

ANÁLISE ESTATÍSTICA DE MEDIDAS EM CIÊNCIAS EXATAS

Vito R. Vanin, Philippe Gouffon, Otaviano Helene

Março 2023

Capítulo 6

Teoria da probabilidade

Começamos aqui a discussão dos fundamentos dos métodos de análise de dados, com a apresentação dos conceitos básicos do quadro formal em que se realiza essa discussão: a teoria da Probabilidade, que estivemos usando de maneira intuitiva. A definição formal de probabilidade é apresentada na seção 7.1, mas para o entendimento deste capítulo é suficiente o conceito de probabilidade da seção 1.1.

As referências básicas são os livros [Feller, Vol.1], [Kendall, Vol.1 5ª Ed. (1987)] e [Kendall, Vol.2 4ª Ed. (1979)]. O livro de Feller contém as noções básicas de probabilidade nos capítulos I, V, VI e IX. O livro de Kendall não trata explicitamente da Teoria da Probabilidade, apenas resume os pontos principais quando necessário. Embora em certos momentos a matemática seja difícil de acompanhar, a estatística e o problema da estimação são sempre o assunto, tornando-o bastante adequado, portanto, a nossos objetivos. Além disso, aborda os problemas de uma maneira mais conforme aos hábitos dos pesquisadores em ciências exatas, mas é uma leitura difícil e requer um nível razoável de familiaridade com a Estatística para acompanhá-lo. De todo modo, é uma obra de referência para assuntos específicos.

6.1 Espaço amostral e evento

Na teoria da probabilidade, *espaço amostral* e *evento* são conceitos primitivos, no mesmo sentido dos de ponto e reta na Geometria Euclidiana. Uma moeda lançada ao ar pode cair *cara*, *coroa*, em pé ou sumir naquele buraco imprevisto do piso. Pode-se, entretanto, convencionar de maneira muito aceitável que

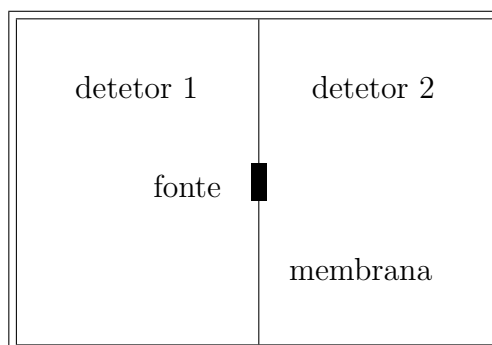


Figura 6.1: Diagrama esquemático de dois detectores de partícula α . Alfas provenientes da fonte podem resultar no disparo do detetor da esquerda, da direita ou de nenhum dos dois.

os únicos resultados possíveis são cara e coroa, o que resume o espaço amostral ao conjunto $\Omega = \{cara; coroa\}$, com grande simplificação da teoria que procura descrever a observação de um conjunto de eventos sucessivos e sem comprometer sua aplicabilidade. Podemos rediscutir este mesmo exemplo por meio de um experimento imaginário, que idealiza uma situação típica em um laboratório.

Imagine dois detectores gasosos de partículas α separados por uma membrana muito fina, com a fonte radioativa colocada sobre a membrana (figura 6.1), de modo que só um detetor dispara quando a fonte emite uma partícula. Na definição do espaço amostral, poderia ser preciso levar em conta que uma partícula não dispara nenhum detetor quando é capturada na própria fonte ou na membrana, mas desconsiderar as raras ocorrências deste fenômeno permitirá simplificar o problema, porque o espaço amostral, que corresponde ao conjunto de todos os eventos que podem ser observados, irá resumir-se a $\Omega = \{deteção\ pelo\ detetor\ direito; deteção\ pelo\ detetor\ esquerdo\}$. O que fizemos aqui ao ignorar o evento *nenhum detetor disparou* é muito comum, quase sempre os espaços amostrais que adotamos são um pouco idealizados.

Um *evento composto* corresponde a um conjunto de eventos elementares. Imaginemos que, ao invés de dois detectores de α e uma fonte de partículas α , o arranjo seja formado por 3 detectores de raios gama e uma fonte de ^{60}Co , que após seu decaimento emite dois gamas de energias g e G em cascata simultaneamente, isto é, em um intervalo de tempo menor que a resolução temporal do sistema. Para simplificar, suporemos que o conjunto dos detectores cubra

Tabela 6.1: Representação esquemática dos eventos possíveis na detecção dos dois gamas de energia g e G do ^{60}Co , com o sistema idealizado de detectores descrito nesta seção. A 1ª coluna relaciona os números com que os eventos descritos nas outras colunas serão identificados ao longo do texto.

Detetor	a	b	c
Evento			
1	$g G$.	.
2	.	$g G$.
3	.	.	$g G$
4	g	G	.
5	g	.	G
6	.	g	G
7	G	g	.
8	G	.	g
9	.	G	g

completamente o espaço em torno da fonte e seja totalmente opaco aos raios gama emitidos por ela. A tabela 6.1 relaciona os eventos que podem ser observados, ou seja, o espaço amostral Ω , neste caso formado por 9 eventos, e atribui um número a cada evento, que servirá de identificação.

Neste exemplo, pode-se definir como \mathbf{A} o evento composto *ambos os gamas observados no mesmo detetor*. Então, $\mathbf{A} = \{1, 2, 3\}$. Outro evento composto \mathbf{B} é *ao menos um gama observado no primeiro detetor*, correspondendo a $\mathbf{B} = \{1, 4, 5, 7, 8\}$.

Se os detetores não são capazes de distinguir as energias dos gamas (isto é, não separam g de G), mas podem discriminar a detecção simultânea de dois raios da detecção de um único, os eventos possíveis são outros, ou seja, o espaço amostral modifica-se e passa a ser formado como apresentado na Tabela 6.2. Note que este espaço amostral simplificado pode decorrer tanto de uma limitação experimental concreta quanto de uma escolha do experimentador, ao qual pode não interessar distinguir os eventos 4, 5 e 6 dos eventos 7, 8 e 9, respectivamente.

Verifica-se, então, que o espaço amostral depende detalhadamente do experimento, inclusive do seu objetivo. Sempre consideraremos que o espaço amostral Ω é dado e que os eventos possíveis são conhecidos. Estas são noções

Tabela 6.2: Representação esquemática dos eventos possíveis na detecção de dois gamas com o sistema de detetores idealizado descrito no texto, na condição em que podem ser separados os eventos correspondentes à detecção de 2 gamas num mesmo detetor da detecção de um gama.

Detetor	a	b	c
Evento			
1	$G G$.
2	.	$G G$.
3	.	.	$G G$
4	G	G	.
5	G	.	G
6	.	G	G

primitivas e, portanto, não definidas, como as noções de ponto e reta na Geometria. À noção de espaço amostral associamos a idéia de um conjunto Ω , e que cada evento é um elemento de Ω . Note que um evento composto também corresponde a um conjunto de eventos elementares, que é, porém, diferente do espaço amostral Ω , porque Ω contém *todos* os eventos que interessam ao problema em consideração.

6.2 Relações entre eventos

Quando ω_i é o i -ésimo evento elementar, o espaço amostral Ω é o conjunto de *todos* eles,

$$\Omega = \{\omega_i\} \quad .$$

Um evento composto \mathbf{A} é, portanto, um certo conjunto dos ω_i tal que

$$\mathbf{A} = \{\omega_i\} \subseteq \Omega \quad .$$

Diz-se que \mathbf{A} ocorre (ou que se observa \mathbf{A}) quando o experimento estatístico resulta em um dos $\omega_i \in \mathbf{A}$. Define-se:

- evento impossível, \emptyset , como o conjunto vazio.
- evento composto \mathbf{A} e \mathbf{B} , $\mathbf{A} \cap \mathbf{B}$, como o conjunto dos eventos ω_j tais que $w_j \in \mathbf{A}$ e $\omega_j \in \mathbf{B}$, ilustrado pela figura 6.2.

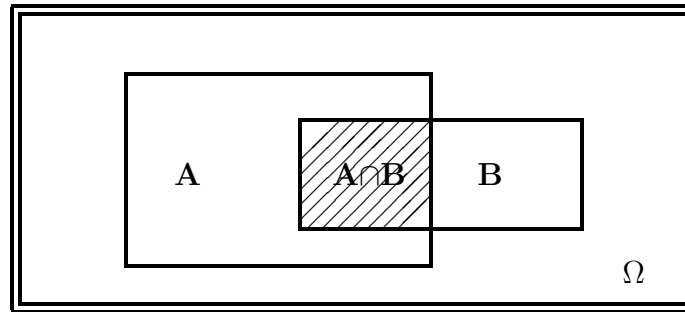


Figura 6.2: Diagrama de Venn, que ilustra o evento composto $\mathbf{A} \cap \mathbf{B}$, hachurado na figura.

- evento composto \mathbf{A} ou \mathbf{B} , $\mathbf{A} \cup \mathbf{B}$, como o conjunto dos eventos ω_j para os quais ao menos uma das duas afirmações $\omega_j \in \mathbf{A}$ e $\omega_j \in \mathbf{B}$ é verdadeira.
- evento complementar de \mathbf{A} , \mathbf{A}^C , como o conjunto dos eventos de Ω que não pertencem a \mathbf{A} , com a propriedade imediata “se $\omega_i \in \mathbf{A} \Leftrightarrow \omega_i \notin \mathbf{A}^C$ ”.

Diagramas como o da figura 6.2 servem para ilustrar as relações entre eventos compostos.

Q6.1 *Ilustre, com diagramas semelhantes ao da figura 6.2: o evento $\mathbf{A} \cup \mathbf{B}$; dois eventos tais que $\mathbf{A} \cap \mathbf{B} = \emptyset$; os eventos \mathbf{A} e \mathbf{A}^C .*

Dois eventos \mathbf{A} e \mathbf{B} são *mutuamente exclusivos* quando $\mathbf{A} \cap \mathbf{B} = \emptyset$, por definição, que pode ser estendida a qualquer número de eventos $\mathbf{A}, \mathbf{B}, \mathbf{C}, \dots$, quando é exigido $\mathbf{A} \cap \mathbf{B} = \mathbf{A} \cap \mathbf{C} = \mathbf{B} \cap \mathbf{C} = \dots = \emptyset$. Decorre da definição que \mathbf{A} e \mathbf{A}^C são mutuamente exclusivos.

Para exemplificar, voltemos ao exemplo da tabela 6.1. O espaço amostral é $\Omega = \{1, 2, 3, 4, 5, 6, 7, 8, 9\}$, o evento *ambos os gamas observados no mesmo detetor* é $\mathbf{A} = \{1, 2, 3\}$ e o evento *ao menos um gama é observado no primeiro detetor* é $\mathbf{B} = \{1, 4, 5, 7, 8\}$. Assim, os eventos compostos “ambos os gamas observados no mesmo detetor **ou** ao menos um gama é observado no primeiro detetor” e “ambos os gamas observados no mesmo detetor **e** ao menos um gama é observado no primeiro detetor” são $\mathbf{A} \cup \mathbf{B} = \{1, 2, 3, 4, 5, 7, 8\}$ e $\mathbf{A} \cap \mathbf{B} = \{1\}$. Como este último conjunto *não* é vazio, \mathbf{A} e \mathbf{B} *não* são mutuamente exclusivos.

Um evento \mathbf{A} pode implicar no evento \mathbf{B} sem que \mathbf{B} implique em \mathbf{A} , bastando que \mathbf{A} esteja contido em \mathbf{B} . Em termos de álgebra de conjuntos, basta que $\mathbf{A} \subset \mathbf{B}$ mas $\mathbf{B} \not\subset \mathbf{A}$. Podemos definir $\mathbf{B} - \mathbf{A}$ como a ocorrência de \mathbf{B} sem que \mathbf{A} ocorra.

