1 Elementos de Probabilidade e Estatística

Neste documento abordaremos de forma resumida os principais tópicos estudados em cursos de Probabilidade e Estatística. A idea é fornecer elementos para que o leitor entenda alguns algoritmos em Inteligência Artificial que fazem uso das teorias aprendidas nessas áreas.

1.1 Probabilidade

Toda a teoria desenvolvida em probabilidade está fundamentada em três conceitos: experimento aleatório, espaço amostral e evento.

1. Um **experimento aleatório** ε (ou não determinístico) satisfaz as seguintes propriedades:

   a) pode ser repetido indefinidamente sob condições que não se alteram;

   b) o resultado de cada experimento realizado não pode ser previsto com certeza, mas é possível descrever o conjunto de todos os resultados possíveis.

2. Para cada experimento aleatório ε, definimos o **espaço amostral** S como sendo o conjunto de todos os resultados possíveis de ε.

3. Um **evento** A (relativo a um espaço amostral S, associado a um experimento ε) é um conjunto de resultados possíveis, ou seja, é um subconjunto de S.

- **Exemplo 1:** Considere a tarefa de jogar um dado e observar o número mostrado na face de cima. Descreva ε e S neste caso.

◇ **Solução:** O experimento aleatório ε neste caso é o lançamento do dado e o espaço amostral é o conjunto formado por todos os números que podem resultar desse lançamento, ou seja, $S = \{1, 2, 3, 4, 5, 6\}$.

- **Exemplo 2:** Ainda considerando o lançamento de um dado, forneça dois exemplos de eventos associados a este experimento.

◇ **Solução:** Um evento A_1 pode ser "um número par ocorre" e, neste caso, temos que $A_1 = \{2, 4, 6\}$. Um evento A_2 pode ser "a face 5 ocorre" e, neste caso, temos que $A_2 = \{5\}$.

1.1.1 Definição de probabilidade

Seja ε um experimento e S um espaço amostral associado a ε. A cada evento A, associamos um número real $P(A)$ denominado probabilidade de A e que satisfaz as seguintes propriedades:

a) $0 \leq P(A) \leq 1$,

b) $P(S) = 1$,

c) se A e B são dois eventos mutuamente excludentes, ou seja, não ocorrem conjuntamente, então $P(A \cup B) = P(A) + P(B)$. Esta identidade diz que a probabilidade de A ou B ocorrerem é igual a soma de suas probabilidades individuais.

Suponha que um espaço amostral S é finito e possui k resultados igualmente prováveis (ou verossímeis) e considere A um evento associado a S que possui r desses resultados. Então,

$$P(A) = \frac{r}{k}.$$

- **Exemplo 3:** Um dado é lançado e todos os resultados se supõem igualmente verossímeis. O evento A ocorrerá se, e somente se, um número maior do que 4 aparecer na face de cima após o lançamento. Descreva S, A e calcule $P(A)$.

◇ **Solução:** $S = \{1, 2, 3, 4, 5, 6\}$, $A = \{5, 6\}$ e, portanto,

$$P(A) = \frac{\text{número de resultados desejados}}{\text{número total de possíveis resultados}} = \frac{2}{6} = \frac{1}{3}.$$

1.1.2 Probabilidade condicionada e independência de eventos

Sejam A e B dois eventos associados a um experimento aleatório ε. A probabilidade condicionada do evento B quando A tiver ocorrido, denotada por $P(B \mid A)$ é definida por

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

- **Exemplo 4:** Dois dados equilibrados são lançados, registrando-se o resultado como (x_1, x_2), onde x_i é o resultado do i-ésimo dado ($i = 1, 2$). O espaço amostral desse esperimento pode ser representado por uma lista contendo 36 resultados igualmente prováveis:

$$S = \left\{ \begin{array}{cccc} (1,1), & (1,2), & \dots & (1,6), \\ (2,1), & (2,2), & \dots & (2,6), \\ \vdots & & & \vdots \\ (6,1), & (6,2), & \dots & (6,6) \end{array} \right\}.$$

Considerando os eventos $A = \{(x_1, x_2) \mid x_1 + x_2 = 10\}$ e $B = \{(x_1, x_2) \mid x_1 > x_2\}$, calcule $P(B \mid A)$.

