe distribuição invariante, teremos que $p_j = lim_{t\to\infty}P_j(t)$ e portanto $lim_{t\to\infty}P_j^{'}(t) = 0$ para todo j. Então a partir da equação (14) obtemos

$$p_1 = \frac{\lambda_0}{\mu_1}p_0$$

A partir da equação (15) fazendo $j = 1$ obtemos

$$0 \quad = \quad \lambda_0 p_0 + \mu_2 p_2 - (\lambda_1 + \mu_1)p_1$$

e a partir dali,

$$p_2 = \frac{\lambda_0\lambda_1}{\mu_1\mu_2}p_0$$

Deixamos como exercicio para o leitor encontrar p_3.

Como um segundo exercicio, assuma que

$$p_j = \frac{\lambda_0\lambda_1...\lambda_{j-1}}{\mu_1\mu_2...\mu_j}p_0$$

e mostre que

$$p_{j+1} = \frac{\lambda_0\lambda_1...\lambda_{j-1}\lambda_j}{\mu_1\mu_2...\mu_j\mu_{j+1}}p_0$$

Temos encontrado a distribuição invariante do processo assumindo que ela existe. Se esse é o caso, teremos que $p_j > 0$ e $\sum_{j=0}^{\infty} p_j = 1$.

Mas $\sum_{j=0}^{\infty} p_j = p_0 \left[1 + \sum_{j=1}^{\infty} \frac{\lambda_0\lambda_1...\lambda_{j-1}}{\mu_1\mu_2...\mu_j}\right]$. A partir dali, vemos que para que exista distribuição invariante precisamos a condição

$$\sum_{j=1}^{\infty} \frac{\lambda_0\lambda_1...\lambda_{j-1}}{\mu_1\mu_2...\mu_j} < \infty$$

.

Sobre essa condição temos que

$$p_0 = \left[1 + \sum_{j=1}^{\infty} \frac{\lambda_0\lambda_1...\lambda_{j-1}}{\mu_1\mu_2...\mu_j}\right]^{-1}$$

4.4 Alguns Exemplos Importantes

Apresentamos nesta seção alguns modelos de filas. Não temos a intenção de abordar profundamente os modelos pois isso faz parte de uma segunda disciplina. O leitor interessado em mais detalhes pode consultar Magalhães (1996).

Modelo de Fila M/M/1. Assuma um sistema de serviços em que chegadas acontecem de acordo a um processo de Poisson com taxa λ. O sistema possui um unico servidor com tempo de atendimento com distribuição exponencial de parámetro μ. A notação M/M/1 é usada na literatura universal, dada a natureza Markoviana dos processos de chegada e saida respectivamente. Assume-se que os tempos entre chegadas são i.i.d com parametro λ e os tempos de serviço são i.i.d com parametro μ

Defina X_t: Número de "pessoas" no sistema (sendo atendidas ou esperando por serviço). Neste caso temos $\lambda_j = \lambda$ para $j = 0, 1, 2, ...$, $\mu_0 = 0$ e $\mu_j = \mu$ para $j = 1, 2, ...$

A condição $\sum_{j=1}^{\infty} \frac{\lambda_0\lambda_1...\lambda_{j-1}}{\mu_1\mu_2...\mu_j} < \infty$ neste caso é $\sum_{j=1}^{\infty} (\frac{\lambda}{\mu})^j < \infty$

Tal condição é satisfeita se e somente se $\lambda < \mu$. Em termos reais essa condição estabelece que **a taxa de chegada precisa ser estritamente menor que a taxa de serviço**. Caso contrário não existe distribuição invariante o que quer dizer que o tamanho da fila cresce indefinidamente.

Se $\lambda < \mu$, temos que

$$\begin{aligned} 1+\sum_{j=1}^{\infty}\frac{\lambda_0\lambda_1...\lambda_{j-1}}{\mu_1\mu_2...\mu_j} &= 1+\sum_{j=1}^{\infty}(\frac{\lambda}{\mu})^j \\ &= \frac{\mu}{\mu-\lambda} \end{aligned}$$

Então

$$p_0 = \frac{\mu-\lambda}{\mu}$$

e a partir dali obtemos

$$p_j = (\frac{\lambda}{\mu})^j\frac{\mu-\lambda}{\mu}$$

Observar que $\{p_0, p_1, ...,\}$ é a função de probabilidade geométrica.

