Capítulo II

Tópicos em Teoria de Probabilidade

"Como podemos falar em leis do acaso?
Não seria o acaso uma antítese da lei?"
Bertrand Russel

II.1 - INTRODUÇÃO

Neste capítulo serão apresentados alguns dos conceitos e ferramentas necessários à compreensão dos algoritmos randômicos. Será dada uma atenção especial ao estudo das variáveis aleatórias discretas. A máquina RAM probabilística ou máquina de Turing probabilística, analogamente aos modelos determinísticos, trabalham apenas com valores inteiros. Assim, o valor da solução gerada e o tempo de processamento são quantidades enumeráveis e podem ser representados convenientemente por variáveis aleatórias discretas.

Um *espaço de probabilidade* consiste de um conjunto (possivelmente infinito) de resultados de um experimento, cada um com uma probabilidade de ocorrência associada. Mais formalmente, um espaço de probabilidade será definido em termos de um *espaço amostral* com uma estrutura algébrica (álgebra ou σ-álgebra) e uma *medida de probabilidade* imposta sobre este espaço. A partir destes conceitos define-se probabilidade condicional, eventos independentes, variáveis aleatórias discretas e as principais distribuições de probabilidade associadas.

Maiores detalhes sobre os resultados aqui apresentados (definições, demonstrações, etc) podem ser encontrados em Meyer[1983] ou James[1981].

II.2 - ESPAÇO DE PROBABILIDADE

Ao executar um experimento, especifica-se não somente o procedimento a ser realizado, mas também, aquilo que se deseja observar. Define-se então um *espaço amostral* Ω como sendo o conjunto de todos os resultados possíveis de um dado experimento. Um subconjunto $A \subseteq \Omega$ será chamado *evento*. Se A contiver apenas um elemento então A será chamado *evento elementar*. Como exemplo, considere o seguinte experimento: jogar um dado e observar o resultado. Neste caso, o conjunto $\Omega = \{1,2,3,4,5,6\}$ representará o espaço amostral associado aos possíveis resultados do experimento e os subconjuntos $A_1 = \{2,4,6\}$ e $A_2 = \{1,3,5\}$ são eventos indicando o subconjunto dos números pares e ímpares respectivamente. Além disso, cada um dos valores de Ω representam eventos elementares.

Na realização de um experimento deve-se determinar, em algumas situações, se os objetos envolvidos no experimento são *equilibrados* ou *viciados*. Por exemplo, uma moeda ou dado podem estar viciados se algum de seus eventos elementares ocorrer com maior probabilidade em relação aos demais. Caso contrário, se os resultados do experimento ocorrem com mesma freqüência os objetos estarão equilibrados.

É comum considerar situações onde o espaço amostral Ω é apenas finito ou enumerável (pode-se ter também Ω infinito e não-enumerável). Um conjunto X será *enumerável* quando for finito ou quando existir uma bijeção $f:N\rightarrow X$ (onde $N=\{1,2,3..\}$ representa o conjunto dos números naturais). Quando existir a bijeção e X for infinito então X será chamado *infinito enumerável.* Representa-se por 2^{Ω}, o conjunto de todas as partes de Ω. As vezes pode ser interessante apenas a geração de uma subcoleção de eventos distintos de 2^{Ω}. Entretanto, nem todas as possíveis subcoleções de 2^{Ω} conduzem a um espaço de probabilidade bem definido (James[1981]). A definição de álgebra (ou σ-álgebra) apresentada a seguir irá determinar que tipo de subcoleções de 2^{Ω} conduzirão a uma caracterização precisa de espaço de probabilidade.

Definição II.1: (Álgebra)

Seja Ψ uma coleção de subconjuntos de Ω não vazio. Diz-se que Ψ define uma *álgebra* de subconjuntos de Ω se os seguintes axiomas forem satisfeitos:

(i) $\Omega\in\Psi$

(ii) Se $A\in\Psi$ então $A^c\in\Psi$ (onde A^c é o conjunto complementar de A em relação a Ω).

(iii) Se $A\in\Psi$ e $B\in\Psi$ então $A\cup B\in\Psi$. •

As propriedades básicas e intuitivas acima (conceitos primitivos) serão essenciais ao desenvolvimento da teoria do cálculo de probabilidades. Para exemplificá-las considere um espaço amostral Ω com 4 elementos.

Exemplo II.1: (Álgebra)

Seja 2^{Ω} o conjunto de todas as partes de $\Omega=\{0,1,2,3\}$ e duas subcoleções Ψ_1 e Ψ_2 como descritas a seguir:

2^{Ω} = {∅; {0}; {1}; {2}; {3}; {0,1}; {0,2}; {0,3}; {1,2}; {1,3}; {2,3}; {0,1,2}; {0,1,3}; {1,2,3}; {0,1,2,3}}

Ψ_1 = {∅; {0}; {1}; {0,1}; {2,3}; {0,2,3}; {1,2,3}; {0,1,2,3}}

Ψ_2 = {∅; {0}; {1}; {0,1}; {1,3}; {2,3}; {0,2,3}; {1,2,3}; {0,1,2,3}}

Note que a subcoleção Ψ_1 define uma álgebra de conjuntos já que as propriedades (i), (ii) e (iii) se verificam. Ao contrário, a subcoleção Ψ_2 não define uma álgebra já que o complementar de $\{1,3\}$ não pertence a Ψ_2. •

Se Ψ é uma álgebra de subconjuntos de Ω então as seguintes propriedades são satisfeitas:

P1) $\emptyset\in\Psi$;

P2) Sejam $A_1, A_2, \ldots, A_n \in \Psi$ e $n \geq 0$ um inteiro qualquer. Então $\bigcup_{i=1}^{n} A_i \in \Psi$ e $\bigcap_{i=1}^{n} A_i \in \Psi$.

