Intro à Probabilidade

Eduardo Horta

Aula 9 (continuação)

Transformações de variáveis aleatórias

Transformações de variáveis aleatórias

- Consideremos uma variável aleatória qualquer, digamos X.
- Lembre que X é uma *função* com domínio Ω e contradomínio $\mathbb{R}$
- Isto é, para cada possível resultado ω de um experimento aleatório com espaço amostral Ω, tem-se que X assume o valor real $X(\omega)$.
- Assim, se $h\colon \mathbb{R} \to \mathbb{R}$ é uma função qualquer, vemos que a aplicação $Z\colon \Omega \to \mathbb{R}$ definida por

$$Z(\omega) = h\big(X(\omega)\big)$$

  é também uma variável aleatória: se o resultado do experimento for ω, então Z assume o valor real $h(X(\omega))$.
- É costumeiro denotar a variável aleatória Z acima definida por $h(X)$, embora formalmente a notação mais precisa seja $h \circ X$; de fato, Z é precisamente a composição das funções X e h.

Transformações de variáveis aleatórias: exemplo

- Considere o seguinte experimento: medir a temperatura às 8am de amanhã, na estação meteorológica do Centro Histórico de Porto Alegre, utilizando um termômetro digital com resolução de 1 centésimo de grau (em graus Celsius).
- Um modelo probabilístico para esse experimento é o seguinte: tomamos Ω como sendo o conjunto de todas as configurações climáticas possíveis (isto é, o conjunto de todas as combinações possíveis de temperatura, pressão, umidade relativa do ar, etc), munido de uma dada medida de probabilidade[1] $\mathbb{P}$.
- Vamos denotar por X a variável aleatória "temperatura medida pelo termômetro", isto é, $X(\omega)$ = "temperatura (em $^\circ$C até a 2^{a} casa decimal) correspondente à configuração climática ω".

[1] A estrutura probabilística a ser utilizada pode advir, por exemplo, de conhecimentos observacionais ou teóricos prévios. A questão é que, segundo algum critério específico, o modelo de probabilidade seja adequado para descrever o fenômeno em questão.

Transformações de variáveis aleatórias: exemplo

- Digamos que vamos reportar a temperatura a ser medida para um site meteorológico norte-americano. Nesse caso, deveremos fazer a conversão de graus Celsius para graus Fahrenheit.
- A transformação de $^\circ$C para $^\circ$F é dada pela regra

$$h(t) = 1.8t + 32, \qquad t \geq -273.$$

- No experimento acima, podemos portanto definir a variável aleatória Z = "temperatura em $^\circ$F" por

$$Z(\omega) = h(X(\omega)).$$

  Em particular, não precisamos dispor de um termômetro que faça a medição em $^\circ$F, pois o resultado de uma medição obtida por um tal aparelho coincidiria com $h(X)$.

Transformações de variáveis aleatórias

- Em geral, se X e Z são como acima, não é imediato obter-se a FDA de Z mesmo quando a FDA de X é conhecida. A dificuldade surge do fato de que a função h pode não ser injetiva.
- Vejamos alguns casos de fácil solução:

1. Se $h(u) = a + bu$, onde $a \in \mathbb{R}$ e $b > 0$, então para qualquer $z \in \mathbb{R}$ vale que

$$\begin{aligned} F_Z(z) &= \mathbb{P}(Z \leq z) \\ &= \mathbb{P}(h(X) \leq z) \\ &= \mathbb{P}(a + bX \leq z) \\ &= \mathbb{P}(X \leq b^{-1}(z-a)) \\ &= F_X\left(\frac{z-a}{b}\right) \end{aligned}$$

Transformações de variáveis aleatórias

2. Se $h(u) = a + bu$, onde $a \in \mathbb{R}$ e $b < 0$ (isto é, $b = -|b|$), então para qualquer $z \in \mathbb{R}$ vale que

$$
\begin{aligned}
F_Z(z) &= \mathbb{P}(Z \leq z) \\
&= \mathbb{P}(h(X) \leq z) \\
&= \mathbb{P}(a - |b|X \leq z) \\
&= \mathbb{P}(-X \leq |b|^{-1}(z - a)) \\
&= \mathbb{P}(X \geq b^{-1}(z - a)) \\
&= \mathbb{P}(X > b^{-1}(z - a)) + \mathbb{P}(X = b^{-1}(z - a)) \\
&= 1 - F_X\left(\frac{z - a}{b}\right) + p_X\left(\frac{z - a}{b}\right)
\end{aligned}
$$

Transformações de variáveis aleatórias

- Em geral, quando a transformação h é estritamente crescente ou estritamente decrescente, é possível obter a FDA $F_{h(X)}$ a partir de F_X sem maiores dificuldades.
- No caso geral, são necessárias técnicas que não veremos na disciplina. Apenas para ilustrar, considere uma variável aleatória N com FMP

  $$p_N(n) = \frac{1}{2^n}, \qquad n \in \{1, 2 \ldots\}$$

  e $p_N(n) = 0$ se $n \notin \mathbb{N}$. **Exercício**: tente encontrar a FDA da variável aleatória

  $$Z = \sin\left(\frac{\pi(N-1)}{4}\right).$$

Transformações de variáveis aleatórias: caso discreto

- Suponha agora que X é uma variável aleatória *discreta* com FMP p_X, e seja S_X o conjunto de possíveis valores assumidos por X:

$$p_X(x) > 0 \iff x \in S_X$$

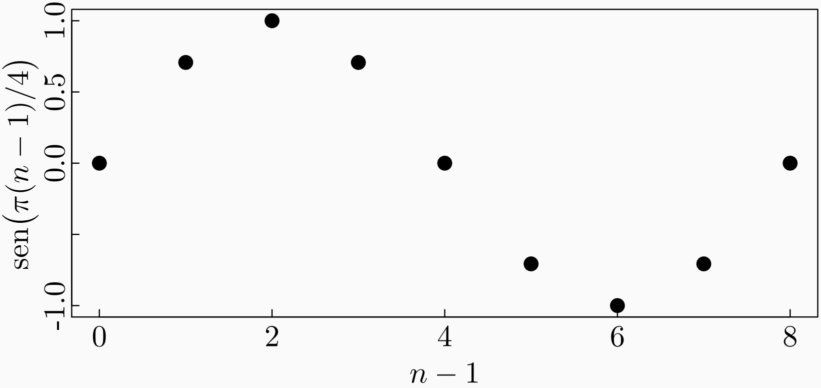
- **Fato**: Se $h\colon \mathbb{R} \to \mathbb{R}$ é uma dada função, então a variável aleatória $Z = h(X)$ é, também, discreta:

$$p_Z(z) > 0 \iff \mathbb{P}\big(h(X) = z\big) > 0 \iff z = h(x) \text{ para algum } x \in S_X$$

Transformações de variáveis aleatórias: caso discreto (exemplo)

- **Exemplo**: Como a função $h(n) = \sin\left(4^{-1}\pi(n-1)\right)$ é periódica (pois $h(n) = h(n+8)$), vemos que o conjunto de possíveis valores da variável aleatória $Z = \sin\left(4^{-1}\pi(N-1)\right)$ é dado por

$$S_Z = \{-1, -\sin(\pi/4), 0, \sin(\pi/4), 1\}$$



Transformações de variáveis aleatórias: caso discreto (exemplo)

- Nesse caso, vemos que vale a igualdade de eventos

$$(Z = 0) \quad = \quad (N = 1) \cup (N = 5) \cup (N = 9) \cup (N = 13) \cup \cdots$$

- Semelhantemente,

$$(Z = -1) \quad = \quad (N = 7) \cup (N = 15) \cup (N = 23) \cup \cdots$$

etc.

- Para obter a FMP de Z, portanto, basta considerarmos os eventos $(Z = z)$ como como acima, com $z \in S_Z$: por exemplo,

$$p_Z(0) = p_N(1) + p_N(5) + p_N(9) + \cdots = \sum_{n=0}^{\infty} p_N(4n+1),$$

isto é, temos

$$p_Z(0) = \sum_{n=0}^{\infty} \frac{1}{2^{4n+1}} = \frac{1}{2} \sum_{n=0}^{\infty} \left(\frac{1}{16}\right)^n = \frac{16}{30}.$$

Aula 10: correção da lista de exercícios

Aula 11

Transformações de múltiplas variáveis aleatórias

Transformações de múltiplas variáveis aleatórias

- Consideremos agora duas variáveis aleatórias *discretas* X e Y.
- Lembre que X e Y são *funções* com domínio Ω e contradomínio $\mathbb{R}$
- Isto é, para cada possível resultado ω de um experimento aleatório com espaço amostral Ω, tem-se que X e Y assumem valores reais $X(\omega)$ e $Y(\omega)$, respectivamente.
- Assim, se $h\colon \mathbb{R}^2 \to \mathbb{R}$ é uma função qualquer, vemos que a aplicação $Z\colon \Omega \to \mathbb{R}$ definida por

  $$Z(\omega) = h(X(\omega), Y(\omega))$$

  é também uma variável aleatória (necessariamente discreta).
- **Notação**: $Z = h(X, Y)$.

Transformações de múltiplas variáveis aleatórias

- Evidentemente, podemos estender a noção acima para o caso em que há mais do que duas variáveis aleatórias envolvidas:
- Se $X_1, X_2, \ldots, X_n$ são n variáveis aleatórias (onde $n \in \mathbb{N}$) então, dada uma função $h\colon \mathbb{R}^n \to \mathbb{R}$ qualquer, também é uma variável aleatória a função $Z\colon \Omega \to \mathbb{R}$ definida por

$$Z(\omega) = h\big(X_1(\omega), X_2(\omega), \ldots, X_n(\omega)\big).$$

Transformações de múltiplas variáveis aleatórias: exemplo

- Considere o seguinte experimento: em um estudo nutricional, serão medidos o peso (em kg) e a altura (em cm) de 10 recém-nascidos (RN) em uma maternidade.
- Vamos considerar que o procedimento de coleta de dados adotado é o de amostragem aleatória simples.
- Para cada $i = 1, 2, \ldots, 10$, sejam X_i e Y_i respectivamente as variáveis aleatórias "peso do i-ésimo RN na amostra (em kg)" e "altura do i-ésimo RN na amostra (em m)".
- Nesse caso, podemos considerar Ω como sendo o conjunto de todas as possíveis amostras de tamanho $n = 10$ da população sob consideração.

