Capítulo 7

Valor Esperado

Temos ainda um elemento central e fundamental para a teoria da probabilidade que precisamos abordar. Este é o conceito de *esperança* ou *valor esperado*.

Intuitivamente, o valor esperado de uma variável aleatória X é a generalização do conceito de média ponderada dos valores que X assume ou equivalentemente, de modo mais geométrico, o centro de massa dos valores que X assume.

Como motivação considere X é uma variável simples dada por

$$X = \sum_{k=1}^{n} a_k \cdot \mathbb{1}_{A_k},$$

com $A_1, A_2, \ldots, A_n$ disjuntos. Então a média ponderada $m(X)$ dos valores que X assume é

$$m(X) = \sum_{k=1}^{n} a_k \, \mathbf{P}(A_k) = \int X \, \mathrm{d}\mathbf{P}.$$

o que motiva a definição:

7.1 Definição (Valor Esperado ou Esperança) *Se X é uma variável aleatória definida em* $(\Omega, \mathcal{F}, \mathbf{P})$, *seu* ***valor esperado***, ***esperança*** *ou* ***média*** *é definido como*

$$\mathbf{E}[X] := \int_{\Omega} X d\mathbf{P}$$

Usaremos aqui a mesma nomenclatura usada em integrais de Lebesgue, e diremos que X é **integrável** se $\mathbf{E}|X| < \infty$ (o que é equivalente a dizer que $\mathbf{E}[X^+] < \infty$ e $\mathbf{E}[X^-] < \infty$ e de modo mais geral falaremos em esperança de X se X for quasi-integrável, isto é se $\mathbf{E}[X^+]$ ou $\mathbf{E}[X^-]$ for finita.

Agora podemos retomar o exemplo de variáveis aleatórias simples de posse da definição da esperança

7.2 Exemplo (Variáveis Aleatórias Simples) *Seja X uma variável aleatória X dada por*

$$X = \sum_{k=1}^{n} a_k \cdot \mathbb{1}_{A_k},$$

com $A_1, A_2, \ldots, A_n$ disjuntos. Então

$$\mathbf{E}[X] = \int X \, d\mathbf{P} = \sum_{k=1}^{n} a_k \, \mathbf{P}(A_k).$$

Se além disso, os valores $a_1, \ldots, a_k \in \mathbb{R}$ são distintos, então $A_k = X^{-1}(\{a_k\}) = \{X = a_k\}$, e

$$\mathbf{E}[X] = \sum_{k=1}^{n} a_k \, \mathbf{P}(X = a_k).$$

Podemos restringir o exemplo anterior é obter um dos exemplos mais simples e também um dos mais úteis.

7.3 Exemplo (A Esperança da Função Indicadora) *Seja $\mathbb{1}_A$ a indicadora de um eventos $A \in \mathcal{F}$. $\mathbb{1}_A$ é uma variável aleatória simples, e pelo exemplo anterior, o seu valor esperado é dado por*

$$\mathbf{E}[\mathbb{1}_A] = \mathbf{P}(A).$$

Ao ser definida como a integral de Lebesgue, a esperança herda uma série de propriedades importantes, que listaremos a seguir.

7.4 Teorema *Suponha que X, Y sejam variáveis aleatórias integráveis. Então*

1. *(Linearidade)* $\mathbf{E}[aX + bY] = a\mathbf{E}[X] + b\mathbf{E}[Y]$ *para quaisquer números reais* a, b.

2. *Se* $X \equiv b \in \mathbb{R}$, *constante, então* $\mathbf{E}[X] = b$ *(escrevemos também que* $\mathbf{E}[b] = b$*).*

3. *Se* $X \geq 0$ *então* $\mathbf{E}[X] \geq 0$.

4. *(Monotonicidade) Se* $X \geq Y$ *então* $\mathbf{E}[X] \geq \mathbf{E}[Y]$.

5. *(Desigualdade triangular)* $|\mathbf{E}[X]| \leq \mathbf{E}[|X|]$

6. *Se* $X_n \uparrow X$ *então* $\mathbf{E}[X] = \lim_{n\to\infty} \mathbf{E}[X_n]$

Demonstração. *Essas propriedades são herança direta das propriedades das integrais de Lebesgue, e não serão demonstradas.*

Mostraremos apenas a propriedade 2, que é típica de medidas de probabilidade.

Tome então uma variável aleatória X *constante igual a* b. *Ou seja,* $X(\omega) = b$ *para todo* ω. *Essa variável é uma variável simples, dada por*

$$X = b \cdot \mathbb{1}_\Omega,$$

e portanto

$$\mathbf{E}[b] = \mathbf{E}[X] = b \cdot \mathbf{P}(\Omega) = b.$$

7.5 Exemplo *Se uma esfera é colorida de modo que 90% de sua área seja vermelha e 10% seja azul, então existe um modo de inscrever um cubo de modo que os 8 vértices estejam tocando pontos vermelhos da esfera.*

Sejam $X_1, \ldots, X_8$ *as variáveis aleatórias indicadoras do fato de cada vértice ser vermelho ou não. Então, claramente* $X_1 + \cdots + X_8$ *é o número de vértices vermelhos. Por linearidade da esperança,*

$$\mathbf{E}[X_1 + \cdots + X_8] = \mathbf{E}[X_1] + \cdots + \mathbf{E}[X_8] = 8 \cdot 0.9 = 7.2.$$

Para que o número médio de vértices seja 7.2, deve haver algum cubo com mais de 7 vértices vermelhos, e portanto deve ter 8 vértices vermelhos.

Esse é um exemplo da utilização de técnicas probabilísticas para provar fatos determinísticos.

Observe que não exigimos que as variáveis sejam independentes. Verifique que de fato isso não é necessário. ◁

Temos também a seguinte caracterização do operador esperança.