Uma última definição, nesta seção: *espaço amostral discreto* é aquele que tem um número finito de eventos, como o do exemplo, ou cujos eventos ω podem ser arranjados numa sequência enumerável $\omega_1, \omega_2, \omega_3, \omega_4, \omega_5, \omega_6, \omega_7, \omega_8, \omega_9, \omega_{10} \dots$, mesmo que seja infinita.

6.3 Probabilidade – regras e propriedades

A probabilidade associada a um evento \mathbf{A} , $P(\mathbf{A})$, é uma grandeza mensurável numa escala contínua. $P(\mathbf{A})$ é, portanto, um número real não negativo. Considera-se que é possível comparar probabilidades associadas a dois eventos e diz-se que

\mathbf{A} é mais provável que \mathbf{B} se $P(\mathbf{A}) > P(\mathbf{B})$;

se $P(\mathbf{A}) = P(\mathbf{B})$, então esses eventos são *equiprováveis*.

A probabilidade é uma função aditiva dos eventos. Assim, quando \mathbf{A} e \mathbf{B} são mutuamente exclusivos (ou seja, se $\mathbf{A} \cap \mathbf{B} = \emptyset$), então

$$P(\mathbf{A} \cup \mathbf{B}) = P(\mathbf{A}) + P(\mathbf{B}) \quad .$$

Em particular,

$$P(\mathbf{A}) = \sum_i P(\omega_i \in \mathbf{A}) \quad .$$

Para fixar a escala numérica da probabilidade, convencionou-se atribuir o valor 1 à certeza e o valor 0 à impossibilidade. Então,

$$0 \leq P(\mathbf{A}) \leq 1 \quad , \quad \forall \mathbf{A}.$$

Em resumo, os axiomas são:

$$P(\mathbf{A}) \geq 0 \quad , \quad \forall \mathbf{A} \tag{6.1}$$

$$\sum_i P(\omega_i) = P(\Omega) = 1 \tag{6.2}$$

$$P(\mathbf{A} \cup \mathbf{B}) = P(\mathbf{A}) + P(\mathbf{B}) \quad , \quad \text{se } \mathbf{A} \cap \mathbf{B} = \emptyset \quad . \tag{6.3}$$

Deduz-se desses axiomas, para $\mathbf{A} \subset \Omega$ e $\mathbf{B} \subset \Omega$:

$$P(\mathbf{A} \cup \mathbf{B}) = P(\mathbf{A}) + P(\mathbf{B}) - P(\mathbf{A} \cap \mathbf{B}) \quad (6.4)$$

$$P(\emptyset) = 0$$

$$P(\mathbf{A}^C) = 1 - P(\mathbf{A}) \quad (6.5)$$

$$P(\mathbf{A}) = P(\mathbf{A} \cap \mathbf{B}) + P(\mathbf{A} \cap \mathbf{B}^C) \quad (6.6)$$

A relação (6.5) vem da normalização para 1 da soma das probabilidades de todos os elementos do espaço amostral, axioma (6.2). Já a (6.4), que difere do axioma (6.3) porque permite lidar com eventos compostos em geral, decorre de que $P(\mathbf{A})$ e $P(\mathbf{B})$ contam, ambos, os elementos da intersecção dos dois conjuntos, como a figura 6.2 ajuda a entender; essa dupla contagem precisa ser corrigida pela subtração da probabilidade associada à intersecção. A relação (6.6) é consequência da (6.4) ao levarmos em conta que $(\mathbf{A} \cap \mathbf{B})$ e $(\mathbf{A} \cap \mathbf{B}^C)$ são mutuamente exclusivos, o que, por sua vez, vem de $\mathbf{B} \cap \mathbf{B}^C = \emptyset$.

Note que a relação da fórmula (6.4) pode ser generalizada. Dados k eventos \mathbf{B}_i , em geral

$$\sum_{i=1}^k P(\mathbf{B}_i) \geq P\left(\bigcup_{i=1}^k \mathbf{B}_i\right) \quad ,$$

sendo que vale a igualdade se todos os eventos \mathbf{B}_i forem mutuamente exclusivos, ou seja

$$\mathbf{B}_i \cap \mathbf{B}_j = \emptyset \quad , \quad \forall i \neq j \quad .$$

Dizemos que o conjunto dos \mathbf{B}_i forma uma *partição* de Ω se, além de serem mutuamente exclusivos, eles cobrem todo o espaço amostral, o que resulta em

$$\sum_{i=1}^k P(\mathbf{B}_i) = 1 \quad \text{se} \quad \bigcup_{i=1}^k \mathbf{B}_i = \Omega \quad .$$

De uma partição $\mathbf{B}_i, i = 1, 2, \dots, k$, é possível criar novas partições de Ω com qualquer evento \mathbf{A} , porque \mathbf{A} e \mathbf{A}^C sempre definem uma outra partição de Ω . Definindo os eventos

$$\mathbf{C}_i = \mathbf{A} \cap \mathbf{B}_i \quad , \quad i = 1, 2, \dots, k \quad \text{e}$$

$$\mathbf{C}_{k+1} = \mathbf{A}^C \quad ,$$

o conjunto $\{\mathbf{C}_j, j = 1, 2, \dots, k + 1\}$ forma uma partição de Ω . Decorre, então,

$$\sum_{i=1}^{k+1} P(\mathbf{C}_i) = 1 \quad \text{ou}$$

$$\sum_{i=1}^k P(\mathbf{C}_i) = \sum_{i=1}^k P(\mathbf{A} \cap \mathbf{B}_i) = 1 - P(\mathbf{A}^C) = P(\mathbf{A}) \quad . \quad (6.7)$$

Esta última relação expressa a *lei da probabilidade total*.

6.4 Probabilidade condicional

No exemplo da detecção de 2 raios gama por 3 detetores da seção 6.1, considere a situação em que os gamas de energias g e G são indistinguíveis, quando o espaço amostral é aquele representado na Tabela 6.2. É possível determinar as relações entre as probabilidades dos diferentes eventos quando se estabelecem condições acerca deles, ou seja, impõe-se uma certa exigência e deseja-se saber quais são as probabilidades dos eventos que satisfazem essa exigência. Os 3 detetores serão identificados como na tabela (\mathbf{a} , \mathbf{b} , \mathbf{c}), e os eventos representados na ordem

| gamas det. por \mathbf{a} | gamas det. por \mathbf{b} | gamas det. por \mathbf{c} | .

A questão consiste em determinar a probabilidade do detetor \mathbf{a} ser disparado quando o detetor \mathbf{b} disparou. O conjunto dos eventos que correspondem à detecção em \mathbf{b} é $\mathbf{B} = \{2, 4, 6\}$, e o conjunto dos eventos que correspondem ao disparo de \mathbf{a} é $\mathbf{A} = \{1, 4, 5\}$. Os eventos correspondentes ao fenômeno “ \mathbf{a} disparou” sujeitos à condição “ \mathbf{b} disparou” estão contidos no conjunto $\mathbf{A} \cap \mathbf{B}$, a intersecção dos dois, neste caso igual ao evento $\{4\}$. Chama-se de *probabilidade condicional de \mathbf{A} dado \mathbf{B}* à razão

$$P(\mathbf{A} \mid \mathbf{B}) = \frac{P(\mathbf{A} \cap \mathbf{B})}{P(\mathbf{B})} \quad , \quad (6.8)$$

com a ressalva

$$P(\mathbf{B}) > 0 \quad .$$

No exemplo, $P(\mathbf{A} \mid \mathbf{B}) = P(4)/[P(2) + P(4) + P(6)]$. Assim, a probabilidade condicional é a probabilidade do evento que corresponde a ambas

características dentro de um novo espaço amostral reduzido, que contém apenas eventos que obedecem à condição formulada. A expressão (6.8) acima corresponde à definição formal de probabilidade condicional e não se aplica aos casos em que $P(\mathbf{B}) = 0$.

Pode-se também calcular a probabilidade condicional de \mathbf{B} dado \mathbf{A} ,

$$P(\mathbf{B} \mid \mathbf{A}) = \frac{P(\mathbf{B} \cap \mathbf{A})}{P(\mathbf{A})} \quad ,$$

que pode ser reescrita na forma

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

A fórmula (6.8) pode ser reescrita como

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

Uma vez que $\mathbf{A} \cap \mathbf{B} \equiv \mathbf{B} \cap \mathbf{A}$, deduz-se que

$$P(\mathbf{A} \mid \mathbf{B})P(\mathbf{B}) = P(\mathbf{B} \mid \mathbf{A})P(\mathbf{A}) \quad , \quad (6.9)$$

que é uma versão do *Teorema de Bayes*, que discutiremos novamente no capítulo seguinte.

6.5 Independência estatística

Consideraremos as expressões *independência estatística*, *independência estocástica* ou, simplesmente, *independência*, como sinônimos. O evento \mathbf{A} é independente de \mathbf{B} se a ocorrência de \mathbf{B} não interfere na *probabilidade* de ocorrência de \mathbf{A} . Usando a linguagem de probabilidade condicional da seção anterior, se

$$P(\mathbf{A} \mid \mathbf{B}) = P(\mathbf{A}) \quad \rightarrow \quad \mathbf{A} \text{ é independente de } \mathbf{B} \quad . \quad (6.10)$$

Usando a fórmula (6.8), constata-se que, para que isso seja verdadeiro, é necessário e suficiente que

$$P(\mathbf{A} \cap \mathbf{B}) = P(\mathbf{A}) \cdot P(\mathbf{B}) \quad . \quad (6.11)$$

De fato, a expressão acima é uma outra maneira de definir independência estatística, preferível à da relação (6.10), que se refere à probabilidade condicional, porque é mais simétrica. Veja que é com essa expressão que se calcula

a probabilidade associada a “eventos independentes”, que é interpretada como “se a ocorrência de **A** nada tem a ver com a ocorrência de **B**, a chance de **A** e **B** ocorrerem é o produto das chances deles ocorrerem independentemente”. Assim, independência tem, na Teoria da Probabilidade, exatamente o sentido intuitivo conferido ao termo.

Detalhes:

- Como **A** independente de **B** \Leftrightarrow **B** independente de **A**, verifica-se que, se $P(\mathbf{A} \mid \mathbf{B}) = P(\mathbf{A})$, então $P(\mathbf{B} \mid \mathbf{A}) = P(\mathbf{B})$.
- A propriedade associativa não vale, ou seja,

$$\left. \begin{array}{l} \mathbf{A} \text{ independente de } \mathbf{B} \text{ e} \\ \mathbf{B} \text{ independente de } \mathbf{C} \text{ e} \\ \mathbf{A} \text{ independente de } \mathbf{C} \end{array} \right\} \text{ não implica em } \mathbf{A} \cap \mathbf{B} \text{ independente de } \mathbf{C}.$$

É possível, agora, precisar o significado estatístico de dados independentes em uma medida. Se E_j é o evento resultante da primeira observação e E_k é o da segunda, a observação foi repetida em condições idênticas de maneira independente (ou o dado foi obtido independentemente nas mesmas condições) se

$$P(E_j \cap E_k) = P(E_j) \cdot P(E_k) \quad .$$

Para exemplificar o conteúdo destas duas últimas seções, voltamos ao exemplo da seção 6.1 e atribuímos probabilidades aos eventos relacionados na tabela 6.1. Vamos escolher as probabilidades de duas maneiras diferentes, que estão discriminadas na tabela 6.3 e identificadas por *I* e *II*, correspondentes a eventos equiprováveis e à existência de duas categorias de eventos, respectivamente.