P3) Se $A\in\Psi$ e $B\in\Psi$ então $A-B\in\Psi$ (onde $A-B=A\cap B^c$).

As propriedades P1, P2 e P3, cujas demonstrações são deixadas como exercício, decorrem diretamente da definição de álgebra.

Definição II.2: (σ-álgebra)

Seja Ψ uma coleção de subconjuntos de Ω não-vazio. Então Ψ define uma σ-álgebra de subconjuntos de Ω se, além das propriedades (i) e (ii) a propriedade (iv) descrita abaixo também for satisfeita, ou seja:

iv) Se $A_i \in \Psi$, para i=1,2,... então $\bigcup_{i=1}^{\infty} A_i \in \Psi$. •

Seja Ψ uma σ-álgebra de subconjuntos de Ω. A propriedade seguinte (decorrente de (i), (ii) e (iv)) garante que a interseção enumerável de eventos de Ψ também pertence a Ψ.

P4) Se $A_i \in \Psi$, para i=1,2,... então $\bigcap_{i=1}^{\infty} A_i \in \Psi$.

De maneira geral, uma σ-álgebra é fechada[1] para uma quantidade enumerável de aplicações das operações *união, interseção* e *complementar*.

Uma vez definida uma álgebra (ou σ-álgebra) sobre Ω pode-se definir agora uma *medida de probabilidade* sobre Ψ.

Definição II.3: (Medida de probabilidade)

Considere uma função *Pr(.)* definida em uma álgebra Ψ satisfazendo os seguintes postulados ou axiomas (Kolmogorov[1933]):

a) $Pr(A) \geq 0, \ \forall A \in \Psi$.

b) $Pr(\Omega)=1$.

c) (aditividade-finita) Se $A_1, A_2, .., A_n \in \Psi$ são mutuamente exclusivos (i.e. $A_i \cap A_j = \emptyset$, $\forall\ i \neq j$) então:

$$\Pr\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} \Pr(A_i)$$

Note que $Pr: \Psi \rightarrow [0,1]$. Diz-se neste caso que *Pr(.)* define uma *medida de probabilidade sobre* Ψ ou simplesmente uma *probabilidade em* Ψ.

Caso o número de eventos em Ψ seja infinito e Ψ defina uma σ-álgebra com eventos mutuamente exclusivos pode-se substituir (c) pelo seguinte axioma:

c') (σ-aditividade) $\Pr\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \Pr(A_i)$. •

Seja *Pr(.)* uma medida de probabilidade imposta sobre uma σ-álgebra ψ. Se $A \in \Psi$ as seguintes propriedades são satisfeitas:

P5) $\Pr(A^c) = 1 - \Pr(A)$. Note, como conseqüência que $\Pr(\emptyset) = 1 - \Pr(\Omega)$.

P6) $0 \leq Pr(A) \leq 1$.

P7) $A_1 \subseteq A_2 \Rightarrow \Pr(A_1) \leq \Pr(A_2)$.

P8) $\Pr\left(\bigcup_{i=1}^{n} A_i\right) \leq \sum_{i=1}^{n} \Pr(A_i)$, p/ algum *n* inteiro positivo.

P9) $\Pr\left(\bigcup_{i=1}^{\infty} A_i\right) \leq \sum_{i=1}^{\infty} \Pr(A_i)$.

A demonstração das propriedades *P5,..,P9* (deixadas como exercício) decorrem diretamente dos axiomas acima (definição II.3).

Tem-se então a seguinte definição de espaço de probabilidade:

Definição II.4: (Espaço de probabilidade)

O trio *(Ω, ψ, Pr(.))* define um *espaço de probabilidade* se e somente se:

---

[1] Um conjunto X é fechado sobre uma operação $\oplus$ se e somente se $a \oplus b \in X$, $\forall$ a, b $\in$ X.

(a) Ω for um espaço amostral não-vazio.
(b) ψ é uma σ-álgebra de subconjuntos de Ω.
(c) *Pr(.)* é uma medida de probabilidade em ψ. •

Os conceitos e definições apresentados nas seções seguintes relacionam eventos e probabilidades de um espaço de probabilidade:

II.3 - PROBABILIDADE CONDICIONAL

Nesta seção, discute-se algumas das relações existentes entre os eventos de um dado experimento. Define-se probabilidade condicional e fórmula de Bayes.