A partir dessa observação concluimos que se definirmos

N: Numero de usuarios no sistema (sendo servido ou esperando por serviço) quando o sistema está em equilibrio, temos que

$$\begin{aligned} E(N) &= \sum_{j=0}^{\infty} j\frac{\mu-\lambda}{\mu}(\frac{\lambda}{\mu})^j \\ &= \frac{\lambda}{\mu-\lambda} \end{aligned}$$

A prova desse resultado é baseado no seguinte:

Sabemos que se $0 < q < 1$, $\sum_{k=0}^{\infty} q^k = \frac{1}{1-q}$. Desde que a funzccão $g(q) = \frac{1}{1-q}$ é derivável no intervalo considerado, tem-se que $\frac{d}{dq}(\sum_{k=0}^{\infty} q^k = \sum_{k=0}^{\infty} kq^{k-1} = \sum_{k=1}^{\infty} kq^{k-1} = \frac{d}{dq}(\frac{1}{1-q} = \frac{1}{(1-q)^2}$. Agora é só fazer $q = \frac{\lambda}{\mu}$.

Observe que a distribuição invariante e $E(N)$ dependem apenas de μ e λ. Na prática λ é conhecida ou estimada. Podemos dimensionar de maneira ótima o sistema procurando um valor de μ visando atingir um determinado objetivo. Por exemplo, suponha que $\lambda = 5$ chegadas por hora. Se $\mu = 6$ serviços por hora, teremos que $E(N) = 5$. Se $\mu = 7$, $E(N) = \frac{5}{2}$. Quer dizer, aumentando a capacidade do servidor para atender mais um usuário por hora, reduzimos $E(N)$ em 50%.

Um outro critério para dimensionar o sistema è o tempo de permanência, a ser denotad por T_P, de um usuário no sistema. Esse tempo é a soma do tempo de espera, a ser denotado por T_E, para iniciar o serviço e o seu proprio tempo de serviço, a ser denotado por T_S.

Calculemos a distribuição de T_E. Observe se um novo usuário chega ao sistema quando n usuários já estão no sistema, o seu tempo de espera será $T_E = T_1 + T_2 + ... + T_n$ com T_1, T_2, . . ., T_n i.i.d de acordo a uma distribuição exponencial com parametro μ. Ou seja T_E tem distribuição gamma com parametros n e μ. Então $E(T_E/N = n) = \frac{n}{\mu}$ Então

$$\begin{aligned} E(T_E) &= E(E(T_E/N) \\ &= E(\frac{N}{\mu}) \\ &= \frac{1}{\mu}\frac{\lambda}{\mu - \lambda} \\ &= \frac{\lambda}{\mu(\mu - \lambda)} \end{aligned}$$

Então

$$\begin{aligned} E(T_P) &= E(T_E) + E(T_S) \\ &= \frac{\lambda}{\mu(\mu - \lambda)} + \frac{1}{\mu} \\ &= \frac{1}{\mu - \lambda} \end{aligned}$$

Deixamos como exercicio para o leitor, provar que na verdade T_P tem distribuição exponencial com parametro $\mu - \lambda$. Para encontrar tal distribuição use

$$P(T_P \leq t) = \sum_{n=0}^{\infty} P(N = n)P(T_P \leq t/N = n).$$

Modelo de Fila M/M/c. A diferença com o caso anterior é que neste caso temos c servidores cada um com taxa de serviço igual a μ. Igual que no caso anterior, defina X_t: Número de ”pessoas” no sistema (sendo servidas ou esperando por serviço).