Transformações de múltiplas variáveis aleatórias: exemplo

- Seja $h\colon \mathbb{R}_+^2 \to \mathbb{R}$ definida por

$$h(x, y) = \frac{x}{y^2}, \qquad x > 0, y > 0.$$

- Para uma pessoa com peso x (em kg) e altura y (em m), o valor $h(x, y)$ é dito seu *índice de massa corporal*.
- Sendo assim, as variáveis aleatórias dadas por

$$Z_i \overset{\text{def}}{=} h(X_i, Y_i), \qquad i = 1, 2, \ldots, 10$$

  são tais que Z_i representa o “índice de massa corporal do i-ésimo RN na amostra”.

Transformações de múltiplas variáveis aleatórias: exemplo

- Seja ainda a função $g\colon \mathbb{R}^{10} \to \mathbb{R}$ definida por

$$g(u_1, \ldots, u_{10}) = \frac{u_1 + \cdots + u_{10}}{10}, \qquad u_i \in \mathbb{R}.$$

- Nesse caso, temos que a variável aleatória

$$\bar{X} \stackrel{\text{def}}{=} g(X_1, \ldots, X_{10}) = \frac{X_1 + \cdots + X_{10}}{10}$$

representa o peso médio (amostral) ao nascer.

- Podemos definir $\bar{Y}$ e $\bar{Z}$ de forma análoga.

FMP da soma de duas variáveis aleatórias discretas

- Considere duas variáveis aleatórias discretas X e Y com FMP p_X e p_Y, respectivamente, e seja Z a variável aleatória definida por

$$Z(\omega) = X(\omega) + Y(\omega), \qquad \omega \in \Omega.$$

- Claramente, Z é também uma variável aleatória discreta[2].
- Vamos obter a FMP de Z: sejam $x \in S_X$ e $y \in S_Y$ possíveis valores assumidos por X e Y, respectivamente.
- Pondo $z = x + y \in S_Z$, vemos que

$$p_Z(z) = \sum_{u \in S_Y} \mathbb{P}(Z = z \,|\, Y = u)\mathbb{P}(Y = u),$$

pela Lei da Probabilidade Total.

---

[2] Talvez não tão claramente assim!

FMP da soma de duas variáveis aleatórias discretas

- Como temos

$$\begin{aligned}\mathbb{P}(Z = z \,|\, Y = u) &= \frac{\mathbb{P}\{(X + Y = z) \cap (Y = u)\}}{\mathbb{P}(Y = u)} \\ &= \frac{\mathbb{P}\{(X + u = z) \cap (Y = u)\}}{\mathbb{P}(Y = u)} \\ &= \frac{\mathbb{P}\{(X = z - u) \cap (Y = u)\}}{\mathbb{P}(Y = u)} \\ &= p_{X|Y}(z - u \,|\, u),\end{aligned}$$

podemos concluir que

$$p_Z(z) = \sum_{u \in S_Y} p_{X|Y}(z - u \,|\, u) p_Y(u).$$

FMP da soma de duas variáveis aleatórias discretas

- Em particular, *se X e Y forem indepenentes*, então $p_{X|Y}(\cdot|y) = p_X(\cdot)$ e

$$p_Z(z) = \sum_{u \in S_Y} p_X(z-u)p_Y(u).$$

- Define-se a **convolução** $p * q$ de duas funções massa de probabilidade p e q pela expressão

$$p * q(z) \stackrel{\text{def}}{=} \sum_{\{u:\ q(u)>0\}} p(z-u)q(u), \qquad z \in \mathbb{R}.$$

- Pelo argumento acima, se X e Y são variáveis aleatórias independentes, então a convolução $p_X * p_Y$ é igual à FMP da variável aleatória $X+Y$: nesse caso vale que $p_X * p_Y = p_{X+Y}$.

Valor Esperado

Valor Esperado

- Considere uma variável aleatória discreta X em um dado espaço amostral Ω, este munido de uma medida de probabilidade $\mathbb{P}$.
- **Definição**: o **valor esperado** de X é o *número* $\mathbb{E}X \in \mathbb{R}$ dado por

$$\mathbb{E}X \stackrel{\text{def}}{=} \sum_{x \in S_X} x\, p_X(x).$$

- O valor esperado também é chamado, dentre outros nomes, de: **esperança**, **esperança matemática**, **média (populacional)**, **valor médio**.

Valor Esperado: interpretação física

- O valor esperado tem uma interpretação física bastante interessante: considere uma régua de 5in (12.7cm), e suponha que dispomos de 6 barrinhas de chumbo (numeradas de 0 a 5) tais que a soma de suas massas seja igual a 1 unidade de massa (por exemplo, 1oz).
- Digamos que a barrinha 0 tem massa $p(0)$, que a barrinha 1 tem massa $p(1)$, e assim por diante. Como cada $p(n)$ corresponde à massa de um objeto físico, temos $p(n) \geq 0$ para cada $n = 0, 2, \ldots, 5$.
- Além disso, por suposição a soma das massas está normalizada em 1 unidade de massa: isto é, $\sum_{n=0}^{5} p(n) = 1$.