◇ **Solução:** De acordo com a definição dos eventos, temos que $A = \{(4,6),(5,5),(6,4)\}$ (3 elementos) e $B = \{(2,1),(3,1),(3,2),\ldots,(6,5)\}$ (15 elementos). Portanto, as probabilidades de cada evento ocorrer são dadas, respectivamente, por $P(A) = 3/36$ e $P(B) = 15/36$. Podemos calcular a probabilidade condicionada $P(B \mid A)$ de duas formas:

a) diretamente, pela consideração da probabilidade de B em relação ao espaço amostral reduzido de A, ou

b) empregando a definição de probabilidade condicionada dada acima, onde $P(A \cap B)$ e $P(A)$ são calculadas em relação ao espaço amostral original S.

Para calcular $P(B \mid A)$ diretamente, note que o espaço amostral reduzido (o conjunto A) contém 3 elementos e apenas 1 deles satisfaz que $x_1 > x_2$. Portanto, segue que $P(B \mid A) = 1/3$. Para utilizar a definição de probabilidade condicionada, devemos perceber que apenas o elemento $(6,4)$ de S cumpre ao mesmo tempo as condições dadas nos eventos A e B. Portanto, $P(A \cap B) = 1/36$ e como $P(A) = 3/36$, segue que

$$P(B \mid A) = \frac{1/36}{3/36} = \frac{1}{3},$$

concordando com o resultado obtido pelo cálculo direto.

Dizemos que A e B são eventos independentes se, e somente se,

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

ou seja, a probabilidade de ambos ocorrerem é igual ao produto de suas probabilidades individuais.

- **Exemplo 5:** Suponha que joguemos dois dados e definamos os eventos A e B da seguinte forma:

$$\begin{aligned} A &= \{\text{o primeiro dado mostra um número par na face de cima}\}, \\ B &= \{\text{o segundo dado mostra um número ímpar na face de cima}\}. \end{aligned}$$

Calcule $P(A \cap B)$.

◇ **Solução:** Evidentemente A e B são eventos independentes, pois saber que B ocorreu não fornece qualquer informação sobre a ocorrência de A. Assim, basta fazer

$$P(A \cap B) = P(A)P(B) = \frac{3}{6} \cdot \frac{3}{6} = \frac{1}{4}.$$

1.2 Variáveis aleatórias

Considere ε um experimento aleatório ϵ e S um espaço amostral associado a ε. Uma variável aleatória é uma função X que associa a cada elemento $s \in S$ um número real $X(s)$. Note que variáveis aleatórias representam por números os resultados de experimentos aleatórios. No entanto, sabemos que os resultados de um experimento aleatório podem não ser numéricos.

Neste documento vamos focar em variáveis aleatórias unidimensionais, que podem ser divididas em discretas e contínuas. As principais propriedades destes dois tipos de variáveis serão discutidas brevemente a seguir.

1.3 Variável aleatória discreta

Uma variável aleatória X será discreta se o número de valores possíveis para X for finito ou infinito enumerável. Isso significa que os valores assumidos por X podem ser colocados em uma lista. No caso finito, a lista acaba, e no caso infinito enumerável, a lista continua indefinidamente.

- **Exemplo 6:** Considere o experimento de lançar uma moeda duas vezes e o evento de contar o número de caras que surgem em cada lançamento. Descreva o espaço amostral S, a variável aleatória X e os valores que X pode assumir.