É possível verificar se os eventos: $\mathbf{A} = \{\text{o detetor } \mathbf{a} \text{ detetou } g\}$ e $\mathbf{B} = \{\text{o detetor } \mathbf{b} \text{ detetou } G\}$, são independentes ou não. A Tabela 6.4 foi construída com esse finalidade. A independência está ligada à propriedade definida na expressão (6.11). Com as probabilidades da tabela 6.4, $P(\mathbf{A} \cap \mathbf{B})$ é igual a $P(\mathbf{A}) \cdot P(\mathbf{B})$ no caso *I*, mas não no caso *II*. Portanto, os eventos **A** e **B** são independentes quando todos os eventos da Tabela 6.3 são equiprováveis, mas não quando a probabilidade de dois gamas serem detetados pelo mesmo detetor for diferente da probabilidade deles serem detetados por detetores diferentes.¹ Outra maneira de testar a independência entre **A** e **B** é calcular a probabilidade

¹Isso ocorre quanto a emissão dos dois gamas em uma mesma direção tem probabilidade diferente da emissão em direções diferentes, o que é relativamente comum na natureza.

Tabela 6.3: Representação esquemática dos eventos possíveis na detecção de dois gamas de energias g e G do ^{60}Co , com o sistema idealizado de detectores descrito na seção 6.1. Note que a condição de normalização das probabilidades no caso I dá $\delta = 1/9$ e, no caso II , $3\chi + 6\phi = 1$. Desejamos $\chi \neq \phi$ para que os dois casos sejam diferentes.

evento	detetor			probabilidade no caso	
	a	b	c	I	II
1	$g G$.	.	δ	χ
2	.	$g G$.	δ	χ
3	.	.	$g G$	δ	χ
4	g	G	.	δ	ϕ
5	g	.	G	δ	ϕ
6	.	g	G	δ	ϕ
7	G	g	.	δ	ϕ
8	G	.	g	δ	ϕ
9	.	G	g	δ	ϕ

condicional dada pela expressão (6.8): a independência entre \mathbf{A} e \mathbf{B} implica em $P(\mathbf{A} \mid \mathbf{B}) = P(\mathbf{A})$, equação (6.10), o que, consultando a tabela 6.4, é verdade no caso I , mas não no II .

Q6.2 Mostre que os eventos: $\mathbf{A} = \{\text{os dois gamas foram detetados no mesmo detetor}\}$ e $\mathbf{B} = \{\text{o detetor } \mathbf{b} \text{ disparou}\}$, não são independentes em qualquer dos casos.

Q6.3 Mostre que os eventos: $\mathbf{A} = \{\text{os dois gamas foram detetados no mesmo detetor}\}$ e $\mathbf{B} = \{\text{o detetor } \mathbf{b} \text{ detetou } g\}$, são independentes em ambos os casos. (Não esqueça da importante condição de normalização das probabilidades, que resulta em $\chi + 2\phi = 1/3$.)

6.6 Variáveis aleatórias

Define-se variável aleatória de maneira análoga a uma função. Relembrando, uma função corresponde a uma regra que associa cada elemento de um domínio aos elementos de um contradomínio. A figura 6.3 ilustra a correspondência

Tabela 6.4: Sequência de cálculo das probabilidades que permitem verificar se \mathbf{A} e \mathbf{B} são estatisticamente independentes. Os valores referem-se aos eventos $\mathbf{A} = \{\text{detetor } \mathbf{a} \text{ disparou com } g\}$ e $\mathbf{B} = \{\text{o detetor } \mathbf{b} \text{ disparou com } G\}$.

evento	probabilidade do evento no caso	
	I	II
$\mathbf{A} = \{1, 4, 5\}$	$3\delta = 1/3$	$\chi + 2\phi = 1/3$
$\mathbf{B} = \{2, 4, 9\}$	$3\delta = 1/3$	$\chi + 2\phi = 1/3$
$\mathbf{A} \cap \mathbf{B} = \{4\}$	$\delta = 1/9$	ϕ
$P(\mathbf{A}) \cdot P(\mathbf{B})$	$1/9$	$1/9$
$P(\mathbf{A} \mathbf{B})$	$1/3$	$\phi / (\chi + 2\phi) = 3\phi$

entre x e $\text{sen}(x)$, para uns poucos valores de x .

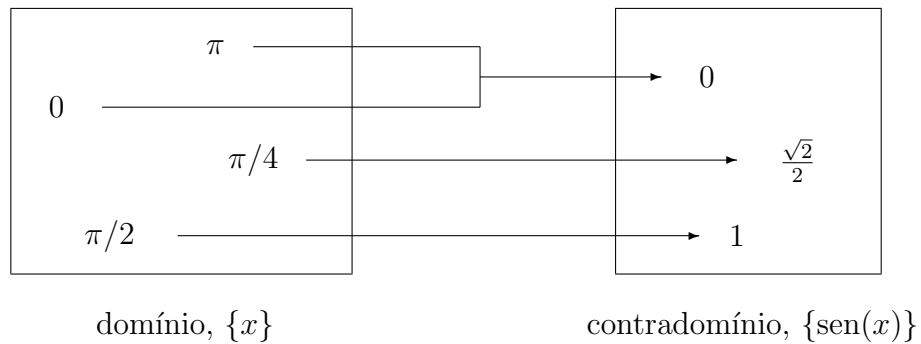


Figura 6.3: Representação pictórica de uma parte da função $\text{sen}(x)$. À esquerda, o domínio da função, \mathbf{R} , e à direita o contradomínio, $\{y \in \mathbf{R}, -1 \leq y \leq 1\}$.

Uma variável aleatória é o análogo estatístico de uma função, ou melhor, é uma função definida sobre o espaço amostral. A figura 6.4 ilustra a definição de variável aleatória, com o exemplo de um sistema de dois átomos de spin 1/2 em que a variável aleatória é o número de átomos com spin em uma certa direção.

[Feller] comenta que o nome “função aleatória” seria mais adequado que variável aleatória. Afinal, o que se deseja mesmo é definir uma relação ou regra que forneça uma característica do evento, exatamente como se faz para

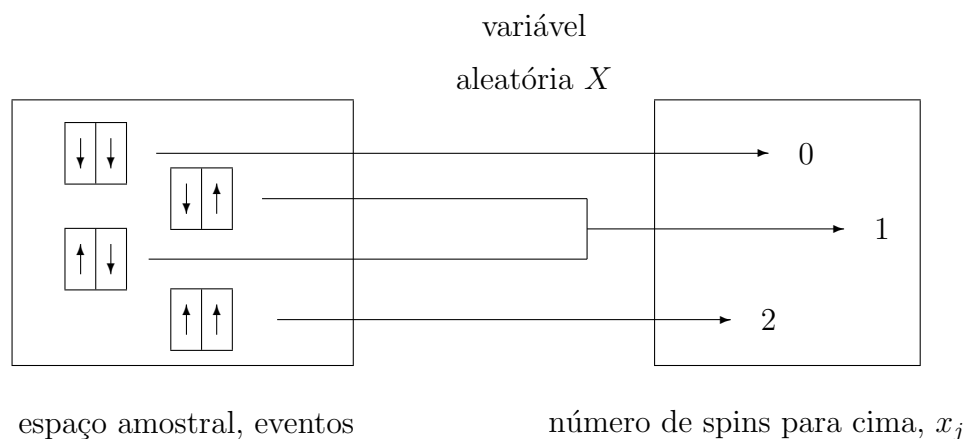


Figura 6.4: Representação pictórica da variável aleatória, para um sistema consistindo de dois átomos de spin $1/2$ em que estamos interessados no número de spins que apontam numa certa direção.

as funções matemáticas no sentido comum. O uso, porém, consagrou o nome *variável aleatória*.

Designando a variável aleatória por X , define-se a função de probabilidade pela relação

$$P(X = x_j) = f(x_j) \quad ,$$

onde o evento $X = x_j$ é o evento composto por todos os elementos do espaço amostral para os quais a variável aleatória X assume o valor numérico x_j e $P(\text{evento})$ é a probabilidade associada ao evento.

Pode-se, então, dizer que a variável aleatória X assume os valores $x_1, x_2, \dots, x_j, \dots$ com probabilidades $f(x_1), f(x_2), \dots, f(x_j), \dots$ sem mais nenhuma referência ao espaço amostral *original*, um novo espaço amostral sendo formado por $x_1, x_2, \dots, x_j, \dots$. Quando se especifica uma função de probabilidade, ela sempre se refere a esse *novo* espaço amostral, cujos elementos são números reais; a função de probabilidade simplesmente fornece a probabilidade associada a esses *novos* eventos.

Em geral, não se necessita nem de tantos detalhes nem tanta formalidade, de modo que só faremos a distinção entre a variável aleatória e os valores numéricos que ela pode assumir neste começo ou quando houver risco de confusão.

Nesta seção, estivemos subentendendo um espaço amostral discreto, isto

é, com uma quantidade finita ou pelo menos enumerável de elementos. As funções densidade de probabilidade são análogas às funções de probabilidade, mas se aplicam a espaços amostrais contínuos, conforme a discussão informal dos capítulos anteriores.

6.7 Distribuição cumulativa de probabilidade

Quando $P(m)$ é a função de probabilidade da variável aleatória discreta m , define-se a distribuição cumulativa de probabilidade pela relação

$$\mathcal{P}(n) = \sum_{m \leq n} P(m) \quad , \quad (6.12)$$

ou seja, $\mathcal{P}(n)$ é a probabilidade da variável aleatória assumir um valor menor ou igual ao argumento da função, n . A definição para variáveis contínuas é análoga: a função distribuição cumulativa de probabilidade é

$$\mathcal{F}(x) = \int_{-\infty}^x f(x') dx' \quad , \quad (6.13)$$

com interpretação idêntica à das distribuições de variável discreta. No último caso, a f.d.p. pode ser calculada da distribuição cumulativa por meio da relação

$$f(x) = \frac{d\mathcal{F}}{dx} \quad .$$

Como a probabilidade de obter um resultado qualquer é normalizada a 1,

$$0 \leq \mathcal{F}(x) \leq 1 \quad ,$$

e como a probabilidade de um evento é sempre positiva ou nula,

$$\frac{d\mathcal{F}}{dx} \geq 0 \quad ,$$

com condições análogas no caso de variáveis discretas.

No desenvolvimento formal da Probabilidade e da Estatística, usam-se muito as distribuições cumulativas. Neste livro, as funções densidade, que parecem refletir mais diretamente as observações experimentais, foram usadas sempre que possível.