Valor Esperado: interpretação

- Suponha agora que posicionamos a barrinha 0 na marca de 0cm, a barrinha 1 na marca de 1cm, etc. Então *o centro de massa da régua com as barrinhas assim posicionadas é dado por*

$$\mu = \sum_{n=0}^{5} n\, p(n).$$

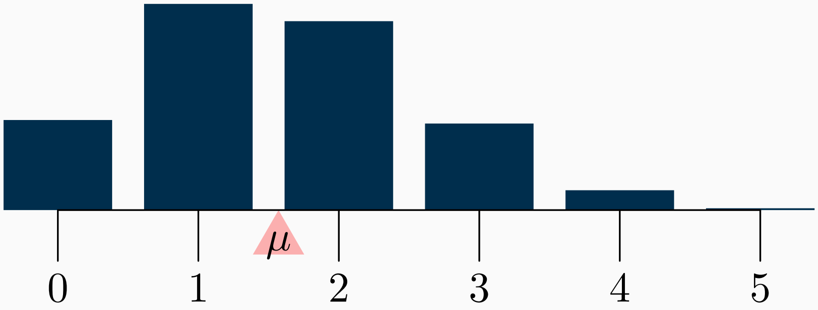
- Quer dizer, o centro de massa do objeto físico descrito acima coincide com o valor esperado de uma variável aleatória X tal que $\mathbb{P}(X = n) = p(n)$ para $n = 0, \ldots, 5$.

Valor Esperado: interpretação

- Assim, se (por exemplo) a massa da barrinha n for

$$p(n) = \binom{5}{n} 10^{-5} \pi^n (10-\pi)^{5-n}, \qquad n = 0, 1, \ldots, 5,$$

então (fazendo a conta!) obtemos $\mu \approx 1.570796$.



- **Importante**: nem sempre o valor esperado de uma variável aleatória está bem definido[3].
- Por exemplo, considere uma variável aleatória X com FMP dada por

$$p_X(n) = \begin{cases} 6/(\pi n)^2, & n \in \mathbb{N} \\ 0, & n \notin \mathbb{N}. \end{cases}$$

- Então, pela definição de valor esperado, temos

$$\mathbb{E}X = \sum_{n \in \mathbb{N}} n \frac{6}{(\pi n)^2} = \frac{6}{\pi^2}\left(1 + \frac{1}{2} + \frac{1}{3} + \cdots\right) = +\infty$$

já que a *série harmônica* $\sum_{n=1}^{\infty} 1/n$ diverge.

[3] Mas não vamos nos preocupar com isso agora. Sempre que escrevermos $\mathbb{E}X$, estaremos assumindo que este valor seja finito (exceto no exemplo dado aqui, é claro).

Valor Esperado: exemplo

- Um apostador chegou a um cassino com apenas $\$1$ pila. Ele decidiu participar do seguinte jogo: a cada rodada, uma moeda equilibrada é lançada. Se o resultado for **cara**, o jogador recebe $\$1$ pila. Se for **coroa**, o jogador deve pagar $\$1$ pila para a banca.
- Felizmente, nosso apostador é um sujeito prudente: ele decidiu que não vai se endividar (ou seja, se ele ficar sem dinheiro ele para de jogar) e também decidiu que vai participar de apenas 3 rodadas do jogo.
- Sendo assim, denotando por X o saldo do apostador depois de encerrada a sua participação no jogo, os possíveis[4] desfechos com os quais ele pode se deparar são os seguintes:

ω	$\circ$	$\bullet\circ\circ$	$\bullet\circ\bullet$	$\bullet\bullet\circ$	$\bullet\bullet\bullet$
$X(\omega)$	0	0	2	2	4
$\mathbb{P}\{\omega\}$	$1/2$	$1/8$	$1/8$	$1/8$	$1/8$

[4] Faça um desenho!

Valor Esperado: exemplo

- Quer dizer, ao final do jogo, o apostador terá $\$0$ pilas com probabilidade $5/8$; ele terá $\$2$ pilas com probabilidade $2/8$; ou $\$4$ pilas com probabilidade $1/8$.
- Sendo assim, o saldo esperado do apostador nesse jogo é

$$\mathbb{E}X = \left(0 \times \frac{5}{8}\right) + \left(2 \times \frac{2}{8}\right) + \left(4 \times \frac{1}{8}\right) = 1.$$

- Ou seja, “em média” o apostador nem perde nem ganha ao participar desse jogo.

Valor Esperado: "long term averages"

- Uma interpretação usual dada ao valor esperado de uma variável aleatória X é a de que $\mathbb{E}X$ pode ser visto como a "média de longo prazo" da variável X que observaríamos se repetíssemos o experimento um grande número de vezes.
- Para ilustrar, considere um espaço amostral finito $\Omega = \{\omega_1, \omega_2, \ldots, \omega_M\}$. De acordo com a interpretação acima, se repetirmos o experimento um grande número de vezes (digamos, n repetições), então esperamos observar o resultado ω_k em aproximadamente $n\,\mathbb{P}\{\omega_k\}$ dessas n repetições (e, consequentemente, observaríamos o valor $X(\omega_k)$ com essa mesma frequência).

Valor Esperado: “long term averages”

- Sendo assim, somando todas as nossas observações dos valores de X obtidos em cada repetição, teríamos

$$\sum_{k=1}^{M} X(\omega_k)\, n\mathbb{P}\{\omega_k\} \approx n\mathbb{E}X$$

- Ou seja, esperamos que para valores de n suficientemente grandes deva valer que a média dos valores observados em cada repetição do experimento deva estar próxima de $\mathbb{E}X$.
- **Importante**: Apesar do argumento das “long term averages” ser considerado intuitivo, é importante que, por ora, nos restrinjamos à definição formal de $\mathbb{E}X$:

$$\mathbb{E}X = \sum_{x \in S_X} x\, p_X(x).$$

- O fato de que as “long term averages” aproximam-se de $\mathbb{E}X$ aparece como um *Teorema* em Probabilidade Axiomática: voltaremos a falar sobre isso!