◇ **Solução:** Denotando a ocorrência de cara e coroa, respectivamente, por C_a e C_o, o espaço amostral é formado por todas as possibilidades que podem surgir nos dois lançamentos da moeda, ou seja, $S = \{C_aC_a, C_aC_o, C_oC_a, C_oC_o\}$. Assim, a variável aleatória X é o número de caras obtidas ao final do experimento. Portanto, X pode assumir os valores: $x_1 = 0$, $x_2 = 1$ e $x_3 = 2$.

A cada possível resultado x_i de uma variável aleatória X associaremos um número real $p(x_i) = P(X = x_i)$, denominado probabilidade de x_i. O conjunto de pares ordenados $(x_i, p(x_i))$, com $i = 1, 2, \ldots$ é chamado de distribuição de probabilidade de X. Quando tratamos de variáveis aleatórias discretas em inteligência artificial, as distribuições de probabilidade mais comuns são a Binomial e a Poisson, que veremos a seguir.

1.3.1 Distribuição binomial

A distribuição binomial é a distribuição de probabilidade de uma variável aleatória discreta que fornece o número de vezes que um evento ocorre dentro de n repetições de um experimento que só possui dois resultados possíveis: 1 (considerado sucesso) com probabilidade p e 0 (considerado fracasso) com probabilidade $1 - p$. Este experimento é chamado de processo de Bernoulli. Sejam A o evento de se obter sucesso e X o número de vezes em que A ocorreu nas n repetições do experimento. Então, a probabilidade de A ocorrer k vezes é dada por

$$P(X = k) = C_n^k \, p^k (1 - p)^{n-k}, \quad \text{com} \quad k = 0, 1, 2, \ldots, n,$$

onde, C_n^k é quantidade de vezes que podemos selecionar k em n sem levar em consideração a ordem da escolha, isto é,

$$C_n^k = \frac{n!}{k!(n-k)!}.$$

Na fórmula acima as probabilidades são multiplicadas porque todas as repetições são independentes.

- **Exemplo 7:** Suponha que uma válvula eletrônica, instalada em determinado circuito, tenha probabilidade 0.2 de funcionar mais do que 500 horas. Se testarmos 20 válvulas, qual será a probabilidade de que delas, exatamente 3 funcionem mais que 500 horas?

◇ **Solução:** Sendo X o número de válvulas que funcionam mais de 500 horas, devemos calcular $P(X = 3)$. Então, usando a fórmula da distribuição binomial, segue que

$$P(X = 3) = C_{20}^3 \, (0.2)^3 (0.8)^{20-3} = 0.205.$$

- **Exemplo 8:** Considere o experimento de lançar um dado e seja A o evento em que a face voltada para cima após o lançamento é igual a 1, com probabilidade $p = 1/6$. Sabendo que o experimento foi repetido 30 vezes e que X é o número de vezes em que A ocorreu nas repetições, construa o gráfico da distribuição de probabilidade associada à variável X.

◇ **Solução:** Para cada $k = 0, 1, 2, \ldots, 30$, considerando $n = 30$ e $p = 1/6$, utilizamos a fórmula da distribuição binomial dada anteriormente para calcular $p(x_k) = p(k) = P(X = k)$. Então, munidos dos pares ordenados $(x_k, p(x_k))$, podemos construir o gráfico da distribuição de probabilidade associada à variável X como mostra a Figura 1. Embora tenhamos considerado a face 1 neste exemplo, obteríamos o mesmo gráfico

