Notas sobre Probabilidade Discreta

por

Roberto Imbuzeiro M. F. de Oliveira

IMPA

Índice

1 Introdução **1**
1.1 Andamento . . . 1
1.1.1 20 de março de 2007 . . . 1
1.2 Referências . . . 1

2 Definições básicas do caso discreto **2**
2.1 Conceitos básicos . . . 2
2.2 Exemplos . . . 3
2.3 Probabilidades condicionais . . . 6
2.4 Partições e probabilidades condicionais . . . 8
2.5 Independência . . . 8

3 Variáveis aleatórias **11**
3.1 Definição . . . 11
3.2 Distribuição de uma variável aleatória . . . 11
3.3 Distribuições novas a partir de antigas . . . 12
3.4 Independência . . . 13
3.5 Somas de variáveis aleatórias independentes . . . 14

4 Valores esperados, momentos e desigualdades **16**
4.1 Valores esperados e momentos . . . 16
4.1.1 A desigualdade de Jansen e as normas L_p . . . 18
4.2 Variância e covariância . . . 19
4.3 A desigualdade de Chebyshev e concentração . . . 21
4.4 Aplicação a aproximações por polinômios . . . 23

5 Interpretação das probabilidades condicionais **25**
5.1 Probabilidades e esperanças condicionais . . . 25
5.1.1 Informação e aproximação: definindo probabilidades condicionais . . . 25
5.1.2 Informação e aproximação: o caso geral . . . 27

Capítulo 1: Introdução

Muitos dos conceitos relevantes em Probabilidade é já são interessantes no caso discreto, em que as tecnicalicades de Teoria da Medida são desnecessárias e as idéias envolvidas se tornam transparentes. Além disso, as distribuições discretas são freqüentemente encontradas em aplicações de Probabilidade.

Estas duas razões sugerem que um curso introdutório de Probabilidade dispense especial atenção a este caso particular da teoria. No entanto, a duração do curso de mestrado em Probabilidade do IMPA exige que se cumpra a parte "dif'ıcil"do programa sem muita demora. Estas notas sucintas se propõem a complementar este curso através de um estudo paralelo da Probabilidade discreta através de exercícios. Apresentaremos uma boa parte do curso neste caso particular, indicaremos a correspondência natural que existe entre conceitos de Probabilidade discreta e de Medida e, por fim, exibiremos as limitações do caso discreto e a forma pela qual elas são superadas pela teoria axiomática de Kolmogorov.

Sugestões e correções devem ser enviadas para `rimfo@impa.br`.

1.1 Andamento

Estas notas são um trabalho em andamento que será constantemente atualizado. Manteremos aqui uma lista das modificações mais relevantes e das seções já prontas.

1.1.1 20 de março de 2007

Só os Capítulos 2 e 3 estão razoavelmente prontos (mas ainda não revistos). O Capítulo 4 já está algo legível e pode ser estudado preliminarmente. O Capítulo 5 ainda está totalmente caótico.

1.2 Referências

As principais referências destas notas são:

1. Kai Lai Chung, "A Course in Probability Theory Revised";
2. William Feller, "An Introduction to Probability Theory and Its Applications, Volume1";
3. Barry James. "Probabilidade: um curso em nível intermediário".

Capítulo 2: Definições básicas do caso discreto

2.1 Conceitos básicos

Aproximadamente um sexto dos lançamentos de um dado resulta no número 4. As brasileiras têm em média $2,5$ filhos. Há uma chance de 22% de que a economia chilena cresca mais do que a indiana no ano de 2007. Cada uma destas afirmações pode ser vagamente interpretada de uma das seguintes duas maneiras:

- *Avaliação de risco:* Há um conjunto de possibilidades para o que pode vir a acontecer. A cada possibilidade atribui-se uma medida numérica do risco de sua ocorrência.
- *Freqüência:* Olhamos para uma série de circunstâncias repetidas. Para cada repetição, observamos a ocorrência de um dado evento e calculamos a fração de vezes em que o evento acontece.

Grosso modo, a definição de probabilidade que veremos a seguir captura a primeira interpretação acima. Um teorema fundamental chamado de Lei dos Grandes Números nos permitirá dizer que, ao menos em alguma situações, podemos recuperar a segunda interpretação de forma precisa.

Nossa definição (provisória) de probabilidade tem dois ingredientes.

Definição 2.1. *Um* espaço de probabilidade discreto *é um par* $(\Omega, \mathbb{P})$ *cujos dois elementos são:*

1. *Um* espaço amostral Ω*, que é o conjunto de possíveis "acontecimentos"*Ω*, e que é um conjunto finito ou enumerável.*
2. *Uma* medida de probabilidade *(ou* distribuição*)* $\mathbb{P}(\cdot)$*, que atribui a cada elemento* $\omega \in \Omega$ *uma probabilidade ("valor de risco")* $\mathbb{P}(\omega) \in [0,1]$*. Exigiremos sempre que a soma das probabilidades seja* 1*, isto é:*
$$\sum_{\omega\in\Omega} \mathbb{P}(\omega) = 1.$$

Hipótese 2.2. *Todos os espaçõs de probabilidade neste capítulo são espaços discretos.*

A definição acima induz uma função sobre os subconjuntos de Ω (isto é, o conjunto $\mathcal{P}(\Omega)$ das *partes* de Ω). Esta nova função também será chamada de $\mathbb{P}$.

$$\begin{array}{rcccc} \mathbb{P} & : & \mathcal{P}(\Omega) & \to & [0,1] \\ & & A & \mapsto & \mathbb{P}(A) = \sum_{\omega\in A} \mathbb{P}(\omega) \end{array} \tag{2.1}$$

Note que, com esta definição, $\mathbb{P}(\{\omega\})$ é o valor de $\mathbb{P}(\omega)$ definido anteriormente.

Exercício 2.1. *Prove que a função* $\mathbb{P}$ *sobre* $\mathcal{P}(\Omega)$ *satisfaz as seguintes propriedades:*

1. $\mathbb{P}(\emptyset) = 0$
2. $\mathbb{P}(\Omega) = 1$

3. *se* $A_1, A_2, \dots \subset \Omega$ *são conjuntos disjuntos 2 a 2,* $\mathbb{P}\left(\cup_{n=1}^{+\infty} A_n\right) = \sum_{n=1}^{+\infty} \mathbb{P}(A_n)$. *(A probabilidade da união de conjuntos disjuntos é a soma das probabilidades dos conjuntos individuais.)*

Mostre a seguinte recíproca: se $\mathbb{P} : \mathcal{P}(\Omega) \to [0,1]$ *satisfaz estas três propriedades, então* $\mathbb{P}(\omega) = \mathbb{P}(\{\omega\})$ *é uma medida de probabilidade no sentido da Definição 2.1.*

Observação 2.3. *Os elementos* $A \in \mathcal{P}(\Omega)$ *são ocasionalmente chamados de* eventos.

Exercício 2.2. *Prove também as seguintes propriedades.*

1. *Se* $A \subset B \subset \Omega$, $\mathbb{P}(A) \leq \mathbb{P}(B)$;
2. *Se* $A_1, A_2 \subset \Omega$, $\mathbb{P}(A_1) = \mathbb{P}(A_1 \cap A_2) + \mathbb{P}(A_1 \backslash A_2)$;
3. *para todo* $A \in \mathcal{P}(\Omega)$, $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$, *onde* A^c *é o complementar de* A *em* Ω.
4. Inclusão-exclusão: *Para quaisquer conjuntos* $A_1, A_2 \subset \Omega$:

$$\mathbb{P}(A_1 \cup A_2) = \mathbb{P}(A_1) + \mathbb{P}(A_2) - \mathbb{P}(A_1 \cap A_2).$$

5. Inclusão-exclusão generalizada: *Para quaisquer conjuntos* $A_1, A_2, \dots, A_n \subset \Omega$:

$$\mathbb{P}(\cup_{i=1}^{n} A_i) = \sum_{k=1}^{n} (-1)^{k+1} \sum_{S \subset \{1,\dots,n\} : |S| = k} \mathbb{P}(\cap_{i \in S} A_i),$$

onde $|S|$ *é a cardinalidade de* S. *[Este é um item mais difícil. Uma estratégia para resolvê-lo é usar indução em* n. *O caso* $n = 2$ *é o item anterior. Para* $n \geq 3$, *considere* $\mathbb{P}(B_{n-1} \cup A_n)$ *onde* $B_{n-1} = \cup_{i=1}^{n-1} A_i$. *Note que pelo item anterior:*

$$\mathbb{P}(B_{n-1} \cup A_n) = \mathbb{P}(A_n) + \mathbb{P}(B_{n-1}) - \mathbb{P}\left(B'_{n-1}\right),$$

onde $B'_{n-1} = \cup_{i=1}^{n-1}(A_i \cap A_n)$. *Agora aplique a hipótese indutiva.]*

2.2 Exemplos

Alguns exemplos básicos de espaços de probabilidade são apresentados a seguir. O leitor deve certificar-se de que cada um corresponde de fato a um espaço de probabilidade.

Exercício 2.3 (Espaços produto)**.** *Para* $1 \leq i \leq n$, *sejam* $(\Omega_i, \mathbb{P}_i)$ *espaços de probabilidade (discretos). Defina um novo espaço* $(\Omega, \mathbb{P})$ *mediante o produto cartesiano*

$$\Omega = \Omega_1 \times \Omega_2 \times \dots \Omega_n$$

e tomando

$$\mathbb{P}(\omega) = \mathbb{P}_1(\omega_1)\, \mathbb{P}_1(\omega_2) \dots \mathbb{P}_n(\omega_n),\ \omega = (\omega_1, \dots, \omega_n) \in \Omega. \tag{2.2}$$

Mostre que $(\Omega, \mathbb{P})$ *é um espaço de probabilidade (o espaço produto) e que (2.2) é equivalente a*

$$\mathbb{P}(A_1 \times \dots \times A_n) = \mathbb{P}_1(A_1)\, \mathbb{P}_1(A_2) \dots \mathbb{P}_n(A_n),\ A_1, \dots, A_n \subset \Omega.$$

Exemplo 2.4 (Bernoulli). *Neste caso* $\Omega = \{0,1\}$ *correspondendo ao cara/coroa de uma moeda. Fixamos um número* $p \in [0,1]$ *e dizemos que* $\mathbb{P}(1) = p$, $\mathbb{P}(0) = 1-p$. *Esta distribuição é chamada de Bernoulli com parâmetro* p (Be_p)

Exercício 2.4 (Produto de Bernoullis). *Escolha* $p \in [0,1]$ *e* $n \in \mathbb{N}$. *Seja* $\Omega = \{0,1\}^n$ *e* $\mathbb{P}(\omega) = p^{|\omega|}(1-p)^{n-|\omega|}$, *onde* $|\omega| = \sum_{i=1}^n \omega_i$. *Mostre que este espaço é o produto de* n *espaços* $\Omega_i = \{0,1\}$ *com medida* $\mathbb{P}_i = Be_p$.

Exemplo 2.5 (Um dado). *Neste caso* $\Omega = \{1,2,3,4,5,6\}$ *correspondendo às faces de um dado. Definimos* $\mathbb{P}(\omega) = 1/6$ *para cada* $\omega \in \Omega$.

Exemplo 2.6 (Distribuição uniforme). *Generalizando o exemplo anterior,* Ω *é um conjunto discreto dado e* $\mathbb{P}(\omega) = 1/|\Omega|$ *para cada* $\omega \in \Omega$, *onde* $|\Omega|$ *é a cardinalidade do conjunto* Ω. *Esta é a distribuição uniforme sobre* Ω $(Unif_\Omega)$

Exercício 2.5. *Mostre que neste caso* $\forall A \subset \Omega$, $\mathbb{P}(A) = |A|/|\Omega|$.