Valor Esperado: “long term averages”

- Uma ilustração: suponha que o apostador do exemplo anterior voltará ao cassino todos os dias ao longo de 10 anos, sempre munido de apenas $\$1$ pila, e sempre adotando aquela mesma estratégia de jogo. Então em algumas ocasiões ele sairá do cassino com $\$4$ pilas, em outras com $\$2$ pilas, e em outras sairá sem dinheiro.
- A heurística acima nos diz que o *saldo* médio do apostador ao longo desses 10 anos deve ser, aproximadamente, $\$1$ pila.

Aula 12

Valor Esperado de transformações de variáveis aleatórias

Valor Esperado de transformações de variáveis aleatórias

- Consideremos agora o caso em que temos uma variável aleatória Z da forma

$$Z = h(X),$$

onde $h\colon \mathbb{R} \to \mathbb{R}$ é alguma função e X é uma variável aleatória discreta.
- Evidentemente, podemos computar o valor esperado de Z a partir da definição:

$$\mathbb{E}Z = \sum_{z \in S_Z} z\, p_Z(z).$$

- Todavia, conforme vimos anteriormente, mesmo quando temos uma expressão algébrica para p_X, nem sempre é imediato obter, a partir daí, uma expressão algébrica para p_Z.

Valor Esperado de transformações de variáveis aleatórias

- O seguinte resultado é importantíssimo, pois permite que calculemos o valor esperado de $Z = h(X)$ sem termos de nos preocupar em obter $p_{h(X)}$.

Teorema (Law of the Unconscious Statistician) Seja X uma variável aleatória discreta, e seja $h\colon \mathbb{R} \to \mathbb{R}$. Então

$$\mathbb{E}h(X) = \sum_{x \in S_X} h(x)\, p_X(x).$$

Law of the Unconscious Statistician: exemplo

- **Exemplo**: Considere uma variável aleatória X com função massa de probabilidade dada por

$$p_X(n) = \begin{cases} {}^{1}\!/_{2} - {}^{n^2}\!/_{4}, & \text{se } n \in \{-1, 0, 1\} \\ 0, & \text{caso contrário} \end{cases}.$$

- Seja ainda a variável aleatória $Z = |X|$, isto é, $Z = h(X)$ onde h é a função *módulo* (ou: *valor absoluto*).
- Para computarmos o valor esperado de Z diretamente a partir da definição, precisamos antes obter a FMP p_Z.

Law of the Unconscious Statistician: exemplo

- Não é difícil ver que Z assume tão somente os valores 0 ou 1: de fato, temos a igualdade de eventos

$$(Z = 0) \equiv (X = 0), \qquad (Z = 1) \equiv (X = -1) \cup (X = 1),$$

donde $p_Z(0) = {}^1\!/_2$ e $p_Z(1) = p_X(-1) + p_X(1) = {}^1\!/_2$.

- Logo,

$$\mathbb{E}Z = \sum_{z=0}^{1} z \cdot p_Z(z) = \left(0 \cdot \frac{1}{2}\right) + \left(1 \cdot \frac{1}{2}\right) = \frac{1}{2}.$$

- Por outro lado, utilizando a *Law of the Unconscious Statistician*, ficamos com

$$\mathbb{E}Z = \sum_{x=-1}^{1} h(x) \cdot p_X(x) = \left(|-1| \cdot \frac{1}{4}\right) + \left(|0| \cdot \frac{1}{2}\right) + \left(|1| \cdot \frac{1}{4}\right) = \frac{1}{2}.$$

Valor Esperado de transformações de variáveis aleatórias

- Abaixo está reenunciado o Teorema em uma versão mais geral:

Teorema (Law of the Unconscious Statistician) Sejam $X_1, X_2, \ldots, X_n$ variáveis aleatórias discretas com FMP conjunta $p_{\boldsymbol{X}}$, e seja $h\colon \mathbb{R}^n \to \mathbb{R}$. Então

$$\mathbb{E}h(X) = \sum_{x_1 \in S_{X_1}} \cdots \sum_{x_n \in S_{X_n}} h(x_1, \ldots, x_n)\, p_X(x_1, \ldots, x_n).$$

Em particular, para duas variáveis aleatórias discretas X e Y, vale que

$$\mathbb{E}h(X, Y) = \sum_{x \in S_X} \sum_{y \in S_Y} h(x, y)\, p_{X,Y}(x, y).$$

Propriedades do valor esperado

Valor esperado: propriedades

Proposição: sejam X e Y variáveis aleatórias discretas. Então valem as seguintes propriedades:

1. Se $p_X(1) = \alpha$ e $p_X(0) = 1 - \alpha$, então tem-se $\mathbb{E}(X) = \alpha$. Em particular, $\mathbb{E}(1) = 1$.