Em determinadas situações, é possível que se tenha um conhecimento prévio de alguns dos eventos associados a um experimento. Por exemplo, suponha que duas moedas equilibradas sejam jogadas e que o resultado obtido seja observado. O espaço amostral associado a este experimento será $\Omega = \{KK, CK, KC, CC\}$ onde *K* e *C* representam *cara* e *coroa* respectivamente. Suponha, entretanto, que se deseje determinar a probabilidade de duas *caras* ocorrerem sabendo que pelo menos uma *cara* tenha ocorrido. Neste caso, esta informação adicional (garantindo a existência de pelo menos uma cara), exclui a possibilidade de 2 coroas, reduzindo portanto, o conjunto dos resultados observáveis no experimento. Para responder perguntas deste tipo faz-se uso da definição de probabilidade condicional.

Definição II.5: (Probabilidade condicional)

Seja $(\Omega, \psi, Pr(.))$ um espaço de probabilidade. Se $B \in \psi$ e $Pr(B)>0$, a *probabilidade condicional* de *A* dado *B* é definida por:

$$\Pr(A|B) = \frac{\Pr(A \cap B)}{\Pr(B)}, \quad A \in \Psi. \quad \bullet$$

No exemplo acima, pode-se fazer $A = \{KK\}$ (duas caras ocorrem) e $B = \{KK, KC, CK\}$ (pelo menos uma cara ocorre). Assim, conclui-se diretamente que $Pr(A|B) = (1/4)/(3/4) = 1/3$.

Como observado a seguir, o conceito de probabilidade condicional poderá ser útil na determinação da probabilidade de interseção de 2 ou mais eventos.

Proposição P10: (Teorema da Multiplicação ou Probabilidade composta)

Considere (Ω, ψ, Pr) um espaço de probabilidade. Além disso, seja *n* é inteiro positivo e $A_1, A_2, \ldots, A_n \in \Psi$. Então:

$$\Pr(A_1 \cap A_2 \cap ... \cap A_n) = \Pr(A_1).\Pr(A_2 \mid A_1).\Pr(A_3 \mid A_2 \cap A_1).....\Pr(A_n \mid A_1 \cap ... \cap A_{n-1}) \quad \bullet$$

Exemplo II.2: (Probabilidade Composta)

Para exemplificar uma aplicação da propriedade P10 considere que 3 cartas de um baralho devam ser selecionadas sem reposição. Qual a probabilidade que exatamente 3 reis sejam retirados? Para resolver este problema considere A_i o evento "tirar um rei na i-ésima remoção". Se A representa o evento "tirar 3 reis", tem-se que $A = A_1 \cap A_2 \cap A_3$. Utilizando a fórmula de probabilidade composta conclui-se que: $\Pr(A) = \Pr(A_1).\Pr(A_2 \mid A_1).\Pr(A_3 \mid (A_1 \cap A_2) =$ (4/52) .(3/51).(2/50) = 1/5525. O que ocorre se as cartas forem selecionadas com reposição? •

Em determinadas situações deseja-se particionar o espaço amostral em um subconjunto enumerável de eventos:

Definição II.6: (Partição)

Seja $A_1, A_2,...$ um conjunto enumerável de eventos pertencentes a ψ. Os eventos $A_1, A_2,...$ definem uma *partição* do espaço amostral Ω, quando:

a) $A_i \cap A_j = \varnothing,\ \forall\ i \neq j$.

b) $\bigcup_i A_i = \Omega$ •

Para todo evento $B \in \psi$, tem-se $B = \bigcup_i (B \cap A_i)$ (vide Figura II.1). Como os A_i's são disjuntos então $B \cap A_i$ também serão disjuntos e:

$$\Pr(B) = \sum_i \Pr(B \cap A_i) = \sum_i \Pr(A_i).\Pr(B \mid A_i)\cdot$$

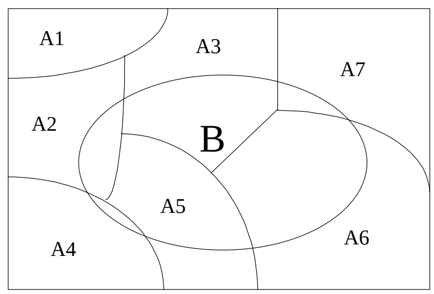

Figura II.1: Partição de um espaço amostral

Mais formalmente, tem-se a seguinte proposição:

Proposição P11: (Probabilidade absoluta)

Seja *(Ω, ψ, Pr(.))* um espaço de probabilidade. Se a seqüência enumerável de eventos aleatórios $A_1, A_2,...$ (pertencentes a *ψ*) formar uma partição de *Ω* então:

$$\Pr(B) = \sum_{i=1}^{\infty} \Pr(A_i).\Pr(B \mid A_i), \quad \forall\ B \in \Psi.$$ •

Este tipo de resultado é interessante quando a probabilidade de um evento $B \in \psi$ não for diretamente determinada. Neste caso, a probabilidade de *B* é obtida em função de eventos com probabilidade conhecida.