Exercício 2.6 (Produtos de distribuições uniformes são uniformes). *O produto de* n *espaços finitos* $(\Omega_i, \mathbb{P}_i = Unif_{\Omega_i})$ *é* $(\Omega, \mathbb{P} = Unif_\Omega)$.

Exemplo 2.7 (Distribuição geométrica). *A técnica de datação por carbono* 14 *é baseada no chamado decaimento: cada átomo transforma-se espontâneamente em outro tipo de átomo ao longo do tempo*[1]. *Nosso espaço* Ω *corresponderá ao número de segundos que um átomo escolhido de carbono* 14 *demora para decair:* $\Omega = \mathbb{N} = \{1,2,3,\dots\}$. *A probabilidade* $\mathbb{P}$ *terá a forma de "decaimento exponencial discreto"determinada pela seguinte fórmula:*

$$\mathbb{P}(\{k, k+1, k+2, \dots\}) = (1-p)^{k-1}, \tag{2.3}$$

onde $p \in [0,1)$ *é um parâmetro que depende das propriedades do carbono-*14 [2]. *De modo geral, a distribuição determinada pela fórmual acima é chamada de* geométrica com parâmetro p (Geo_p).

Exercício 2.7. *Mostre que há uma única função* $\mathbb{P} : \Omega \to [0,1]$ *compatível com (2.3) e que ela é dada por* $\mathbb{P}(\omega) = p(1-p)^{\omega-1}$. *Calcule também a* meia-vida H, *isto é, o menor* k *tal que*

$$\mathbb{P}(\{k, k+1, k+2, \dots\}) \leq 1/2.$$

Exemplo 2.8 (Binomial). *Recordamos a definição do coeficiente binomial* $\binom{n}{k}$:

$$\binom{n}{k} \equiv \frac{n!}{k!(n-k)!} \quad (k, n \in \mathbb{N}, 0 \leq k \leq n).$$

A distribuição binomial com parâmetros $n \in \mathbb{N}$, $p \in [0,1]$ $(Bin_{n,p})$ *é a probabilidade sobre* $\Omega = \{0,1,2,\dots,n\}$ *dada por*

$$\binom{n}{k} p^k (1-p)^{n-k}, k \in \Omega.$$

[1] Ver `http://en.wikipedia.org/wiki/Carbon-14`.

[2] Por que não podemos tomar $p \geq 1$?

Exercício 2.8 (Produto de Bernoullis e Binomial.)**.** *Volte ao Exercício 2.4. Considere os eventos* $E_k \equiv \{\omega : |\omega| = k\}$*. Prove que cada* E_k *é a união disjunta de* $\binom{n}{k}$ *eventos*

$$F_S \equiv \{\omega : \forall 1 \leq i \leq k,\, \omega_i = 1 \text{ se } i \in S \text{ ou } 0 \text{ se } i \notin S\}.$$

Mostre que $\mathbb{P}(F_S) = p^k(1-p)^{n-k}$ *e que* $\mathbb{P}(E_k) = Bin_{n,p}(k)$*. Esta conexão entre o produto de Bernoullis e a distribuição Binomial será elucidada quando falarmos de variáveis aleatórias.*

Exercício 2.9 (Apresentando a distribuição Poisson)**.** *Fixe* $\lambda > 0$ *e considere (para* $n \geq \lambda$*) a distribuição* $\mathbb{P}_n = Bin_{n,p_n}$*, com* $p_n = \lambda/n$*. Note que* $\mathbb{P}_n(k)$ *está definido para todo* $0 \leq k \leq n$ *inteiro. Prove o seguinte limite para todo* $k \in \mathbb{N}$

$$\lim_{n \to +\infty} \mathbb{P}_n(k) \equiv Po_\lambda(k) = e^{-\lambda}\frac{\lambda^k}{k!}.$$

Po_λ *é uma probabilidade sobre* $\mathbb{N}$ *conhecida como* Poisson com parâmetro $\lambda > 0$*. Este exercício mostra que "binomiais convergem para Poisson"; um resultado mais forte será provado bem mais adiante.*

Exemplo 2.9 (Retirando bolas de urnas com ou sem reposição)**.** *Imagine uma urna com bolas numeradas de 1 a* n *das quais* $k \leq n$ *bolas* $a_1, \dots, a_k$ *são retiradas sucessivamente.*

Para definirmos as distribuições abaixo, seja $[b] = \{1, dots, b\}$ *(b natural).* S^R *é o conjunto de funções de* S *em* R *e* S^R_{inj} *é o subconjunto de funções injetivas. Descrevemos duas situações possíveis.*

1. *Cada bola retirada é reposta. Se definimos* $\omega : [k] \to [n]$ *via* $\omega(i) = a_i$*, então* ω *é um elemento do espaço* $\Omega = [n]^{[k]}$*. Se* $\mathbb{P} = Unif_{[n]^{[k]}}$*, este caso é conhecido como retirada de bolas* com reposição.

2. *Cada vez que uma bola é retirada, ela não é reposta na urna, de modo que na i-ésima retirada restam na urna as bolas* $\{1, \dots, n\} \backslash \{a_1, \dots, a_{i-1}\}$*. Se definimos* $\omega : [k] \to [n]$ *via* $\omega(i) = a_i$*, então* ω *é um elemento do espaço* $\Omega = [n]^{[k]}_{inj}$ *Se* $\mathbb{P} = Unif_{[n]^{[k]}_{inj}}$*, temos o que se chama de* retirada de bolas com reposição.

Exercício 2.10. *Prove que* $[n]^{[k]}$ *e* $[n]^k$ *têm uma bijeção natural onde cada* $\omega \in [n]^{[k]}$ *corresponde ao vetor* $(\omega(1), \dots, \omega(n))$*. Logo* $Unif_{[n]^{[k]}}$ *corresponde naturalmente á medida produto sobre* $[n]^k$ *(Exercício 2.6).*

Exercício 2.11. *Considere o caso de uma urna com* n *bolas da qual* $k = n$ *bolas são tiradas sem reposição. Mostre que neste caso* $\Omega = S_n$*, o conjunto das permutações de* $[n]$*. Agora considere o conjunto das permutações com pontos fixos, isto é, que mapeiam algum* $i \in [n]$ *nele mesmo.*

$$F_n = \{\omega \in S_n : \forall i \in [n],\, \omega(i) = i\}.$$

Seja $\mathbb{P}_n = Unif_\Omega$*. Seguindo os passos abaixo, provaremos que*

$$\mathbb{P}_n(F_n) = \sum_{k=1}^{n} \frac{(-1)^{k+1}}{k!} \to 1 - e^{-1}.$$

1. *Defina* $E_i = \{\omega \in S_n : \omega(i) = i\}$*. Mostre que* $F_n = \cup_i E_i$*.*

2. *Mostre que para todos* $1 \leq i_1 < \cdots < i_k \leq n$ $\mathbb{P}(E_{i_1} \cap \cdots \cap E_{i_k}) = (n-k)!/n! = 1/k!\binom{n}{k}$ *[Dica: seja* σ *uma permutação de* $[n]\backslash\{i_1, \ldots, i_k\}$ *(que tem* $n-k$ *elementos). Defina* $\omega = \omega_\sigma$ *com a permutaçã com* $\omega(i_j) = i_j$ *e* $\omega(u) = \sigma(u)$ *para* $u \in [n]\backslash\{i_1, \ldots, i_k\}$*. Prove que a cada* $\omega \in E_{i_1} \cap \cdots \cap E_{i_k}$ *corresponde um* σ *como acima e vice-versa. Isto permite contar os elementos da intersecção.]*

3. *Aplique a fórmula da Inclusão-Exclusão Generalizada aos* E_i*'s.*

2.3 Probabilidades condicionais

Suponha que temos um espaço de probabilidade $(\Omega, \mathbb{P})$ correspondendo por exemplo a uma carta de um baralho com disribuição uniforme. Estas cartas estão particionadas em quatro conjuntos correspondendo aos naipes e queremos saber se a carta $\omega \in \Omega$ escolhida é de copas . De início, tudo o que podemos dizer é que, se C é o evento copas,

$$\mathbb{P}(C) = \frac{1}{4}.$$

Equivalentemente, o "risco"atribuído a C é de 25%. Suponha, no entanto, que recebemos a informação de que a carta ω escolhida é preta, isto é, $\omega \in P$. Neste caso, é necessário atualizar nossa medida de risco: como todas as cartas de copas são vermelhas, devemos passar a atribuir risco 0 ao dado evento. Isto é, *cndicionado ao evento* P, o evento C em probabilidade 0. Por outro lado, se descobrimos que ω é vermelha ($\omega \in V = P^c$), então continua existindo um risco de que $\omega \in C$; como metade das cartas vermelhas pertence a V, parece natural dizer que *condicionado a* V, a probabilidade de $\omega \in C$ é $1/2$.

A probabilidade condicional pode ser vista como uma formalização da idéia de que probabilidades devem ser atualizadas cada vez que informação nova é recebida. Mostraremos mais adiante que as fórmulas abaixo representam uma atualização ótima das medidas de risco de acordo com uma certa medida natural de qualidade. Segue que as fórmulas abaixo não são arbitrárias; pelo contrário, sua escolha é bem motivada.

Definição 2.10. *Seja* $(\Omega, \mathbb{P})$ *um espaço de probabilidade discreto e* $A \subset \Omega$ *um evento com* $\mathbb{P}(A) > 0$*. A probabilidade condicional de* $\omega \in \Omega$ *dado* A *é definida pela fórmula.*

$$\mathbb{P}(\omega \mid A) = \begin{cases} \frac{\mathbb{P}(\omega)}{\mathbb{P}(A)}, & \omega \in A; \\ 0, & \omega \in A^c. \end{cases} \tag{2.4}$$

Observação 2.11. *Note que para todo evento* B*,* $\mathbb{P}(B \mid A) = \mathbb{P}(B \cup A)/\mathbb{P}(A)$ *[Exercício]. Ocasionalmente falaremos de* $\mathbb{P}(B \mid A)$ *para* $\mathbb{P}(A) = 0$*; neste caso, a probabilidade condicional pode ser definida de maneira arbitrária, pois seu valor quase nunca fará diferença.*

Exercício 2.12. *Formalize o problema do baralho descrito acima e mostre que* $\mathbb{P}(C \mid P) = 0$*,* $\mathbb{P}(C \mid V) = 1/2$*.*

Exercício 2.13 (Regra da probabilidade total.)**.** *Se* $A_1, A_2, \ldots$ *é uma partição de* Ω*,*

$$\forall B \subset \Omega, \, \mathbb{P}(B) = \sum_i \mathbb{P}(B \mid A_i).$$

Exercício 2.14 (Regra de Bayes.)**.** *Se* $\mathbb{P}(A), \mathbb{P}(B) > 0$,

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

Exercício 2.15. *Considere* $\Omega = [n] = \{1, \dots, n\}$ *com a medida uniforme. Suponha que* n *é divisível por 4. Seja* $P \subset [n]$ *o sub-conjunto dos pares e* $I = P^c$ *o sub-conjunto dos ímpares e* Q *o sub-conjunto dos números divisíveis por* 4*. Calcule* $\mathbb{P}(Q), \mathbb{P}(Q \mid P), \mathbb{P}(Q \mid I), \mathbb{P}(P \mid Q)$.

Exercício 2.16 (Falta de memória da distribuição geométrica)**.** *Consideramos agora* $(\Omega, \mathbb{P}) = (\mathbb{N}, Geo_p)$ *como no Exemplo 2.7. Considere um evento* $M_k \equiv \{k, k+1, k+2, \dots\}$*. Mostre que a distribuição condicional de* $\mathbb{P}(\cdot \mid M_k)$ *é dada por*

$$\mathbb{P}(i + k - 1 \mid M_k) = p(1-p)^{i-1}, i \in \mathbb{N} \text{ e } \mathbb{P}(j \mid M_k) = 0, j < k.$$

Em particular, mostre que a meia-vida da distribuição condicional é $k - 1 + H$*, onde* H *é a meia-vida de* Geo_p*. Intuitivamente, isto quer dizer que se o átomo não decaiu até o tempo* k*, o tempo que falta para o decaimento tem a mesma distribuição que tinha originalmente: o átomo* não se lembra de quanto tempo já passou.