6.8 Funções de duas variáveis aleatórias, densidades condicional e marginal

Considere duas variáveis aleatórias X_1 e X_2 definidas sobre um espaço amostral Ω , cujos eventos correspondem a pares de valores (x_1, x_2) . Em princípio, a probabilidade de se obter um evento dentro do intervalo de larguras dx_1 e dx_2 depende tanto de x_1 quanto de x_2 , ou seja, a f.d.p. conjunta f depende das duas variáveis, de modo que

$$dP(x_1 \leq X_1 \leq x_1 + dx_1, x_2 \leq X_2 \leq x_2 + dx_2) = f(x_1, x_2) dx_1 dx_2 \quad . \quad (6.14)$$

Assim, a relação entre as funções distribuição cumulativa \mathcal{F} e densidade é

$$\mathcal{F}(x_1, x_2) = \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} f(x'_1, x'_2) dx'_1 dx'_2 = P(X_1 \leq x_1, X_2 \leq x_2) \quad ,$$

equivalente a

$$f(x_1, x_2) = \frac{\partial^2 \mathcal{F}}{\partial x_1 \partial x_2} \quad .$$

Define-se a f.d.p. condicional de X_1 dado que X_2 assumo o valor fixo x_2 de maneira parecida com a probabilidade condicional em variáveis discretas (fórmula 6.8), ou seja, por meio da função $g(x_1)$ que resulta da imposição da restrição a X_2 no espaço amostral Ω inteiro,

$$g(x_1 | x_2 = k) = \frac{f(x_1, k) dk}{\int_{-\infty}^{\infty} f(x'_1, k) dx'_1 dk} = \frac{f(x_1, k)}{\int_{-\infty}^{\infty} f(x'_1, k) dx'_1} \quad , \quad (6.15)$$

onde o denominador, no membro do meio, é a probabilidade de observar a variável X_2 no intervalo $k \leq x_2 \leq k + dk$, que pode ser nula; como não há sentido em definir probabilidade condicional para uma condição impossível, convencionou-se que g é nula quando esse denominador é nulo.

Chama-se f.d.p. marginal de X_1 à função $h(x_1)$ que resulta de ignorar o valor assumido por X_2 ,

$$h(x_1) = \int_{-\infty}^{\infty} f(x_1, x'_2) dx'_2 \quad . \quad (6.16)$$

Marginal tem, aqui, o sentido de *estar à margem*, ou seja, é a projeção da densidade conjunta sobre um dos eixos, sobre uma margem das coordenadas. Há também uma densidade marginal de X_2 ,

$$w(x_2) = \int_{-\infty}^{\infty} f(x'_1, x_2) dx'_1 \quad .$$

Note que, com a definição acima, a f.d.p. condicional de x_1 dado x_2 , relação (6.15), pode ser reescrita como a razão entre a densidade conjunta e a marginal de x_2 com a condição $x_2 = k$,

$$g(x_1 | x_2 = k) = \frac{f(x_1, x_2)}{w(x_2)} \Big|_{x_2=k},$$

o que ressalta a forma da restrição.

Quando X_1 é estatisticamente independente de X_2 , o resultado acima é o mesmo para qualquer k , de modo que

$$g(x_1 | x_2 = k) = h(x_1) \quad ,$$

onde identifica-se que o resultado independente de x_2 corresponde à densidade marginal $h(x_1)$ da fórmula (6.16). Assim,

$$f(x_1, x_2) = h(x_1) w(x_2) \quad \text{se } x_1 \text{ independente de } x_2 \quad .$$

6.9 Valor esperado de uma variável aleatória

Certos parâmetros associados à f.d.p. permitem descrevê-la, ao menos qualitativamente. Assim, a média é uma medida da *localização* da f.d.p. e a variância, da *dispersão*. Esta seção formaliza e generaliza os conceitos associados a essas grandezas.

Define-se *valor esperado* da variável aleatória X à grandeza

$$E(X) = \langle x \rangle = \int_{-\infty}^{\infty} x f(x) dx \quad , \quad (6.17)$$

condicionado a que a integral convirja *absolutamente*, isto é, que a integral

$$\int_{-\infty}^{\infty} |x| f(x) dx$$

exista; caso contrário, diz-se que X não tem valor esperado finito.

Exemplo 6.1

A f.d.p. de t de Student (fórmula 3.25), quando há apenas um grau de liberdade,

$$\psi(t) = \frac{1}{\pi(t^2 + 1)}$$

é um exemplo importante de inexistência do valor médio, por falta de convergência absoluta da integral que define o valor esperado, equação (6.17).

É o comportamento aleatório da variável que faz com que ela não siga um padrão de simetria, necessário para anular a integral de uma função antissimétrica; ao repetir a observação um certo número de vezes, é improvável obter a mesma quantidade de valores positivos e negativos, e as distâncias à origem também vão variar ao acaso. Esse comportamento imprevisível da média pode ser observado em uma simulação: sorteie uma centena de conjuntos $\{t_i, i = 1 \dots N\}$, com $N \approx 10$, e veja como as médias aritméticas desses conjuntos variam irregularmente, devido à presença ocasional de um dado de módulo excepcionalmente alto em um ou alguns conjuntos. Repita a simulação com $N \approx 100$ e note que o comportamento continua irregular.

Essa função aparece em vários outros problemas em uma forma um pouco mais genérica, com parâmetros de posição e largura θ e Γ ,

$$f(x; \theta, \Gamma) = \frac{\Gamma}{\pi((x - \theta)^2 + \Gamma^2)} \quad (6.18)$$

e tem vários nomes, os mais comuns função de Cauchy ou de Lorentz, o último associado ao fenômeno da ressonância.

Toda função $\phi(x)$ define uma nova variável aleatória, que podemos chamar de $\phi(X)$, cujo valor esperado é, simplesmente,

$$E[\phi(X)] = \int_{-\infty}^{\infty} \phi(x) f(x) dx \quad , \quad (6.19)$$

sujeito também à convergência absoluta da integral. Um caso particular que ocorre frequentemente é

$$\phi(X) = aX \quad ,$$

quando se obtém

$$E[\phi(X)] = aE(X) \quad ,$$

ou seja, a é simplesmente um fator de escala.

Um resultado imediato da linearidade da operação matemática com que se determina o valor esperado é que

$$E(X_1 + X_2 + \dots + X_N) = E(X_1) + E(X_2) + \dots + E(X_N) \quad . \quad (6.20)$$

Q6.4 Calcule o valor médio (ou valor esperado) do número n de sucessos quando os eventos seguem a função de probabilidade Binomial. Estenda o resultado acima para variáveis discretas e considere que

$$n = \sum_{i=1}^N X_i \quad ,$$

onde X_i é 1 se a i -ésima observação é um sucesso e 0, se não é um sucesso (portanto, $X_i = 0$ com probabilidade q e $X_i = 1$ com probabilidade p , onde $p + q = 1$, na notação do capítulo 2).

Sugestão: calcule primeiro $E(X_i)$ e use o fato de todas as observações serem repetidas em condições idênticas.

6.10 Momentos de uma função densidade de probabilidade

Como comentamos na seção 6.6, aqui não vamos ser rigorosos em distinguir a variável aleatória X da variável x que representa seus valores possíveis, uma vez que não há possibilidade de confusão nas definições que seguem.

Define-se o *momento de ordem r de uma variável aleatória x* , μ'_r , como o valor esperado de x^r ,

$$\mu'_r = E(x^r) = \langle x^r \rangle = \int_{-\infty}^{\infty} x^r f(x) dx \quad . \quad (6.21)$$

Assim, μ'_1 é a média,

$$\mu = \mu'_1 = E(x) = \langle x \rangle = \int_{-\infty}^{\infty} x f(x) dx \quad ,$$

habitualmente simbolizado simplesmente por μ na literatura especializada. Nos capítulos 1 e 2, o valor esperado $E(X)$ foi denotado por $\langle x \rangle$ ou x_0 e, agora, mudamos para μ ; isso não significa nenhuma mudança conceitual, mas apenas uma escolha de notação mais adequada ao restante do capítulo, porque as fórmulas ficam pesadas com muitos subscritos e símbolos $\langle \rangle$ aninhados. Usamos tantas notações diferentes para o valor esperado: $E(X) = \mu = \mu' = E(x) = \langle x \rangle = x_0$, todas muito comuns nos livros, porque cada uma dá melhor aparência ou clareza a alguma fórmula ou definição. Enfim, essa

multiplicidade de símbolos também vem da importância desse parâmetro, que é uma medida da *localização* de f , comumente uma *medida* da grandeza, que é o que buscamos, mas, infelizmente, nunca poderemos conhecer.

Define-se *momento central de ordem r de uma variável aleatória x* , μ_r , como o valor esperado de $(x - \mu)^r$,

$$\mu_r = E[(x - \mu)^r] = \int_{-\infty}^{\infty} (x - \mu)^r f(x) dx \quad . \quad (6.22)$$

O primeiro desses momentos é nulo, porque é a média da diferença com a própria média. Já o segundo corresponde à *variância*, σ_0^2 ,

$$\sigma_0^2 = \mu_2 = E[(x - \mu)^2] = \langle (x - \mu)^2 \rangle \quad . \quad (6.23)$$

O desvio padrão σ_0 é **definido** como a raiz quadrada positiva da variância,

$$\sigma_0 = +\sqrt{\mu_2} \quad ,$$

que é uma medida da *largura* da f.d.p..

Q6.5 Mostre que a variância iguala o valor esperado de x^2 (μ'_2 ou $E(X^2)$) menos o quadrado da média (μ^2 ou $E(X)^2$),

$$\sigma^2 = \mu'_2 - \mu^2 = E(X^2) - E(X)^2 \quad . \quad (6.24)$$

Outras grandezas que servem para caracterizar as f.d.p.s são o *parâmetro de assimetria* β ,

$$\beta = \frac{\mu_3}{(\mu_2)^{3/2}} \quad , \quad (6.25)$$

e a *curtose* κ ,

$$\kappa = \frac{\mu_4}{(\mu_2)^2} - 3 \quad . \quad (6.26)$$

Como β tem o sinal de μ_3 , uma f.d.p. unimodal com β positivo quase sempre indica que a f.d.p. tem uma cauda para a direita e, portanto, tem

$$\text{média} > \text{mediana} > \text{moda} \quad .$$

Já $\beta < 0$ indica uma f.d.p. com cauda para a esquerda, para a qual

$$\text{média} < \text{mediana} < \text{moda} \quad .$$

$\beta = 0$ implica em uma f.d.p. simétrica em torno da média. Já a curtose não tem interpretação simples, e em muitos textos e funções de computador, é definida sem a constante numérica; essa propriedade é facilmente testada calculando a curtose da distribuição normal, que é nula com a definição da fórmula (6.26) e, claro, igual a 3, sem essa constante.

Pode-se mostrar que duas f.d.p.s contínuas e deriváveis infinitamente que têm mesmos momentos são idênticas. Com um número pequeno de dados, verifica-se que duas funções cujos momentos de ordem mais baixa são iguais são muito parecidas, de maneira que usualmente os 4 primeiros momentos — ou grandezas associadas: μ , σ , β e κ — caracterizam bastante bem a função.

Se $\{x_i, i = 1, \dots, N\}$ é o conjunto de dados de uma medida, estimam-se os momentos centrais μ_r por meio das estatísticas

$$m_r = \frac{1}{N} \sum_{i=1}^N (x_i - \langle x \rangle)^r .$$

Se a média $\langle x \rangle$ é desconhecida, ela pode ser estimada por

$$\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i .$$

As variâncias dessas estatísticas se relacionam com os diversos momentos da f.d.p. e podem ser calculadas. A Tabela 6.5 apresenta as variâncias no caso geral e também no importante caso de dados que seguem a distribuição gaussiana.