Exemplo II.3: (Probabilidade Absoluta)

Para ilustrar este resultado suponha que, em um lote de 100 peças, 20 sejam defeituosas e 80 sejam *não* defeituosas. Deseja-se retirar 2 peças em seguida e sem reposição. Qual a probabilidade da segunda peça ser defeituosa? Para resolver este problema considere os seguintes eventos A={a primeira peça é defeituosa } e B={a segunda peça é defeituosa}. Utilizando a fórmula da probabilidade absoluta tem-se:

$$\Pr(B) = \Pr(B \mid A).\Pr(A) + \Pr(B \mid A^c).\Pr(A^c) = \frac{19}{99}.\frac{1}{5} + \frac{20}{99}.\frac{4}{5} = \frac{1}{5}$$ •

Considere agora o seguinte resultado. Da expressão de probabilidade condicional (Definição II.5) tem-se que:

$$\Pr(A|B) = \frac{\Pr(A \cap B)}{\Pr(B)}, \quad A \in \Psi.$$

De maneira análoga, pode-se concluir que $Pr(A \cap B)=Pr(B|A).Pr(A)$. Além disso, da expressão de probabilidade absoluta: $Pr(B)=Pr(B|A).Pr(A)+Pr(B|A^c).Pr(A^c)$. Substituindo $Pr(A \cap B)$ e $Pr(B)$ na expressão de probabilidade condicional chega-se à seguinte igualdade, também conhecida como *fórmula de Bayes*:

$$\Pr(A \mid B) = \frac{\Pr(B \mid A).\Pr(A)}{\Pr(B \mid A).\Pr(A) + \Pr(B \mid A^c).\Pr(A^c)}$$

O exemplo seguinte ilustra uma aplicação da fórmula de Bayes:

Exemplo II.4: (Fórmula de Bayes)

Considere duas moedas onde uma é equilibrada e outra é viciada. A moeda viciada sempre retorna *cara*. Neste experimento uma moeda será selecionada aleatóriamente e jogada duas vezes. Deseja-se saber qual a probabilidade da moeda selecionada ser viciada sempre que forem observadas duas caras. Suponha que 3 eventos sejam considerados ao final deste experimento: A = {uma moeda viciada é selecionada}, A^c={uma moeda equilibrada é selecionada} e B = {duas caras são observadas}. Assim, aplicando-se diretamente a fórmula de Bayes tem-se:

$$\Pr(A \mid B) = \frac{\Pr(B \mid A).\Pr(A)}{\Pr(B \mid A).\Pr(A) + \Pr(B \mid A^c).\Pr(A^c)} = \frac{1.(1/2)}{1.(1/2) + (1/4).(1/2)} = \frac{4}{5}$$

Logo, a probabilidade que uma moeda viciada seja retirada, caso duas caras sejam observadas, será igual a 80%. •

Suponha que os eventos $A_1, A_2, \ldots$ (pertencentes a ψ) definam uma partição de Ω e $B \in \Omega$ seja um evento qualquer. De maneira geral, a fórmula de Bayes pode ser representada através da seguinte expressão:

$$\Pr(A_i \mid B) = \frac{\Pr(A_i).\Pr(B \mid A_i)}{\sum_j \Pr(A_j).\Pr(B \mid A_j)}$$

A fórmula de Bayes é útil quando as probabilidades de cada um dos eventos A_i e as probabilidades de B dado A_i (*p/* $i=1,2,\ldots$) são conhecidas sem que se conheça diretamente a probabilidade do evento *B*.

II.4 - INDEPENDÊNCIA ENTRE EVENTOS

É comum situações onde dois ou mais eventos associados a um experimento ocorram independentemente um do outro. Neste caso, a ocorrência ou não de um evento qualquer *B* não irá influenciar na ocorrência ou não de outro evento A. Formalmente, considere a seguinte definição de eventos independentes:

Definição II.7: (Eventos Independentes)

Seja $(\Omega, \psi, Pr(.))$ um espaço de probabilidade. Dois eventos aleatórios *A* e *B* pertencentes a ψ são *independentes* se e somente se $\Pr(A \cap B) = \Pr(A).\Pr(B)$. Em outras palavras, se $Pr(A) \neq 0$ então $Pr(B|A) = Pr(B)$. •

Exemplo II.5: (Independência entre eventos)

Suponha que um dado equilibrado seja jogado 2 vezes consecutivas. Considere que os seguintes eventos estejam associados a este experimento:

A = {o primeiro dado retorna par}
B = {o segundo dado retorna 5 ou 6}

Se x e y representam os valores obtidos no primeiro e segundo dados respectivamente então $\Omega = \{(x,y): 1 \le x \le 6 \text{ e } 1 \le y \le 6\}$. É fácil ver neste caso que: $Pr(A) = 18/36 = 1/2$ e $Pr(B) = 12/36 = 1/3$. Ainda $Pr(A \cap B) = 6/36 = 1/6$. Segue então que $Pr(A \cap B) = Pr(A).Pr(B) = 1/6$. Conclui-se portanto que A e B são eventos independentes. Note ainda que $Pr(B|A) = 12/36 = 1/3$, ou seja, $Pr(B|A) = Pr(B) = 1/3$ (a ocorrência do evento A não influencia em nada a probabilidade de B). •

Definição II.8: (Eventos *dois*-a-*dois* independentes)

Uma coleção de eventos $A_1, A_2, A_3, \ldots, A_n$ pertencentes a um espaço amostral Ω é dita *dois-a-dois independente* se e somente se $Pr(A_i \cap A_j) = Pr(A_i).Pr(A_j), \forall\, 1 \le i \le j \le n$.