Neste caso temos

$$\lambda_j = \lambda \qquad \text{para } j = 0, 1, 2, ...$$

e

$$\mu_j = j\mu \qquad \text{para } j = 0, 1, ..., c-1 \qquad \text{e } c\mu \text{ para } j \geq c$$

Temos então que para $j = 1, 2, ..., c-1$,

$$p_j = \frac{\lambda^j}{j!\mu^j} p_0$$

e para $j \geq c$,

$$p_j = \frac{\lambda^c}{c!\mu^c} (\frac{\lambda}{c\mu})^{j-c}$$

Então

$$\begin{aligned} \sum_{j=1}^{\infty} p_j &= \sum_{j=1}^{c-1} \frac{\lambda^j}{j!\mu^j} p_0 + \frac{\lambda^c}{c!\mu^c} \sum_{j=c}^{\infty} (\frac{\lambda}{c\mu})^{j-c} p_0 \\ &= p_0 \left[\sum_{j=1}^{c-1} \frac{\lambda^j}{j!\mu^j} + \frac{\lambda^c}{c!\mu^c} \sum_{j=0}^{\infty} (\frac{\lambda}{c\mu})^j \right] \\ &= p_0 \left[\sum_{j=1}^{c-1} \frac{\lambda_j}{j!\mu^j} + \frac{\lambda^c}{c!\mu^c} \frac{c\mu}{c\mu - \lambda} \right] \end{aligned}$$

A partir dessa última expressão temos

$$p_0 = \left[1 + \sum_{j=1}^{c-1} \frac{\lambda^j}{j!\mu^j} + \frac{\lambda^c}{c!\mu^c} \frac{c\mu}{c\mu - \lambda} \right]^{-1}$$

O leitor deve ter observado que para existir a distribuição invariante precisamos assumir que $\sum_{j=0}^{\infty} (\frac{\lambda}{c\mu})^j < \infty$ e essa condição é equivalenta a $\lambda < c\mu$. Essa última condição quer dizer que a taxa total de serviço precisa ser estritamente maior que a taxa de chegada.

Exemplo. Suponha $\lambda = 20$ chegadas por hora e $\mu = 8$ serviços por hora. O mínimo valor para c é 3 e para essess valores temos $p_0 = 0,04494$ e a partir de p_0 obtemos os valores para as outras probabilidades. Alguns valores estão representados na Tabela 2 a seguir

Table 2: Tabela 2. Distribuição Invariante na Fila M/M/3,$\lambda = 20$, $\mu = 8$

0	0,04494
1	0,11235
2	0,14044
3	0,11703
4	0,09753
5	0,08128
6	0,06773
7	0,05644
8	0,04703
9	0,03920
10	0,03266

A partir dos valores calculados mostrados na tabela, algumas conclusões podemos obter com respeito ao sistema e tomar decisões sobre se podemos melhora-lo. Observe, por exemplo, que a probabilidade de um usuário ao chegar ao sistema tenha que esperar por serviço (entrar na fila) é $\sum_{j=3}^{\infty} p_j = 0,70227$. A probabilidade de ter pelo menos 10 pessoas no sistema, ou seja pelo menos 7 pessoas esperando é $\sum_{j=10}^{\infty} p_j = 0,19603$. Tais probabilidades parecem ser muito grandes. Podemos aumentar o valor de c ou de μ. Deixamos para o leitor fazer os calculos com $\mu = 9$ e $c = 3$ e com $\mu = 8$ e $c = 4$. Compare os resultados.

Um outro critério para avaliar o sistema é atraves do tempo medio de permanência no sistema. Ao igual que no caso do modelo de fila M/M/1, $T_P = T_E + T_S$, onde T_E é o tempo de espera para iniciar o serviço de um novo usuário e T_S o seu tempo de serviço. Calculemos o tempo mèdio de permanência de um novo usuário dado que existem n pessoas no sistema no instante da chegada.

Observe que se $n < c$, existe pelo menos um servidor livre e portanto o novo usuário não precisa esperar. Nesse caso seu tempo de permanência será apenas seu tempo de serviço. Isto é, para $n < c$, $E(T_P/N = n) = \frac{1}{\mu}$.

Se $n > c$, para que o serviço do novo usuário seja iniciado, o sistema precisa ter um servidor livre. Isso acontece quando $n - c + 1$ usuários são atendidos. Ou seja, dado que $N = n$, o tempo de espera, $T_P = T_1 + T_2 + ... + T_{n-c+1}$. T_1, T_2, . . ., T_{n-c+1} são independentes e identicamente distribuidos de acordo a uma exponencial com parametro $c\mu$ (porque?). O tempo de serviço do novo usuario tem, ao igual que no caso anterior, distribuição exponencial com parametro μ. Então

$$\begin{aligned} E(T_P/N = n) &= \frac{n-c+1}{c\mu} + \frac{1}{\mu} \\ &= \frac{n+1}{c\mu} \end{aligned}$$

No exemplo anterior, suponha que existam 8 pessoas no instante que voce chega ao sistema. O seu tempo médio de permanência no sistema será $E(T_P/N = 8) = \frac{8+1}{3\mu} = \frac{9}{24} = 0,375$

Quer dizer que se existem 8 pessoas no sistema no instante da sua chegada (5 na fila), você permanecerá no sistema por 22,5 minutos.