Em muitos casos usa-se probabilidades condicionais para definir uma medida $\mathbb{P}$ implicitamente. Abaixo vemos alguns exemplos.

Exemplo 2.12. *Tem-se um saco com* n *moedas. Uma moeda é escolhida aleatoriamente e joga-se cara/coroa com ela, obtendo* 1 *ou* 0*. Nosso espaço amostral será dado por* $\Omega = [n] \times \{0, 1\}$*, correspondendo ao par moeda/resultado, e cada elemento de* Ω *é um par* $\omega = (k, b)$*.*

Considere os eventos $E_k = \{k\} \times \{0, 1\}$ *correspondentes á escolha da* k*-ésima moeda. Sejam* $F_b = [n] \times \{b\}$ *os eventos correspondentes ao valor cara/coroa. Nossa regra para definir probabilidades em* Ω *é a seguinte.*

1. $\mathbb{P}(E_k) = 1/n$ *para cada* $k \in [n]$ *(ou seja, as moedas são equiprováveis);*
2. $\mathbb{P}(F_1 \mid E_k) = 1 - \mathbb{P}(F_0 \mid K = k) = p_k$*, onde* $p_k \in (0, 1)$ *(a* k*-ésima moeda tem probabilidade* p_k *de dar cara).*

Isto define unicamente uma medida sobre Ω *dada por*

$$\mathbb{P}((k, b)) = \frac{p_k b + (1 - p_k)(1 - b)}{n}.$$

[Exercício.]

Exercício 2.17. *Suponha que* $p_1 > \cdots > p_n$*. Calcule* $\mathbb{P}(F_b)$ *e* $\mathbb{P}(E_k \mid F_b)$ *e mostre que*

$$\mathbb{P}(E_k \mid F_1) \text{ decresce com } k.$$

Ou seja: se o resultado do lançamento é cara, as moedas com proabilidade alta de cara são as mais prováveis (segundo a probabilidade condicional).

Exemplo 2.13. *Voltamos ao cenário do Exemplo 2.7. Agora temos dois tipos de átomo e observamos o decaimento de um deles. Formalmente,*

$$\Omega = \{0, 1\} \times \mathbb{N}$$

correspondendo a pares "(átomo,tempo do átomo)". Se $A_b = \{b\} \times \mathbb{N}$ *e* $D_k \equiv \{0, 1\} \times \{k\}$*, definimos:*

1. *$\mathbb{P}(A_i) = 1/2$ (os átomos são equiprováveis);*

2. *$\mathbb{P}(D_k \mid A_i) = p_i(1-p_i)^{k-1}$, onde $p_i \in (0,1)$ (o decamento do i-ésimo átomo tem distribuição Geo_{p_i}).*

Isto também define uma probabilidade sobre Ω [Exercício].

Exercício 2.18. *Calcule $\mathbb{P}(D_k)$ e $\mathbb{P}(A_i \mid D_k)$. Se $p_1 = 1/2$ e $p_0 = 1/3$, para quais k a probabilidade de A_0 condicionada a D_k é maior (isto é, quando o átomo 0 tem probabilidade condicional maior)?*

[Outros exercícios: os das seções 1.1 e 1.2 do Barry James que não envolvem explicitamente σ-álgebras ou conjuntos não discretos.]

2.4 Partições e probabilidades condicionais

Acima falamos que probabilidades condicionais podem ser vistas como uma forma geral de atualizar a medida de risco de acordo com alguma informação nova recebida. De modo geral, receber informação sobre $\omega \in \Omega$ significa saber que ω está em algum subconjunto $A \subset \Omega$. Se $\mathcal{F} = \{A_1, \ldots, A_n\}$ é uma partição de Ω, podemos imaginar que a informação recebida é $\mathcal{F}(\omega) = A_i$ a que ω pertence. Isto leva a uma definição de probabilidade condicionada a $\mathcal{F}$ como uma função.

Definição 2.14. *Se Ω é um conjunto com partição $\mathcal{F}$ e $\mathbb{P}$ é uma distribuição sobre Ω,*

$$\mathbb{P}(\cdot \mid \mathcal{F}) : (\omega, B) \in \Omega \times \mathcal{F} \mapsto \mathbb{P}(B \mid \mathcal{F}(\omega))$$

é a probabilidade condicional sobre $\mathcal{F}$. Aqui $\mathcal{F}(\omega)$ é o (único) elemento de $\mathcal{F}$ a que ω pertence.

A função $\mathbb{P}(\cdot \mid \mathcal{F})$ "condensa"todas as probabilidades condicionais $\mathbb{P}(B \mid A_i)$ em um único objeto. Veremos mais adiante de que forma isto é útil.

Exercício 2.19. *Reformule a regra no Exercício 2.13 como $\mathbb{P}(B) = \sum_\omega \mathbb{P}(\omega)\mathbb{P}(B \mid \mathcal{F}(\omega))$.*

Exercício 2.20. *Se $\mathcal{F} = \{A, A^c\}$, $\mathbb{P}(B \mid \mathcal{F})(\omega) = \mathbb{P}(B \mid A)$ se $\omega \in A$ e $\mathbb{P}(B \mid A^c)$ em caso contrário.*

2.5 Independência

Se probabilidades condicionais representam atualizações na avaliação de risco, independência significa que saber se A ocorreu *não altera a avaliação de risco de B*. Isto sugere que A e B são independentes quando $\mathbb{P}(B \mid A) = \mathbb{P}(B)$, o que equivale pela regra de Bayes a $\mathbb{P}(A \mid B) = \mathbb{P}(A)$ quando $\mathbb{P}(B), \mathbb{P}(A) > 0$. Para evitar esta úlima condição, toma-se em geral uma definição ligeiramente diferente (porém equivalente).

Definição 2.15. *Dizemos que eventos A, B são independentes quanto $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$. Em geral, $n \geq 2$ eventos $A_1, \ldots, A_n$ são ditos independentes se para todas as escolhas de $1 \leq i_1 < i_2 < \cdots < i_k \leq n$,*

$$\mathbb{P}(A_{i_1} \cap \cdots \cap A_{i_k}) = \mathbb{P}(A_{i_1}) \ldots \mathbb{P}(A_{i_k}).$$

Exercício 2.21. *$A_1, \ldots, A_n$ são independentes sse $B_1, \ldots, B_n$ o são, onde cada B_i é A_i ou A_i^c.*

Uma definição um pouco mais geral é dada por

Definição 2.16. *Dizemos que $n \geq 2$ partições $\mathcal{F}_1, \ldots, \mathcal{F}_n$ são independentes se para toda escolha de $F_i \in \mathcal{F}_i$,*

$$\mathbb{P}(F_1 \cap \cdots \cap F_n) = \mathbb{P}(F_1) \ldots \mathbb{P}(F_n).$$

Exercício 2.22. *Se cada $\mathcal{F}_i = \{A_i, A_i^c\}$, a definição acima equivale à independência de $A_1, \ldots, A_n$.*

Exercício 2.23. *Qualquer subconjunto de uma família de eventos/partições independentes é ele próprio independente.*

Exercício 2.24. *$\mathcal{F}_1$ e $\mathcal{F}_2$ são independentes sse $\mathbb{P}(F_2 \mid \mathcal{F}_1)(\omega) \equiv \mathbb{P}(F_2)$ para todo $F_2 \in \mathcal{F}_2$ e $\omega \in \Omega$ com $\mathbb{P}(\omega) > 0$.*

Exercício 2.25. *Os $\mathcal{F}_i$ são independentes sse o seguinte ocorre: sempre que A_i é a união de eventos em $\mathcal{F}_i$ para cada i, então $A_1, \ldots, A_n$ são independentes. Mostre que isto implica que se $\mathcal{G}_1$ é outra partição de Ω e $\mathcal{F}_1$ refina $\mathcal{G}_1$*[3]*, então $\mathcal{G}_1, \mathcal{F}_2, \ldots, \mathcal{F}_n$ são independentes. [Dica: cada $G \in sG_1$ é a união de elementos de $\mathcal{F}_1$.]*

Exercício 2.26. *Seja $\{\mathcal{F}_{i,j} : 1 \leq i \leq n, 1 \leq j \leq m_i\}$ uma família de partições independentes. Defina*

$$\wedge_{j=1}^{m_i} F_{i,j} \equiv \{F_{i,1} \cap \ldots F_{i,m_i} : \forall 1 \leq j \leq m_i \, F_{i,j} \in \mathcal{F}_{i,j}\} \;\; (1 \leq i \leq n).$$

Mostre que as $\wedge_{j=1}^{m_i} F_{i,j}$'s também são partições independentes.

O exemplo mais simples de partições independentes é o dado por espaços produto (Exercício 2.3). Seja $\Omega = \Omega_1 \times \ldots \Omega_n$ com uma medida produto $\mathbb{P}$. Para cada $1 \leq i \leq n$, considere a partição $\mathcal{F}_i$ de Ω onde elementos são separados pela i-ésima coordenada.

$$\mathcal{F}_i \equiv \{F_{i,\eta_i} \equiv \{\omega = (\omega_j)_{j=1}^n \in \Omega : \omega_i = \eta_i\} : \eta_i \in \Omega_i\}.$$

As partições assim construídas são independentes [Exercício]. Grosso modo, qualquer outra família de partições independentes tem comportamento semelhante a este exemplo. Ressaltamos, no entanto, que nem sempre eventos independentes vêm de espaços produto. Veja por exemplo o exercício a seguir.

Exercício 2.27. *Tome $\Omega = [n]$ com medida $\mathbb{P} = Unif_{[n]}$. Determine os valores de n para os quais os seguintes eventos são independentes:*

$$P = \{k \in [n] : k \text{ par}\},$$

$$M = \{k \in [n] : k \geq n/2\}.$$

O exercício seguinte faz outra ressalva importante.

[3] Isto é, todo $F \in \mathcal{F}_1$ está contido em algum $G \in \mathcal{G}_1$.

Exercício 2.28. *Sejam* $\Omega = \Omega_1 \times \Omega_2 \times \Omega_3$ *com* $\Omega_i = \{0, 1\}$. *Seja* $\mathbb{P}$ *dada por*

$$\mathbb{P}\left((\omega_1, \omega_2, \omega_3)\right) = \begin{cases} \frac{1}{4}, & \omega_3 = \omega_1 + \omega_2 \mod 2; \\ 0 & \omega_3 \neq \omega_1 + \omega_2 \mod 2 \end{cases}$$

Cheque que isto é de fato uma medida de probabilidade. Agora construa $\mathcal{F}_1, \mathcal{F}_2, \mathcal{F}_3$ *como no caso de espaços produto. Prove que* $\mathcal{F}_1, \mathcal{F}_2, \mathcal{F}_3$ não são independentes, *mas que qualquer par delas é. Isto mostra que a independência de três eventos não é conseqüência da independência dois-a-dois.*

Exercício 2.29. *Sejam* $A_1, A_2 \subset \Omega$ *independentes com probabilidades* p_1, p_2 *(resp.). Prove que* $\mathbb{P}\left(A_1 \cup A_2\right) = p_1 + p_2 - p_1 p_2$. *Generalize este resultado via Inclusão-Exclusão generalizada para* $A_1, \ldots, A_n$ *independentes.*

Exercício 2.30. *Seja* $\Omega = \{0, 1\}^n$ *com a medida produto* Be_p *(Exercício 2.4). Use o exercício acima para calcular a probabilidade dos* $\omega \in \Omega$ *com exatamente uma coordenada igual a* 1. *[Resposta:* $np(1 - p)^{n-1}$.*]*

Capítulo 3: Variáveis aleatórias

3.1 Definição

Quase todos os problemas interessantes em Probabilidade envolvem o conceito de variável aleatória.