7.6 Teorema (Caracterização da Esperança) *A esperança é o único operador* $\mathbf{E} : \mathcal{L}^1(\Omega) \to \mathbb{R}$ *que satisfaz:*

- *Linearidade. Para* $a, b \in \mathbb{R}$, $\mathbf{E}[aX + bY] = a\mathbf{E}[X] + b\mathbf{E}[Y]$.
- *Continuidade. Se* $X_n \to X$ *então* $\mathbf{E}[X_n] \to \mathbf{E}[X]$
- *Relação com a probabilidade. Para cada evento* A, $\mathbf{E}[\mathbb{1}_A] = \mathbf{P}(A)$.

Demonstração. *Veja que*

$$\mathbf{E}_1[X] := \int_\Omega X d\mathbf{P}$$

tem as propriedades pedidas. Assim temos trivialmente a existência.

Vamos provar a unicidade mostrando a coincidência de um operador $\mathbf{E}$ *com as propriedades listadas e* $\mathbf{E}_1$.

O item 3. e a linearidade nos permite mostrar a coincidência para funções simples.

Se $\mathrm{X} = \sum_i a_i \mathbb{1}_{A_i}$ *então* $\mathbf{E}_1[\mathrm{X}] = \sum_i a_i \mathbf{P}(A_i) = \mathbf{E}[\mathrm{X}]$.

Provaremos agora para funções positivas. Dado $\mathrm{X} > 0$, *temos que existem funções simples* X_n *com* $\mathrm{X}_n \uparrow \mathrm{X}$. *Por continuidade*

$$\mathbf{E}_1[\mathrm{X}] = \lim \mathbf{E}_1[\mathrm{X}_n] = \lim \mathbf{E}[\mathrm{X}_n] = E[\mathrm{X}]$$

Para funções quaisquer. Consideramos a decomposição $\mathrm{X} = \mathrm{X}^+ + \mathrm{X}^-$ *e usamos a linearidade*

ntes de darmos continuidade, vamos estabelecer uma nomenclatura que será central deste ponto em diante.

7.7 Definição *Dada um espaço de probabilidade* $(\Omega, \mathcal{F}, \mathbf{P})$, *diremos que uma afirmativa* $\mathbb{A}$ *ocorre* ***quase-certamente*** *se*

$$\mathbf{P}(\{\omega \in \Omega : \omega \text{ satisfaz a afirmativa } \mathbb{A}\}) = 1.$$

De notaremos isso por "$\mathbb{A}$ *ocorre q.c.".*

Em particular, dadas duas variáveis aleatórias X, Y *definidas em* Ω, *temos que* $X = Y$ *q.c., se* $\mathbf{P}(X = Y) = 1$. *Ou ainda que* $X \le Y$ *q.c se* $\mathbf{P}(X \le Y) = 1$.

Integrais de Lebesgue ignoram eventos de medida nula, e isso é mostrado no seguinte resultado.

7.8 Proposição *Dadas variáveis aleatórias* X *e* Y, *vale que*

1. *Se* $X = 0$ *q.c então* $\mathbf{E}[X] = 0$;
2. *Se* $X = Y$ *q.c e* X *é integrável ou* $\mathbf{E}[X] = \pm\infty$, *então* $\mathbf{E}[Y] = \mathbf{E}[X]$;
3. *Se* $X \le Y$ *q.c. e* Y *é integrável, então* $\mathbf{E}[X] \le \mathbf{E}[Y]$ ($\mathbf{E}[X]$ *pode ser* $-\infty$) .
4. *Se* $X \ge 0$ *e* $\mathbf{E}[X] = 0$, *então* $X = 0$ *q.c..*

Demonstração.

1. *Suponha que* X *é simples. Temos então que*

$$X = \sum_{k=1}^{n} a_k \mathbb{1} A_k,$$

e como $\mathbf{P}(X = 0) = 1$, *então* $\mathbf{P}(A_k) = 0$ *sempre que* $a_k \neq 0$. *Com isso, temos que* $\mathbf{E}[X] = 0$.

Para X *não negativa, tome* $(X_n)_{n\geq 1}$ *uma sequência não-decrescente de variáveis aleatórias simples e não-negativas tal que* $X_n \uparrow X$. *Como* $0 \leq X_n \leq X$ *e* $\mathbf{P}(X = 0) = 1$, *temos que* $X_n = 0$ *q.c e portanto* $\mathbf{E}[X_n] = 0$. *Assim, pelo Teorema da Convergência Monótona,* $\mathbf{E}[X] = \lim_{n\to\infty} \mathbf{E}[X_n] = 0$.

Para X *qualquer, basta notar que se* $X = 0$ *q.c, então* $X^+ = 0$ *q.c. e* $X^- = 0$ *q.c..*

$\boxed{2}$ *Exercício (aplique o item anterior em* $X - Y$*)*

$\boxed{3}$ *Tome* X, Y *variáveis aleatórias tais que* $X \leq Y$ *q.c. e* Y *integrável. Defina*

$$X^* = X \cdot \mathbb{1}_{X\leq Y} + Y \cdot \mathbb{1}_{X>Y}.$$

Como $\mathbf{P}(X > Y) = 0$, *então* $X^* = X$ *q.c., e além disso* $X^* \leq Y$ *pontualmente.*

Se X^* *é integrável, então a monotonicidade da esperaça nos dá que* $\mathbf{E}[X^*] \leq \mathbf{E}[Y]$, *e pelo item anterior* $\mathbf{E}[X] = \mathbf{E}[X^*] \leq \mathbf{E}[Y]$.

Caso contrário, temos $(X^*)^+ \leq Y^+$, *de modo que* $\mathbf{E}(X^*)^+ \leq \mathbf{E}[Y^+] < \infty$ *e* $\mathbf{E}[(X^*)^-] = \infty$, *de modo que* $\mathbf{E}[X^*] = -\infty$.

Agora basta usar o item anterior para X^+ *e* X^- *para concluir o resultado.*

$\boxed{4}$ *Dado* $\epsilon > 0$, *defina* $Y_\epsilon = \epsilon \mathbb{1}_{\{X>\epsilon\}}$ *e observe que* $Y_\epsilon \leq X$ *para todo* $\epsilon > 0$. *Segue que*

$$\epsilon \mathbf{P}(X > \epsilon) = \mathbf{E}[Y_\epsilon] \leq \mathbf{E}[X] = 0,$$

e portanto $\mathbf{P}(X > \epsilon) = 0$ *para todo* $\epsilon > 0$.