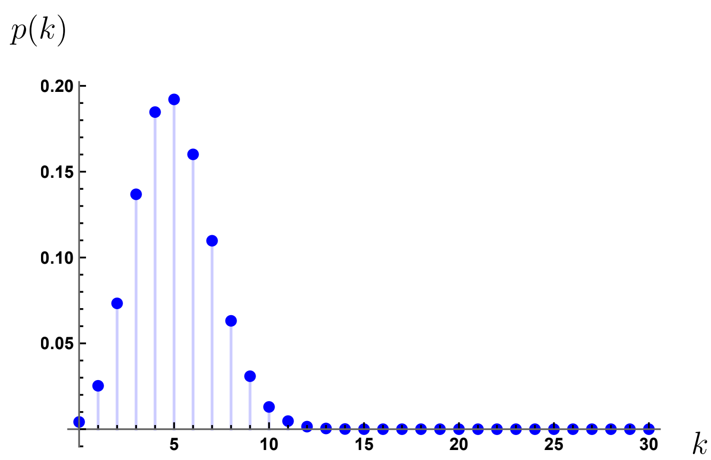

Figura 1: Distribuição binomial com $n = 30$ e $p = 1/6$.

se definíssemos como sucesso o surgimento de qualquer outra das cinco faces do dado.

1.3.2 Distribuição de Poisson

Se X é uma variável aleatória que segue a distribuição de Poisson, a probabilidade de X assumir qualquer valor inteiro é dada por

$$P(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}, \quad \text{onde} \quad k = 0, 1, 2, \ldots,$$

onde λ é um parâmetro que indica o número médio de ocorrências de um evento durante um intervalo de tempo. É possível provar que esta distribuição de probabilidade é o caso limite de uma distribuição binomial quando n é muito grande e p é muito pequeno.

- **Exemplo 9:** Foi observado que, em média, um comprador entra em uma determinada loja a cada 15 segundos. Qual é a probabilidade de que, em um intervalo de 1 minuto, nenhum comprador entre na loja?

◇ **Solução:** Se o número médio de compradores que entram na loja em 15 segundos é igual a 1, então o número médio de compradores que entrarão na loja no intervalo de 1 minuto é igual a 4. Logo, $\lambda = 4$. Neste exemplo X é a variável aleatória que descreve o número de compradores que entram na loja no intervalo de tempo dado. Portanto, a probabilidade de que nenhum comprador entre na loja em 1 minuto é dada por

$$p(0) = P(X = 0) = \frac{4^0 e^{-4}}{0!} \approx 0.018.$$

A Figura 2 mostra o gráfico da distribuição de Poisson deste exemplo para $\lambda = 4$ e k variando de 0 a 14. Podemos perceber que á medida que k aumenta a probabilidade de X valer k diminui consideravelmente. Note, por exemplo, que a chance de 14 compradores entrarem na loja no intervalo de 1 minuto é praticamente igual a zero.

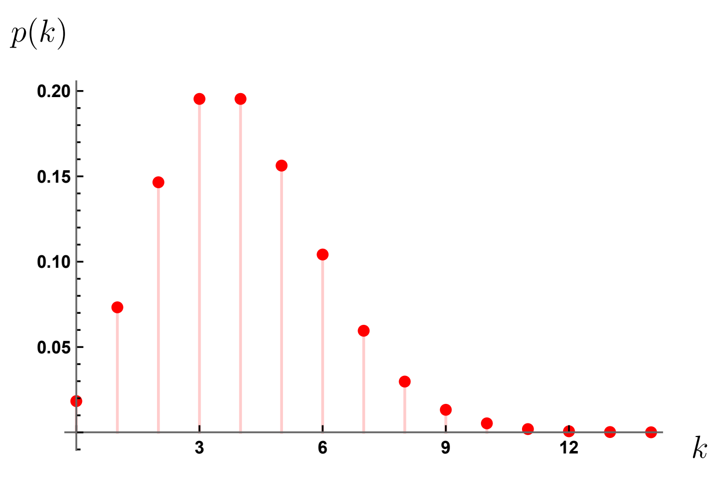

Figura 2: Distribuição de Poisson com $\lambda = 4$.

1.4 Variável aleatória contínua

Uma variável aleatória é chamada contínua, se existir uma função f, denominada função de densidade de probabilidade de X (fdp), que satisfaz as seguintes condições:

a) $f(x) \geq 0$ para todo x,

b) $\int_{-\infty}^{+\infty} f(x)dx = 1$,

c) para quaisquer a, b, com $-\infty < a < b < +\infty$, teremos $P(a \leq X \leq b) = \int_a^b f(x)dx$.