Definição 3.1. *Seja* $(\Omega, \mathbb{P})$ *um espaço de probabilidade discreto. Uma função* $X : \Omega \to \Theta$ *(onde* Θ *é algum outro conjunto) é chamada de* variável aleatória *(ou* v.a.*).*

Intuitivamente[1], uma variável aleatória corresponde a algum tipo de *informação* obtido ou desejado sobre o elemento $\omega \in \Omega$. Se por exemplo $(\Omega, \mathbb{P})$ corresponde aos valores de ações numa bolsa de valores – ou seja, cada $\omega \in \Omega$ é um vetor com preços de ações diferentes em momentos diferentes – $X = X(\omega)$ pode ser o preço das ações de uma dada empresa ao final do pregão de um dia fixo. Podemos imaginar duas situações: ou queremos *estimar* X, ou pretendemos usar X como base para estimar uma outra quantidade $Y = Y(\omega)$ (por exemplo, o preço da mesma ação em outro dia).

A maior parte dos exemplos de variáveis aleatórias que consideraremos terá valores em $\mathbb{N}$, $\mathbb{Z}$, $\mathbb{R}$ ou $\mathbb{R}^d$; neste último caso, elas também serão chamadas de *vetores aleatórios*.

3.2 Distribuição de uma variável aleatória

Note que a definição de v.a. não envolve a medida de probabilidade $\mathbb{P}$. A probabilidade entra em cena quando percebemos que $\mathbb{P}$ e X induzem uma medida de probabilidade no contradomínio Θ.

Definição 3.2. *Sejam* $(\Omega, \mathbb{P})$ *e* X *como acima. Considere o conjunto (enumerável) dado por*

$$X(\Omega) \equiv \{X(\omega) \,:\, \omega \in \Omega\} \subset \Theta.$$

A distribuição de X *é a probabilidade* $\mathbb{P}_X$ *sobre* $X(\Omega)$ *dada por*

$$\mathbb{P}_X(\theta) \equiv \mathbb{P}\left(X^{-1}(\theta)\right) \quad (\theta \in \Theta).$$

Os conjuntos $X^{-1}(\theta)$ *e* $X^{-1}(A)$ *(*$A \subset X(\Omega)$*) são normalmente representados pelas expressões* $\{X = \theta\}$ *e* $\{X \in A\}$*, respectivamente. A probabilidade* $\mathbb{P}_X$ *pode ser estendida para todo* $E \subset \Theta$ *pela fórmula*

$$\mathbb{P}_X(E) = \mathbb{P}(X \in E) \equiv \mathbb{P}(X \in X(\Omega) \cap E) = \sum_{\omega : X(\omega) \in E} \mathbb{P}(\omega).$$

A última parte da definição é um abuso da definição de proabilidade discreta, já que Θ pode não ser enumerável. No entanto, como $X(\Omega)$ é sempre enumerá vel, isto não causará problemas.

[1] Esta intuição será esclarecida na Seção **??**

Exercício 3.1. *Seja Θ um conjunto finito ou enumerável e μ uma medida de probabilidade sobre Θ. Prove que existe uma v.a. $X : \Omega \to \Theta$ e uma distribuição $\mathbb{P}$ sobre Ω tal que $\mathbb{P}_X = \mu$. [Dica: o exercício é trivial!]*

Observação 3.3. *Podemos definir distribuições condicionais: $\mathbb{P}_Y(y \mid A) = \mathbb{P}(Y = y \mid A)$.*

Exercício 3.2 (Falta de memória de v.a.'s geométricas)**.** *Reformule o Exercício 2.16 da seguinte maneira: se X tem distribuição Geo_p, então para todo k a distribuição de $X - k + 1$ condicionada a $X \geq k$ também é Geo_p. Em outras palavras:*

$$\mathbb{P}(X = x + k - 1 \mid X \geq k) = p(1-p)^{x-1}.$$

3.3 Distribuições novas a partir de antigas

Apresentaremos agora alguns resultados gerais e específicos sobre distribuições de v.a.'s. Como em muitos problemas trataremos de variáveis aleatórias definidas a partir de outras v.a.'s, é conveniente começar por um resultado para esta situação.

Exercício 3.3. *Seja $X : \Omega \to \Theta_1$ uma v.a. e $f : \Theta_1 \to \Theta_2$ uma função. Defina a composição $f(X) \equiv f \circ X$. Mostre que $f(X)$ é uma v.a. e que*

$$\forall A \subset \Theta_2,\ \mathbb{P}_{f(X)}(A) = \mathbb{P}_X\left(f^{-1}(A)\right).$$

Um caso particular especialmente importante é o de soma de variáveis aleatórias.

Exercício 3.4 (Soma de variáveis aleatórias)**.** *Seja $X = (X_1, X_2, \ldots, X_n)$ uma v.a. com valores em $\mathbb{R}^d$ e defina $f(x_1, x_2, \ldots, x_n) = \sum_{i=1}^n x_i$. Mostre que $S_n \equiv f(X) = X_1 + X_2 + \cdots + X_n$ e deduza do exercício anterior que*

$$\mathbb{P}(S_n = z) = \mathbb{P}\left(\bigcup_{(z_1,z_2,\ldots,z_n)\in X(\Omega)\,:\,z_1+\cdots+z_n=z} (\{X_1 = z_1\} \cap \{X_2 = z_2\} \cap \cdots \cap \{X_n = z_n\})\right).$$

Mostre que os eventos nesta união são disjuntos e deduza:

$$\mathbb{P}(S_n = z) = \sum_{(z_1,z_2,\ldots,z_n)\in X(\Omega)\,:\,z_1+\cdots+z_n=z} \mathbb{P}\left(\bigcap_{i=1}^n \{X_i = z_i\}\right).$$

Os próximos exercícios consideram outras situações que não as de soma.

Exercício 3.5. *Suponha que $(\Omega, \mathbb{P})$ é um espaço finito com $\mathbb{P}$ uniforme (Exemplo 2.6). Seja $X : \Omega \to \Theta$ tal que para um certo inteiro s, todo $\theta \in \Theta$ tem exatamente s pré-imagens por X. Prove que $\mathbb{P}_X = Unif_\Theta$.*

Exercício 3.6. *Seja $[n] = \{1, \ldots, n\}$. Definamos a quantidade $\binom{n}{k}$ como o número de subconjuntos de $[n]$ com cardinalidade k. Definamos também $n!$ como o número de permutações de $[n]$. Nosso objetivo é provar* sem usar as fórmulas de $\binom{n}{k}$ e $n!$ *que*

$$\forall 0 \leq k \leq n,\ \binom{n}{k} = \frac{n!}{(n-k)!k!}.$$

Para isso devemos seguir os seguintes passos.

1. *Seja* $\Omega = S_n$ *o conjunto de permutações de* $[n]$ *e* $\Theta = \{S \in \mathcal{P}([n]) : |S| = k\}$.
2. *Ponha medida* $\mathbb{P} = Unif_{S_n}$ *sobre* S_n.
3. *Defina a v.a.* $X : \pi \in S_n \mapsto \{\pi(1), \ldots, \pi(k)\}$.
4. *Fixe* $S \in \Theta$. *Escolha* $\pi \in X^{-1}(S)$. *Prove que qualquer outro* $\eta \in X^{-1}(S)$ *é da forma*

$$\eta = \sigma_1 \circ \sigma_2 \circ \pi,$$

   onde σ_1 *é permutação de* $[k]$ *(estendida a* $i > k$ *via* $\sigma_1(i) = i$*) e* σ_2 *é permutação de* $[n]\backslash[k]$ *(estendida a* $j \leq k$ *via* $\sigma_2(j) = j$*).*

5. *Prove agora que para cada par* (σ_1, σ_2) *deste tipo há exatamente um* η *como acima. Mostre que há* $k!(n-k)!$ *destes pares e deduza que*

$$\mathbb{P}(X = S) = \frac{k!(n-k)!}{n!}.$$

6. *Por outro lado, use o exercício anterior para mostrar que* $\mathbb{P}(X = S) = 1/|\Theta|$ *e conclua a prova.*

Exercício 3.7. *Dados* $1 \leq k \leq n$, *considere* $\Omega = [n] \times [n-1] \times \cdots \times [n-k+2] \times [n-k+1]$ *com a medida uniforme. Defina* $X_1(\omega) = \omega_1$ *e para cada* $2 \leq i \leq k$:

$$X_i(\omega) = \text{o } \omega_i\text{-ésimo elemento de } [n]\backslash\{X_1(\omega), \ldots, X_{i-1}(\omega)\}.$$

Tome $\mathbb{P} = Unif_\Omega$, *considere a "função aleatória"* $X : \Omega \to [n]^{[k]}$ *que para cada* $\omega \in \Omega$ *é dada por*

$$\begin{array}{cccc} X(\omega) & : & [k] & \to & [n] \\ & & t & \mapsto & X_t(\omega) \end{array}.$$

Mostre que X *tem a distribuição de "k bolas retiradas sem reposição de uma urna com n bolas"(Exemplo 2.9), isto é,* X *é uniformemente distribuída sobre as funções injetivas* $[n]^{[k]}_{inj}$. *[Dica: use o Exercício 2.6 para mostrar que* $\mathbb{P}$ *é uma distribuição produto.]*

Exercício 3.8. *Considere novamente "k bolas retiradas sem reposição de uma urna com n bolas"(Exemplo 2.9), isto é:* $\Omega = [n]^{[k]}_{inj}$ *e* $\mathbb{P} = Unif_\Omega$. *Seja* $S \subset [k]$ *um conjunto com s elementos e tome* X *como a restrição a* S:

$$X : \omega \in [n]^{[k]}_{inj} \mapsto'' \omega\,|_S : i \in S \mapsto \omega(i)".$$

Mostre que $\mathbb{P}_X = Unif_{[n]^S_{inj}}$. *Este fato tem a seguinte interpretação: se só olhamos para s das k bolas retiradas , a distribuição observada é a mesma de s bolas tiradas sem reposição de uma urna com n bolas.*

3.4 Independência

Grosso modo, eventos são independentes quando qualquer subgrupo deles não dá informação alguma a respeito dos eventos restantes. A definição de independência de variáveis aleatórias significa algo semelhante e de fato é equivalente à independência das partições correspodentes a cada v.a., conforme o exercício abaixo.

Exercício 3.9. *Cada v.a.* $X : \Omega \to \Theta$ *gera uma partição* $\mathcal{F}_X$ *de* Ω*:*

$$\mathcal{F}_X = \{X^{-1}(x) \,:\, x \in X(\Omega)\}.$$

Por outro lado, se $\mathcal{F}$ *é uma partição, existe uma v.a.* X *com* $\mathcal{F}_X = \mathcal{F}$*.*

Definição 3.4. *Sejam* $X_i : \Omega \to \Theta_i$, $1 \leq i \leq n$ *v.a.'s. Elas são* independentes *se alguma das seguintes condições equivalentes é satisfeita:*

1. *A distribuição do vetor* $(X_1, \ldots, X_n)$ *sobre* $\Theta = \times_{i=1}^n X_i(\Omega_i)$ *é uma medida produto.*

2. *Para todos* $x_i \in X_i(\Omega_i)$*,*

$$\mathbb{P}_{(X_1,\ldots,X_n)}\left((x_1, \ldots, x_n)\right) = \mathbb{P}\left(\cap_i \{X_i = x_i\}\right) = \prod_{i=1}^n \mathbb{P}_{X_i}\left(x_i\right) = \prod_{i=1}^n \mathbb{P}\left(X_i = x_i\right);$$

3. *Para todos* $A_i \in \mathcal{P}(X_i(\Omega_i))$*,*

$$\mathbb{P}_{(X_1,\ldots,X_n)}\left(A_1 \times \cdots \times A_n\right) = \prod_{i=1}^n \mathbb{P}_{X_i}\left(A_i\right);$$

4. *As partições* $\mathcal{F}_{X_i}$ *são independentes.*

Exercício 3.10. *Prove a equivaência.*

Os dois próximos exercícios podem ou ser resolvidos diretamente, ou via os resultados sobre partições independentes (p. ex. Exercício 2.25,Exercício 2.26).