2. Se $X = \mathbb{I}_A$, onde $A \subseteq \Omega$, então $\mathbb{E}(X) = \mathbb{P}(A)$.

3. $\mathbb{E}(aX + bY) = a\mathbb{E}(X) + b\mathbb{E}(Y)$, para quaisquer $a, b \in \mathbb{R}$. Em particular, $\mathbb{E}(a) = a$ para qualquer constante $a \in \mathbb{R}$.

4. Se X é não-negativa, então $\mathbb{E}X \geq 0$. Em particular, se $X(\omega) \geq Y(\omega)$ para todo $\omega \in \Omega$, então $\mathbb{E}X \geq \mathbb{E}Y$.

5. Se X e Y são independentes, então $\mathbb{E}(XY) = (\mathbb{E}X)(\mathbb{E}Y)$.

Valor esperado: alguns comentários

- É usual denotar o valor esperado também pelos símbolos μ ou μ_X (e, claro, as notações $\mathbb{E}X$ e $\mathbb{E}(X)$ significam a mesma coisa).
- Aqui fica evidenciada a possível ambiguidade na notação $h(X)$ para funções de variáveis aleatórias: anteriormente, escrevemos $Z = h(X)$ para denotar uma nova v.a. dada por $Z(\omega) = h(X(\omega))$.
- Todavia, no caso $h = \mathbb{E}$, estamos de fato lidando com um *operador*: uma função cujo domínio é um conjunto de variáveis aleatórias e cujo contradomínio é a reta:

$$\mathbb{E}\colon \mathcal{L}^1 \to \mathbb{R},$$

onde $\mathcal{L}^1$ é o conjunto formado por todas as variáveis aleatórias para as quais "faz sentido" falar em seu valor esperado[5].

[5] Mais adiante veremos que a noção de valor esperado se estende (sob certas condições) a variáveis aleatórias que *não são* discretas. Aqui, quando dizemos "todas as variáveis aleatórias", estamos nos referindo a todas as variáveis aleatórias *em um dado espaço amostral*.

Variância e desvio padrão

Variância e desvio padrão

- Considere uma variável aleatória discreta X com função massa de probabilidade p_X.
- **Definição**: a **variância de** X é o número $\mathrm{Var}(X) \in \mathbb{R}$ definido por

$$\mathrm{Var}(X) \stackrel{\mathrm{def}}{=} \mathbb{E}\big\{\big(X - \mathbb{E}X\big)^2\big\}.$$

- Escrevendo $\mu = \mathbb{E}X$ e utilizando a propriedade 3 acima, vemos que (expandindo o quadrado na definição)

$$\mathrm{Var}(X) = \mathbb{E}\big\{X^2 - 2\mu X + \mu^2\big\} = \mathbb{E}(X^2) - 2\mu\mathbb{E}X + \mu^2 = \mathbb{E}(X^2) - \mu^2.$$

- Resumidamente:

$$\mathrm{Var}(X) = \mathbb{E}(X^2) - (\mathbb{E}X)^2.$$

Variância e desvio padrão

- **Importante**: Pela Law of the Unconscious Statistician, podemos computar a variância de X pela fórmula

$$\mathrm{Var}(X) = \sum_{x \in S_X} (x - \mu)^2 \cdot p_X(x),$$

onde escrevemos $\mu = \mathbb{E}X$.
- Analogamente,

$$\mathrm{Var}(X) = \sum_{x \in S_X} x^2 \cdot p_X(x) \quad - \quad \mu^2.$$

Variância e desvio padrão: comentários

- É usual denotar a variância de uma variável aleatória X pelos símbolos σ^2 ou σ_X^2.
- **Definição**: nas mesmas condições acima, o **desvio padrão de** X é definido por

$$\sigma_X = \sqrt{\mathrm{Var}(X)}.$$

- Da mesma forma que o valor esperado admite uma interpretação física (sendo este análogo ao centro de massa de um sistema físico em uma certa configuração), a variância corresponde ao momento inercial com respeito a uma rotação ao redor do centro de massa de um sistema físico em uma certa configuração (distribuição) de massa.

Propriedades da variância

Variância e desvio padrão: propriedades

Proposição: sejam X e Y variáveis aleatórias discretas. Então valem as seguintes propriedades:

1. Se $X = a$ com probabilidade 1, onde $a \in \mathbb{R}$ é uma constante, então $\mathrm{Var}(X) = 0$ (isto é, uma constante não varia!).

2. $\mathrm{Var}(aX + b) = a^2 \mathrm{Var}(X)$, para quaisquer $a, b \in \mathbb{R}$.

3. Se X e Y são independentes, então

$$\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y).$$

Covariância e correlação

Covariância e correlação: preliminares

- Dadas variáveis aleatórias X e Y, vimos que a noção de *independência* entre X e Y captura a ideia de que as duas variáveis não se influenciam mutuamente (do ponto de vista epistêmico): se X e Y são independentes, obter conhecimento sobre o valor de, digamos, X não nos traz nenhuma informação a respeito da distribuição de probabilidade de Y.
- Por **exemplo**, considere o seguinte modelo para dois lançamentos consecutivos de uma moeda equilibrada: o espaço amostral é dado por $\Omega = \{\bullet\bullet, \bullet\circ, \circ\bullet, \circ\circ\}$ com

$$\mathbb{P}\{\bullet\bullet\} = \mathbb{P}\{\bullet\circ\} = \mathbb{P}\{\circ\bullet\} = \mathbb{P}\{\circ\circ\} = \frac{1}{4}.$$

- Considere ainda as variáveis aleatórias X e Y representando, respectivamente, o "número de **caras** no primeiro lançamento" e o "número de **caras** no segundo lançamento"; isto é, $X = \mathbb{I}_{\{\bullet\bullet, \bullet\circ\}}$ e $Y = \mathbb{I}_{\{\bullet\bullet, \circ\bullet\}}$.