Fazendo $\epsilon \to 0$, *vemos que* $\mathbf{P}(X > 0) = 0$, *e como* $\mathbf{P}(X \geq 0) = 1$, *o resultado segue.* □

7.1 Calculando a Esperança

Nesta seção veremos como a esperança de uma variável aleatória pode ser determinada a partir de sua distribuição. Para isso, vamos seguir a estratégia de três passos, mas de um ponto de vista bem específico.

Tome agora uma variável aleatória X qualquer, e $\phi : \mathbb{R} \to \mathbb{R}$ uma função mensurável e simples e não-negativa, dada por

$$\phi(x) = \sum_{k=1}^{n} a_k \mathbb{1}_{A_k}(x),$$

$A_1, \ldots, A_n \in \mathcal{B}(\mathbb{R})$ disjuntos.

Temos assim que

$$\phi(X) = \sum_{k=1}^{n} a_k \mathbb{1}_{A_k}(X) = \sum_{k=1}^{n} a_k \mathbb{1}_{X \in A_k},$$

que é uma variável aleatória simples.

Segue que

$$\mathbf{E}[\phi(X)] = \sum_{k=1}^{n} a_k \mathbf{P}(X \in A_k) = \sum_{k=1}^{n} a_k \mathbf{P}_X(A_k) = \int_{\mathbb{R}} \phi(x)\, \mathrm{d}\mathbf{P}_X(x).$$

Dada agora uma função $\varphi : \mathbb{R} \to \mathbb{R}$ mensurável e não negativa, tome uma sequencia não-decrescente de funções simples e mensuráveis $\phi_n : \mathbb{R} \to \mathbb{R}$, $n \geq 1$, não-negativas e tais que $\phi_n(x) \uparrow \varphi(x)$, para todo $x \in \mathbb{R}$.

Pelos mesmos argumentos anteriores, $\phi_n(X), n \geq 1$ é uma sequência de variáveis aleatórias simples e não-negativas, tais que $\phi_n(X) \uparrow \varphi(X)$.

Segue do Teorema da Convergência Monótona, que

$$\mathbf{E}[\varphi(X)] = \lim_{n\to\infty} \mathbf{E}[\phi_n(X)] = \lim_{n\to\infty} \int_{\mathbb{R}} \phi_n(x)\, \mathrm{d}\mathbf{P}_X(x) = \int_{\mathbb{R}} \varphi(x)\, \mathrm{d}\mathbf{P}_X(x).$$

Dada agora uma função $\varphi : \mathbb{R} \to \mathbb{R}$ mensurável, podemos escrever $\varphi = \varphi^+ - \varphi^-$, e aplicando o que descobrimos até agora, vemos que

$$\mathbf{E}[\phi(X)] = \mathbf{E}[\phi^+(X)] - \mathbf{E}[\phi^-(X)] = \int_{\mathbb{R}} \varphi^+(x)\, \mathrm{d}\mathbf{P}_X(x) - \int_{\mathbb{R}} \varphi^-(x)\, \mathrm{d}\mathbf{P}_X(x) = \int_{\mathbb{R}} \varphi(x)\, \mathrm{d}\mathbf{P}_X(x).$$

Com isso mostramos que

7.9 Teorema (Teorema da Mudança de Variáveis) *Dada uma variável aleatória X definida em um espaço de probabilidade $(\Omega, \mathcal{F}, \mathbf{P})$ com distribuição $\mathbf{P}_X$, e uma função mensurável $\varphi : \mathbb{R} \to \mathbb{R}$, então*

$$\mathbf{E}[\varphi(X)] = \int_{\mathbb{R}} \varphi(x)\, \mathrm{d}\mathbf{P}_X(x).$$

Ou seja, se qualquer um dos lados da equação acima estiver bem definido, o outro também está e a igualdade vale.

Em particular

$$\mathbf{E}[X] = \int_{\mathbb{R}} x\, \mathrm{d}\mathbf{P}_X(x)$$

Uma consequência direta do resultado anterior é a seguinte.

7.10 Corolário (Distribuição Determina a Esperança) *Dado* X *integrável então se* $\mathrm{X} \stackrel{d.}{=} \mathrm{Y}$ *então* Y *é integrável e* $\mathbf{E}[\mathrm{X}] = \mathbf{E}[\mathrm{Y}]$.

Agora que sabemos que a esperança de uma variável aleatória é determinada unicamente pela sua distribuição, falta apenas entender como surgem aquelas expressões para variáveis discretas e absolutamente contínuas.

Variáveis Aleatórias Discretas

Vamos lembrar que uma variável aleatória discreta é aquela que assumi uma quantidade enumerável de valores. Ou seja, existe um conjunto $S = \{x_1, x_2, \ldots\}$ tal que $\mathbf{P}(X \in S) = 1$.

Com isso, temos que a distribuição $\mathbf{P}_X$ é totalmente definida pelos valores de $\mathbf{P}(\{x_k\}) = \mathbf{P}(X = x_k)$.

De fato, para qualquer $A \in \mathcal{B}(\mathbb{R})$

$$\mathbf{P}_X(A) = \sum_{k : x_k \in A} \mathbf{P}_X(\{x_k\}).$$

Perceba também que

$$X = \sum_{k=1}^{\infty} x_k \, \mathbb{1}_{A_k},$$

onde $A_k = \{X = x_k\} \in \mathcal{F}$.