Para $\mu = 9$, $\lambda = 20$ e $c = 3$, o valor de p_0 é igual a 0,07846. Os outros valores de p_j estão representados na tabela seguinte:

Table 3: Tabela 3. Distribuição Invariante na Fila M/M/3

0	0,07846
1	0,17435
2	0,19372
3	0,14350
4	0,10629
5	0,07874
6	0,05830
7	0,04320
8	0,03200
9	0,02370
10	0,01756

O leitor pode fazer comparações dos dois sistemas de serviço e a partir dali, optar por um ou outro sistema.

Fila M/M/c/b. No modelo de fila M/M/c, assumimos a capacidade de fila infinita. Uma modificação deste modelo é assumir a capacidade da fila finita (sala de espera finita). Um exemplo típico deste modelo é um sistema 0800. Nesse tipo de sistemas tem-se c atendentes e b salas de espera (as desagradáveis musiquinhas que escutamos quando ligamos para um sistema 0800). Quando todos os atendentes estão ocupados e todos os "atendentes musicais" também estãoocupados e uma chegada acontece, essa chamada é perdida. Então o espaço de estados deste processo de chegadas e saídas é igual a $E = \{0, 1, 2, ..., b + c\}$. Os correspondentes parâmetros são:

$$\lambda_j = \begin{cases} \lambda & \text{para } j = 0, 1, ..., b + c - 1, \\ 0 & \text{para } j = b + c \end{cases}$$

e

$$\mu_j = \begin{cases} j\mu & \text{para } j = 0, 1, ..., c - 1, \\ c\mu & \text{para } j = c, c + 1, ..., c + b \end{cases}$$

A distribuição invariante neste caso é dada por

$$p_j = \begin{cases} \frac{\lambda^j}{i!\mu^j} p_0 & \text{para } j = 0, 1, ..., c - 1, \\ \\ \frac{\lambda^c}{c!\mu^c} (\frac{\lambda}{c\mu})^{j-c} & \text{para } j = c, c + 1, ..., c + b \end{cases}$$

O valor de p_0, então é dado por

$$[1 + \sum_{j=1}^{c-1} \frac{\lambda^j}{j!\mu^j} + \frac{\lambda^c}{c!\mu^c} \sum_{j=c}^{b+c} (\frac{\lambda}{c\mu})^{j-c}]^{-1}$$

Abordaremos a seguir um problema interessante do ponto de vista prático. Suponha uma linha de produção que precisa de N máquinas para seu pleno funcionamento. Os tempos de funcionamento dessas máquinas são independentes e identicamente distribuidos de acordo a uma exponencial com parâmetro λ. No instante da falha de uma máquina, ela é enviada para a oficina que conta com c equipes e cada equipe conserta uma máquina com tempo distribuido de acordo a uma exponencial de parâmetro μ. A linha dispõe, além disso de r máquinas reserva de tal maneira que ao falhar uma máquina, ela é substituida por uma outra da reserva (se houver disponível).

Definamos X_t: Número de máquinas na oficina. Encontraremos a distribuição invariante deste processo de nascimento e morte (chegadas e saídas da máquinas na oficina). Dados os valores de N e c, a partir da distribuição invariante, poderemos escolher valores de r de tal forma que o sistema seja apropriadamente dimensionado.

O espaço de estados deste processo de nascimento e morte é $E = \{0, 1, ..., N + r\}$. Os parâmetros são

$$\lambda_j = \begin{cases} N\lambda & \text{para } j = 0, 1, ..., r - 1, \\ (N - j + r)\lambda & \text{para } j = r, r - 1, ..., N + r \end{cases}$$

e

$$\mu_j = \begin{cases} j\mu & \text{para } j = 0, 1, ..., c - 1, \\ c\mu & \text{para } j = c, c + 1, ..., N + r \end{cases}$$

Para fixar idéias, suponha $N = 10$, $\lambda = 0,005$, $\mu = 0,02$ e $r = 2$. De acordo a esses parâmetros os tempos médios de funcionamento e conserto das máquinas são 2.000 e 50 horas respectivamente.

