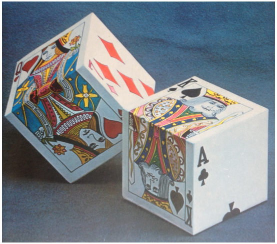
Fundamentos Estatísticos para Ciência dos Dados
V.A.'s I.I.D. e Transformções de v.a.'s

Renato Martins Assunção

DCC, UFMG



Material do capítulo 6 das Notas de Aula

Eventos independentes

- Eventos A e B, ambos contidos em Ω
- são eventos independentes se

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$$

- Equivalentemente: A e B são independentes se, e somente se,

$$\mathbb{P}(A|B) = \mathbb{P}(A)$$

- Podemos estender este conceito de independência para v.a.'s ao invés apenas de eventos.

Independência de V.A.'s

- Intuitivamente, duas v.a.'s X e Y são v.a.'s independentes se saber o valor de uma v.a. não muda as probabilidades associadas com os possíveis valores da outra v.a.
- v.a.'s são representações matemáticas das colunas da matriz de dados.
- Considere k dessas colunas como instâncias das v.a.'s $X_1, X_2, \ldots, X_k$.
- Estamos interessados nos valores dessas variáveis *numa mesma linha da tabela de dados*.
- Para $\omega \in \Omega$, observamos os valores das v.a.'s $X_1(\omega), X_2(\omega), \ldots, X_k(\omega)$.
- Todos são valores medidos no mesmo resultado ω.

Independência de V.A.'s

- Por causa disso, é possível que os valores $X_1(\omega), X_2(\omega), \ldots, X_k(\omega)$ estejam associados.
- Um valor dando alguma informação sobre os demais.
- Na maioria das vezes será de fato assim.
- Mas existem situações em que as v.a.'s não estão associadas.

Um exemplo

- $\omega =$ um indivíduo escolhido ao acaso de certa população humana.
- Sejam $X_1(\omega)$ o seu nível de colesterol LDL (colesterol ruim),
- $X_2(\omega)$ um indicador binário de que o indivíduo é obeso,
- $X_3(\omega)$ um indicador binário de que o indivíduo é fumante,
- $X_4(\omega)$ um indicador binário de que seu primeiro nome começa com uma das letras $A, \ldots, M$ ou se começa com $N, \ldots, Z$,
- Intuitivamente, podemos esperar que as variáveis X_1, X_2 e X_3 não sejam independentes umas das outras.
- É difícil imaginar como X_4 pode estar associadas com X_1, X_2 e X_3.

Outro exemplo

- n lançamentos de uma moeda.
- ω a sequência dos n lançamentos: uma n-upla com C (cara) ou $\tilde{C}$ (coroa)
- O resultado ω é uma n-upla com C (cara) ou $\tilde{C}$ (coroa)
- Seja $X_i(\omega)$ uma v.a. binária indicando se o i-ésimo lançamento foi cara ($X_i(\omega) = 1$) ou coroa ($X_i(\omega) = 0$).
- Os lançamentos não guardam qualquer relação com os resultados prévios ou futuros.
- Intuitivamente, $X_1, X_2, \ldots, X_n$ seriam v.a.s independentes.

Definição

- Sejam $X_1, X_2, \ldots, X_n$ v.a.'s medidas num mesmo espaço de probabilidade $(\Omega, \mathcal{A}, \mathbb{P})$.
- As v.a.'s $X_1, X_2, \ldots, X_n$ são v.a.s independentes se

$$\mathbb{P}(X_1 \in B_1, X_2 \in B_2, \ldots, X_n \in B_n) = \mathbb{P}(X_1 \in B_1)\mathbb{P}(X_2 \in B_2)\ldots\mathbb{P}(X_n \in \tag{1}$$

  para todo conjunto B_i da reta real, $i = 1, \ldots, n$.
- Se as variáveis não forem independentes, dizemos que elas são dependentes.

Informalmente...

- As v.a.'s $X_1, X_2, \ldots, X_n$ são independentes se, e somente se, quaisquer eventos determinados por qualquer grupo de variáveis distintas formam eventos independentes.
- Por exemplo, se as v.a.'s são independentes então:
  - os eventos $[X_1 \leq 2]$ e $[X_2 > 4]$ são eventos independentes;
  - $[X_1 > 4]$ e $[X_2 > 4]$ são independentes também, mesmo que o número 4 apareça nos dois eventos;
  - $[X_1 \leq 2]$, $[X_2 > 4$ e $X_3 > 7]$ e $[X_4 < 0$ ou $X_5 > 10]$ são eventos independentes;

Relevância...