Para simplificar nossa vida, suponha que $X \geq 0$. Ou seja $x_k \geq 0$ para todo $k \geq 1$. Nestas condições, as variáveis

$$X_n = \sum_{k=1}^{n} x_k \, \mathbb{1}_{A_k},$$

são simples e portanto

$$\mathbf{E}[X_n] = \sum_{k=1}^{n} x_k \, \mathbf{P}(A_k) = \sum_{k=1}^{n} x_k \, \mathbf{P}(X = x_k).$$

Além disso, temos que $X_n \uparrow X$ e o Teorema da Convergência Monótona nos diz que

$$\mathbf{E}[X] = \lim_{n \to \infty} \mathbf{E}[X_n] = \sum_{k=1}^{\infty} x_k \, \mathbf{P}(X = x_k).$$

Deixamos como exercício o caso onde X não assume apenas valores positivos.

Fica assim (quase) demonstrado então o seguinte resultado.

7.11 Proposição *Para uma variável aleatória discreta X positiva ou integrável, assumindo valores em $\{x_1, x_2, \ldots\}$, temos que*

$$\mathbf{E}[X] = \sum_{k=1}^{\infty} x_k \, \mathbf{P}(x = x_k), \tag{7.1}$$

de modo que a série à direita é absolutamente convergente sempre que X for integrável.

Variáveis Absolutamente Contínuas

Vamos recordar que uma variável aleatória é absolutamente contínua se existe uma função mensurável $f : \mathbb{R} \to \mathbb{R}^+$, conhecida como **função densidade de probabilidade de X** (ou simplesmente densidade de X) tal que

$$\mathbf{P}_X(A) = \mathbf{P}(X \in A) = \int_A f(x)\, \mathrm{d}\mu,$$

onde a integral à direita é a integral de Lebesgue (feita em relação à medida de Lebesgue da reta).

Para entender quem é $\mathbf{E}[\varphi(X)]$ devemos primeiro entender como calcular

$$\int_{\mathbb{R}} \varphi(x)\, \mathrm{d}\mathbf{P}_X(x).$$

Para isso, tome primeiro uma função simples

$$\phi(x) = \sum_{k=1}^{n} a_k \mathbb{1}_{A_k},$$

e note que

$$\begin{aligned}
\int_{\mathbb{R}} \phi(x)\, \mathrm{d}\mathbf{P}_X(x) &= \sum_{k=1}^{n} a_k \mathbf{P}_X(A_k) \\
&= \sum_{k=1}^{n} a_k \int_{A_k} f(x)\, \mathrm{d}\mu \\
&= \sum_{k=1}^{n} a_k \int_{\mathbb{R}} \mathbb{1}_{A_k}(x) f(x)\, \mathrm{d}\mu \\
&= \int_{\mathbb{R}} \sum_{k=1}^{n} a_k \mathbb{1}_{A_k}(x) f(x)\, \mathrm{d}\mu \\
&= \int_{\mathbb{R}} \phi(x) f(x)\, \mathrm{d}x.
\end{aligned}$$

Tomando uma sequência de funções simples $\phi_n, n \geq 1$ não negativas, com $\phi_n \uparrow \varphi \geq 0$, concluímos pelo Teorema da Convergência Monótona que

$$\int_{\mathbb{R}} \varphi(x)\, \mathrm{d}\mathbf{P}_X(x) = \lim_n \int_{\mathbb{R}} \phi_n(x)\, \mathrm{d}\mathbf{P}_X(x) = \lim_n \int_{\mathbb{R}} \phi_n(x) f(x)\, \mathrm{d}x = \int_{\mathbb{R}} \varphi(x) f(x)\, \mathrm{d}x.$$

E separando uma função mensurável φ em parte positiva e negativa, concluímos o seguinte resultado.

7.12 Proposição *Para variável aleatória absolutamente contínua X com função densidade de probabilidade $f : \mathbb{R} \to \mathbb{R}^+$, e uma função mensurável $\phi : \mathbb{R} \to \mathbb{R}$ vale que*

$$\mathbf{E}[\varphi(X)] = \int_{\mathbb{R}} \varphi(x) f(x)\, \mathrm{d}x, \tag{7.2}$$

onde um lado da equação está bem definido se, e somente se, o outro também está.

7.2 Funções de Vetores Aleatórios

Nesta seção falaremos brevemente sobre o cálculo da esperança para funções de vetores aleatórios.

Mas especificamente, dado um vetor aleatório $X = (x_1, \ldots, X_n)$ definido em um espaço de probabilidade $(\Omega, \mathcal{F}, \mathbf{P})$ e uma função mensurável $\varphi : \mathbb{R}^n \to \mathbb{R}$, queremos entender como calcular $\mathbf{E}[\varphi(X_1, \ldots, X_n)]$.

Vamos olhar primeiro para a distribuição $\mathbf{P}_X$ do vetor aleatório X. Lembre-se que, de modo análogo a variáveis aleatórias, $\mathbf{P}_X$ é uma medida de probabilidade em $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ dada por

$$\mathbf{P}_X(A) = \mathbf{P}((X_1, \ldots, X_n) \in A),$$

para $A \in \mathcal{B}(\mathbb{R}^n)$.

De posse da distribuição $\mathbf{P}_X$ podemos recuperar as **distribuições marginais** das variáveis X_k, fazendo

$$\begin{aligned}
\mathbf{P}_{X_k}(B) &= \mathbf{P}(X_k \in B) \\
&= \mathbf{P}(X_k \in B; X_j \in \mathbb{R}; j \neq k) \\
&= \mathbf{P}((X_1, \ldots, X_n) \in \mathbb{R}^{k-1} \times B \times \mathbb{R}^{n-k}) \\
&= \mathbf{P}_X(\mathbb{R}^{k-1} \times B \times \mathbb{R}^{n-k}).
\end{aligned}$$

A relação entre $\mathbf{P}_X$ e as marginais $\mathbf{P}_{X_k}$ pode ser complicada, e em geral precisamos de informações adicionais sobre o vetor para descreve-la com mais precisão. Uma das situações onde esta relação está bem definida é no caso onde as variáveis $X_1, \ldots, X_n$ são independentes. Deixaremos os detalhes para a lista de exercícios.

Voltando ao nosso problema, não é difícil mostrar, usando os mesmos argumentos usados para mostrar o Teorema 7.9, que vale o seguinte resultado.