Definição II.9: (Eventos mutuamente independentes)

Uma coleção de eventos $A_1, A_2, A_3, \ldots, A_n$ contidos em um espaço amostral Ω é *mutuamente independente* (ou simplesmente *independente*) se, e somente se, para qualquer subcoleção $A_{i_1}, A_{i_2}, \ldots, A_{i_k}$ (onde $2 \le k \le n$ e $1 \le i_1 \le i_2 \le \ldots \le i_k \le n$) a seguinte igualdade for satisfeita:

$$\Pr(A_{i_1} \cap A_{i_2} \cap \ldots \cap A_{i_k}) = \Pr(A_{i_1}).\Pr(A_{i_2})\ldots\Pr(A_{i_k}). \quad \bullet$$

Obviamente, se uma coleção de eventos é mutuamente independente, ela será também dois-a-dois independente, entretanto, a recíproca não é verdadeira como pode ser observado no exemplo seguinte:

Exemplo II.6: (Independência entre eventos)

Para ilustrar os conceitos acima considere o experimento onde duas moedas equilibradas são jogadas e 3 eventos distintos são observados.

A_1 = {a primeira moeda é *cara*},
A_2 = {a segunda moeda é *cara*},
A_3 = {as duas moedas são distintas}.

É fácil ver que $Pr(A_1) = Pr(A_2) = Pr(A_3) = 1/2$, $Pr(A_1 \cap A_2) = Pr(A_1).Pr(A_2)$, $Pr(A_1 \cap A_3) = Pr(A_1).\ Pr(A_3)$ e $Pr(A_2 \cap A_3) = Pr(A_2).Pr(A_3)$ sendo portanto *dois-a-dois* independentes. Entretanto, $Pr(A_1 \cap A_2 \cap A_3) \neq Pr(A_1).Pr(A_2).Pr(A_3)$. Note que, $Pr(A_1 \cap A_2 \cap A_3) = 0$ e $Pr(A_1).Pr(A_2).Pr(A_3) = 1/8$, logo os eventos A_1, A_2 e A_3 não são mutuamente independentes! •

II.5 - VARIÁVEIS ALEATÓRIAS DISCRETAS:

Em probabilidade, associa-se freqüentemente valores numéricos aos possíveis resultados ou eventos de um experimento. Estes resultados podem ser representados convenientemente através de variáveis aleatórias. Aqui, será dada uma atenção especial às variáveis aleatórias discretas. Mais formalmente, considere a seguinte definição:

Definição II.10: (Variável Aleatória)

Seja ε um experimento e *(Ω,ψ,Pr(.))* um espaço de probabilidade associado. Uma função *X* que associa a cada elemento $w \in \Omega$ um número real *X(w)* é denominada *variável aleatória.* Se $X:\Omega \rightarrow R_X$ é variável aleatória e o contradomínio R_X for finito ou infinito enumerável então *X* será uma *variável aleatória discreta.* Note por exemplo que, se R_X é infinito enumerável então $R_X=\{x_1,x_2,...\}$, onde $x_i \in R$ para $i=1,2,...$. •

Exemplo II.7: (Variáveis Aleatórias)

(a) Novamente, considere um experimento onde *duas* moedas equilibradas são jogadas. Suponha ainda que se deseje observar o número de caras obtidos neste experimento. Como *Ω = {KK, CK, KC, CC},* uma variável aleatória X pode ser obtida fazendo-se *X(KK) = 2, X(CK)= X(KC) = 1 e X(CC) = 0.* Uma outra variável *Y* pode ser gerada sobre o mesmo experimento fazendo-se, por exemplo, *Y(KK) = Y(CC) = 1* e *Y(KC) = Y(CK) = 0,* ou seja, a imagem de *Y* retorna 1 se as moedas forem e iguais e retorna 0 caso contrário.

(b) Em outro experimento, um dado equilibrado é jogado e seu resultado observado. O experimento deve ser repetido até que o número 6 seja obtido. Suponha que este problema seja resolvido por um computador onde cada escolha aleatória consuma uma unidade de tempo (p. ex., 1 segundo) . Neste caso, pode-se definir uma variável aleatória *X* representando o tempo total de processamento até que a propriedade P, desejada pelo algoritmo, seja satisfeita (neste caso, sortear o número 6). É fácil ver neste exemplo que $R_X=\{1,2,3,...\}$, ou seja, $x_i=i$ para $i=1,2,3....$ •

Definição II.11: (Função Densidade de Probabilidade)

Seja X uma variável aleatória discreta associada a um número infinito enumerável de valores $x_1,x_2,x_3,....$ A cada número x_i será associado um número $Pr(x_i) = Pr(X= x_i)$, denominado probabilidade de x_i. Os valores $Pr(x_i)$, para $i=1,2,...$ deverão satisfazer às seguintes condições:

a) $Pr(x_i) \geq 0$, para $i = 1,2,3,...$

b) $\sum_{I=1}^{\infty} \Pr(x_i) = 1$

A função $Pr:R_X \rightarrow [0,1]$ é denominada *função densidade de probabilidade.* A coleção $[x_i, Pr(x_i)]$ é algumas vezes chamada de *distribuição de probabilidade.* •