6.11 Momentos de funções de várias variáveis

As covariâncias são os momentos das funções densidade de probabilidade de duas ou mais variáveis aleatórias que, junto com as médias e variâncias, formam o conjunto das grandezas estatísticas de maior interesse. Se X_1 e X_2 são variáveis aleatórias com f.d.p. $f(x_1, x_2)$, define-se momento central de ordem r em X_1 e ordem s em X_2 por meio da expressão

$$\begin{aligned} \mu_{rs} &= E[(x_1 - x_{1,0})^r (x_2 - x_{2,0})^s] = \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x_1 - x_{1,0})^r (x_2 - x_{2,0})^s f(x_1, x_2) dx_1 dx_2 , \end{aligned}$$

Tabela 6.5: Variâncias das estatísticas correspondentes aos momentos de ordens mais baixas de uma função densidade de probabilidade. No caso geral, elas estão em função dos vários momentos e, no caso normal, reduzem-se a potências da variância, σ^2 . Note que não se costuma usar m_2 para estimar a variância e o desvio padrão, mas sim a fórmula (1.10), que tem $N - 1$ no denominador e não N , de modo que a diferença com a estimativa não tendenciosa usual é muito pequena quando o número de dados é grande.

estatística		variância multiplicada por N	
nome	símbolo	qualquer f.d.p.	normal
média	\bar{x} ,	$\mu_2 = \sigma^2$	σ^2
variância	m_2	$\mu_4 - \mu_2^2$	$2\sigma^4$
desvio padrão	$\sqrt{m_2}$	$(\mu_4 - \mu_2^2)/(4\mu_2)$	$\sigma^2/2$
3º momento central	m_3	$\mu_6 - \mu_3^2 - 6\mu_4\mu_2 + 9\mu_2^3$	$6\sigma^6$
4º momento central	m_4	$\mu_8 - \mu_4^2 - 8\mu_5\mu_3 + 16\mu_3^2\mu_2$	$96\sigma^8$
assimetria	$\beta = \frac{m_3}{m_2^{3/2}}$		6
curtose	$\kappa = \frac{m_4}{m_2^2} - 3$		24
coef. de correlação	$\rho = \frac{\text{cov}(x_1, x_2)}{\sigma_1 \sigma_2}$		$\approx (1 - \rho^2)^2$

onde $x_{1,0}$ e $x_{2,0}$ são os valores esperados das variáveis X_1 e X_2 . Chama-se covariância ao momento de ordem 1 nas duas variáveis,

$$\text{cov}(x_1, x_2) = \mu_{11} = E[(x_1 - x_{1,0})(x_2 - x_{2,0})] \quad . \quad (6.27)$$

Em funções de mais do que duas variáveis, a covariância corresponde a uma extensão dessa definição, onde integra-se em todas as variáveis,

$$\text{cov}(x_i, x_j) = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} (x_i - x_{i,0})(x_j - x_{j,0}) f(x_1, \dots, x_n) dx_1 \dots dx_n \quad . \quad (6.28)$$

A desigualdade de Schwarz fornece a relação

$$\text{var}(x_i) \text{var}(x_j) \geq (\text{cov}(x_i, x_j))^2 \quad , \quad (6.29)$$

onde $\text{var}(x_i) = \sigma_i^2$ e $\text{var}(x_j) = \sigma_j^2$ são as variâncias em x_i e x_j , respectivamente. Quando as variáveis aleatórias são independentes, a covariância é nula. A desigualdade acima mais este último fato sugere o uso do parâmetro adimensional

$$\rho = \frac{\text{cov}(x_i, x_j)}{\sigma_i \sigma_j} \quad , \quad (6.30)$$

chamado de *coeficiente de correlação*, para medir a interrelação estatística entre duas variáveis aleatórias contínuas. A desigualdade (6.29) acima implica em

$$-1 \leq \rho \leq 1 \quad . \quad (6.31)$$

Assim, se as variáveis são estatisticamente independentes, $\rho = 0$.² Os valores $\rho = 1$ e $\rho = -1$ indicam variáveis completamente correlacionadas, ou seja, $X_2 = \alpha X_1 + \beta$ com α e β constantes, $\alpha > 0$ se $\rho = 1$ e $\alpha < 0$ se $\rho = -1$.

A variância do coeficiente de correlação entre variáveis com distribuição normal está na tabela 6.5, mas deve ser usada com muita cautela, especialmente quando o número de dados é pequeno — $N \approx 30$ deve ser considerado pequeno quando $|\rho| \approx 1$. Também a dependência com o quadrado de ρ exige cuidado na substituição da estimativa na fórmula da variância. Por exemplo, se a estimativa for $\hat{\rho} = 0,8$ com $N = 16$, a substituição de ρ por sua estimativa sugere $\sigma_\rho \approx 0,09$, que indica que $\rho = 0,4$ é pouco provável, mas no teste da hipótese $\rho = 0,4$, a substituição do valor hipotético dá $\sigma_\rho \approx 0,21$, que não permite descartar essa hipótese com um nível de significância razoável.

²A recíproca não é verdadeira, embora sejam raros os casos práticos em que a correlação é nula entre grandezas estatisticamente relacionadas.

6.12 Função geratriz

Define-se função geratriz da função de probabilidade de variável discreta como

$$G(t) = \sum_{n \in \Omega} p_n t^n \quad , \quad (6.32)$$

em que p_n é a probabilidade de observar-se o evento \tilde{n} ($n \in \mathbf{Z}$) e Ω é o espaço amostral.

Como exemplo, a função geratriz da f.p. Binomial é

$$G_{N,p}(t) = \sum_{n=0}^N \binom{N}{n} p^n (1-p)^{N-n} t^n = [pt + (1-p)]^N \quad , \quad (6.33)$$

calculada diretamente com o binômio de Newton.

As probabilidades podem ser recuperadas da função geratriz por meio das suas derivadas calculadas em $t = 0$,

$$p_k = \frac{1}{k!} \left. \frac{d^k G}{dt^k} \right|_{t=0} \quad , \quad (6.34)$$

para *todos* os elementos do espaço amostral. Daí vem o nome de *geratriz* e a propriedade de que a uma função geratriz corresponde uma única função de probabilidade.

Q6.6 Prove o resultado (6.34) acima a partir da definição da função geratriz, eq. (6.32).

Note as propriedades elementares da Função Geratriz:

$$G(0) \geq 0 \quad ;$$

$$G(1) = 1 \quad \text{e}$$

$$\frac{dG}{dt} \geq 0 \quad \text{para } 0 \leq t \leq 1 \quad .$$

Q6.7 (a). Mostre que se pode calcular o valor esperado, $\langle n \rangle$, a partir da função geratriz G por meio da expressão

$$\langle n \rangle = \left. \frac{dG}{dt} \right|_{t=1} \quad . \quad (6.35)$$

- (b). Mostre que a média do número de sucessos na f.p. Binomial é $\langle n \rangle = Np$, a partir da expressão acima.
- (c). Calcule a função geratriz da f.p. de Poisson e use-a, junto com a propriedade (6.35), para determinar o valor esperado.

A questão acima mostra que a relação entre a média e a primeira derivada da função geratriz é simples, relação (6.35). Em geral, os demais momentos da f.p. também podem ser calculados desta função com certa facilidade, o que é uma das suas utilidades. O momento de ordem r , μ'_r , com uma definição similar a da fórmula (6.21), mas adaptada para uma distribuição discreta, relaciona-se com as derivadas da função geratriz de ordens $1, 2, \dots, r$ no ponto $t = 1$,

$$\mu'_r = F \left\{ \left. \frac{dG}{dt} \right|_{t=1}, \dots, \left. \frac{d^r G}{dt^r} \right|_{t=1} \right\} .$$

A partir desses momentos, pode-se determinar os momentos centrais. Por exemplo, da relação

$$\left. \frac{d^2 G}{dt^2} \right|_{t=1} = \sum_n n(n-1)p_n t^{n-2} \Big|_{t=1} = \sum_n n^2 p_n - \sum_r n p_n = \langle n^2 \rangle - \langle n \rangle ,$$

pode-se calcular a variância usando a identidade (6.35) acima,

$$\sigma^2 = \langle n^2 \rangle - \langle n \rangle^2 = \left. \frac{d^2 G}{dt^2} \right|_{t=1} + \left. \frac{dG}{dt} \right|_{t=1} - \left(\left. \frac{dG}{dt} \right|_{t=1} \right)^2 . \quad (6.36)$$

Exemplo 6.2

Pode-se calcular a variância de uma distribuição binomial por meio das equações (6.36) e (6.33). A aplicação da equação (6.33) dá

$$\begin{aligned} \frac{dG(t)}{dt} &= N[pt + (1-p)]^{N-1} \cdot p \\ \frac{d^2 G(t)}{dt^2} &= N(N-1)[pt + (1-p)]^{N-2} \cdot p^2 . \end{aligned}$$

Substituindo $t = 1$ nas derivadas acima e usando a eq. (6.36), obtém-se

$$\sigma^2 = N(N-1)p^2 + Np - (Np)^2 = Np(1-p) ,$$

resultado que já obtido na seção 2.1.

6.13 Soma de variáveis aleatórias

É comum que a grandeza observada seja uma soma de variáveis aleatórias, com a forma

$$S_N = \sum_{i=1}^N X_i \quad . \quad (6.37)$$

Vamos buscar a solução desse problema em etapas, começando com o caso $N = 2$. A fim de evitar fórmulas carregadas com sub-índices, chamamos as variáveis aleatórias de V e W , de modo que

$$S_2 = S = V + W \quad ,$$

onde tanto V quanto W podem assumir qualquer valor inteiro não-negativo. Terminaremos esta seção com a soma da equação (6.37) com qualquer N e fica para a próxima seção o caso mais difícil, em que também N é uma variável aleatória.

Começamos por mostrar que a função geratriz de S é o produto das geratrizes de V e W , $G_V(t)$ e $G_W(t)$,

$$G_S(t) = G_V(t) \cdot G_W(t) \quad . \quad (6.38)$$

Para ajudar nesta demonstração, montamos a Tabela 6.6, que apresenta a relação entre as probabilidades associadas aos diversos valores de S , V e W .

Chamando de v_i a probabilidade que a variável V assuma o valor i , $i = 0, 1, 2, \dots$, a função geratriz de V , $G_V(t)$, é

$$G_V(t) = \sum v_i t^i \quad .$$

Do mesmo modo, se w_i é a probabilidade associada a W para os mesmos valores de i , a função geratriz de W é $G_W(t)$,

$$G_W(t) = \sum w_i t^i \quad .$$

O produto das funções geratrizes de V e W é

$$G_V(t) \cdot G_W(t) = \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} v_i w_j t^{i+j} \quad . \quad (6.39)$$

Ao isolar o termo que multiplica a potência n de t , que é

$$\sum_{i=0}^{\infty} \sum_{j=0}^{\infty} v_i w_j \delta_{n,i+j} = \sum_{i=0}^n v_i w_{n-i} \quad ,$$

Tabela 6.6: Relação entre as probabilidades de $S = n$, s_n , dada pela soma $S = V + W$, com as probabilidades associadas às variáveis aleatórias V e W , v_i e w_j respectivamente, onde i e j podem assumir qualquer valor inteiro não-negativo.

S	$P(S)$
0	$s_0 = v_0 w_0$
1	$s_1 = v_0 w_1 + v_1 w_0$
2	$s_2 = v_0 w_2 + v_1 w_1 + v_2 w_0$
3	$s_3 = v_0 w_3 + v_1 w_2 + v_2 w_1 + v_3 w_0$
	\vdots
n	$s_n = \sum_{i=0}^n v_i w_{n-i}$
	\vdots

ele pode ser identificado com a probabilidade de $S = n$, s_n , conforme detalhado na Tabela 6.6. Por isso, pode-se reescrever o duplo somatório da relação (6.39) como uma série de potências de t^n da seguinte maneira:

$$\sum_{i=0}^{\infty} \sum_{j=0}^{\infty} v_i w_j t^{i+j} = \sum_{n=0}^{\infty} s_n t^n = G_S(t) \quad ,$$

que, substituída na equação (6.39), prova a relação (6.38).