7.13 Teorema *Dado um vetor aleatório $X = (X_1, \ldots, X_n)$ definido em um espaço de probabilidade $(\Omega, \mathcal{F}, \mathbf{P})$ com distribuição $\mathbf{P}_X$, e uma função mensurável $\varphi : \mathbb{R}^n \to \mathbb{R}$, então*

$$\mathbf{E}[\varphi(X_1, \ldots, X_n)] = \int_{\mathbb{R}^n} \varphi(x_1, \ldots, x_n)\, \mathrm{d}\mathbf{P}_X(x_1, \ldots, x_n).$$

Ou seja, se qualquer um dos lados da equação acima estiver bem definido, o outro também está e a igualdade vale.

No caso em que as variáveis $X_1, \dots, X_n$ são discretas, assumindo valores em $S_1, \dots, S_n$ respectivamente, então o vetor é também simples, assumindo valores em $S_1 \times \cdots \times S_n$, e a variável $\phi(X_1, \dots, X_n)$ é discreta, assumindo valores $\varphi(x_1, \dots, x_n) \in \phi(S_1 \times \cdots \times S_n)$ com probabilidade $\mathbf{P}(X_1 = x_1, \dots, X_n = x_n)$.

Vale então o seguinte resultado.

7.14 Proposição *Dadas variáveis aleatórias $X_1, \dots, X_n$ discretas assumindo valores em $S_1, \dots, S_n$, respectivamente, e uma função mensurável $\varphi : \mathbb{R}^n \to \mathbb{R}$, então*

$$\mathbf{E}[\varphi(X_1, \dots, X_n)] = \sum_{x_1 \in S_1} \cdots \sum_{x_n \in S_n} \varphi(x_1, \dots, x_n) \, \mathbf{P}(X_1 = x_1, \dots, X_n = x_n),$$

quando os dois lados estiverem bem definidos.

O caso onde $X_1, \dots, X_n$ são absolutamente contínuas é um pouco mais complicado, pois não garante que o vetor X seja absolutamente contínuo. Ou seja, a existência de densidades $f_1, \dots, f_n : \mathbb{R} \to \mathbb{R}$ não garante a existência de uma função densidade $f : \mathbb{R}^n \to \mathbb{R}^+$ tal que

$$\begin{aligned} \mathbf{P}_X(A) &= \mathbf{P}((X_1, \dots, X_n) \in A) \\ &= \int_{\mathbb{R}^n} f(x_1, \dots, x_n) \, d\lambda_n(x_1, \dots, x_n) \\ &= \int_{\mathbb{R}} \cdots \int_{\mathbb{R}} f(x_1, \dots, x_n) \, dx_1 \cdots dx_n, \end{aligned}$$

para $A \in \mathcal{B}(\mathbb{R}^n)$, onde λ_n é a medida de Lebesgue em $\mathbb{R}^n$.

7.15 Observação *Para $X = (X_1, \dots, X_n)$ é um vetor aleatório absolutamente contínuo, também é usual dizer que as variáveis $X_1, \dots, X_n$ são conjuntamente absolutamente contínuas.* ◁

Não vamos entrar em maiores detalhes sobre esse tipo de vetor, mas podemos mostrar que

7.16 Proposição *Para um vetor aleatório absolutamente contínuo X com função densidade de probabilidade $f : \mathbb{R}^n \to \mathbb{R}^+$, e uma função mensurável $\phi : \mathbb{R}^n \to \mathbb{R}$ vale que*

$$\mathbf{E}[\varphi(X)] = \int_{\mathbb{R}} \cdots \int_{\mathbb{R}} \varphi(x_1, \dots, x_n) f(x_1, \dots, x_n) \, dx_1 \cdots dx_n, \tag{7.3}$$

onde um lado da equação está bem definido se, e somente se, o outro também está.

7.3 Variância, Covariância e Momentos de uma Variável Aleatória

A esperança mede, como já discutimos, o comportamento médio de uma variável aleatória. Uma média, ponderada pelas probabilidades, dos diversos valores assumidos pela variável. Porém, essa a média falha em captar a dispersão dos valores e suas probabilidades.

Para entender isso, vamos ver um exemplo simples.

7.17 Exemplo *Tome X e Y variáveis aleatórias discretas tais que*

$$\mathbf{P}(X=-1)=\mathbf{P}(X=1)=\frac{1}{2},$$

e

$$\mathbf{P}(Y=-100)=\mathbf{P}(Y=-50)=\mathbf{P}(Y=50)=\mathbf{P}(Y=100)=\frac{1}{4}.$$

Calculando, encontramos que $\mathbf{E}[X]=\mathbf{E}[Y]=0$, mas a variável aleatória Y assume valores muito mais dispersos em torno da média. ◁

Uma das alternativas que temos para medir essa dispersão é medir uma espécie de "distância média" (usado aqui de forma livre) entre a variável e sua média. É isso que motiva as próximas definições.

É isso que motiva as próximas definições.

7.18 Definição (Momentos, Variância e Covariância) *Se X é uma variável aleatória em $(\Omega, \mathcal{F}, \mathbf{P})$:*

- *Dado $n \in \mathbb{N}^*$ e X tal que X^n seja integrável então*

  $$m_k := \mathbf{E}[X^k] \quad M_k := \mathbf{E}[|X|^k] \quad \text{para } k = 1, \ldots, n$$

  *são denominados o k-ésimo **momento** e k-ésimo **momento absoluto** de X respectivamente.*

- *Se X é quadrado integrável, então*

  $$\mathbf{Var}[X] := \mathbf{E}[(X-\mathbf{E}[X])^2]$$

  *é a variância de X. O número $\sigma := \sqrt{\mathbf{Var}[X]}$ é denominado **desvio padrão** de X.*

- *Se X e Y são quadrado integráveis definimos a covariância de X e Y como*

  $$\mathbf{Cov}[X,Y] := \mathbf{E}\left[(X-\mathbf{E}[X])(Y-\mathbf{E}[Y])\right]$$

  As variáveis X e Y são ditas não correlacionadas se $\mathbf{Cov}[X,Y]=0$.