A partir dos resultados gerais encontrados anteriormente, temos que:

$p_1 = \frac{\lambda_0}{\mu_1} p_0 = \frac{0,005}{0,02} p_0 = 0,25 p_0,$

$p_2 = \frac{\lambda_0 \lambda_1}{\mu_1 \mu_2} p_0 = 0,03125 p_0,$

$p_3 = \frac{\lambda_0 \lambda_1 \lambda_2}{\mu_1 \mu_2 \mu_3} p_0 = 0,00390625 p_0,$

$p_4 = \frac{\lambda_0 \lambda_1 \lambda_2 \lambda_3}{\mu_1 \mu_2 \mu_3 \mu_4} p_0 = 0,00048825 p_0.$

Os valores de p_5,. . ., p_{12} são extremamente pequenos. O leitor pode calcula-los. Considerando os valores de p_5, . . ., p_{12} desprezíveis temos que $p_0 = 0,778210$, $p_1 = 0,194552$, $p_2 = 0,024319$, $p_3 = 0,0030708$, $p4 = 0,00037996$.

Observe que quando o processo estiver em equilibrio, teremos com probabilidade 0,778210, a linha de produção em pleno funcionamento e duas máquinas disponíveis para substituir alguma das máquinas que eventualmente venham falhar. Com probabilidade 0,194552, teremos uma máquina na oficina, a linha em pleno funcionamento e uma máquina disponível para substituir uma máquina que possa falhar. Podemos observar também que a probabilidade de ter no máximo duas máquinas na oficina é 0,9971. Com essa probabilidade teremos a linha de produção em pleno funcionamento. Observe também que a probabilidade de que a oficina esteja vazia é $p_0 = 0,778210$. Do ponto de vista do administrador da linha de produção essa ultima probabilidade pode ser alta ou baixa. Se essa probabilidade não for apropriada, pode-se mudar o número de equipes na oficina ou o número de máquinas reserva. Esse tipo de analise torna-se mais interessante quando N é grande. Isto é quando assume valores mais apropriados do ponto de vista prático. Convidamos ao leitor abordar o problema quando $N = 100$, $r = 10$, $c = 4$ e os mesmos valores para λ e μ. Nesses casos é conveniente elaborar um programa computacional para a abordagem do problema (Veja exercicio ...)

4.5 Exercícios

4.5.1. a) Seja $\{X_t : t \geq 0\}$ um processo de Poissom com taxa de ocorrência λ. Suponha que cada chegada é registrada com probabildade p, independentemente das outras chegadas. Seja $\{Y_t : t \geq 0\}$ o processo de chegadas registradas. Prove que $\{Y_t : t \geq 0\}$ é um processo de Poissom com taxa λp,
b) Fregueses chegam a uma loja de acordo a um processo de Poisson com taxa $\lambda = 20$ fregueses por hora. A probabilidade de que cada freguês faça uma compra é 0,4. Seja Y_t: número de compras realizadas até o instante t (em horas). Calcule $P(Y_3 = 8, Y_5 = 12)$.

4.5.2. Dizemos que dosis processos $\{X_t : t \geq 0\}$ e $\{Y_t : t \geq 0\}$ são independentes se para quaisquer l e m e $0 < t_1 < t_2 < ... < t_l$; $0 < s_1 < s_2 < ... < s_m$, os vetores $(X_{t_1}, ..., X_{t_l})$ e $(Y_{s_1}, ..., Y_{s_m})$ são independentes. Sejam $\{X_t : t \geq 0\}$ e $\{Y_t : t \geq 0\}$ dois processos de Poisson independentes com taxas λ_1 e λ_2 respectivamente. Para cada t defina $Z_t = X_t + Y_t$. Prove que o processo $\{Z_t : t \geq 0\}$ é um processo de Poisson com taxa $\lambda = \lambda_1 + \lambda_2$.