- Basta sabermos a distribuição de cada v.a. individualmente, $\mathbb{P}(X_i \in B_i)$, para obtermos as probabilidades envolvendo todas as v.a.'s.
- Isto é, ao invés de especificar a distribuição conjunta das v.a.'s

$$\mathbb{P}(X_1 \in B_1, X_2 \in B_2, \ldots, X_n \in B_n) , \tag{2}$$

- nós *afirmamos* que ela é igual ao produto das distribuições individuais das v.a.'s,

$$\mathbb{P}(X_1 \in B_1)\mathbb{P}(X_2 \in B_2)\ldots\mathbb{P}(X_n \in B_n) . \tag{3}$$

I.I.D.

- Sejam as v.a.'s $X_1, X_2, \ldots, X_n$ medidas num mesmo espaço de probabilidade $(\Omega, \mathcal{A}, \mathbb{P})$.
- Dizemos que as v.a.'s são independentes e identicamente distribuídas quando:
  - elas forem todas independentes
  - e tiverem, cada uma delas, a mesma distribuição de probabilidade.
- Dizemos que elas são i.i.d.

Amostra aleatória

- $X_1, X_2, \ldots, X_n$ v.a.'s i.i.d.
- Elas têm a mesma distribuição de probabilidade de uma v.a. X com distribuição acumulada $\mathbb{F}$ e densidade $f(x)$ (caso contínuo) ou função de probabilidade $\mathbb{P}(X = x_i)$.
- Dizemos que o vetor aleatório $(X_1, \ldots, X_n)$ é uma *amostra aleatória* da v.a. X.
- Alternativamente, dizemos que o vetor é uma amostra de $\mathbb{F}$ ou de $f(x)$ ou de $\mathbb{P}(X = x_i)$.

Como saber que sâo independentes?

- Em princípio verificando a definição.
- No caso de X_1 e X_2, deveríamos verificar que

$$\mathbb{P}(X_1 \in B_1, X_2 \in B_2) = \mathbb{P}(X_1 \in B_1)\mathbb{P}(X_2 \in B_2)$$

  para todo par de conjuntos B_1 e B_2 da reta real.
- Na prática, existem duas maneiras:
  - Por suposição (ou seja, não "verificamos" coisa nenhuma, assumimos que são independentes)
  - Verificando matematicamente

Independência por suposição

- Refletindo sobre as condições do problema específico nós *assumimos* que as v.a.'s são independentes.
- Assumimos que X_4 (se o primeiro nome começa com $A, \ldots, M$ ou não) é independente do nível de colesterol X_1.
- Obtemos um modelo de distribuição $\mathbb{P}(X_i \in B_i)$ para cada coluna-variável separadamente.
- A seguir, *afirmamos* que $\mathbb{P}(X_1 \in B_1, X_4 \in B_4) = \mathbb{P}(X_1 \in B_1)\mathbb{P}(X_4 \in B_4)$.
- E quando isto não bater com a realidade?
- E quando a suposição de independência não for válida?
- Existem métodos para testarmos se duas ou mais variáveis são independentes (MAIS TARDE NO CURSO).

Independência por dedução

- Neste caso, a probabilidade conjunta $\mathbb{P}(X_1 \in B_1, X_2 \in B_2, \ldots, X_n \in B_n)$ é fornecida.
- De alguma forma, temos um modelo para as probabilidades envolvendo todas as variáveis.
- Obtemos cada probabilidade individual $\mathbb{P}(X_i \in B_i)$ e então mostramos que a probabilidade conjunta é igual ao produto $\mathbb{P}(X_1 \in B_1)\mathbb{P}(X_2 \in B_2) \ldots \mathbb{P}(X_n \in B_n)$

Por exemplo,...

- Duas moedas lançadas em sequência com resultados $X_1(\omega)$ e $X_2(\omega)$.
- Suponha que

ω	$\mathbb{P}(\omega)$
CC	θ^2
$C\tilde{C}$	$\theta(1-\theta)$
$\tilde{C}C$	$(1-\theta)\theta$
$\tilde{C}\tilde{C}$	$(1-\theta)^2$
Total	1

- Então obtemos as distribuições individuais (marginais):

$$\begin{aligned}
\mathbb{P}(X_1 = 1) &= \mathbb{P}(X_1 = 1 \text{ e } X_2 \in \{0,1\}) && (4)\\
&= \mathbb{P}(X_1 = 1, X_2 = 0) + \mathbb{P}(X_1 = 1, X_2 = 1) && (5)\\
&= \theta(1-\theta) + \theta\theta \qquad \text{pela probabilidade conjunta} && (6)\\
&= \theta\,(1-\theta+\theta) = \theta && (7)
\end{aligned}$$

- Temos $\mathbb{P}(X_1 = 0) = 1 - \mathbb{P}(X_1 = 1) = 1 - \theta$.
- Analogamente, $\mathbb{P}(X_2 = x)$ para $x = 0, 1$
- Descobrimos que $X_1 \sim X_2$ (possuem a mesma distribuição).
- Também verificamos que são independentes:

$$\mathbb{P}(X_1 = x_1, X_2 = x_2) = \mathbb{P}(X_1 = x_1)\mathbb{P}(X_2 = x_2)$$

para toda combinação de x_1 e x_2 em $\{0, 1\}$.