Esse resultado pode ser aplicado ao caso da soma da fórmula (6.37). O objetivo é determinar a f.p. de y , que chamaremos $h(y)$. Suporemos também que todos os X_i têm mesma distribuição e podem assumir qualquer valor inteiro não-negativo.

$$S_N = \sum_{i=1}^N X_i \quad , \quad N = \text{constante} \quad .$$

Quando as variáveis somadas têm todas a mesma função geratriz, a aplicação repetida do resultado (6.38) dá

$$G_{S_N}(t) = [G_X(t)]^N \quad . \quad (6.40)$$

Dispondo da função geratriz, pode-se usar as propriedades apresentadas na seção 6.12 para calcular todas as grandezas de interesse — média, desvio-padrão, probabilidade de obter o valor k em uma medição. A próxima seção fará bom uso do resultado que obtivemos.

Exemplo 6.3

Considere duas variáveis n e m com distribuições binomiais, cujos números de ensaios e probabilidades de sucesso são N_1, p_1 e N_2, p_2 , respectivamente. A função geratriz da soma dessas variáveis, $\ell = n + m$, é o produto das funções geratrizes de n e m :

$$G(t) = G_{N_1, p_1}(t)G_{N_2, p_2}(t) \quad .$$

Substituindo a função geratriz da distribuição binomial, equação (6.33), obtém-se

$$G(t) = (p_1 t + 1 - p_1)^N (p_2 t + 1 - p_2)^M \quad .$$

Desse resultado se conclui que, se $p_1 = p_2 = p$, então $G(t)$ é a função geratriz de uma binomial com parâmetros $M + N$ e p ; portanto, ℓ obedece a uma distribuição binomial. Se $p_1 \neq p_2$ então $G(t)$ não pode ser escrita na forma da equação (6.33) e, portanto, ℓ não obedece a uma distribuição binomial.

6.14 Soma de um número aleatório de variáveis aleatórias

Considere um espectro multicanal, em que o número total N de eventos detectados obedece a uma f.p. conhecida. Ao considerar um canal i isoladamente, verifica-se que cada evento observado cai nesse canal com probabilidade p e em qualquer outro canal diferente de i com probabilidade $q = 1 - p$. Nessa situação, o número de eventos y observado no canal i é uma soma de variáveis aleatórias em que o número de termos da soma (N) também é uma variável aleatória,

$$y = \sum_{i=1}^N X_i \quad , \quad (6.41)$$

em que $X_i = 1$ com probabilidade p e $X_i = 0$ com probabilidade $q = 1 - p$. Este problema é parecido com o da seção anterior, mas ao contrário do N fixo ali adotado, aqui N também é uma variável aleatória.³ Ao final, vamos encontrar

³Note que y não tem distribuição binomial somente porque N não é um número previamente determinado, uma vez que todas as demais condições são preenchidas: todos os

uma fórmula geral para variáveis aleatórias do tipo da equação (6.41), mas preferimos construir este resultado a partir de um exemplo particular para facilitar a compreensão da sua dedução.

Vamos definir

f_j : probabilidade de observar o resultado $X = j$, que supomos iguais para todos os X_i .

g_k : probabilidade de observar um número total de eventos $k = N$.

As funções geratrizes dessas frequências f_j e g_k são, então,

$$\phi(t) = \sum_{j=0}^{\infty} f_j t^j \quad \text{e} \quad (6.42)$$

$$\gamma(s) = \sum_{k=0}^{\infty} g_k s^k \quad . \quad (6.43)$$

Note que, na definição da geratriz, a variável usada é muda, assim tanto faz usar t , s , ou qualquer outra letra — entender isso com clareza é fundamental para compreender a dedução que segue.

A probabilidade de obter $y = M$ é h_M ,

$$h_M = \sum_{n=0}^{\infty} P(y = M \mid N = n) P(N = n) \quad . \quad (6.44)$$

A probabilidade do resultado $N = n$ (segundo fator da expressão acima) é, simplesmente, g_n . Já a probabilidade condicional de $y = M$ quando $N = n$ (primeiro fator) está ligada ao resultado da seção anterior e vamos simbolizá-la por

$$c_M^{(n)} = P(y = M \mid N = n) \quad , \quad (6.45)$$

cujas função geratriz será

$$\psi(t) = \sum_{M=0}^{\infty} c_M^{(n)} t^M \quad . \quad (6.46)$$

Na seção anterior, calculamos essa geratriz ψ a partir da *geratriz* de um X da soma, que, neste caso, é a função ϕ . Assim, conforme a relação (6.40), obtem-se

$$\psi(t) = [\phi(t)]^n \quad , \quad (6.47)$$

eventos têm mesma probabilidade de sucesso (cair no canal i) e a ocorrência de um evento não interfere no outro.

6.14. SOMA DE UM NÚMERO ALEATÓRIO DE VARIÁVEIS ALEATÓRIAS 201

com ϕ da expressão (6.42) porque, neste caso, $N = n$ define o número fixo de termos no somatório.

Reescrevendo a fórmula (6.44), chega-se a

$$h_M = \sum_{n=0}^{\infty} c_M^{(n)} g_n \quad ,$$

cujas função geratriz é

$$\beta(t) = \sum_{M=0}^{\infty} h_M t^M = \sum_{M=0}^{\infty} \left[\sum_{n=0}^{\infty} c_M^{(n)} g_n \right] t^M \quad .$$

Os somatórios podem ser rearranjados de modo a obter

$$\beta(t) = \sum_{n=0}^{\infty} g_n \left[\sum_{M=0}^{\infty} c_M^{(n)} t^M \right] \quad ,$$

que tem a vantagem de isolar, entre os parênteses retos, [], a função geratriz das probabilidades de obter M no somatório de n termos, equação (6.46), de modo que

$$\beta(t) = \sum_{n=0}^{\infty} g_n \psi(t) = \sum_{n=0}^{\infty} g_n [\phi(t)]^n = \gamma[\Phi(t)] \quad . \quad (6.48)$$

Esse resultado foi obtido usando a relação (6.47) e identificando que $\phi(t)$ faz as vezes de s na expressão (6.43). Assim, a função geratriz das probabilidades associadas à variável aleatória y definida pela fórmula (6.41) é a função geratriz associada à variável N (número de termos do somatório), mas com um *argumento* igual à função geratriz associada à variável X .

No caso típico dos espectros multicanais, X é uma variável binomial, cujo valor pode ser 0 com probabilidade q e 1, com probabilidade p , de modo que a geratriz é a da equação (6.33) com $N = 1$,

$$\phi(t) = (1 - p) + pt \quad .$$

Já a variável N obedece a uma f.p. de Poisson de média N_0 , com função geratriz

$$\gamma(t) = \exp\{N_0(t - 1)\} \quad (\text{da questão Q6.7-c}) \quad (6.49)$$

Reunindo esses resultados na expressão (6.48),

$$\phi(t) = \exp\{N_0(1 - p + pt - 1)\} = \exp\{N_0 p(t - 1)\} \quad ,$$

obtém-se um resultado que é exatamente da mesma forma da equação (6.49), com N_0 substituído por N_0p . Portanto,

$$h(y) = \frac{(N_0p)^y e^{-N_0p}}{y!} .$$

Assim, no caso particular do número de contagens em um canal de um espectro multicanal em que o número de contagens total N não é prefixado (com *tempo* prefixado, por exemplo, como é comum em experimentos), mostramos que ele se distribui como uma Poisson de média $y_o = N_0p$ (número total de contagens vezes a probabilidade de cada contagem cair no canal particular). Portanto, o desvio padrão do número de contagens no canal é $\sigma_y = \sqrt{y_o}$, independente da probabilidade p ser pequena ou grande.

6.15 Função característica – cumulantes

Para uma variável contínua x com f.d.p. $f(x)$, a função com propriedades similares à geratriz da equação (6.32) é a *Função Característica*, $\phi(t)$, que é a transformada de Fourier de $f(x)$, como já definimos no capítulo 2 e repetimos abaixo:

$$\phi(t) = \int_{-\infty}^{\infty} \exp(itx) f(x) dx . \quad (2.23)$$

Desde que a exponencial complexa tem módulo 1, a convergência absoluta dessa integral é assegurada, porque $f(x)$ é normalizada. No capítulo 2, também discutimos o uso da função característica na dedução da distribuição da média de dados gaussianos, seção 2.6, e também mostramos que se recupera f por meio da relação:⁴

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \exp(-itx) \phi(t) dt . \quad (2.24)$$

Nesta seção, vamos explorar outros aspectos da função característica que serão úteis na demonstração do Teorema Central do Limite na seção 6.16.

Q6.8 *Mostre que a função característica da f.d.p. de Cauchy*

$$f(x) = \frac{1}{\pi(1+(x-\theta)^2)} \quad \text{é} \quad \phi(t) = \exp(i\theta t - |t|) .$$

⁴Ao contrário da integral que define a função característica, a integral que recupera f a partir de ϕ pode não convergir separadamente nos dois extremos; se isso acontecer, deve-se calcular o valor principal de Cauchy da integral, algo a que estamos, aliás, habituados a fazer com transformadas de Fourier.

Q6.9 Mostre que a função característica da soma de duas variáveis aleatórias estatisticamente independentes u e v , $s = u+v$, é o produto das funções características dos termos da soma, $\phi_s(t) = \phi_u(t) \cdot \phi_v(t)$.

Q6.10 Considere uma grandeza x com a f.d.p. de Cauchy da fórmula da questão anterior e que $\{x_1, x_2\}$ é uma medida dessa grandeza. Mostre que a f.d.p. de $\hat{x} = \frac{x_1+x_2}{2}$ é a mesma de cada um dos dados do conjunto.

Uma função aparentada à função característica é a *Função Geratriz dos Cumulantes*,

$$FGC(t) = \ln[\phi(t)] \quad . \quad (6.50)$$

Os coeficientes κ_r da expansão em série de potências de (it) dessa função,

$$\ln[\phi(t)] = \sum_{r=0}^{\infty} \kappa_r \frac{(it)^r}{r!} \quad , \quad (6.51)$$

são chamados *cumulantes*, que se relacionam com os momentos:

$$\begin{aligned} \kappa_1 &= \mu'_1 = \mu = \text{média} \quad , \\ \kappa_2 &= \mu'_2 - (\mu'_1)^2 = \text{variância} \quad , \\ \kappa_3 &= \mu'_3 - 3\mu'_2 \cdot \mu'_1 + 2(\mu'_1)^3 \quad \text{etc.} \end{aligned}$$

Os cumulantes são uma alternativa aos momentos na descrição das distribuições, com duas vantagens: *i*) o cumulante de uma soma de variáveis aleatórias é a soma dos cumulantes de cada uma das variáveis⁵ e *ii*) a gaussiana tem apenas dois cumulantes não nulos, número mínimo que uma f.d.p. que têm variância finita pode ter.

Já sabíamos que a média e a variância têm a propriedade *i*), que caracteriza um cumulante: é uma grandeza que se *acumula* na adição de variáveis aleatórias. Infelizmente, porém, os cumulantes de ordem maior que 2 são muito complexos, o que dificulta seu uso em geral.

A propriedade *ii*) vem da função característica da gaussiana, já calculada no capítulo 2, fórmula (2.25), da qual obtém-se

$$\ln[\phi(t)] = x_o(it) + \sigma_o^2 \frac{(it)^2}{2!} \quad . \quad (6.52)$$

⁵Essa propriedade decorre do fato que a função característica da soma de duas variáveis aleatórias é o produto das suas respectivas funções características.

Nenhuma outra f.d.p. tem FGC tão simples. É por essa razão que a gaussiana serve frequentemente como aproximação das outras f.d.p.s em situações limites.