Exercício 3.11 (Agrupar v.a.'s não destrói a independência)**.** *Sejam* $\{X_{i,j} \,:\, 1 \leq i \leq n, 1 \leq j \leq m_i\}$ *v.a.'s independentes. Considere os vetores* $Y_i = (X_{i,j})_{j=1}^{m_i}$*. Mostre que eles também são independentes. [Dica/obs: na notação do Exercício 2.26,* $\mathcal{F}_{Y_i} = \wedge_{j=1}^{m_i} \mathcal{F}_{X_{i,j}}$*.]*

Exercício 3.12 (Tomar funções das v.a.'s não destrói a independência)**.** *Sejam* $X_i : \Omega \to \Theta_i$*,* $1 \leq i \leq n$ *v.a.'s independentes e* f_i *funções definidas nos espaços apropriados. Então as v.a.'s* $Y_i = f_i(X_i)$ *são independentes.[Dica/obs: na notação do Exercício 2.25, cada* $\mathcal{F}_{Y_i}$ *é refinada por* $\mathcal{F}_{X_i}$*.]*

3.5 Somas de variáveis aleatórias independentes

Exercício 3.13 (Somas de v.a.'s produto; convoluçoes discretas)**.** *Se as* $X_i : \Omega \to \mathbb{R}^d$ *(*$1 \leq i \leq n$*) são independentes,*

$$\forall x = (x_1, \ldots, x_n) \in (\mathbb{R}^d)^n \; \mathbb{P}_{(X_1,\ldots,X_n)}\left(x\right) = \prod_{i=1}^n \mathbb{P}_{X_i}\left(x_i\right).$$

Aplique este resultado junto com o Exercício 3.4para mostrar que neste caso:

$$\mathbb{P}\left(S_n = z\right) = \sum_{(z_1,z_2,\ldots,z_n) \in \,:\, z_1 + \cdots + z_n = z} \mathbb{P}\left(X_1 = z_1\right) \mathbb{P}\left(X_2 = z_2\right) \ldots \mathbb{P}\left(X_n = z_n\right).$$

Em particular, se $X_1, X_2, \ldots, X_n$ tomam valores em $\mathbb{Z}$:

$$\begin{aligned}\forall z \in \mathbb{Z},\, \mathbb{P}(S_n = z) &= \sum_{(z_1, z_2, \ldots, z_{n-1}) \in \mathbb{Z}^{n-1}} (\mathbb{P}(X_1 = z_1) \times \mathbb{P}(X_2 = z_2) \times \cdots \times \\ &\quad \times \mathbb{P}(X_{n-1} = z_{n-1}) \times \mathbb{P}(Z_n = z - z_1 - z_2 - \cdots - z_{n-1})).\end{aligned}$$

No caso $n = 2$, a operação que leva os vetores infinitos $(\mathbb{P}(X_i = z_i))_{z_i \in \mathbb{Z}}$ em $(\mathbb{P}(S_2 = z))_{z \in \mathbb{Z}}$ é chamada de convolução discreta. Para $n > 2$, temos convoluç oes iteradas.

Exercício 3.14 (A soma de Bernoullis produto é Binomial)**.** *Considere o espaço-produto de Be_p's discutido no Exercício 2.4: isto é, $\Omega = \{0,1\}^n$ e $\mathbb{P}(\omega) = p^{|\omega|}(1-p)^{n-|\omega|}$, onde $|\omega| = \sum_i \omega_i$. Defina $\Pi_i : \omega \mapsto \omega_i$ como a função que leva ω na sua i-ésima coordenada. Note que a soma das Π_i's é*

$$S(\omega) = \sum_{i=1}^{n} \Pi_i(\omega) = |\omega|.$$

Use o resultado anterior para mostrar que

$$\mathbb{P}_S(z) = \mathbb{P}(S = z) = \begin{cases} \binom{n}{k} p^k (1-p)^{n-k}, & k \in \{0, \ldots, n\}; \\ 0 & \textit{caso contrário.} \end{cases}$$

Isto é, S tem distribuição $Bin_{n,p}$ (Exemplo 2.8). [Obs: de que forma isto elucida o Exercício 2.8.]

Exercício 3.15 (A soma de Poissons produto é Poisson)**.** *Considere $\Omega_1, \Omega_2, \ldots, \Omega_n = \mathbb{N}$ onde cada Ω_i recebe medida Po_{λ_i}. Construa o produto $(\Omega, \mathbb{P})$, defina Π_i como no exercício anterior e considere $S_j = \sum_{i \leq j} \Pi_i$. Mostre por indução que cada S_j tem distribuição $Po_{\sum_{i \leq j} \lambda_i}$.*

Capítulo 4: Valores esperados, momentos e desigualdades

4.1 Valores esperados e momentos

Nesta seção definiremos o valor esperado de uma variável aleatória com valores reais e algumas outras quantidades daí derivadas. Primeiro começamos com v.a.'s especialmente simples.

Definição 4.1. *Seja $X : \Omega \to \mathbb{R}$ uma v.a. . Dizemos que X é a* função indicadora *(ou* característica*) de $A \subset \Omega$ se $X(\omega) = 1$ quando $\omega \in A$ e $X(\omega) = 0$ quando $\omega \in A^c$. Neste caso escreveremos X como $X = \mathbb{I}_A$.*

Exercício 4.1. *Se $X : \Omega \to \mathbb{R}$, $\mathbb{P}X = Be_p$ para algum $p \in [0,1]$ (cf. Exemplo 2.4) se e somente se existe $A \subset \Omega$ com $\mathbb{P}(X = \mathbb{I}_A) = 1$. Neste caso, $p = \mathbb{P}(A)$.*

Exercício 4.2. *Mostre que toda $X : \Omega \to \mathbb{R} \cup \{\pm\infty\}$ é uma combinação linear de funções simples. Mais exatamente,*

$$X = \sum_{x \in X(\Omega)} x \, \mathbb{I}_{X^{-1}(x)}.$$

A definição de valor esperado $\mathbb{E}[X]$ pode ser escrita da seguinte maneira: se $X = \mathbb{I}_A$, $\mathbb{E}[X] = \mathbb{P}(A)$; para outras X, a definição se estende por linearidade:

$$X = \sum_{j=1}^{+\infty} c_j \mathbb{I}_{A_j} \Rightarrow \mathbb{E}[X] = \sum_{j=1}^{+\infty} c_j \mathbb{P}(A_j).$$

No entanto, esta extensão pode apresentar problemas de convergência quando Ω é infinito. Evitamos esta dificuldade com uma definição em duas partes. Começamos com v.a.'s com valores não-negativos.

Definição 4.2. *Seja $X : \Omega \to [0, +\infty]$ uma v.a. com valores não negativos. O* valor esperado *(ou* esperança*) de X, simbolizado por $\mathbb{E}[X]$, é dado por*

$$\mathbb{E}[X] \equiv \sum_{\omega \in \Omega} X(\omega) \mathbb{P}(\omega),$$

com a convenção de que "$0.\infty = 0$". O valor esperado também pertence a $[0, +\infty]$. Ocasionalmente representaremos $\mathbb{E}[X]$ como uma integral:

$$\mathbb{E}[X] = \int_\Omega X \, d\mathbb{P}.$$

Observação 4.3. *Suponha que $\Omega = \{0,1\}$, $\mathbb{P} = Be_p$ (cf. Exemplo 2.4) e $X = +\infty . \mathbb{I}_{\{1\}}$, i.e. $X(0) = 0$ e $X(1) = +\infty$. Então $\mathbb{E}[X] = 0$ se $p = 0$ e $\mathbb{E}[X] = +\infty$ em caso contrário.*

Observação 4.4. *Se Ω é finito e $X \geq 0$, $\mathbb{E}[X] < +\infty$ se e somente se $\mathbb{P}(X = +\infty) = 0$. Quando Ω é infinito ainda é verdade que $\mathbb{E}[X] < +\infty \Rightarrow \mathbb{P}(X = +\infty) = 0$, mas a recíproca é falsa (ex: $\Omega = \mathbb{N}$, $\mathbb{P}_X = Geo_{1/2}$ como em Exemplo 2.7 e $X(\omega) = 3^\omega$ para $\omega \in \mathbb{N}$).*

Exercício 4.3. *Mostre que, na situação acima:*

$$\mathbb{E}[X] = \sum_{x\in X(\Omega)} x\,\mathbb{P}(X=x) = \sum_{x\in X(\Omega)} x\,\mathbb{P}_X(x).$$

Deduza que se $(\Theta, \mathbb{Q})$ *é outro espaço de probabilidade discreto e* $Y : \Theta \to [0, +\infty]$ *com* $\mathbb{Q}_Y = \mathbb{P}_X$ *(isto é,* X *e* Y *têm a mesma distribuição), então* $\mathbb{E}[X] = \mathbb{E}[Y]$.

Exercício 4.4. *Se* $X, Y : \Omega \to \mathbb{R} \cup \{\pm\infty\}$ *e* $X \leq Y$ *sempre,* $\mathbb{E}[X] \leq \mathbb{E}[Y]$.

Agora definiremos $\mathbb{E}[X]$ para X qualquer.

Definição 4.5. *Seja* $X : \Omega \to [-\infty, +\infty]$ *uma v.a. com valores reais (possivelmente divergentes).* X *é dita* integrável *se* $\mathbb{E}[|X|] < +\infty$ *segundo a Definição 4.2. Se* X *é integrável, o* valor esperado *(ou* esperança*) de* X*, simbolizado por* $\mathbb{E}[X]$*, é dado por*

$$\mathbb{E}[X] \equiv \mathbb{E}[\max\{X, 0\}] - \mathbb{E}[\max\{-X, 0\}],$$

onde os dois valores esperados do lado direito são definidos como antes.

Exercício 4.5. *Use o Exercício 4.4 para mostrar que* $\mathbb{E}[|X|] < +\infty$ *implica que* $\mathbb{E}[\max\{X, 0\}] < +\infty$*,* $\mathbb{E}[\max\{-X, 0\}] < +\infty$*. Mais ainda, mostre que neste caso*

$$\mathbb{E}[X] = \sum_{\omega\in\Omega} X(\omega)\mathbb{P}(\omega)$$

onde a série é absolutamente convergente. Por fim, prove que os fatos no Exercício 4.3 pernamecem válidos sob a definição geral sempre que X *for integrável (o que é equivalente a* Y *integrável).*

Exercício 4.6. *Se* X *toma valores em* $\mathbb{N} \cup \{0\}$*,* $\mathbb{E}[X] = \sum_{n=1}^{+\infty} \mathbb{P}(X \geq n)$.

Exercício 4.7. *Mostre que*

1. $\mathbb{P}_X = Be_p$ *(cf. Exemplo 2.4)* $\Rightarrow \mathbb{E}[X] = p$;
2. $\mathbb{P}_X = Geo_p$ *(cf. Exemplo 2.7)* $\Rightarrow \mathbb{E}[X] = 1/p$;
3. $\mathbb{P}_X = Bin_{n,p}$ *(cf. Exemplo 2.8)* $\Rightarrow \mathbb{E}[X] = pn$ *[Dica: Escreva*

   $$G(a, b) = (a + b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k}.$$

   Derivando termo a termo em a*, mostre que*

   $$\mathbb{E}[X] = a\frac{\partial G}{\partial a}(a, b)\Bigg]_{(a,b)=(p,1-p)}$$

   e calcule $\mathbb{E}[X]$ *daí.];*
4. $\mathbb{P}_X = Po_\lambda$ *(cf. Exercício 2.9)*$\Rightarrow \mathbb{E}[X] = \lambda$.