Outro caso, dependente

- Duas moedas lançadas em sequência com resultados $X_1(\omega)$ e $X_2(\omega)$.
- Suponha que

ω	$\mathbb{P}(\omega)$
CC	$\theta\sqrt{\theta}$
$C\tilde{C}$	$\theta(1-\sqrt{\theta})$
$\tilde{C}C$	$\theta(1-\theta)$
$\tilde{C}\tilde{C}$	$(1-\theta)^2$
Total	1

- Repetindo os cálculos anteriores: $\mathbb{P}(X_1 = 1) = \theta$ e $\mathbb{P}(X_2 = 1) = \theta(1-\theta+\sqrt{\theta})$ mas

$$\mathbb{P}(X_1 = 1, X_2 = 1) = \theta\sqrt{\theta} \neq \mathbb{P}(X_1 = 1)\mathbb{P}(X_2 = 1) = \theta^2(1-\theta+\sqrt{\theta}) \, .$$

- Não são independentes (basta mostrar que numa configuração específica o produto não vale).

Transformação de V.A.s

- Seja $X \sim U(0,1)$, v.a. uniforme em (0,1).
- Densidade é $f(x) = 1$ para $x \in (0,1)$.
- Seja $Y = X^2$, o quadrado com lado aleatório X.
- Y também é uma v.a.?
- Qual sua distribuição?
- Duas coisas...

Descobrindo experimentalmente

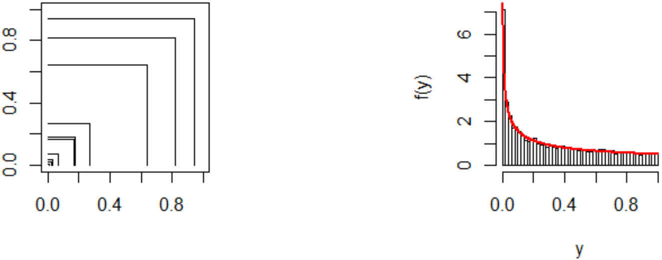


Figura: Simule X muitas vezes, obtenha $Y = X^2$ e faça histograma dos valores de Y.

Um problema em seguros

- Seguro de vida: paga 200 mil reais a beneficiários (esposa e filhos) no momento de morte de um indivíduo que tem 30 anos.
- Suponha um mundo sem inflação.
- O indivíduo vai pagar um único valor, chamado prêmio, no instante de assinatura do contrato.
- Quanto deve ser este prêmio? Qual seria um valor just para ele?
- Dinheiro tem valor no tempo...o valor do amanhã.

O valor do amanhã

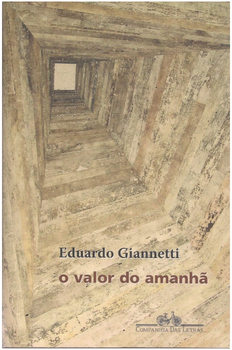


Figura: Eduardo Gianetti

Outro livro altamente recomendado

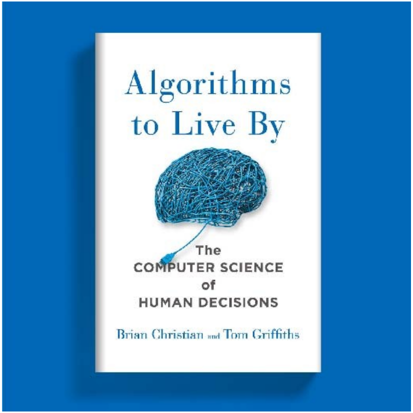


Figura: Sem relação com nosso assunto: A fascinating exploration of how insights from computer algorithms can be applied to our everyday lives, helping to solve common decision-making problems and illuminate the workings of the human mind.