Observações:

- ⋄ Como $f(x) \geq 0$, $P(a \leq X \leq b)$ representa a área sob a curva do gráfico de f, entre $x = a$ e $x = b$.

- ⋄ A distribuição de probabilidade de uma variável aleatória discreta é uma fdp. Neste caso, as condições a) e b) são dadas, respectivamente, por $p(x_i) \geq 0$ para todo i e $\sum_i p(x_i) = 1$.

1.4.1 Função de distribuição acumulada

A função de distribuição acumulada $F(x)$ de uma variável aleatória contínua, com fdp f é dada por

$$F(x) = \int_{-\infty}^{x} f(x)dx.$$

No caso de uma variável aleatória discreta, esta definição fica assim

$$F(x) = \sum_i p(x_i),$$

onde o somatório é estendido a todos os índices i que satisfazem à condição $x_j \leq x$.

- **Exemplo 10:** Seja X uma variávela aleatória discreta que assume os valores 0, 1 e 2, com probabilidades 1/3, 1/6 e 1/2, respectivamente. Construa o gráfico da função $F(x)$.

- ⋄ **Solução:** Aplicando a definição de $F(x)$ para variáveis discretas, obtemos

$$F(x) = \begin{cases} 0 & \text{se } x < 0, \\ 1/3 & \text{se } 0 \leq x < 1, \\ 1/2 & \text{se } 1 \leq x < 2, \\ 1 & \text{se } x \geq 2. \end{cases}$$

O gráfico de $F(x)$ está indicado na Figura 3. Note que $F(x)$ é contínua à direita.

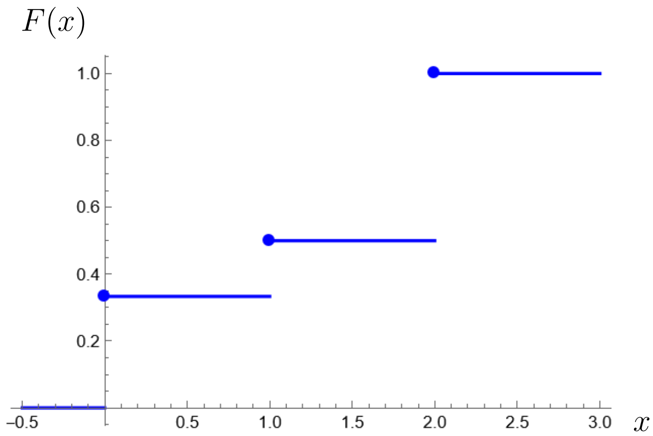

Figura 3: Função de distribuição acumulada de X discreta.

- **Exemplo 11:** Suponha que X seja uma variável aleatória contínua com fdp dada por

$$f(x) = \begin{cases} 2x, & \text{se } 0 < x < 1, \\ 0, & \text{para quaisquer outros valores.} \end{cases}$$

Construa o gráfico da função $F(x)$.

◇ **Solução:** Novamente aplicando a definição, temos

$$F(x) = \begin{cases} 0 & \text{se } x \leq 0, \\ \int_0^x 2s\ ds = x^2 & \text{se } 0 < x \leq 1, \\ 1 & \text{se } x \geq 1. \end{cases}$$

O gráfico de $F(x)$ está indicado na Figura 4.