Exercício 4.8. *Seja $L_1 = L_1(\Omega, \mathbb{P})$ o espaço vetorial cujos elementos são as v.a.s integráveis $X : \Omega \to \mathbb{R}$. Mostre que $\mathbb{E}[\cdot]$ é um operador linear sobre este espaço. Isto é, se $\alpha \in \mathbb{R}$ e X, $Y \in L_1$, então $\alpha X + Y \in L_1$ e $\mathbb{E}[\alpha X + Y] = \alpha \mathbb{E}[X] + \mathbb{E}[Y]$. [Se $X, Y, \alpha \geq 0$, o requerimento de estar em L_1 pode ser eliminado.]*

Observação 4.6. *Todas as definições acima têm análogos para $X = (X_1, \ldots, X_d) : \Omega \to \mathbb{R}^d$. $\mathbb{E}[X]$ é o vetor cujas coordenadas são os $\mathbb{E}[X_i]$ (se estes valores estão definidos). Se $\mathbb{E}[|X_i|] < +\infty$ para cada i, dizemos que X é integrável.*

Definiremos agora os momentos de X.

Definição 4.7. *Para $p \in [0, +\infty)$ com $\mathbb{E}[|X|^q] < +\infty$, o q-ésimo momento de X é dado por $\mathbb{E}[X^q]$. Se $p \in [1, +\infty)$, a "norma" L_p de X é $\|X\|_p \equiv (\mathbb{E}[|X|^p])^{1/p}$.*

Exercício 4.9. *Mostre que $\|X\|_p = 0$ sse $\mathbb{P}(X = 0) = 1$.*

4.1.1 A desigualdade de Jansen e as normas L_p

Definição 4.8. *Seja $K \subset \mathbb{R}^d$ convexo. Dizemos que a função $\Psi : K \to \mathbb{R}$ é convexa se para todos $x, y \in K$ e $\alpha \in [0, 1]$:*

$$\Psi(\alpha x + (1-\alpha)y) \leq \alpha \Psi(x) + (1-\alpha)\Psi(y).$$

Lema 4.9 (Desigualdade de Jansen)**.** *Se $K \subset \mathbb{R}^d$ é convexo, $X : \Omega \to K$ é integrável e $\Psi : K \to \mathbb{R}$ é convexa e contínua,*

$$\Psi(\mathbb{E}[X]) \leq \mathbb{E}[\Psi(X)].$$

Prova: [Esboço.] O primeiro passo é provar que $\forall n \in \mathbb{N}$, $\forall x_1, \ldots, x_n \in K$ e $\forall \alpha_1, \ldots, \alpha_n \in [0, +\infty)$ com $\sum_{i=1}^n \alpha_i = 1$,

$$\Psi(\sum_{i=1}^n \alpha_i x_i) \leq \sum_{i=1}^n \alpha_i \Psi(x_i).$$

De fato, se tomamos $x = x_n$, $y = \frac{\sum_{i=1}^{n-1} \alpha_i x_i}{\sum_{i=1}^{n-1} \alpha_i}$ e $\alpha = \alpha_n$, deduzimos que

$$\begin{aligned} \Psi\left(\sum_{i=1}^n \alpha_i x_i\right) &= \Psi(\alpha x + (1-\alpha)y) \\ &\leq \alpha \Psi(x) + (1-\alpha)\Psi(y) \\ &= \alpha_n \Psi(x_n) + \left(\sum_{i=1}^{n-1} \alpha_i\right) \Psi\left(\frac{\sum_{i=1}^{n-1} \alpha_i x_i}{\sum_{i=1}^{n-1} \alpha_i}\right) \end{aligned}$$

e o resto do resultado segue por indução. Suponha agora que a imagem de $X : \Omega \to \mathbb{R}^d$ contém finitos pontos $x_1, \ldots, x_n$, isto é

$$X = \sum_{i=1}^n x_i \, \mathbb{I}_{\{X = x_i\}}.$$

Aplicando o resultado de convexidade acima com $\alpha_i = \mathbb{P}(X = x_i)$, deduzimos que

$$\begin{aligned} \Psi(\mathbb{E}[X]) &= \Psi(\sum_{i=1}^{n} x_i \, \mathbb{P}(X = x_i)) \\ &\leq \sum_{i=1}^{n} \mathbb{P}(X = x_i) \, \Psi(x_i) \\ &= \mathbb{E}[\Psi(X)]. \end{aligned}$$

Em geral, a imagem de X é um conjunto enumerável $\{x_i\}_{i=1}^{+\infty}$. Mas este caso é um limite do primeiro [Exercício.] $\square$

Exercício 4.10. *Aplicando Jansen a* $\Psi(x) = |x|$ *(norma euclideana), mostre que para toda* $X : \Omega \to \mathbb{R}^d$ $|\mathbb{E}[X]| \leq \mathbb{E}[|X|]$.

Exercício 4.11. *Sejam* $q > p > 0$. *Aplicando Jansen a* $\Psi(x) = x^{q/p}$ $(x \geq 0)$*, mostre que para toda* $X : \Omega \to \mathbb{R}$ $\|X\|_p \leq \|X\|_q$. *Em particular,* $\mathbb{E}\left[X^2\right] \geq (\mathbb{E}[|X|])^2 \geq (\mathbb{E}[X])^2$.

Exercício 4.12. *O exercício anterior mostra que* $\|X\|_p$ *é função crescente de* p. *Em particular, existe o limite*

$$\|X\|_\infty \equiv \lim_{p \to +\infty} \|X\|_p \in [0, +\infty].$$

Prove que

$$\|X\|_\infty = \sup\{|X(\omega)| \, : \, \omega \in \Omega, \mathbb{P}(\omega) > 0\}.$$

Observação 4.10. *A chamada* desigualdade de Hölder *implica que, se* X *e* Y *são v.a.'s sobre o mesmo espaço* Ω *e* $1 \leq p, q \leq +\infty$ *satisfazem* $p^{-1} + q^{-1} = 1$,

$$\|XY\|_1 \leq \|X\|_p \|Y\|_q.$$

Em particular, se $p = q = 2$ *temos a desigualdade de Cauchy-Schwartz:*

$$\|XY\|_1 \leq \|X\|_2 \|Y\|_2$$

.

4.2 Variância e covariância

Apresentamos agora duas quantidades essenciais: a variância e a covariância.

Definição 4.11. *Seja* $X : \Omega \to \mathbb{R}$ *dada com* X^2 *integrável (logo, pelo Exercício 4.11* X *é integrável). A quantidade*

$$\mathbb{V}(X) \equiv \mathbb{E}\left[(X - \mathbb{E}[X])^2\right]$$

é chamada a variância *de* X. *Ela pode ser equivalentemente escrita como* $\mathbb{V}(X) = \mathbb{E}\left[X^2\right] - (\mathbb{E}[X])^2$ *[Exercício.].*

Observação 4.12. *A variância é sempre não negativa (Exercício 4.11).* $\mathbb{V}(X) = 0$ *sse* $X = \mathbb{E}[X]$ *com probabilidade* 1 *(Exercício 4.9).*

Definição 4.13. *Sejam $X, Y : \Omega \to \mathbb{R}$ com X^2, Y^2 integráveis. A* covariância *de X e Y é a quantidade dada por:*

$$\mathbb{C}(X,Y) \equiv \mathbb{E}\left[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])\right].$$

Equivalentemente, $\mathbb{C}(X,Y) = \mathbb{E}[XY] - \mathbb{E}[X]\,\mathbb{E}[Y]$ [Exercício].

Observação 4.14. $\mathbb{V}(X) = \mathbb{C}(X,X)$.

Exercício 4.13. *Usando os resultados do Exercício 4.7, mostre que:*

1. $\mathbb{P}_X = Be_p$ *(cf. Exemplo 2.4)* $\Rightarrow \mathbb{V}(X) = p(1-p)$*;*
2. $\mathbb{P}_X = Geo_p$ *(cf. Exemplo 2.7)* $\Rightarrow \mathbb{V}(X) = 1/p^2 - 1/p$ *[Dica: calcule $\Delta = \mathbb{E}[X(X-1)]$ e determine $\mathbb{V}(X)$ a partir de Δ. Note que*

$$\begin{aligned}\Delta &= p\sum_{k=2}^{+\infty} k(k-1)(1-p)^{k-1} \\ &= p(1-p)\frac{d^2}{dp^2}\left(\sum_{k=0}^{+\infty}(1-p)^k\right) \\ &= p(1-p)\frac{d^2}{dp^2}\left(\frac{1}{p}\right) = \frac{2-2p}{p^2}.\end{aligned}$$

   ];
3. $\mathbb{P}_X = Bin_{n,p}$ *(cf. Exemplo 2.8)* $\Rightarrow \mathbb{V}(X) = p(1-p)n$ *[Dica: Como antes, é melhor calcular $\Delta = \mathbb{E}[X(X-1)]$ antes. Escreva*

$$G(a,b) = (a+b)^n = \sum_{k=0}^{n}\binom{n}{k}a^k b^{n-k}.$$

   Derivando termo a termo em a, mostre que

$$\mathbb{E}[X(X-1)] = a^2\frac{\partial^2 G}{\partial a^2}(a,b)\Big]_{(a,b)=(p,1-p)}$$

   e calcule $\Delta = \mathbb{E}[X(X-1)]$ daí.];
4. $\mathbb{P}_X = Po_\lambda$ *(cf. Exercício 2.9)*$\Rightarrow \mathbb{V}(X) = \lambda$ *[Dica: mais uma vez vale a pena começar calculando $\Delta = \mathbb{E}[X(X-1)]$. Neste caso,*

$$\Delta = \sum_{k\geq 0} e^{-\lambda}k(k-1)\frac{\lambda^k}{k!} = \lambda^2\sum_{k\geq 2} e^{-\lambda}\frac{\lambda^{k-2}}{(k-2)!}.$$

   .]