4.5.3. (Continuação). Se $\{X_t : t \geq 0\}$ e $\{Y_t : t \geq 0\}$ são dois processos de Poisson independentes, encontre a distribuição de X_t dado que $X_t + Y_t = n$.

4.5.4 Veiculos chegam a um estacionamento de acordo a um processo de Poisson com taxa $\lambda = 4$ veiculos por minuto.
a)Qual a probabilidade de que em 40 minutos cheguem no maximo 150 veiculos?,
b) Se X_t é o número de veiculos que chegam até o instante t (em minutos), encontre $P(X_4 = 15, X_6 = 21, X_10 = 30)$

4.5.5. Aproxime a probabilidade calculada no Exercicio 4.4.4 .

4.5.6. Um estacionamento tem duas entradas. Pela primeira porta, entram veiculos de acordo a um processo de Poisson com taxa $\lambda_1 = 4$ veiculos por minuto e pela segunda porta, entram veiculos com taxa $\lambda_2 = 5$ veiculos por minuto. Calcule a probabilidade de que em uma hora entrem no máximo 500 veiculos no estacionamento. É necessário fazer alguma suposição para abordar o problema?. Calcule também um valor aproximado dessa probabilidade.

4.5.7. Considere os processos de Poisson homogeneos $\{X_{1t} : t \geq 0\}$, $\{X_{2t} : t \geq 0\}$, . . ., $\{X_{kt} : t \geq 0\}$ independentes com taxas λ_1, λ_2, . . ., λ_k . Para cada t, defina $X_t = X_{1t} + X_{2t} + ... + X_{kt}$. Prove que $\{X_t : t \geq 0\}$ é um processo de Poisson homogeneo.

4.5.8 Existem cinco portas de entrada ao campus da UFMG. Pelas portas 1, 2, 3, 4 e 5 entram veiculos com taxas $\lambda_1 = 10$, $\lambda_2 = 4$, $\lambda_3 = 4$, $\lambda_4 = 8$ e $\lambda_5 = 5$ veiculos por minuto respectivamente. Calcule a probabilidade de que em 5 minutos entrem ao campus pelo menos 150 e não mais de 200 veiculos. Aproxime essa probabilidade.

4.5.9. Considere um processo de Poisson homogeneo com taxa λ.
a)Para s e t, com $0 < s < t$, Calcule $Cov(X_s, X_t)$,
b) Calcule $E(X_t/X_s)$.

4.5.10. Considere uma máquina cujo tempo de funcionamento é exponencial com média igual a 5000 horas. Quando a máquina falha é enviada para a oficina e o tempo de conserto é exponencial com média igual a 100 horas. Se no instante $t = 0$ a máquina está funcionando, calcule a probabilidade de que no instante t ela esteja funcionando.

4.5.11. Seja $\{X_t : t \geq 0\}$ um processo de nascimento puro. Assuma que:

$$P(\text{um evento ocorra em } (t, t+h]/X_t = impar) = \lambda_1 h + o(h),$$

$$P(\text{um evento ocorra em } (t, t+h]/X_t = par) = \lambda_2 h + o(h).$$

Assuma que $X_0 = 0$ e calcule $P(X_t = par)$.

4.5.12. a) No processo de nascimento e morte, a partir da expressão para p_2, mostre que

$$p_3 = \frac{\lambda_0 \lambda_1 \lambda_2}{\mu_1 \mu_2 \mu_3} p_0$$

b) Suponha que

$$p_r = \frac{\lambda_0 \lambda_1 \lambda_2 ... \lambda_{r-1}}{\mu_1 \mu_2 \mu_3 ... \mu_r} p_0$$

e prove que

$$p_{r+1} = \frac{\lambda_0\lambda_1\lambda_2...\lambda_{r-1}\lambda_r}{\mu_1\mu_2\mu_3...\mu_r\mu_{r+1}}p_0$$

4.5.13. Prove que no modelo M/M/1, o tempo de permanência, de um usuário, no sistems tem distribuição exponencial com parametro $\mu - \lambda$.

4.5.14. Suponha que $g(t)$ é a taxa de falha condicional de um equipamento no instante t, dado que não falhou até o instante t. Isto é, a probabilidade do equipamento falhar no intervalo $(t, t+h]$ dado que não falhou até o instante t é $g(t)h + o(h)$ quando $h \downarrow 0$. Suponha que $g(t)$ é positiva e continua em $(0, \infty)$ e encontre uma expressão para

$$F(t) = P(\text{falhar em algum instante } \tau,\ \tau \leq t),$$

em termos de $g(t)$.