- *o **coeficiente de correlação** de X e Y não constantes por*

  $$\rho(X,Y) = \frac{\mathbf{Cov}(X,Y)}{\sigma_X \sigma_Y};$$

7.19 Proposição (Propriedades da Variância e Covariância) *Dadas* X *e* Y *variáveis aleatórias integráveis. Então:*

1. $\mathbf{Var}[X] \geq 0$

2. $\mathbf{Var}[X] = \mathbf{E}[X^2] - \mathbf{E}[X]^2$.

3. $\mathbf{Var}[X] = 0 \iff X = \mathbf{E}[X]$ *quase certamente.*

4. *A aplicação* $f : \mathbb{R} \to \mathbb{R}$, $x \longmapsto \mathbf{E}[(X-x)^2]$ *tem um mínimo em* $\mathbf{E}[X]$ *com* $f(\mathbf{E}[X]) = \mathbf{Var}[X]$.

5. $\mathbf{Var}[aX + b] = a^2\mathbf{Var}[X]$ *para* $a, b \in \mathbb{R}$.

6. $\mathbf{Cov}[X, Y] = \mathbf{E}[XY] - \mathbf{E}[X]\mathbf{E}[Y]$.

7. $\mathbf{Var}[X + Y] = \mathbf{Var}[X] + \mathbf{Var}[Y] + 2\mathbf{Cov}[X, Y]$

8. $\mathbf{Cov}[X, X] = \mathbf{Var}[X]$;

Demonstração. 1 *Segue direto da definição e das propriedades da esperança, uma vez que* $(X - \mathbf{E}[X])^2 \geq 0$;

2 *Segue da linearidade da esperança. De fato*

$$\begin{aligned}\mathbf{Var}[X] &= \mathbf{E}[(X - \mathbf{E}[X])^2] \\ &= \mathbf{E}[X^2 - 2X\mathbf{E}[X] + \mathbf{E}[X]^2] \\ &= \mathbf{E}[X^2] - 2\mathbf{E}[X]^2 + \mathbf{E}[X]^2 \\ &= \mathbf{E}[X^2] - \mathbf{E}[X]^2.\end{aligned}$$

3 *Se* $\mathbf{E}[(X - \mathbf{E}[X])^2] = \mathbf{Var}[X] = 0$, *então* $(X - \mathbf{E}[X])^2 = 0$ *q.c., e o resultado segue.*

4 *Imediato.*

5 $\mathbf{Var}[aX+b] = \mathbf{E}[(aX+b)^2] - \mathbf{E}[aX+b]^2 = a^2\mathbf{E}[X^2] + 2ab\mathbf{E}[X] + b^2 - (a^2\mathbf{E}[X]^2 + 2ab\mathbf{E}[X] + b^2) = a^2\mathbf{Var}[X]$.

6 *Segue da linearidade da esperança que*

$$\begin{aligned}\mathbf{Cov}[X, Y] &= \mathbf{E}[(X - \mathbf{E}[X])(Y - \mathbf{E}[Y])] \\ &= \mathbf{E}[XY - X\mathbf{E}[Y] - Y\mathbf{E}[X] + \mathbf{E}[X]\mathbf{E}[Y]] \\ &= \mathbf{E}[XY] - \mathbf{E}[X]\mathbf{E}[Y].\end{aligned}$$

$\boxed{7}$ *Segue da linearidade da esperança e dos itens anteriores que*

$$\begin{aligned}\mathbf{Var}(X+Y) &= \mathbf{E}[(X+Y)^2] - \mathbf{E}[X+Y]^2 \\ &= \mathbf{E}[X^2+Y^2+2XY] - (\mathbf{E}[X]^2+\mathbf{E}[Y^2]+2\mathbf{E}[X]\mathbf{E}[Y]) \\ &= \mathbf{Var}[X]+\mathbf{Var}[Y]+2\mathbf{Cov}[X,Y].\end{aligned}$$

$\boxed{8}$ *Segue direto da definição.* □

O item $\boxed{4}$ nos fornece uma descrição geométrica: imagine que queremos minimizar o erro quadrático médio de uma estimativa x da variável X. Nesse caso a estimativa que minimiza esse erro é a esperança e a variância é o menor valor do erro quadrático médio.

7.20 **Exemplo** *Seja* $X \in \mathcal{L}^2(\mathbf{P})$ *integrável e* $Y = aX + b$, *com* $a,b \in \mathbb{R}$ *não nulos. Temos então que* $\mathbf{Var}(X) < \infty$, $\mathbf{E}[Y] = a\mathbf{E}[X] + b$ *e que* $\mathbf{Var}[Y] = \mathbf{Var}[aX + b] = a^2\mathbf{Var}[X]$.

Da mesma forma temos que $\mathbf{Cov}[X,Y] = \mathbf{E}[(X - \mathbf{E}[X])(aX + b - (a\mathbf{E}[X] + b))] = a\mathbf{Var}[X]$.

Segue que

$$\rho(X,Y) = \frac{a\mathbf{Var}[X]}{\sqrt{a^2(\mathbf{Var}[X])^2}} = \operatorname{sinal}(a) = \begin{cases} 1; & a > 0 \\ -1; & a < 0 \end{cases}.$$

A covariância a correlação servem para medir o "nível de correlação" *entre as variáveis. Nesse exemplo vimos que se* $Y = aX + b$ *então* $\mathbf{Cov}[X,Y]$ *tem o mesmo sinal de* a *e* $\rho(X,Y)$ *é* 1 *ou* -1, *de acordo com o sinal de* a.

7.21 **Teorema (Variáveis Aleatórias Independentes são Não Correlacionadas)** *Sejam* X, Y *são variáveis aleatórias independentes integráveis. Então* (XY) *é integrável e*

$$\mathbf{E}[XY] = \mathbf{E}[X]\mathbf{E}[Y]$$

Em particular, variáveis aleatórias independentes não estão correlacionadas.