Exercício 4.14. $\mathbb{V}(X) = \mathbb{V}(X-c)$ *para qualquer $c \in \mathbb{R}$. Se $\mathbb{E}[X] = 0$ $\mathbb{V}(X) = \mathbb{E}\left[X^2\right]$. Do mesmo modo, $\mathbb{C}(X,Y) = \mathbb{C}(X - c_X, Y - c_Y)$ e $\mathbb{C}(X,Y) = \mathbb{E}[XY]$ se $\mathbb{E}[X] = \mathbb{E}[Y] = 0$*

4.3 A desigualdade de Chebyshev e concentração

Variâncias são freqüentemente mais fáceis de se calcular do que probabilidades exatas de eventos. A desigualdade abaixo mostra que em alguns casos importantes, pode-se estimar probabilidades a partir de variâncias:

Proposição 4.15 (Desigualdade de Chebyshev)**.** *Se* $\mathbb{E}\left[|X|^2\right] < +\infty$*,*

$$\forall \lambda > 0,\ \mathbb{P}\left(|X - \mathbb{E}\left[X\right]| \geq \lambda\right) \leq \frac{\mathbb{V}\left(X\right)}{\lambda^2}.$$

Prova: Suponha sem preda de generalidade que $\mathbb{V}\left(X\right) > 0$. Seja $A \equiv \{|X - \mathbb{E}\left[X\right]| \geq \lambda\}$. Se $\Psi : \mathbb{R} \to [0, +\infty)$ é uma função crescente com $\Psi(\lambda) > 0$, temos que

$$\forall \omega \in \Omega,\ |X(\omega) - \mathbb{E}\left[X\right]| \geq \lambda \Leftrightarrow \Psi(|X - \mathbb{E}\left[X\right]|) \geq \Psi(\lambda) \Leftrightarrow \frac{\Psi(|X - \mathbb{E}\left[X\right]|)}{\Psi(\lambda)} \geq 1.$$

Tomando $\Psi(x) = x^2$, podemos reescrever:

$$A \equiv \{\frac{(X(\omega) - \mathbb{E}\left[X\right])^2}{\lambda^2} \geq 1\}. \tag{4.1}$$

Mas então temos:

$$\forall \omega \in \Omega,\ \frac{(X(\omega) - \mathbb{E}\left[X\right])^2}{\lambda^2} \geq \mathbb{I}_A(\omega).$$

De fato, a desigualdade vale para $\omega \in A$ por conta de (4.1) e para $\omega \in A^c$ porque lado esquerdo é sempre ≥ 0. Tomando valores esperados, vemos que:

$$\frac{\mathbb{V}\left(X\right)}{\lambda^2} = \int \left(\frac{(X(\omega) - \mathbb{E}\left[X\right])^2}{\lambda^2}\right) d\mathbb{P}\left(\omega\right) \geq \int \mathbb{I}_A(\omega)\, d\mathbb{P}\left(\omega\right) = \mathbb{P}\left(A\right).$$

□

De que forma se utiliza este resultado? Consideremos o caso em que $\mathbb{P}_X = Bin_{n,p}$. Neste caso, vimos acima que $\mathbb{E}\left[X\right] = np$, $\mathbb{V}\left(X\right) = p(1-p)n$. Suponha que queremos estimar uma probabilidade do tipo

$$Bin_{n,p}(\{k\,:\,|k - np| \geq \epsilon np\}) = \mathbb{P}\left(|X - np| \geq \epsilon np\right).$$

Usando Chebyshev com $\lambda = \epsilon np$, temos

$$Bin_{n,p}(\{k\,:\,|k - np| \geq \epsilon np\}) \leq \frac{p(1-p)n}{\epsilon^2 n^2 p^2} \leq \frac{1}{\epsilon^2 pn}. \tag{4.2}$$

Equivalentemente,

$$Bin_{n,p}(\{k\,:\,|k - np| < \epsilon np\}) \geq 1 - (\epsilon^2 pn)^{-2}.$$

Isto quer dizer que se pn é "grande", a maior parte da "massa"da distribuição $Bin_{n,p}$ se concentra no intervalo $((1-\epsilon)np, (1+\epsilon)np)$. Em outras palavras, $Y_n \equiv X/np - 1$ está quase sempre no intervalo $(-\epsilon, \epsilon)$. Esta propriedade é um exemplo simples do que se chama de *concentração de medida*: a distribuição $\mathbb{P}_{Y_n}$ está quase toda concentrada num pequeno intervalo ao redor de um valor determinístico 0 cujo tamanho tende a 0 quando $n \to +\infty$. Um resultado semelhante vale sempre que $\mathbb{V}\left(X\right) \ll \mathbb{E}\left[X\right]^2$:

Definição 4.16 (Concentração). *Considere uma seqüencia de distribuições μ_n sobre $\mathbb{R}^d$*[1]. *Dizemos que $\{\mu_n\}$ se concentra em $c \in \mathbb{R}^d$ se para toda bola aberta B centrada em c temos*

$$\lim_{n\to+\infty} \mu_n(B) = 1.$$

Uma seqüência de v.a.'s se concentra em c quando suas distribuições se concentram em c.

Exercício 4.15. *Se $p_n \in [0,1]$ satisfaz $p_n n \to +\infty$, e $\mathbb{P}_{X_n} \equiv Bin_{n,p_n}$, então $Y_n = X_n/pn - 1$ se concentra em 0. Se $\lambda_n \to +\infty$ e $\mathbb{P}_{X_n} \equiv Po_{\lambda_n}$, $Y_n = X_n/\lambda_n$ se concentra em 1.*

Uma pergunta importante é: quando as condições de concentração acima descritas são satisfeitas? Um caso particular é dado por variáveis aleatórias *sem covariância*, que discutimos a seguir.

Definição 4.17. *$\{X_n : \Omega \to \mathbb{R}\}_{n=1}^{+\infty}$ são v.a.'s sem covariância se para todos $i, j \in \mathbb{N}$ distintos $\mathbb{C}(X_i, X_j) = 0$.*

Proposição 4.18. *Para quaisquer v.a.'s $X_1, \ldots, X_n$,*

$$\mathbb{V}\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n \mathbb{V}(X_i) + 2\sum_{1\leq i<j\leq n} \mathbb{C}(X_i, X_j).$$

Em particular, se as X_i's não têm covariância, a variância da soma é a soma das variâncias.

Prova: [Exercício.] $\square$

Teorema 4.19 (Lei fraca dos grandes números.). *Seja $\{X_n : \Omega \to \mathbb{R}\}_{n=1}^{+\infty}$ uma seqüência de v.a.'s sem covariância e cujas variâncias são limitadas por $\sigma^2 < +\infty$ e tais que Então as* médias empíricas centradas*:*

$$C_n \equiv \frac{\sum_{i=1}^n (X_i - \mathbb{E}[X_i])}{n}$$

se concentram ao redor de 0. De fato,

$$\mathbb{P}(|C_n| \geq \epsilon) \leq \frac{\sigma^2}{\epsilon^2 n}.$$

Prova: Basta aplicar a Desigualdade de Chebyshev a nC_n: como não há correlações entre os $(X_i - \mathbb{E}[X_i])$'s

$$\mathbb{V}(nC_n) = \sum_{i=1}^n \mathbb{V}((X_i - \mathbb{E}[X_i])) = \sum_{i=1}^n \mathbb{V}(X_i) \leq \sigma^2 n.$$

Logo

$$\mathbb{P}(|C_n| \geq \epsilon) = \mathbb{P}(|nC_n - \mathbb{E}[nC_n]| \geq n\epsilon) \leq \frac{\sigma^2 n}{\epsilon^2 n^2}$$

e a concentração segue do fato que o lado direito tende a 0 quando $n \to +\infty$ para todo ϵ fixo. $\square$

[1] Mais exatamente, existe $S_n \subset \mathbb{R}^d$ finito ou enumerável tal que μ_n é medida sobre S_n. Neste caso, estendemos μ_n a todo $A \subset \mathbb{R}^d$ como fizemos no caso de v.a.'s (Definição 3.2): $\mu_n(A) \equiv \sum_{\omega\in A\cap S_n} \mu_n(\omega)$.

Exercício 4.16. *No teorema acima, se* $\mu \equiv \lim_{n\to+\infty} \sum_{i=1}^{n} \mathbb{E}\left[X_i\right]/n$ *existe, então as médias* $F_n \equiv \sum_{i\leq n} X_i/n$ *se concentram ao redor de* μ.

No caso $X_i = \mathbb{I}_{A_i}$ para uma seqüência A_i de eventos, podemos interpretar a Lei Fraca da seguinte forma (cf. a introdução do capítulo). As probabilidades $\mathbb{P}(A_i)$ oferecem nossa avaliação dos "riscos"de cada evento A_i. Supondo-se que

$$\mathbb{C}\left(\mathbb{I}_{A_i}, \mathbb{I}_{A_j}\right) = \mathbb{P}(A_i \cup A_j) - \mathbb{P}(A_i)\,\mathbb{P}(A_j) = 0 \tag{4.3}$$

para todo par $i \neq j$, vemos temos $\sigma^2 \leq 1$ e que portanto a Lei Fraca dos Grandes Números nos diz que

$$\text{para } n \text{ grande, } \frac{\sum_{i=1}^{n} \mathbb{I}_{A_i}}{n} \approx \frac{\sum_{i=1}^{n} \mathbb{P}(A_i)}{n} \text{ com probabilidade } \approx 1.$$

Em outras palavras: sob a hipótese (4.3), há um baixo "risco"de que as freqüências com que os A_i's ocorrem se desvie muito do valor esperado, quando olhamos para um número grande de eventos.

A condição (4.3) é chamada de *independência*. O capítulo seguinte contem muitos exemplos de independência; por hora, notamos apenas o seguinte resultado.

Definição 4.20. *Dois eventos* $A, B \subset \Omega$ *são ditos independentes quando* $\mathbb{P}(A \cap B) = \mathbb{P}(A)\,\mathbb{P}(B)$.

Corolário 4.21 (Lei Fraca dos Grandes Números para Eventos Independentes)**.** *Seja* $\{A_i\}_{i=1}^{+\infty}$ *uma seqüência de evento independentes dois-a-dois. Considere*

$$C_n \equiv \frac{\sum_{i=1}^{n}(\mathbb{I}_{A_i} - \mathbb{P}(A_i))}{n}.$$

Então C_n *se concentra em* 0 *quando* $n \to +\infty$. *Mais exatamente,*

$$\forall \epsilon > 0,\, \mathbb{P}\left(\left\{\omega \in \Omega \,:\, \left|\sum_{i=1}^{n} \mathbb{I}_{A_i}(\omega) - \sum_{i=1}^{n} \mathbb{P}(A_i)\right| > \epsilon n\right\}\right) \leq \frac{1}{\epsilon^2 n}.$$

4.4 Aplicação a aproximações por polinômios

Concluímos esta seção com um "bônus": uma prova probabilística do conhecido Teorema de Weierstrass sobre aproximações por polinômios.

Teorema 4.22 (Weierstrass)**.** *Para toda função contínua* $f : [0,1] \to \mathbb{R}$*, existe uma seqüência* $P_n[f]$ *de polinômios tas que* $\lim_{n\to+\infty}(\sup_{x\in[0,1]} |f(x) - P_n[f](x)|) = 0$.

A prova que daremos dá uma expressão explícita para cada $P_n[f]$ e uma cota de aproximação para cada n finito (como veremos num Exercício). A demonstração se baseia em duas observações simples:

1. $Bin_{n,p}$ se concentra quando $n \to +\infty$ (cf. (4.2)); e
2. para qualquer $f : [0,1] \to \mathbb{R}$,

$$P_n[f](x) = \int f(k/n)\, dBin_{n,x}(k) = \sum_{k=0}^{n} \binom{n}{k} f(k/n) x^k (1-x)^{n-k}$$

é um polinômio na variável x [Exercício].