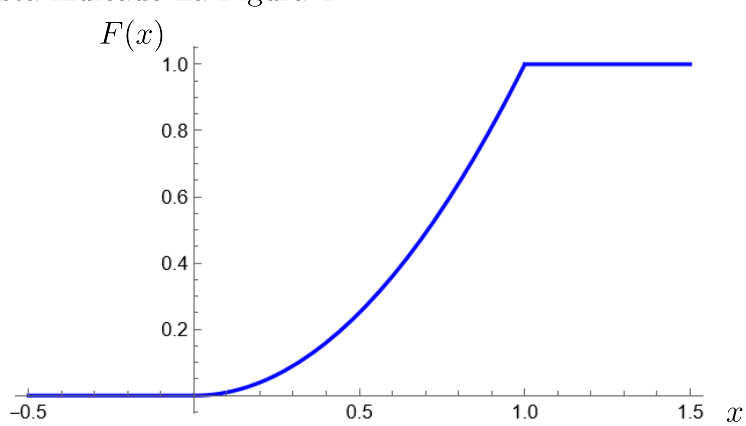

Figura 4: Função de distribuição acumulada de X contínua.

1.4.2 Distribuição normal

Uma variável aleatória contínua terá uma distribuição normal (ou Gaussiana) se sua função densidade de probabilidade (fdp) for dada por

$$f(x) = \frac{1}{\sqrt{2\pi\sigma}}\, e^{-\frac{(x-\mu)^2}{2\sigma^2}},$$

com $-\infty < x < \infty$, $-\infty < \mu < \infty$ e $\sigma > 0$. Os parâmetros μ e σ são chamados de média e desvio padrão, respectivamente. Escrevendo de forma mais resumida, diremos que X tem uma distribuição $N(\mu, \sigma^2)$. No caso em que $\mu = 0$ e $\sigma = 1$, diremos que a variável X possui uma distribuição normal padrão (ou reduzida). A Figura 5 mostra o gráfico da função densidade de probabilidade associada à distribuição normal padrão no intervalo $[-6, 6]$.

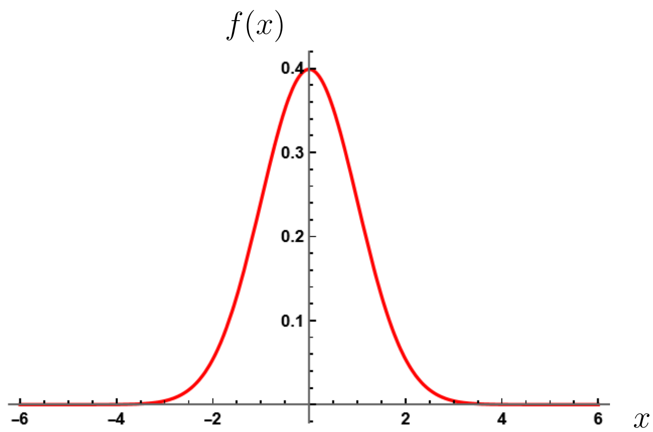

Figura 5: fdp da distribuição normal padrão: $\mu = 0$ e $\sigma = 1$.

Para realizar cálculos de probabilidade utilizando a distribuição normal padrão definimos a função

$$\Phi(s) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{s} e^{-x^2/2} dx.$$

Assim, é fácil perceber que o cálculo $P(a \leq X \leq b)$ será dado por $\Phi(b) - \Phi(a)$. É importante observar que não é possível aplicar o Teorema Fundamental do Cálculo para resolver integrais envolvendo a função Φ. Contudo, métodos numéricos de integração podem ser empregados para calcular essas integrais. Por essa razão, $P(X \leq s)$ é tabelada.

1.5 Esperança e variância de variáveis aleatórias

O valor esperado de uma variável aleatória X, denotado por $E(X)$ é definido da seguinte forma:

1. Se X for discreta, $E(X) = \sum_{i=1}^{\infty} x_i p(x_i)$, onde x_i é um valor possível para X e $p(x_i) = P(X = x_i)$. Claramente a soma será finita se a variável X assumir um número finito de valores.

2. Se X for contínua, $E(X) = \int_{-\infty}^{\infty} x f(x)\, dx$, onde $f(x)$ é a função densidade de probabilidade associada à X.