Demonstração. *Assuma primeiro que as variáveis* X *e* Y *sejam simples, de modo que* X *e* Y *são discretas assumindo uma quantidade finita de valores cada uma.*

Nestas condições XY *também toma apenas um número finito de valores, de modo que* XY *é* *integrável e*

$$\begin{aligned}
\mathbf{E}[XY] &= \sum_{i=1}^{n}\sum_{j=1}^{m} x_i y_j \mathbf{P}[X = x_i, Y = y_j] \\
&= \sum_{i=1}^{n}\sum_{j=1}^{m} x_i y_j \mathbf{P}[X = x_i]\mathbf{P}[Y = y_j] \quad \textit{(pela independência)} \\
&= \left(\sum_{i=1}^{n} x_i \mathbf{P}[X = x_i]\right)\left(\sum_{j=1}^{m} y_i \mathbf{P}[Y = y_j]\right) \\
&= \mathbf{E}[X]\mathbf{E}[Y].
\end{aligned}$$

Agora suponha $X,Y \geq 0$, *e tome uma sequência crescente* $(\phi_n)_{n\geq 0}$ *de funções* $\phi_n : \mathbb{R}^+ \longrightarrow \mathbb{R}^+$ *simples, tal que* $\phi_n(x) \uparrow x$, *para todo* $x \in \mathbb{R}^+$.

Para cada $n \geq 1$ *defina agora* $X_n = \phi_n(X)$ *e* $Y_n = \phi_n(Y)$.

Como X,Y *são independentes,* X_n, Y_n *são independentes e simples, para cada* $n \geq 1$. *Além disso,* $X_n \uparrow X$ *e* $Y_n \uparrow Y$. *Assim, pelo Teorema da Convergência Monótona, segue que*

$$\mathbf{E}[XY] = \lim_{n\to\infty} \mathbf{E}[X_nY_n] = \lim_{n\to\infty} \mathbf{E}[X_n]\mathbf{E}[Y_n] = \mathbf{E}[X]\mathbf{E}[Y] < \infty.$$

Para variáveis quaisquer basta utilizar a decomposição em parte positiva e negativa e observar que as famílias $\{X^+,Y^+\}$, $\{X^+,Y^-\}$, $\{X^-,Y^+\}$ *e* $\{X^-,Y^-\}$ *são formadas por variáveis independentes.* □

Vimos assim que variáveis independentes tem realmente covariância nula. Mas a recíproca não é verdadeira. Vamos deixar esse fato como exercício.

Exercício: Dê um exemplo de variáveis X e Y dependentes, mas com $\mathbf{Cov}[X,Y] = 0$.

(Dica: tente pensar em uma variável X com $\mathbf{E}[X] = 0$ e uma variável Y tal que $Y = 0$ sempre que $X \neq 0$.) □

Uma outra consequência importante é sobre a variância da soma de variáveis independentes.

7.22 Corolário *Se* $X_1, \ldots, X_n$ *são variáveis aleatórias independentes, então*

$$\mathbf{Var}[X_1 + \cdots + X_n] = \mathbf{Var}[X_1] + \cdots + \mathbf{Var}[X_n].$$

Demonstração. *Para* $n = 2$ *temos*

$$\mathbf{Var}[X_1 + X_2] = \mathbf{Var}[X_1] + \mathbf{Var}[X_2] + 2\mathbf{Cov}[X_1,X_2] = \mathbf{Var}[X_1] + \mathbf{Var}[X_2].$$

Os demais casos segue por indução. □

7.23 Proposição *A aplicação* $\mathbf{Cov} : \mathcal{L}^2(\mathbf{P}) \times \mathcal{L}^2(\mathbf{P}) \to \mathbb{R}$ *é uma forma simétrica bilinear positiva semidefinida e* $\mathbf{Cov}[X, Y] = 0$ *se* Y *é constante quase certamente.*

Demonstração. *A demonstração é direta e será deixada como exercício.* □

7.24 Proposição (Desigualdade de Cauchy–Schwarz) *Se* $X, Y \in \mathcal{L}^2(\mathbf{P})$, *então*

$$\mathbf{Cov}[X, Y]^2 \leq \mathbf{Var}[X]\mathbf{Var}[Y].$$

A igualdade é válida se, e somente se, existirem $a, b, c \in \mathbb{R}$, *não todas nulas, e tais que* $aX + bY + c = 0$ *q.c.*

Demonstração. *A desigualdade de Cauchy-Schwarz é verdadeira para qualquer forma bilinear positiva semidefinita e, portanto, em particular para a covariância. Ainda assim, faremos uma demonstração aqui para deixar o texto mais completo e auto-contido.*

Faremos a demonstração supondo que $\mathbf{Var}[Y] > 0$. *O caso no qual* $\mathbf{Var}[Y] = 0$ *será deixada como exercício.*

Dada $\mathbf{Var}[Y] > 0$, *seja* $\theta := -\frac{\mathbf{Cov}[X,Y]}{\mathbf{Var}[Y]}$.

Temos assim $\theta\mathbf{Var}[Y] = -\mathbf{Cov}[X,Y]$, *e portanto*

$$\begin{aligned} 0 &\leq \mathbf{Var}[X + \theta Y]\mathbf{Var}[Y] \\ &= (\mathbf{Var}[X] + 2\theta\mathbf{Cov}[X, Y] + \theta^2\mathbf{Var}[Y])\mathbf{Var}[Y] \\ &= \mathbf{Var}[X]\mathbf{Var}[Y] - \mathbf{Cov}[X, Y]^2. \end{aligned}$$

Note que $\mathbf{Var}[X + \theta Y] = 0$ *se, e somente se,* $X + \theta Y$ *for constante quase certamente.*

Com isso, como supomos $\mathbf{Var}[Y] \neq 0$, *segue que a igualdade na primeira inequação acima vale se, e somente se,* $X + \theta Y$ *for constante quase certamente.* □

7.25 Corolário *Dadas variáveis* $X, Y \in \mathcal{L}^2(\mathbf{P})$ *então*

$$|\rho(X,Y)| \leq 1,$$

com igualdade apenas se existirem constantes $a, b, c \in \mathbb{R}$, *não todas nulas, tais que* $aX + bY + c = 0$.