Note que, se *X* é variável aleatória e $x \in \Re$, o evento *X=x* estará associado ao conjunto $\{w \in \Omega: X(w)=x\}$. Assim:

$$\Pr(X = x) = \sum_{w \in \Omega} \Pr(w)$$

Considere o seguinte exemplo:

Exemplo II.8: (Função densidade de probabilidade)

(a) Considere novamente o exemplo II.7.(a) acima. Como as *duas* moedas envolvidas no experimento são equilibradas, é fácil ver que *Pr(X=2) =1/4, Pr(X=1) =1/2 e Pr(X=0) =1/4.* Na segunda variável conclui-se diretamente que *Pr(Y=1) = Pr(Y=0) =1/2.*

(b) No experimento II7.(b) pode-se fazer, por exemplo, $Pr(X=n) = (1-p)^{n-1}.p$ onde $n=1,2,3,....$ Neste caso, *n* representa o tempo total de execução em segundos do algoritmo randômico (número de "chutes") e *p*=1/6 representa a probabilidade de ocorrência da propriedade P (sortear o número 6). Enquanto o número 6 não for obtido (evento complementar) um novo "chute" será realizado.

Observe ainda que, nestes *dois* experimentos as condições (a) e (b) da Definição II.11 são atendidas (verifique). •

Definição II.12: (Função Densidade de Probabilidade Conjunta)

Suponha que X e Y sejam duas variáveis aleatórias discretas associadas a um espaço amostral Ω. A *função densidade de probabilidade conjunta* $f_{XY}: R_X x R_Y \rightarrow [0,1]$ nas variáveis X e Y é definida por: $f_{XY}(x,y)=Pr(X=x \text{ e } Y=y)$. •

Da definição acima tem-se que, se $y \in R_Y$ é fixo então:

$$\Pr(Y=y)=\sum_x \Pr(X=x \ \ e \ \ Y=y)\cdot$$

Analogamente, se $x \in R_X$ é fixo então:

$$\Pr(X=x)=\sum_y \Pr(X=x \ \ e \ \ Y=y)\cdot$$

Da Definição II.5, de probabilidade condicional, conclui-se que:

$$\Pr(X=x \mid Y=y)=\frac{\Pr(X=x \ \ e \ \ Y=y)}{\Pr(Y=y)}$$

Assim, duas variáveis aleatórias discretas X e Y associadas a um espaço amostral Ω serão independentes se, e somente se, $Pr(X=x \text{ e } Y=y) = Pr(X=x).Pr(Y=y)$, qualquer que seja $x \in R_X$ e $y \in R_Y$. Equivalentemente, X e Y serão independentes se $Pr(X=x|Y=y) = Pr(X=x)$.

A seguir define-se *valor esperado* ou *expectância* de uma variável aleatória:

Definição II.13: (Valor Esperado)

Seja X uma variável aleatória discreta. Se $R_X=\{x_1,x_2,,...\}$ e $Pr(X=x_i)=Pr(x_i)$, para $i=1,2,3....$, o *valor esperado* $E(X)$ será denotado por:

$$E(X)=\sum_{i=1}^{\infty} x_i \Pr(x_i)$$

se a série $\sum_{i=1}^{\infty} x_i \Pr(x_i)$ convergir absolutamente, isto é, se $\sum_{i=1}^{\infty} | x_i | \Pr(x_i) < \infty \cdot$ •

Note que podemos ter $\Sigma_{n\to\infty} a_n = k > 0$ (série convergente) mas $\Sigma_{n\to\infty}|a_n| = +\infty$ (série divergente). Por exemplo, a série S = 1 – 1/2 + 1/3 – 1/4 + ... = log2. Entretanto, a série harmônica S' = S = 1 + 1/2 + 1/3 + 1/4 + ... = +∞ (série divergente). Para maiores detalhes sobre este assunto consulte Lima[1982].

Definição II.14: (Valor esperado de uma função de uma variável aleatória)

Seja X uma variável aleatória discreta assumindo valores $x_1, x_2, x_3,...$ e seja $Y = H(X)$ uma de função de X. O *valor esperado* $E(Y)$ será dado por:

$$E(Y)=E(H(X))=\sum_{i=1}^{\infty} H(x_i).\Pr(x_i)$$

se a série $\sum_{i=1}^{\infty} H(x_i).\Pr(x_i)$ convergir absolutamente. •

Definição II.15: (Valor esperado condicionado)

Seja (X,Y) uma variável aleatória discreta bidimensional. Define-se *valor esperado condicionado* de X para um dado $Y=y_i$ como sendo:

$$E(X \mid y_i) = \sum_{i=1}^{\infty} x_i \ \Pr(x_i \mid y_i)$$

O valor esperado condicionado de Y, para um dado X, é definido analogamente. •

Sejam *X* e *Y* variáveis aleatórias definidas sobre um espaço amostral *Ω*. Algumas propriedades importantes do valor esperado podem ser obtidas com base nas definições acima:

P12) Se *c* é constante então *E(cX+Y) = cE(X)+E(Y).*
P13) *E(X.Y) = E(X).E(Y),* se *X* e *Y* são independentes.