Exemplo 6.4

Vamos usar a técnica desenvolvida nesta seção para mostrar que a média de N dados que obedecem a distribuições uniformes tende a uma gaussiana quando N tende a infinito. Para isso, vamos usar um procedimento similar àquele usado no capítulo 2, na dedução da f.d.p. da média de dados gaussianos.

Considere N dados x_1, x_2, \dots, x_N que obedecem a funções densidade de probabilidade

$$\begin{aligned} g(x_i) &= 1, & 0 \leq x_i \leq 1 \\ &= 0, & x_i < 0 \text{ ou } x_i > 1 \end{aligned}$$

Definindo

$$y_i = \frac{x_i}{N},$$

a f.d.p. de y_i é

$$\begin{aligned} f(y_i) &= N, & 0 \leq y_i \leq \frac{1}{N} \\ &= 0, & y_i < 0 \text{ ou } y_i > \frac{1}{N} \end{aligned}$$

A média dos N dados x_i é

$$y = \frac{x_1 + x_2 + \dots + x_N}{N} = y_1 + y_2 + \dots + y_N,$$

que tem uma f.d.p. que pode ser escrita formalmente como

$$h(y) = \int \dots \int f(y_1)f(y_2)\dots f(y_N)\delta(y - y_1 - y_2 - \dots - y_N) dy_1 dy_2 \dots dy_N,$$

onde a função δ impõe a condição: y igual à soma dos N valores y_i , o que significa que as integrais não são independentes.

A estratégia a seguir é calcular a função característica de y e verificar que seu logaritmo, no limite em que $N \rightarrow \infty$, é igual ao logaritmo da função característica de uma gaussiana.

A função característica de y é

$$\phi_y(t) = \int e^{ity} h(y) dy.$$

Substituindo $h(y)$, obtem-se

$$\phi_y(t) = \int \dots \int e^{ity} f(y_1) f(y_2) \dots f(y_N) \delta(y - y_1 - y_2 - \dots - y_N) dy_1 dy_2 \dots dy_N dy \quad .$$

Integrando em y usando a propriedade da função δ e, depois, distribuindo o termo da soma dos y_i nas integrais de dy_i , temos

$$\phi_y(t) = N \int_0^{1/N} e^{ity_1} dy_1 \quad N \int_0^{1/N} e^{ity_2} dy_2 \dots \quad N \int_0^{1/N} e^{ity_N} dy_N \quad ,$$

onde foi usado que $f(y_i) = N$ entre 0 e $1/N$ e os limites de integração estão explícitos. Cada uma das N integrais acima dá

$$\frac{1}{it} (e^{it/N} - 1)$$

e, como todas têm o mesmo valor, o produto das integrais reduz-se a

$$\phi_y(t) = \left[\frac{N}{it} (e^{it/N} - 1) \right]^N \quad .$$

A exponencial $e^{it/N}$, quando expandida em série, dá

$$\phi_y(t) = \left[\frac{N}{it} \left(1 + \frac{it}{N} - \frac{t^2}{2N^2} - \frac{it^3}{6N^3} \dots - 1 \right) \right]^N$$

ou, ainda,

$$\phi_y(t) = \left(1 + \frac{it}{2N} - \frac{t^2}{6N^2} \dots \right)^N \quad ,$$

que, quando $N \gg 1$, fica

$$\phi_y(t) = 1 + \frac{it}{2} - \left(\frac{t^2}{24N} + \frac{t^2}{8} \right) + O(t^3) \quad .$$

Por outro lado, os termos da expansão em série da função $e^{\frac{it}{2} - \frac{t^2}{24N}}$ em torno de $t = 0$ até ordem 2 são idênticos aos da expressão acima e os termos seguintes semelhantes, com diferenças da ordem de $1/N$. Assim, para $N \gg 1$, pode-se aproximar

$$\phi_y(t) \cong e^{\frac{it}{2} - \frac{t^2}{24N}}$$

e, portanto,

$$\ln \phi_y(t) \cong \frac{it}{2} - \frac{t^2}{24N} \quad .$$

Comparando-se esta última expressão com a expressão para a função geratriz dos cumulantes da gaussiana (eq. (6.52)), concluímos que a f.d.p. da média de N dados distribuídos no intervalo $[0, 1]$ tende a uma gaussiana com média $\frac{1}{2}$ e variância $\frac{1}{12N}$.

6.16 O Teorema Central do Limite

Este é um teorema que admite diversas variantes. Ele estabelece que a f.d.p. de \bar{X}_N tende à normal, mesmo que as diversas variáveis aleatórias tenham f.d.p.s *diferentes*, inclusive com médias e desvios padrão diferentes. Se

$$\begin{aligned} X_1 &\text{ tem média } \mu_1 \text{ e variância } \sigma_1^2, \\ X_2 &\text{ tem média } \mu_2 \text{ e variância } \sigma_2^2, \\ &\vdots \\ X_i &\text{ tem média } \mu_i \text{ e variância } \sigma_i^2, \end{aligned}$$

então, no limite de N tendendo a infinito,

$$\bar{X}_N = \frac{X_1 + X_2 + \dots + X_N}{N} = \frac{X_1}{N} + \frac{X_2}{N} + \dots + \frac{X_N}{N} \quad (6.53)$$

tem f.d.p. **normal** de valor esperado

$$E(\bar{X}_N) = (\mu_1 + \mu_2 + \dots + \mu_N)/N$$

e variância

$$\text{var}(\bar{X}_N) = (\sigma_1^2 + \sigma_2^2 + \dots + \sigma_N^2)/N^2 \quad .$$

O Teorema não estabelece, porém, com que rapidez esse limite é atingido. A lógica da dedução é que, adicionando muitas variáveis aleatórias, os cumulantes de ordem maior que 2 são da ordem de $(\sigma^2)^{3/2}/\sqrt{N}$ e a f.d.p. de \bar{X} fica definida apenas pelos dois primeiros cumulantes e, portanto, é a gaussiana, conforme mostramos na seção anterior.

Apresentaremos uma demonstração na situação onde todas as variáveis X_i tem mesma f.d.p.. Todas as variáveis terão mesma média, que designaremos por μ , e mesma variância, designada por σ^2 , o que resulta em

$$E(\bar{X}_N) = \mu \quad \text{e}$$

$$\text{var}(\bar{X}_N) = \sigma^2/N \quad .$$

A função característica terá a mesma forma para todas as variáveis X_i e a função geratriz dos cumulantes para cada uma delas será, até ordem 3,

$$\text{FGC}_X(t) = \mu(it) + \sigma^2 \frac{(it)^2}{2!} + \kappa_3 \frac{(it)^3}{3!} + \dots \quad .$$

Para mostrar que o cumulante de ordem 3 da soma, κ_3 , pode ser ignorado frente ao segundo, é preciso explicitar de alguma forma essas grandezas numa escala definida. O mais fácil é definir uma variável aleatória normalizada,

$$\eta = \frac{\bar{X}_N - \mu}{\sigma/\sqrt{N}} \quad ;$$

essa transformação garante $E(\eta) = 0$ e $\text{var}(\eta) = 1$. A tese equivalente a provar é, portanto, que

$$\text{fdp}(\eta) = N(0, 1) \quad \textbf{(Tese)},$$

onde $N(x_0, \sigma_0)$ é a f.d.p. gaussiana de média x_0 e desvio padrão σ_0 . A variável aleatória η pode ser escrita

$$\eta = \sum_{i=1}^N \frac{X_i - \mu}{\sqrt{N}\sigma} = \sum_{i=1}^N \xi_i \quad ,$$

com a definição

$$\xi_i = \frac{X_i - \mu}{\sqrt{N}\sigma} \quad .$$

A FGC de cada variável ξ_i é

$$\text{FGC}_{\xi}(t) = \frac{\sigma^2}{2!} \cdot \left(\frac{it}{\sqrt{N}\sigma} \right)^2 + \frac{\kappa_3}{3!} \cdot \left(\frac{it}{\sqrt{N}\sigma} \right)^3 + \dots \quad ,$$

onde o termo linear em t desaparece porque $E(\xi_i) = 0$ e a FGC de ξ_i foi obtida a partir da FGC de X_i jogando o fator que multiplica X_i para a variável t (recorde que $\text{FGC}(t) = \ln(E(e^{itX}))$). Como a função geratriz dos cumulantes de uma soma é a soma das funções geratrizes dos cumulantes, no caso iguais para todos os ξ_i , temos

$$\text{FGC}_{\eta}(t) = N \left\{ \frac{\sigma^2}{2!} \cdot \left(\frac{it}{\sqrt{N}\sigma} \right)^2 + \frac{\kappa_3}{3!} \cdot \left(\frac{it}{\sqrt{N}\sigma} \right)^3 + \dots \right\} ,$$

$$\text{FGC}_\eta(t) = \frac{(it)^2}{2!} + \frac{\kappa_3}{3!} \frac{(it)^3}{\sqrt{N}\sigma^3} + \dots$$

Quando N vai a infinito, como κ_3 e σ são finitos, o termo de ordem 3 da expansão em série vai a 0, obtendo-se então que

$$\text{FGC}_\eta(t) \xrightarrow{N \rightarrow \infty} \frac{(it)^2}{2!},$$

que, por comparação com a FGC obtida a partir do logaritmo da função característica da gaussiana da fórmula (2.25), é a FGC da $N(0, 1)$, provando o teorema. Veja que não só o terceiro cumulante tende a zero, mas também os de ordem superior a 3, de maneira que o teorema vale mesmo que o terceiro cumulante de cada variável aleatória da soma for nulo. A demonstração no caso de soma de variáveis aleatórias diferentes segue o mesmo caminho, sendo que pode-se verificar que a soma dos terceiros cumulantes é menor que $N^{-1/2}$ multiplicado pelo maior dos terceiros cumulantes das várias variáveis envolvidas.

Este teorema justifica utilizar como variável aleatória aproximadamente normal a função

$$\alpha = \frac{\sum_{i=1}^N x_i - \frac{N}{2}}{\sqrt{\frac{N}{12}}} \quad (6.54)$$

onde x_i é uma variável aleatória com f.d.p. uniforme na faixa $[0, 1]$ e N é um número grande (veja o exemplo 6.4). O valor médio e a variância de α são, respectivamente,

$$\langle \alpha \rangle = 0 \quad \text{e} \quad \sigma_\alpha^2 = 1 \quad . \quad (6.55)$$

Assim, a função de variáveis aleatórias da expressão (6.54) tem f.d.p. aproximadamente normal, com média nula e desvio padrão 1, $N(0, 1)$.

Q6.11 Verifique as fórmulas (6.55).

Exemplo 6.5

Em geral, utilizava-se $N = 12$ na obtenção de aleatórios gaussianos pela fórmula (6.54), pela facilidade em extrair a raiz quadrada do denominador. Não é aconselhável, porém, usar este método para fornecer uma grande quantidade de números aleatórios gaussianos em computador,

porque os números gerados estão restritos à faixa $] - 6, +6[$, impondo limites que a gaussiana não tem. Esse truncamento da distribuição chama a atenção para o fato de que valores nas caudas da gaussiana vão aparecer mais raramente do que o esperado e compromete a adequação do algoritmo mesmo para alguns milhares de números aleatórios.

Uma aplicação deste teorema corresponde à determinação da amplitude máxima do ruído eletrônico num certo instante, quando ele provém da superposição de muitas fontes diferentes espalhadas aleatoriamente no tempo (figura 6.5a), cada uma produzindo na saída uma tensão $V(t)$ de forma definida (figura 6.5b).