Sendo X uma variável aleatória (contínua ou discreta), a variância de X, denotada por $V(x)$ é definida como

$$V(X) = E[(X - E(X))^2] = E(X^2) - [E(X)]^2,$$

E a raiz quadrada positiva de $V(X)$ é denominada desvio-padrão de X, sendo denotado por σ_X.

- **Exemplo 12:** O serviço de meteorologia classifica o tipo de céu que é visível, em termos de "graus de nebulosidade". Uma escala de 11 categorias é empregada: 0, 1, 2, ..., 10, onde 0 representa um céu perfeitamente claro, 10 representa um céu completamente encoberto, enquanto os outros valores representam diferentes condições intermediárias. Suponha que tal classificação seja feita em uma determinada estação meteorológica, em um determinado dia e hora. Seja X a variável aleatória que pode assumir um dos 11 valores acima com a seguinte distribuição de probabilidade:

$$p_0 = p_{10} = 0.05, \quad p_1 = p_2 = p_8 = p_9 = 0.15, \quad p_3 = p_4 = p_5 = p_6 = p_7 = 0.06.$$

Calcule o valor esperado, a variância e o desvio-padrão de X.

◇ **Solução:** Como X é uma variável aleatória discreta, o valor esperado é dado por

$$\begin{aligned} E(X) &= 1(0.15) + 2(0.15) + 3(0.06) + 4(0.06) + 5(0.06) \\ &+ 6(0.06) + 7(0.06) + 8(0.15) + 9(0.15) + 10(0.05) \\ &= 5.0. \end{aligned}$$

Para calcular a variância, necessitamos calcular $E(X^2)$:

$$\begin{aligned} E(X) &= 1(0.15) + 4(0.15) + 9(0.06) + 16(0.06) + 25(0.06) \\ &+ 36(0.06) + 49(0.06) + 64(0.15) + 81(0.15) + 100(0.05) \\ &= 35.6. \end{aligned}$$

Logo, $V(X) = E(X^2) - [E(X)]^2 = 35.6 - 25 = 10.6$ e $\sigma_X = 3.25$.

- **Exemplo 13:** Seja X uma variável aleatória contínua com fdp dada por

$$f(x) = \begin{cases} 1 + x, & \text{se} \quad -1 \leq x \leq 0, \\ 1 - x & \text{se} \quad 0 \leq x \leq 1. \end{cases}$$

Calcule $V(X)$.

$\diamond$ **Solução:** Para calcular $V(X)$ necessitamos do cálculo de $E(X)$ e $E(X^2)$. Ocorre que $f(x)$ é uma função par. Logo, $xf(x)$ será ímpar e

$$E(X) = \int_{-1}^{1} xf(x)dx = 0.$$

Além disso,

$$E(X^2) = \int_{-1}^{0} x^2 f(x)dx + \int_{0}^{1} x^2 f(x)dx = \int_{-1}^{0} x^2(1+x)dx + \int_{0}^{1} x^2(1-x)dx = \frac{1}{6}.$$

Portanto, $V(X) = E(X^2) - [E(X)]^2 = 1/6 - 0 = 1/6$.

1.6 Covariância e coeficiente de correlação

Sejam X e Y um par de variáveis aleatórias. O coeficiente de correlação entre X e Y é definido por

$$\rho_{xy} = \frac{E[X - E(X)][Y - E(Y)]}{\sqrt{V(X)V(Y)}}.$$

O numerador da fórmula acima é chamado de covariância entre X e Y e é frequentemente denotado por σ_{xy} ou Cov(X,Y).

Bibliografia

1. Probabilidade e estatística para engenharia e ciências - J. L. Devore - Cengage Learning, 2014.

2. Probabilidade - Aplicações à Estatística - P. L. Meyer - 2ª Edição - LTC, 2011.

3. Probability For the Enthusiastic Beginner - D. Morin, 2016.