4.5.15 Uma linha de produção opera em tempo contínuo (Liga dia 2 de janeiro e desliga 30 de dezembro de cada ano). O número de falhas que acontecem nessa linha de produção segue um processo de Poisson com taxa $\lambda = 2$ falhas por semana. A equipe de inspeção visita a linha e o tempo entre visitas segue uma distribuição exponencial com média igual a 100 horas. A linha sofre danos sérios se mais de 3 falhas acontecem sem serem detectadas. Calcule a probabilidade de que a linha não sofra danos sérios.

4.5.16 **Processo de Poisson não homogeneo** Considere um processo de Poisson em que o parametro λ depende do tempo. Isto é, considere

$$P(X_{t+h} - X_t = 1/X_t = j) = \lambda(t)h + o(h)$$

a) Prove que a probabilidade de não ter ocorrência durante o intervalo $(0, s]$ é $e^{-\int_0^s \lambda(u)du}$.

b) Prove que a probabilidade de k ocorrências durante o intervalo $(0, s]$ é

$$\frac{e^{-\int_0^s \lambda(u)du}(\int_0^s \lambda(u)du)^k}{k!}$$

4.5.17 Um Modelo de crescimento linear com mortes pode ser considerado como um processo de nascimento e morte com $\lambda_n = n\lambda$ e $\mu_n = n\mu$. Encontre a distribuição do número de individuos vivos no instante da primeira morte.

4.5.18. Uma central telefôonica possui m linhas de atendimento. Chamadas chegam de acordo a um processo de Poisson com taxa λ. As chamadas são aceitas se existir uma linha desocupada, caso contrário elas são perdidas. A duração de cada chamada é uma variável aleatória com distribuição exponencial com parametro μ. Os tempos de duração das chamadas são independentes. Encontre a distribuição invariante do processo associado a esse problema.

4.5.19 **Fila opcional.** Clientes chegam a uma estação de serviço de acordo a um processo de Poisson com parametro λ. Um cliente ao chegar na estação entra na fila com probabilidade p se a estação estiver ocupada. O tempo de atendimento tem distribuição exponencial com parametro μ. Formule esse problema como um Modelo de nascimento e morte. Encontre os parametros.

4.5.20. **Processos de Poisson Composto** a) O número de vôos que chegam no aeroporto de Confins segue um processo de Poisson com $\lambda = 10$ vôos por hora. O número de passageiros por vôo é uma variável aleatória com média $\mu = 100$ passageiros e desvio padrão $\sigma = 15$. Encontre o número esperado e a variância do número de passageios que chegam em 4 horas.
b) O número de acidentes sofridos pelos veiculos segurados por uma companhia segue um processo de Poisson com $\lambda = 6$ acidentes por dia. O custo por acidente é uma variável aleatória com média $\mu = R6.000,00$ e desvio padrão $\sigma = R1.000,00$. Encontre a média e a variância das indenizaç oes pagas pela companhia durante um mes.

4.5.21. Seja (X_t, Y_t) um processo estocastico bidimensional. $\{X_t : t \geq 0\}$ e $\{Y_t : t \geq 0\}$ são processos de Poisson independentes com taxas λ_1 e λ_2 respectivamente. No instante $t = 0$ o processo está no estado (k_1, k_2) com $k_1 + k_2 < k$. Qual a probabilidade de que o processo encontre a reta $x + y = k$ no ponto (k_3, k_4)?

4.6 Referências

Allen, A. O. Probability, Statistics, and Queueing Theory, with Computer Science Applications. 2nd edition. Academic Press. N. York. 1990.

Bhattacharya, R. N. ; Waymire, E. C. Stochastic Processes with applications. J. Wiley. N. York. 1990.

Çinlar, E. Introduction to Stochastic Processes. Prentice Hall. N. Jersey. 1975.

Karlin, S.; Taylor, H. M. A first Course in Stochastic Processes. Academic Press. N. York. 1975.

Magalhães, M. N. Introdução a redes de filas. 12o. SINAPE. Caxambu. 1996.