Valor presente

- Seja $\delta = 0.05$ a taxa de juros real (equivalente a 5% ao ano) no futuro.
- O indivíduo vai viver mais T anos a partir da assinaura do contrato.
- Daqui a T anos, a seguradora vai desembolsar 100 mil reais (corrigido por inflação, mas ignore isto)
- Cálculo financeiro básico mostra que o *valor presente* (no instante de assinar o contrato) do benefício é

$$Y = 100 \exp(-\delta T) = 100 \exp(-0.05 T) = 100 \times 0.9512^{T}$$

- Se ele pagar HOJE esta quantidade Y, a seguradora terá recebido o que precisa ara pagar de volta o benefício de 100 mil reais.
- Por exemplo, se ele falecer dentro de 20 anos (com $T = 20$ então), basta ele pagar HOJE $Y = 100 exp(0.05 \times 20) = 36.78$.
- Se ele falecer dentro de $T = 10$ anos, a seguradora deveria cobrar 60.65 mil reaispara conseguir agar o benefício de 100 mil reais.

Valor presente é variável aleatória

- Em resumo, cobre $Y = 100\exp(-\delta T) = 100\exp(-0.05T)$ do seu cliente no momento da assinatura do contrato: o pagamento do benefício está coberto.
- Quanto mais tempo ele demorar para falecer (maior T), menor o valor de Y.
- Mas não sabemos o valor de T, ele é aleatório e só vai ser instanciado em algum momento no futuro.
- Se não temos T, não temos $Y = 100\exp(-0.05T)$.

Valor presente é variável aleatória

- $Y = 100\exp(-0.05T)$ é uma variável aleatória, função matemática da v.a. T.
- Note que $Y = g(T)$ onde $g(t) = 100\exp(-0.05t)$
- O que cobrar do cliente?
- Solução: cobre o valor esperado $\mathbb{E}(Y) = \mathbb{E}\left(100\exp(-0.05T)\right)$.
- Isto é um número fixo, não uma v.a.
- Para alguns indivíduos teremos $Y > \mathbb{E}(Y)$ e para outros teremos $Y < \mathbb{E}(Y)$.

Valor presente atuarial

- Vários indivíduos de 30 anos indexados por $i = 1, 2, \ldots, n$
- Os seus tempos de vida futura $T_1, T_2, \ldots, T_n$ são v.a.'s i.i.d.
- Cada um deveria pagar $Y_1, Y_2, \ldots, Y_n$.
- Cobramos o mesmo valor $\mathbb{E}(Y) = \mathbb{E}\left(100 \exp(-0.05T)\right)$ de todos (se soubermos calculá-lo)
- Para alguns teremos $Y_i > \mathbb{E}(Y)$ e para outros teremos $Y_i < \mathbb{E}(Y)$.
- Mas sabemos que $(Y_1 + Y_2, \ldots + Y_n)/n \approx \mathbb{E}(Y)$
- Ou seja:
  - Eu deveria coletar, $(Y_1 + Y_2, \ldots + Y_n)$, mas isto é impossível ois não conhecemos Y_i na assinatura do contrato
  - Este valor é aproximadamente igual ao que eu de fato coletei: $n\mathbb{E}(Y)$ (o valor $\mathbb{E}(Y)$ de cada um deles)
- Se a carteira de clientes da seguradora for grande, esperamos que os desequilíbrios se cancelem no final.
- $\mathbb{E}(Y)$ é chamado de valor presente atuarial.

Distribuição de $Y = h(X)$

- Como obter a distribuição de probabilidade de $Y = h(X)$ conhecendo-se a distribuição de X?
- Vamos supor que X e Y sejam contínuas.
- Dois métodos:
  1. Inversão de h: obtenha $\mathbb{F}_Y(y)$ e a seguir obtenha a sua derivada $\mathbb{F}'_Y(y) = f_Y(y)$, que é a densidade de Y.
  2. Use o teorema da transformação de v.a.'s.
- O método (1) é mais simples e intuitivo mas só serve para casos univariados, em que Y é função de uma única v.a. X.
- O método (2) é matematicamente mais sofisticado mas permite generalizar para obter a densidade de Y quando ela for da forma $Y = h(X, Z)$, função de duas ou mais v.a.'s.