Covariância e correlação: preliminares

- Nesse caso, é fácil verificar que

$$p_X(0) = p_X(1) = p_Y(0) = p_Y(1) = 1/2$$

  e que $p_{X,Y}(x, y) = 1/4$ para $x, y \in \{0, 1\}$. Disso segue imediatamente que X e Y são independentes.
- Em particular, digamos que a moeda foi lançada uma primeira vez e que observamos o resultado **coroa**. Nesse modelo, essa informação não altera a distribuição de probabilidades da variável aleatória Y: de fato, $p_{Y|X}(y|0) = p_Y(y)$, $y \in \{0, 1\}$ (verifique!).
- Isto é, aqui o conhecimento de que ocorreu **coroa** no primeiro lançamento da moeda *não altera* a probabilidade de ocorrência de **cara** ou **coroa** no segundo lançamento.

Covariância e correlação: preliminares

- Suponha agora que, antes de lançarmos a moeda, a retiramos de uma caixa contendo duas moedas, uma perfeitamente equilibrada e outra tendo **coroa** em ambas as suas faces.
- Nesse caso, podemos propor como espaço amostral o conjunto

$$\Omega = \{e \bullet \bullet, e \bullet \circ, e \circ \bullet, e \circ \circ, \bar{e} \circ \circ\},$$

  onde $e \circ \bullet$ representa o *outcome* “moeda retirada é a equilibrada, **cara** no primeiro lançamento, **coroa** no segundo lançamento”, etc.
- A medida de probabilidade $\mathbb{P}$ que vamos introduzir, baseada na descrição do experimento, é a seguinte:

$$\mathbb{P}\{\bar{e} \circ \circ\} = \frac{1}{2}, \qquad \mathbb{P}\{\omega\} = \frac{1}{8}, \text{ se } \omega \neq \bar{e} \circ \circ$$

- Aqui, as variáveis aleatórias X e Y como definidas no caso anterior *não são independentes*: de fato, vale que $p_{Y|X}(1|1) = {}^1\!/_8 \neq {}^1\!/_4 = p_Y(1)$ (verifique!).

Covariância e correlação

- Uma primeira maneira de descrever o grau de dependência entre duas variáveis aleatórias é através do conceito de *covariância*:
- **Definição**: Se X e Y são variáveis aleatórias discretas, definimos a **covariância entre X e Y** como sendo o número real $\mathrm{Cov}(X,Y) \in \mathbb{R}$ dado por

$$\mathrm{Cov}(X,Y) \stackrel{\mathrm{def}}{=} \mathbb{E}\big\{(X-\mathbb{E}X)\cdot(Y-\mathbb{E}Y)\big\}.$$

- **Importante**: Utilizando as propriedades do valor esperado vistas anteriormente, temos a identidade

$$\mathrm{Cov}(X,Y) = \mathbb{E}(XY) - (\mathbb{E}X)(\mathbb{E}Y).$$

  Exercício: verifique a igualdade acima.

Covariância e correlação: exemplo

- Continuando o exemplo anterior, vamos calcular $\mathrm{Cov}(X, Y)$.
- Utilizando a identidade $\mathrm{Cov}(X, Y) = \mathbb{E}(XY) - (\mathbb{E}X)(\mathbb{E}Y)$, vemos que basta obtermos separadamente os valores de $\mathbb{E}X$, $\mathbb{E}Y$ e $\mathbb{E}(XY)$.
- Além disso, para computar $\mathbb{E}(XY)$, podemos utilizar a *Law of the Unconscious Statistician* (versão geral). Assim não precisamos dar o passo intermediário de obter a FMP da variável aleatória XY.
- Claramente, $\mathbb{E}X = \mathbb{E}Y = 1/4$.

Covariância e correlação: exemplo

- Para $\mathbb{E}(XY)$, temos (usando a LotUS)

$$\mathbb{E}(XY) = \sum_{x=0}^{1}\sum_{y=0}^{1} xy \cdot p_{X,Y}(x,y) = 1 \cdot 1 \cdot p_{X,Y}(1,1) = \frac{1}{8}.$$

- Segue que

$$\mathrm{Cov}(X,Y) = \frac{1}{8} - \frac{1}{4} \cdot \frac{1}{4} = \frac{1}{16}.$$

Aula 13

Covariância: observações

- Conforme visto acima, se X e Y são variáveis aleatórias independentes, então $\mathbb{E}(XY) = (\mathbb{E}X)(\mathbb{E}Y)$ e, em particular, $\mathrm{Cov}(X, Y) = 0$.
- Todavia, a recíproca não é verdadeira: $\mathrm{Cov}(X, Y) = 0$ *não implica independência entre X e Y*!
- Se X e Y são tais que $\mathrm{Cov}(X, Y) = 0$, dizemos que X **e** Y **são não-correlacionadas**.