7.4 Integração com respeito à Distribuição

Nesta seção, definimos integrais em relação às funções de distribuição. Essas integrais, conhecidas como **integrais de Lebesgue- Stieltjes**, podem ser definidas a partir do zero, mas no entanto, elas são simplesmente as integrais com respeito a probabilidades induzidas em $\mathbb{R}$. Nossa principal aplicação ao cálculo de esperanças.

7.26 Definição *Dado uma função de distribuição F em $\mathbb{R}$, existe uma única probabilidade $\mathbf{P}_F$ em $(\mathbb{R},\mathcal{B}(\mathbb{R}))$ tal que $\mathbf{P}_F((a,b]) = F(b) - F(a)$ para todo intervalo $(a,b]$.*

A demonstração desse fato é análoga a construção que fizemos na seção 2.4 da Integral de Lebesgue e foi feita no exercício 2.3.

7.27 Definição (Integração com respeito à distribuição) *Dada F uma função de distribuição em $\mathbb{R}$. Para toda função $g \in \mathcal{L}^1(\mathbf{P}_F)$ definimos a integral de g com respeito a F*

$$\mathbf{E}_F[g] := \int_{\mathbb{R}} g(x)dF(x)$$

onde estamos considerando g como uma variável aleatória no espaço de probabilidade $(\mathbb{R},\mathcal{B}(\mathbb{R}),\mathbf{P}_F)$.

7.28 Proposição *Se $F(t) = \sum p_i \mathbb{1}_{(t_i \leq t)}$ então para toda $g \geq 0$*

$$\mathbf{E}_F[g] = \int_{\mathbb{R}} g dF = \sum_i p_i g(t_i).$$

Demonstração. *Para provar esse fato suponha que $g = \sum_{j=1}^m b_j \mathbb{1}_{B_j}$ é uma função aleatória simples.*

$$\mathbf{E}_F[g] := \sum_{j=1}^m b_j \mathbf{P}_X(B_j) = \sum_{j=1}^m b_j \sum_i p_i \mathbb{1}_{t_i \in B_j} \tag{7.4}$$

$$= \sum_i p_i \sum_{j=1}^m b_j \mathbb{1}_{B_j}(t_i) \tag{7.5}$$

$$= \sum_i p_i g(t_i) \tag{7.6}$$

Para provar para as funções mensuráveis $g \geq 0$ basta utilizar o Teorema da Convergência Monótona. Se $g \geq 0$ e $g_n \uparrow g$

$$\mathbf{E}_F[g] = \lim_{n\to\infty} \mathbf{E}_F[g_n] = \lim_{n\to\infty} \sum_i p_i g_n(t_i) = \sum_i p_i g(t_i)$$

7.29 Proposição *Suponha que F é absolutamente contínua com densidade contínua por partes f. Então se g é positiva e contínua por partes*

$$\mathbf{E}_F[g] = \int_{\mathbb{R}} g dF = \int_{-\infty}^{\infty} g(x)f(x)dx$$

Demonstração. *Para provar esse fato suponha que* $g = \sum_{j=1}^{m} b_j \mathbb{1}_{B_j}$ *é uma função aleatória do tipo escada. Então*

$$\mathbf{E}_F[g] := \sum_{j=1}^{m} b_j \mathbf{P}(B_j) = \sum_{j=1}^{m} b_j \int_{B_j} f(x)dx \tag{7.7}$$

$$= \sum_{j=1}^{m} b_j \int_{-\infty}^{\infty} f(x)\mathbb{1}_{B_j} dx \tag{7.8}$$

$$= \int_{-\infty}^{\infty} f(x) \left[\sum_{j=1}^{m} b_j \mathbb{1}_{B_j}\right] dx \tag{7.9}$$

$$= \int_{-\infty}^{\infty} f(x)g(x)dx \tag{7.10}$$

A demonstração para uma função geral $g > 0$ *segue, aproximando* g *por funções escadas, o que é possível porque* g *é contínua por partes e, em seguida, usando o Teorema de Convergência Monótona.* □

Utilizando as Proposições 7.28 e 7.29 podemos integrar com respeito a uma distribuição que seja combinação de uma discreta e uma absolutamente contínua.

7.30 Teorema (Mudança de Variáveis) *Dada uma variável aleatória* $X \in \mathcal{L}^1$ *e* F_X *a função de distribuição de* X *então*

$$\mathbf{E}[X] = \int_{\mathbb{R}} x \, dF_X(x)$$

e de modo mais geral,

$$\mathbf{E}[g(X)] = \int_{\mathbb{R}} g(x) \, dF_X(x).$$

Demonstração. *Provaremos primeiro para funções simples não negativas* $X = \sum_{i=1}^{n} a_i \mathbb{1}_{A_i}$. *Nesse caso* $\mathbf{E}[X] = \sum_{i=1}^{n} a_i \mathbf{P}(A_i)$. *Por outro lado* $F_X(t) = \sum_i \mathbf{P}(A_i)\mathbb{1}_{a_i \le t}$ *de modo que*

$$\int_{\mathbb{R}} x \, dF_X(x) = \sum_{i=1}^{n} a_i \mathbf{P}(A_i)$$

Deixaremos o restante da demonstração, que segue a estratégia usual de 3 passos, como exercício.

7.31 Exemplo *Se* $X \sim \mathcal{N}(0;1)$ *então* $\mathbf{E}(X) = 0$.

$$\mathbf{E}[X] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x e^{\frac{-x^2}{2}} \mathrm{d}x = -\frac{1}{\sqrt{2\pi}} \int_{*}^{*} e^u du = -\frac{1}{\sqrt{2\pi}} e^{-x^2/2} \Big|_{-\infty}^{\infty} = 0.$$

7.32 Exemplo *Se* $X \sim \mathrm{Exp}(\lambda)$ *então* $\mathbf{E}[X^n] = \frac{n!}{\lambda^n}$. ◁