Método da inversão de h

- Considere $Y = X^2$ onde $X \sim U(0,1)$ (área do quadrado de lado aleatório X).
- Lista de valores possíveis de Y: intervalo $(0,1)$.
- Considere um dos valores possíveis: $y = 0.37$.
- Queremos $\mathbb{F}_Y(0.37) = \mathbb{P}(Y \leq 0.37)$.
- Temos a igualdade de eventos $[Y \leq 0.37] = [X \leq \sqrt{0.37}]$ pois

$$
\begin{aligned}
[Y \leq 0.37] &= \{\omega \in \Omega \text{ tais que } Y(\omega) \leq 0.37\} \\
&= \{\omega \in \Omega \text{ tais que } (X(\omega))^2 \leq 0.37\} \\
&= \{\omega \in \Omega \text{ tais que } X(\omega) \leq \sqrt{0.37}\} \\
&= [X \leq \sqrt{0.37}]
\end{aligned}
$$

- Se $[Y \leq 0.37]$ e $[X \leq \sqrt{0.37}]$ são os mesmos, suas probabilidades também são iguais:

$$\mathbb{F}_Y(0.37) = \mathbb{P}(Y \leq 0.37) = \mathbb{P}(X \leq \sqrt{0.37}) = \mathbb{F}_X(\sqrt{0.37}) = \sqrt{0.37}\,.$$

Método da inversão de h

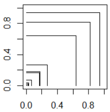
- Refazendo para um valor $y \in (0,1)$ qualquer:

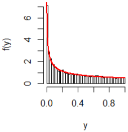
$$\mathbb{F}_Y(y) = \mathbb{P}(Y \leq y) = \mathbb{P}(X \leq \sqrt{y}) = \mathbb{F}_X(\sqrt{y}) = \sqrt{y}\,.$$

- Para obter a densidade $f_Y(y)$ para $y \in (0,1)$

$$f_Y(y) = \frac{d}{dy}\mathbb{F}_Y(y) = \frac{d}{dy}\sqrt{y} = \frac{1}{2\sqrt{y}}\,.$$

- Para $y \notin (0,1)$, temos $f_Y(y) = 0$ pois $\mathbb{F}_Y(y)$ é constante e igual a 0 (para $y < 0$) ou 1 (para $y > 1$).
- O gráfico da densidade $f_Y(y)$ é a curva vermelha abaixo.





Caso geral

- A função de distribuição acumulada $\mathbb{F}_Y(y)$ de $Y = h(X)$ é definida em termos de $\mathbb{F}_X(x)$ da seguinte forma

$$
\begin{aligned}
\mathbb{F}_Y(y) &= \mathbb{P}(Y \leq y) = \mathbb{P}(h(X) \leq y) \\
&= \mathbb{P}\left(X \in \{x \text{ tal que } h(x) \leq y\}\right) \\
&= \mathbb{P}\left(X \in \{x \text{ tal que } x \leq h^{-1}(y)\}\right) \\
&= \mathbb{F}_X(h^{-1}(y))
\end{aligned}
$$

- Veja mais detalhes, exemplos e discussão nas notas de aula.

Teorema da transformação de v.a.'s.

- Seja $Y = h(X)$. Suponha que no suporte de X a função h seja inversível com $g = h^{-1}$ e portanto $x = g(y)$.
- Então a densidade de Y no ponto y é dada por

$$f_Y(y) = f_X(g(y)) \left|\frac{dg(y)}{dy}\right| = f_X(g(y)) \left|\frac{dh^{-1}(y)}{dy}\right| \tag{8}$$

- Exemplo: $Y = h(X) = X^2$ com $X \sim U(0,1)$. Então, $x = g(y) = \sqrt{y}$.
- Pelo teorema, $f_Y(y) = f_X(\sqrt{y}) \times |\frac{d}{dy}\sqrt{y}|$.
- Temos $f_X(x) = 1$ para tdo x e portanto $f_X(\sqrt{y}) = 1$.
- Assim, $f_Y(y) = 1 \times |\frac{d}{dy}\sqrt{y}| = \frac{1}{2\sqrt{y}}$.

Obtendo $\mathbb{E}(Y)$

- Muitas vezes, qeremos apenas $\mathbb{E}(Y)$ onde $Y = h(X)$.
- Temos duas formas de obter esta esperança
- (1): Obtenha $f_Y(y)$ por uma dos métodos anteriores e, a seguir, obtenha

$$\mathbb{E}(Y) = \int y f_Y(y) dy$$

- (2): simplesmente obtenha

$$\mathbb{E}(Y) = \mathbb{E}(h(X)) = \int h(x) f_X(x) dx$$

sem nunca precisar obter $f_Y(y)$.

Exemplo

- $Y = X^2$ onde $X \sim U(0,1)$
- Temos $f_X(x) = 1$ e já obtivemos $f_Y(y) = 1/(2\sqrt{y})$.
- Então, pelo primeiro método:

$$\mathbb{E}(Y) = \int_0^1 y f_Y(y) dy = \int_0^1 y \frac{1}{2\sqrt{y}} dy = \frac{1}{3}$$

- Pelo segundo método:

$$\mathbb{E}(Y) = \int h(x) f_X(x) dx = \int_0^1 x^2 \, 1 \, dx = \frac{1}{3}$$