Covariância: observações

- **Exemplo**: Considere uma variável aleatória X com FMP $p_X(-1) = p_X(0) = p_X(1) = {}^1\!/_3$. Definindo $Y = X^2$, vemos que $\mathbb{E}X = 0$, $\mathbb{E}Y = {}^2\!/_3$ e (usando a LotUS)

$$\mathbb{E}(XY) = \mathbb{E}(X \cdot X^2) = \mathbb{E}(X^3) = \left((-1)^3 \cdot \tfrac{1}{3}\right) + \left(0^3 \cdot \tfrac{1}{3}\right) + \left(1^3 \cdot \tfrac{1}{3}\right) = 0.$$

- Consequentemente, temos

$$\mathrm{Cov}(X, Y) = \mathbb{E}(XY) - (\mathbb{E}X)(\mathbb{E}Y) = 0 - 0 \cdot \tfrac{2}{3} = 0,$$

  isto é, X e Y são não-correlacionadas.
- Todavia, X e Y *não são* independentes: de fato, temos por exemplo

$$p_{Y|X}(1|1) = 1 \neq \tfrac{2}{3} = p_Y(1).$$

Covariância: propriedades:

Proposição: sejam X, Y e Z variáveis aleatórias discretas. Então valem as seguintes propriedades:

1. $\mathrm{Cov}(X, X) = \mathrm{Var}(X)$.
2. $\mathrm{Cov}(X, Y) = \mathrm{Cov}(Y, X)$.
3. $\mathrm{Cov}(aX + Y, Z) = a\mathrm{Cov}(X, Z) + \mathrm{Cov}(Y, Z)$, para qualquer $a \in \mathbb{R}$.
4. $\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\mathrm{Cov}(X, Y)$.

Coeficiente de correlação

Coeficiente de correlação

- A covariância entre duas variáveis aleatórias X e Y captura o *grau de associação (ou dependência) linear* entre elas. Todavia, pode ser difícil de quantificar a intensidade dessa associação a partir do valor de $\mathrm{Cov}(X, Y)$, pois este é um número *não padronizado*.
- Para mitigar essa dificuldade e fornecer um *parâmetro* que capture o grau de associação linear entre X e Y e, ao mesmo tempo, quantifique a intensidade dessa associação, introduz-se *coeficiente de correlação*.
- **Definição**: Se X e Y são variáveis aleatórias discretas com variâncias σ_X^2 e σ_Y^2 respectivamente, definimos o **coeficiente de correlação entre X e Y** como sendo o número real $\rho(X, Y) \in \mathbb{R}$ dado por

$$\rho(X, Y) \stackrel{\text{def}}{=} \frac{\mathrm{Cov}(X, Y)}{\sigma_X \sigma_Y}.$$

Coeficiente de correlação: exemplo

- Vamos continuar o exemplo anterior, onde

$$\Omega = \{e \bullet \bullet, e \bullet \circ, e \circ \bullet, e \circ \circ, \bar{e} \circ \circ\},$$

com

$$\mathbb{P}\{\bar{e} \circ \circ\} = \frac{1}{2}, \qquad \mathbb{P}\{\omega\} = \frac{1}{8}, \text{ se } \omega \neq \bar{e} \circ \circ,$$

- Já calculamos $\operatorname{Cov}(X, Y) = {}^1\!/_{16}$ (onde $X = \mathbb{I}_{\{e\bullet\bullet, e\bullet\circ\}}$ e $Y = \mathbb{I}_{\{e\bullet\bullet, e\circ\bullet\}}$)
- Vamos computar $\rho(X, Y)$. Para isso, falta apenas obter os desvios-padrão σ_X e σ_Y.

Coeficiente de correlação: exemplo

- Nesse caso,

$$\mathrm{Var}(X) = \mathbb{E}(X^2) - (\mathbb{E}X)^2 = \mathbb{E}(X) \cdot (1 - \mathbb{E}(X)) = \tfrac{1}{4} \cdot \tfrac{3}{4} = \tfrac{3}{16},$$

  já que $X = X^2$.
- Semelhantemente, $\mathrm{Var}(Y) = \frac{3}{16}$.
- Segue que a correlação entre X e Y é dada por

$$\rho(X,Y) = \frac{\mathrm{Cov}(X,Y)}{\sigma_X \sigma_Y} = \frac{1/16}{\sqrt{3/16 \cdot 3/16}} = \frac{1}{3}.$$

Coeficiente de correlação: propriedades

Proposição: sejam X e Y variáveis aleatórias discretas. Então valem as seguintes propriedades:

1. $-1 \leq \rho(X, Y) \leq 1$.
2. Suponha que
$$Y = aX + b,$$
onde a e b são constantes reais.
   2.1 Se $a > 0$, então $\rho(X, Y) = 1$.
   2.2 Se $a < 0$, então $\rho(X, Y) = -1$.

- **Comentário**: O item (2) acima admite uma recíproca. Por exemplo, se $\rho(X, Y) = 1$, então Y é da forma $Y = aX + b$ para algum par de constantes $a > 0$ e $b \in \mathbb{R}$, etc.

Coeficiente de correlação: propriedades

- Os itens 2.1 e 2.2 na proposição acima são de fácil verificação. Quanto ao item 1, basta expandir o quadrado em

$$\left(\frac{X - \mathbb{E}X}{\sigma_X} \pm \frac{Y - \mathbb{E}Y}{\sigma_Y}\right)^2 \geq 0$$

para obter as desigualdades desejadas.