De maneira geral, se $X_1, X_2, ..., X_n$ são variáveis aleatórias sobre o espaço amostral *Ω* então valem as seguintes propriedades:

P14) $E(X_1 + X_2 + ... + X_n) = E(X_1) + E(X_2) + ... + E(X_n)$;
P15) $E(X_1.X_2......X_n) = E(X_1).E(X_2).....E(X_n)$, se $X_1, X_2, ..., X_n$ forem mutuamente independentes;
P16) *E(E(X|Y))=E(X)* e *E(E(Y|X))=E(Y).*

As demonstrações das propriedades P12 a P16 são deixadas como exercício. O exemplo seguinte ilustra uma utilização da propriedade (P14):

Exemplo II.9: (Linearidade do valor esperado)
Um navio atraca em um porto marítimo com 40 marinheiros a bordo. Todos saem a noite para comemorar a chegada. Ao voltarem para o navio, todos eles embriagados em função da grande comemoração, escolhem aleatóriamente uma cabine para dormir. Qual o número esperado de marinheiros que conseguem retornar para sua própria cabine?
Seja X_i uma variável aleatória que indica se o *i-ésimo* marinheiro encontrou ou não sua própria cabine. Assim, $X_i = 1$ em caso afirmativo e $X_i = 0$, caso contrário. Desta forma, tem-se que 1/40 é a probabilidade do i-ésimo marinheiro encontrar sua própria cabine. Logo, $E(X_i) = 1.(1/40) + 0.(39/40)$. Da linearidade do valor esperado tem-se que:

$$E\left(\sum_{i=1}^{40} X_i\right) = \sum_{i=1}^{40} E(X_i) = \sum_{i=1}^{40} \frac{1}{40} = 1.$$

Espera-se então que apenas um marinheiro encontre sua própria cabine. •

A *variância* de uma variável aleatória *X*, assim como o *desvio padrão*, representam medidas de variação de *X* em torno de seu valor esperado *E(X).* Mais formalmente:

Definição II.16: (Variância e Desvio Padrão)
Seja *X* uma variável aleatória. A *variância* de X é definida como: $\sigma_X^2 =$ *V(X)* = $E\left[(X-\mu_X)^2\right]$, onde E(X) = μ_X. O *desvio-padrão* de X será igual a $\sigma_X = \sqrt{\sigma_X^2}$. •

São listadas a seguir algumas propriedades da envolvendo a variância de uma ou mais variáveis aleatórias:

P17) Se *c* é constante então $V(cX) = c^2 V(X)$;

P18) Se *c* é constante então $V(X+c)=V(X)$;
P19) $V(X)=E(X^2)-(E(X))^2$;

A demonstração dessas propriedades são deixadas como exercício.

Proposição P20: Sejam $X_1, X_2,.., X_N$ variáveis aleatórias independentes. Se $X = X_1+...+X_N$ então $V(X)=V(X_1)+...+V(X_N)$.

Prova: Considere inicialmente a seguinte notação:

$$\mu_i = E(X_i) \text{ e } \mu = E(X) = \sum_{i=1}^{m} E(X_i) = \sum_{i=1}^{m} \mu_i$$

Assim,

$$V(X) = E[(X-\mu)^2] = E\left[\left(\sum_{i=1}^{n} X_i - \sum_{i=1}^{n} \mu_i\right)^2\right] = E\left[\left(\sum_{i=1}^{n} (X_i - \mu_i)\right)^2\right].$$

Fazendo $Y_i = X_i - \mu_i$ segue da linearidade do valor esperado que:

$$E\left[\left(\sum_{i=1}^{n} Y_i\right)^2\right] = \sum_{i=1}^{n} E(Y_i^2) + 2\sum_{i<j} E(Y_iY_j)$$

Como $X_1, X_2,.., X_n$ são independentes os pares X_i, X_j juntamente com os pares $X_i - \mu_i$ e $X_j - \mu_j$ também o são.

Como *E(XY) = E(X).E(Y)* (proposição P13) então:

$$V(X) = E[(X-\mu)^2] = E\left[(X_i-\mu_i)^2\right] + 2\sum_{i<j} E(X_i-\mu_i).E(X_j-\mu_j)$$

Mas $E(X_i-\mu_i) = E(X_i)-\mu_i = 0.$ Assim: $V(X)=V(X_1)+...+V(X_N)$ •

O resultado seguinte mostra de maneira mais precisa a variabilidade da variável aleatória X em relação a seu valor esperado *E(X).*

Proposição P21: (Desigualdade de Tchebyshev)

Se X é variável aleatória com expectância μ_X e desvio padrão σ_X então:

$$\Pr\left(|X-\mu_X| \geq t\sigma_X\right) \leq \frac{1}{t^2}, \quad \forall\, t \in \Re^+.$$ •

A Figura II.2 ilustra melhor a utilização da desigualdade de Tchebyshev:

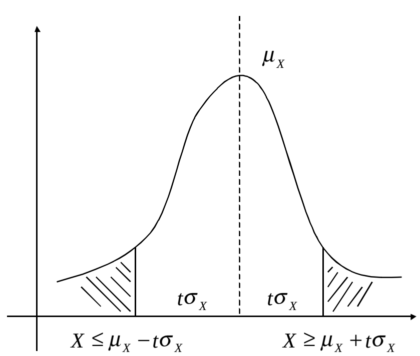

Figura II.2: Desigualdade de Tchebyshev

II.6 - ALGUMAS DISTRIBUIÇÕES DE PROBABILIDADE IMPORTANTES

Apresenta-se a seguir algumas distribuições de probabilidade bastante comuns no estudo das variáveis aleatórias discretas:

Definição II.17: (Distribuição de Bernoulli)

Uma variável aleatória *X* satisfaz a distribuição de Bernoulli com parâmetro *p* se, e somente se, $Pr(X = 0) = 1\text{-}p$ *e* $Pr(X = 1) = p$. É fácil ver neste caso que $E(X) = p$ e $V(X) = p - p^2$ (verifique).

Definição II.18: (Distribuição Binomial)

Seja ε um experimento e A algum evento associado com probabilidade $Pr(A)=p$ (portanto, $p(A^c)=1\text{-}p$). Para *n* repetições de ε, o espaço amostral será formado por todas as seqüências possíveis $(x_1,x_2,...,x_n)$ onde $x_i \in A$ ou $x_i \in A^c$ *p/* $i=1,2,..,n$. Neste caso, a variável aleatória *X*, representando o número total de ocorrências do evento *A*, define uma distribuição binômial com parâmetros *n* e *p*.

Se $X=k$ então:

$$\Pr(X = k) = \binom{n}{k} p^k.(1-p)^{n-k} \qquad \text{onde } k=0,1,2,...,n$$ •

Proposição P22: Se uma variável aleatória *X* com parâmetros *n* e *p* é binomialmente distribuída então $E(X)=n.p$ e $V(X)=n.p(1\text{-}p)$.

Prova: Para provar que $E(X) = n.p$ considere uma variável aleatória Y_i que assume valor 1 sempre que o evento *A* ocorrer na i-ésima repetição do experimento, e assume 0, em caso contrário. Se $X=Y_1+Y_2 + ...+Y_n$, tem-se, da linearidade do valor esperado que $E(X) = E(Y_1)+ E(Y_2)+ ...+E(Y_n)$. Como $E(Y_i) = 1.Pr(Y_i=1) = p$, então segue diretamente que $E(X) = n.p$.

A prova de que $V(X)=n.p(1\text{-}p)$ é deixada ao leitor como exercício. •

Exemplo II.10: (Distribuição Binomial)

Suponha que uma urna contenha um conjunto de bolas azuis (A) e vermelhas (V). Dessas, 20% são azuis e 80% são vermelhas. Considere um experimento onde 3 bolas são selecionadas ao acaso e com reposição. Seja X uma variável aleatória indicando o número de bolas azuis selecionadas. É fácil ver que $\Re_X=\{0,1,2,3\}$ define o conjunto dos possíveis valores de *X*. Qual o valor de $Pr(X=2)$? Observe neste caso que a variável X define uma distribuição binomial com parâmetros $n=3$ e $p=0,2$. Assim:

$$\Pr(X = 2) = \binom{3}{2}(0,2)^2.(0,8) = 0,096.$$

O valor esperado para o número de bolas azuis selecionadas será $E(X)=0,6$ e $V(X) = 0,48$. •

Definição II.19: (Distribuição Geométrica)

Considere um experimento obtido através de uma seqüência de ensaios de Bernoulli. Seja *p* a probabilidade de sucesso e *1-p* a probabilidade de fracasso. Quantos passos (valor esperado) serão executados até um evento com sucesso tenha sido encontrado?

Seja $X=k$ o número de repetições até que um sucesso tenha sido obtido. Assim:

$$\Pr(X = k) = (1-p)^{k-1}.p \quad \text{onde } k=1,2,3,...$$ •

Proposição P23: Seja X uma variável aleatória com distribuição geométrica e parâmetro *p*. Então *E(X)=1/p* e $V(X) = (1-p)/p^2$. •

Prova: Fazendo *q =1-p*, segue da definição de valor esperado que:

$$E(X) = \sum_{i=1}^{\infty} i.p.q^{i-1} = p\sum_{i=1}^{\infty} \frac{d}{di} q^i \cdot$$

Assim, da linearidade da derivada e da derivada do quociente entre duas funções (vide Lima[1976]) conclui-se que:

$$E(X) = p\sum_{i=1}^{\infty} \frac{d}{di}.q^i = p.\frac{d}{di}\left(\sum_{i=1}^{\infty} q^i\right) = p.\frac{d}{di}\left(\frac{1}{1-q}\right) = \frac{1}{p} \cdot$$

De maneira análoga, prova-se que $V(X) = (1-p)/p^2$. •

Exemplo II.10: (Distribuição Geométrica)

Suponha que um dado equilibrado seja jogado e seu resultado observado (exemplo II7(b)). Qual o valor esperado para o número de jogadas até que o número 6 seja obtido? Como *p=1/6, X* define uma variável com distribuição geométrica e probabilidade de sucesso *p = 1/6.* Segue então de proposição P23 que *E(X) = 6* e *V(X) = 30.* •