Prova: [Bernstein] Seja $f : [0,1] \to \mathbb{R}$ contínua. É sabido que qualquer f deste tipo é uniformemente contínua, isto é, o *módulo de continuidade*

$$m(\delta) \equiv \sup\{|f(x) - f(y)| \,:\, x, y \in [0,1],\, |x - y| \leq \delta\} \;\; (\delta > 0)$$

satisfaz $\lim_{\delta \to 0} m(\delta) = 0$. Sabemos também que $\|f\|_\infty = \sup_{x \in [0,1]} |f(x)| < +\infty$. A desigualdade de Jansen implica que

$$|f(x) - P_n[f](x)| \leq \int |f(k/n) - f(x)| \, dBin_{n,x}(k).$$

Fixamos um $\delta > 0$ e dividimos a integral do lado direito em dois termos.

$$\begin{aligned}
\int |f(k/n) - f(x)| \, dBin_{n,x}(k) \quad &= \int_{\{k \,:\, |k - nx| \leq \delta n\}} |f(k/n) - f(x)| \, dBin_{n,x}(k) \\
&+ \int_{\{s \,:\, |s - nx| > \delta n\}} |f(s/n) - f(x)| \, dBin_{n,x}(s).
\end{aligned}$$

Na primeira integral, $|k/n - x| \leq \delta$, logo $|f(k/n) - f(x)| \leq m(\delta)$. Na segunda usamos a cota mais fraca (e sempre válida) $|f(s/n) - f(x)| \leq 2 \sup_{t \in [0,1]} |f(t)| = 2\|f\|_\infty$. Deduzimos que

$$\begin{aligned}
|f(x) - P_n[f](x)| \quad &\leq \int_{\{k \,:\, |k - nx| \leq \delta n\}} m(\delta) \, dBin_{n,x}(k) \\
&+ \int_{\{s \,:\, |s - nx| > \delta x\}} 2\|f\|_\infty \, dBin_{n,x}(k) \\
&\leq \quad m(\delta) + 2\|f\|_\infty Bin_{n,x}\{s \,:\, |s - nx| > \delta n\} \\
((4.2) \text{ com } p = x,\, \epsilon = \delta/x) \quad &\leq \quad m(\delta) + \frac{2\|f\|_\infty x(1 - x)}{\delta^2 n} \\
(\forall 0 \leq x \leq 1,\, x(1 - x) \leq 1/4) \quad &\leq \quad m(\delta) + \frac{\|f\|_\infty}{2\delta^2 n}.
\end{aligned}$$

Esta última cota é uniforme em x e vale para $\delta > 0$ arbitrário, logo

$$\forall \delta > 0,\, \|f - P_n[f]\|_\infty \leq m(\delta) + \frac{\|f\|_\infty}{2\delta^2 n}.$$

A prova se encerra tomando limites em $n \to +\infty$ e $\delta \to 0$ (nesta ordem!). $\Box$

Exercício 4.17. *A prova acima dá uma cota quantitativa para a qualidade da aproximação por $P_n[f]$. Quanto menor o módulo de continuidade $m(\delta)$, melhor a cota. Mostre que se f é Lipschitz com constante $\|f\|_{Lip}$,*

$$\|f - P_n[f]\|_\infty \leq \|f\|_{Lip}\delta + \frac{\|f\|_\infty}{2\delta^2 n}$$

e otimize a escolha de $\delta = \delta_n$ para obter uma cota explícita para cada $n \in \mathbb{N}$.

Capítulo 5: Interpretação das probabilidades condicionais

5.1 Probabilidades e esperanças condicionais

Dissemos no início do capítulo que uma medida de probabilidade corresponde a uma avaliação de risco. Também foi dito que uma variável aleatória corresponde a informação recebida a respeito de um dado espaço amostral. Nesta seção discutiremos de que maneira a informação recebida nos permite calibrar a nossa medida de risco de modo a fazer previsões mais precisas. Isto nos levará a deduzir as regras das chamadas *probabilidades condicionais*.

5.1.1 Informação e aproximação: definindo probabilidades condicionais

Nosso problema nesta seção é *prever* o valor de uma variável aleatória $X : \Omega \to \mathbb{R}$ que suporemos satisfazer $\mathbb{E}\left[|X|^2\right] < +\infty$, com base em alguma informação $I : \Omega \to \Theta$ que dispomos sobre $\omega \in \Omega$. Para isto, escolheremos uma função $f : \Theta \to \mathbb{R}$ de modo a minimizar o *erro médio quadrático*:

$$\Delta(X, f(I)) \equiv \mathbb{E}\left[(X - f(I))^2\right].$$

No caso trivial $I =$constante; a informação que obtemos é inútil. e nosso problem é equivalente a achar $c \in \mathbb{R}$ tal que

$$\mathbb{E}\left[(X - c)^2\right] = \min_{x\in\mathbb{R}} \mathbb{E}\left[(X - c)^2\right].$$

Proposição 5.1. *Para qualquer $x \in \mathbb{R}$ temos*

$$\mathbb{E}\left[(X - x)^2\right] = \mathbb{V}(X) + (\mathbb{E}[X] - x)^2.$$

Logo o problema acima tem uma única solução $c = \mathbb{E}[X]$.

Prova:

$$\begin{aligned}
\mathbb{E}\left[(X - x)^2\right] &= \mathbb{E}\left[X^2 + x^2 - 2xX\right] \\
&= \mathbb{E}\left[X^2\right] + x^2 - 2x\mathbb{E}[X] \\
&= \mathbb{E}\left[X^2\right] + (x - \mathbb{E}[X])^2 - \mathbb{E}[X]^2 \\
&= \mathbb{V}(X) + (x - \mathbb{E}[X])^2.
\end{aligned}$$

□

Consideraremos a seguir o caso $I = \mathbb{I}_A$ para algum $A \subset \Omega$. Isto é, toda a informação que temos sobre $\omega \in \Omega$ é se $\omega \in A$ ou não. Suporemos que $0 < \mathbb{P}(A) < 1$, de modo que A não é nem "impossível"nem "certo"[1]. Procuramos então uma funç ao $f : \{0,1\} \to \mathbb{R}$ tal que

$$\mathbb{E}\left[(X - f(I))^2\right] = \inf_{g:\{0,1\}\to\mathbb{R}} \mathbb{E}\left[(X - g(I))^2\right]. \tag{5.1}$$

Este é um problema geomeétrico no espaço $L_2 = L_2(\Omega, \mathbb{P})$. Este espaço é Hilbert com o produto interno $\langle U, V\rangle \equiv \mathbb{E}[UV]$, ao menos quando identificamos quaisquer $U, U' \in L_2$ com $\mathbb{P}(U = U') = 1$ [Exercício]. De agora em diante faremos esta identificação tacitamente.

[1] Intuitivamente, se um evento sempre ocorre (ou nunca ocorre), ele não nos dá informação alguma sobre a situação em questão.

Lema 5.2. *O subconjunto*

$$E_I \equiv \{g(I) \,:\, g : \{0,1\} \to \mathbb{R}\}$$

é o subespaço linear de L_2 gerado por $\{\mathbb{I}_A, \mathbb{I}_{A^c}\}$.

Prova: De fato,

$$g(I) = g(0)\mathbb{I}_A + g(1)\mathbb{I}_{A^c}$$

sempre está neste espaço, e inversamente qualquer v.a.

$$U = a_1\mathbb{I}_A + a_0\mathbb{I}_{A^c} \in \mathrm{span}\{\mathbb{I}_A, \mathbb{I}_{A^c}\}$$

é dada por $g(I)$ com $g(0) = a_0$, $g(1) = a_1$. $\square$

Segue-se que queremos achar $\alpha, \beta \in \mathbb{R}$ que minimizem

$$\mathbb{E}\left[(X - \alpha\mathbb{I}_A - \beta\mathbb{I}_{A^c})^2\right] = \mathbb{E}\left[(X-\alpha)^2\mathbb{I}_A + (X-\beta)^2\mathbb{I}_{A^c}\right].$$

Afirmamos que há uma única escolha possível para α e β:

$$\begin{aligned} \alpha = \mathbb{E}[X \mid A] &\equiv \frac{\mathbb{E}[X\mathbb{I}_A]}{\mathbb{P}(A)} \\ \beta = \mathbb{E}[X \mid A^c] &\equiv \frac{\mathbb{E}[X\mathbb{I}_A]}{\mathbb{P}(A^c)}. \end{aligned}$$

De fato, temos o seguinte resultado:

Proposição 5.3. *Para qualquer $\alpha \in \mathbb{R}$ temos*

$$\mathbb{E}\left[(X-\alpha)^2\mathbb{I}_A\right] = \mathbb{P}(A)\,\mathbb{E}\left[(X - \mathbb{E}[X \mid A])^2\right] + \mathbb{P}(A)\,(\mathbb{E}[X \mid A] - x)^2$$

e analogamente para A^c.

Prova: Basta seguir os passos da prova de Proposição 5.1. $\square$

Exercício 5.1 (Apresentando a probabilidade condicional)**.** *Mostre que $\mathbb{E}[X \mid A]$ como definido acima satisfaz*

$$\mathbb{E}[X \mid A] = \int_\Omega X(\omega)\, d\mathbb{P}(\omega \mid A),$$

onde para todo $E \subset \Omega$

$$\mathbb{P}(E \mid A) = \mathbb{E}[\mathbb{I}_E \mid A] = \frac{\mathbb{P}(E \mid A)}{\mathbb{P}(A)}$$

ou equivalentemente

$$\mathbb{P}(\omega \mid A) = \frac{\mathbb{P}(\omega)\,\mathbb{I}_A(\omega)}{\mathbb{P}(A)}\;(\omega \in \Omega).$$

A distribuição $\mathbb{P}(\cdot \mid A)$ é a distribuição condicionada a A. $\mathbb{P}(E \mid A)$ é a probabilidade condicional de E dado A.

Juntando todos os resultados anteriores, vemos há uma única função $f : \{0,1\} \to \mathbb{R}$ que satisfaz

$$\mathbb{E}\left[(X - f(I))^2\right] = \inf_{g:\{0,1\}\to\mathbb{R}} \mathbb{E}\left[(X - g(I))^2\right].$$

e ela é dada por

$$f(x) \equiv \begin{cases} \mathbb{E}[X \mid A], & x = 1; \\ \mathbb{E}[X \mid A^c], & x = 0. \end{cases}$$

5.1.2 Informação e aproximação: o caso geral

Suponha agora que $I : \Omega \to \Theta$ é geral. Provaremos que existe uma $f : \Theta \to \mathbb{R}$ tal que

$$\mathbb{E}\left[(X - f(I))^2\right] = \inf_{g:\Theta\to\mathbb{R}} \mathbb{E}\left[(X - g(I))^2\right]. \tag{5.2}$$

Primeiro notamos o seguinte resultado.

Proposição 5.4. *Considere a partição $\mathcal{P}_I$ de Ω induzida pelas imagens inversas dos valores de I:*

$$\mathcal{P}_I \equiv \{I^{-1}(\theta) \,:\, \theta \in I(\Omega)\}.$$

Então para toda $Y : \Omega \to \mathbb{R}$, $Y = f(I)$ para algum $f : \Theta \to \mathbb{R}$ se e somente se

$$Y = \sum_{E\in\mathcal{P}_I} c_E \mathbb{I}_E,$$

com $c_E \in \mathbb{R}$ para cada $E \in \mathcal{P}_I$. Além disso, para cada partição $\mathcal{P}$ de Ω existe um conjunto Θ e uma função $I : \Omega \to \Theta$ tal que $\mathcal{P} = \mathcal{P}_I$

Prova: Exercício. Para a última afirmação, basta tomar $\Theta = \mathcal{P}$ e $I(\omega) = E \in \mathcal{P}$ tal que $\omega \in E$. $\square$

Esta proposição mostra que partições e v.a.'s são em certo sentido equivalentes. Podemos verificar que isto faz sentido quando notamos que a informação que $i = I(\omega)$ dá a respeito de ω é justamente que $\omega \in I^{-1}(i)$. Optamos por lidar com partições a seguir. Se $\mathcal{P}$ é uma partição e

$$L_2(\Omega, \mathcal{P}, \mathbb{P}) = \mathrm{span}\{\mathbb{I}_E \,:\, E \in \mathcal{P}\},$$

então a otimização descrita em (5.2) se torna a busca por $U \in L_2(\Omega, \mathcal{P}, \mathbb{P})$ tal que

$$\mathbb{E}\left[(X - U)^2\right] = \inf_{V\in L_2(\Omega,\mathcal{P},\mathbb{P})} \mathbb{E}\left[(X - V)^2\right]. \tag{5.3}$$

Provaremos o seguinte teorema geral:

Teorema 5.5. *Sejam $X : \Omega \to \mathbb{R}$ com $\mathbb{E}\left[X^2\right] < +\infty$ e $\mathcal{P}$ uma partição de Ω. Então há uma solução $U \in L_2(\Omega, \mathcal{P}, \mathbb{P})$ que é equivalentemente descrita pelas seguintes propriedades:*

1. *U é solução de (5.3);*
2. *para todo $E \in \mathcal{P}$, $\mathbb{E}[U\mathbb{I}_E] = \mathbb{E}[X\mathbb{I}_E]$.*

U é a única solução de cada um destes dois problemas, no sentido de que qualquer outra solução V satisfaz $\mathbb{P}(U = V) = 1$.