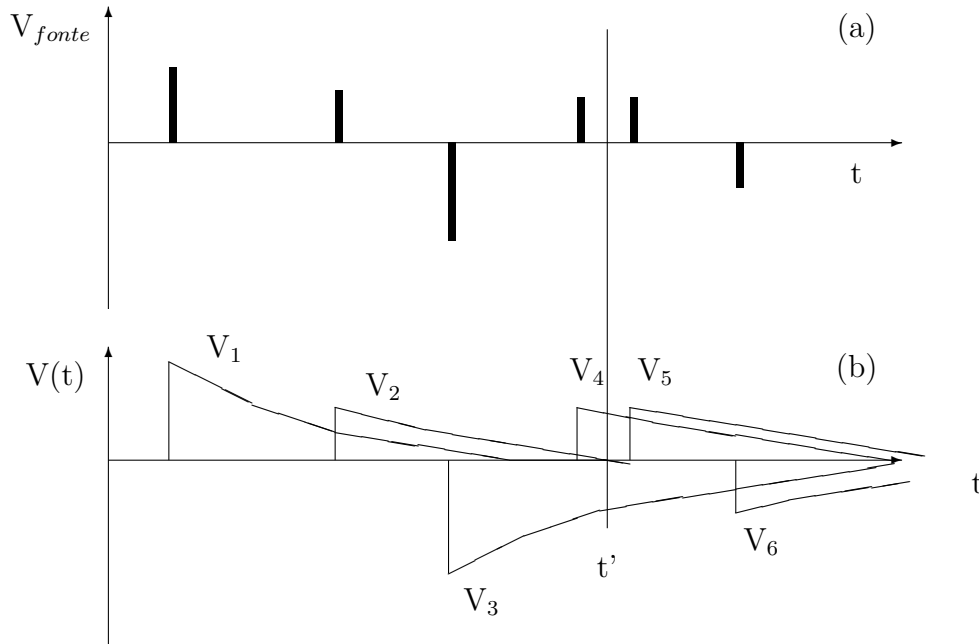


Figura 6.5: **a)** Pulsos rápidos, produzidos por uma fonte de ruído, distribuídos aleatoriamente no tempo. **b)** Componentes da tensão de ruído produzidas pelo circuito eletrônico, correspondentes aos diversos pulsos gerados pela fonte de ruído.

A tensão devida ao ruído será simplesmente a soma das componentes mostradas em **b)**.

Neste caso, para um dado instante t' , a tensão é a superposição de muitas variáveis aleatórias $V_j(t)$, o que torna o Teorema Central do Limite aplicável. Assim, a f.d.p. da voltagem de ruído é gaussiana.

6.17 Desigualdade de Chebyshev

Vamos obter aqui uma desigualdade que permite calcular valores limites das probabilidades associadas a intervalos do tipo

$$|x - x_0| > \ell \quad ,$$

onde $x_0 = \langle x \rangle$, válida para qualquer tipo de f.d.p. $f(x)$. Para simplificar a dedução, vamos supor $x_0 = \langle x \rangle = 0$ e generalizar a fórmula no final.

Da propriedade de normalização de $f(x)$ e com a definição do segundo momento, neste caso $\mu_2 = \mu'_2$, momento central porque a média é nula, tem-se

$$\int_{-\infty}^{\infty} \left(1 - \frac{x^2}{\ell^2}\right) f(x) dx = 1 - \frac{\mu_2}{\ell^2} \quad .$$

A integral pode ser reescrita como a soma de três integrais

$$\int_{-\infty}^{-\ell} \left(1 - \frac{x^2}{\ell^2}\right) f(x) dx + \int_{-\ell}^{\ell} \left(1 - \frac{x^2}{\ell^2}\right) f(x) dx + \int_{\ell}^{\infty} \left(1 - \frac{x^2}{\ell^2}\right) f(x) dx \quad .$$

A primeira e a última dão um resultado negativo ($|x| > \ell$) e a do meio dá um resultado positivo. Portanto, a integral de $-\infty$ a $+\infty$ é menor que a integral de $-\ell$ a ℓ . Assim,

$$1 - \frac{\mu_2}{\ell^2} \leq \int_{-\ell}^{\ell} \left(1 - \frac{x^2}{\ell^2}\right) f(x) dx \leq \int_{-\ell}^{\ell} f(x) dx \quad ,$$

em que a última desigualdade vale porque de $-x^2/\ell^2$ é sempre negativo. A última integral pode ser interpretada como a probabilidade de $|x| < \ell$ e conclui-se que

$$P(|x| < \ell) \geq 1 - \frac{\mu_2}{\ell^2} \quad .$$

Agora, $P(|x| < \ell) = 1 - P(|x| > \ell)$, o que fornece

$$P(|x| > \ell) \leq \frac{\mu_2}{\ell^2} \quad .$$

Se a média é não nula, devemos considerar “ x ” como a diferença entre o ponto e o valor médio, garantindo que μ_2 é o momento central (nesta nova situação, o momento central difere do momento em torno da origem, o que não acontecia com a hipótese de média nula). Assim, *em geral*,

$$P(|x - x_0| > \ell) \leq \frac{\mu_2}{\ell^2} \quad , \tag{6.56}$$

que é a desigualdade de Chebyshev.

Embora no caso prático os limites fornecidos por essa desigualdade sejam pouco precisos (muito grandes com probabilidades elevadas), ela é útil no desenvolvimento teórico da estatística porque é *quase* independente de hipóteses sobre a f.d.p., uma vez que o único requisito é a *existência* da variância. O resultado importante é que, independentemente da forma de $f(x)$, a probabilidade de x diferir de x_o por muitos desvios-padrão é pequena.

Para verificar a baixa precisão, note que se x é uma variável gaussiana, sabemos que

$$P(|x - x_o| > 2\sigma) = 0,045 = 4,5\% \quad ,$$

o que pode ser comparado ao resultado

$$P(|x - x_o| > 2\sigma) \leq \frac{1}{4} = 0,25 = 25\% \quad ,$$

da desigualdade de Chebyshev. A comparação da precisão fica ainda pior para ℓ maior.

6.18 A lei dos grandes números

É possível mostrar que a média de variáveis aleatórias $X_1, X_2, \dots, X_i, \dots$ independentes mas com mesma f.d.p. (portanto, todos X_i têm mesma média μ e variância σ^2 finita) distribui-se mais e mais concentradamente em torno do valor médio μ à medida que o número N de dados tende a infinito.

Definindo

$$\bar{X}_N = \frac{1}{N} \sum_{i=1}^N X_i \quad , \quad (6.57)$$

a lei dos grandes números estabelece que para qualquer $\epsilon > 0$,

$$\lim_{N \rightarrow \infty} P\{|\bar{X}_N - \mu| > \epsilon\} = 0 \quad . \quad (\text{Tese}) \quad (6.57')$$

As hipóteses foram estabelecidas no início da discussão e levam a

$$\text{var}(\bar{X}_N) = \frac{\sigma^2}{N} \quad .$$

Usando a desigualdade de Chebyshev da seção anterior,

$$P\{|\bar{X}_N - \mu| > \epsilon\} \leq \frac{\sigma^2}{N\epsilon^2} \quad .$$

Para um dado ϵ – no caso, a diferença máxima tolerada entre \bar{X} e μ – podemos aumentar N até que P seja tão pequena quanto desejarmos, provando a tese.

EXERCÍCIOS

6.1. Mostre que a média, a mediana e a moda de uma distribuição *normal* são iguais.

6.2. Considere a f.d.p.

$$f(x) = \begin{cases} 0 & x < 0 \\ \lambda e^{-\lambda x} & x \geq 0 \end{cases}$$

onde $\lambda > 0$. Mostre que a média, a mediana e a moda valem $1/\lambda$, $\ln 2/\lambda$ e zero, respectivamente.

6.3. Considere a função densidade de probabilidade de duas variáveis, x_1 e x_2 , centradas em zero ($\langle x_i \rangle = 0$), cujas variâncias são σ_1^2 e σ_2^2 , respectivamente, e a covariância é $\rho\sigma_1\sigma_2$.

(a) Determine a f.d.p. marginal de x_1 , verificando que seu desvio padrão continua sendo σ_1 .

(b) Determine a f.d.p. marginal de x_1 para x_2 fixado em x_F e mostre que seu desvio padrão é $\sqrt{1-\rho^2} \sigma_1$ e determine o novo valor de $\langle x_1 \rangle$.

6.4. Mostre que a função geratriz da distribuição de Poisson com parâmetro a é

$$\alpha(t) = e^{a(t-1)} \quad .$$

6.5. Usando o fato que a função geratriz de duas variáveis aleatórias é o produto das funções geratrizes correspondentes, mostre que a soma de duas variáveis que obedecem a distribuições de Poisson com parâmetros a e b é uma variável aleatória que obedece a uma distribuição de Poisson com parâmetro $a + b$.

6.6. Considere um experimento no qual o número de eventos observados, y , é igual à soma de eventos originários de N fontes independentes,

$$y = \sum_{i=1}^N x_i \quad ,$$

onde x_i obedece a uma distribuição de Poisson com parâmetro a e N obedece também a uma distribuição de Poisson, mas com parâmetro b .

- (a) Mostre que y não obedece a uma distribuição de Poisson.
- (b) Determine o valor esperado de y , $\langle y \rangle$.
- (c) Determine a variância de y e verifique que ela é maior do que $\langle y \rangle$.

6.7. A função geratriz de uma distribuição binomial com parâmetros N e p é

$$(pt + 1 - p)^N \quad .$$

Mostre que no limite em que $p \rightarrow 0$ e $N \rightarrow \infty$ com o produto $Np = a$ finito, a função geratriz acima fica

$$e^{a(t-1)} \quad ,$$

que corresponde a uma distribuição de Poisson.

6.8. A função geratriz da soma $s = v_1 + v_2$ de duas variáveis v_1 e v_2 que obedecem a distribuições de probabilidade binomiais com parâmetros N, p_1 e M, p_2 é

$$G = (p_1t + 1 - p_1)^N (p_2t + 1 - p_2)^M \quad .$$

Mostre que a média da variável s é igual à soma das médias das duas variáveis. Mostre que a variância é a soma das variâncias. (Sugestão: use as equações (6.35) e (6.36) e note que essa função geratriz só corresponde à de uma distribuição binomial se $p_1 = p_2$, veja exemplo 6.3.)

6.9. Simule M dados, onde cada um deles é a média de m dados uniformemente distribuídos entre 0 e 1. Faça isso para $m = 1, 2, 4$ e 10, histogramando as M simulações. (Use M da ordem de 100 ou mais.) Visualmente, verifique que os histogramas vão ficando cada vez mais parecidos com uma distribuição gaussiana à medida em que m aumenta. Repita a simulação, mas agora com dados que obedecem a distribuições binomiais com parâmetros $p = 0, 1$ e $N = 10, 100$ e 1000. Simule o número de sucessos, n , e calcule o valor $p_{est} = n/N$. Histograme os valores obtidos de p_{est} . Verifique, ainda visualmente, que os histogramas também tendem a uma gaussiana.

Bibliografia

- [Arfken] Mathematical Methods for Physicists, G.Arfken & H.Weber, Academic Press, 4ª edição (1995)
- [Bard] Nonlinear Parameter Estimation, Yonathan Bard, Academic Press (1974)
- [Benzécri] Histoire et Préhistoire de l'Analyse des Données, J.P.Benzécri, Ed. Bordas, Paris 1982
- [Bevington] Data Reduction and Error Analysis for the Physical Sciences, P.Bevington, McGraw-Hill, 1969
- [Birge] The calculation of errors by the method of least squares, Raymond T. Birge, Phys Rev *40 (1932) 207-227*
- [Conover] Practical Nonparametric Statistics, W.J.Conover, John Wiley & Sons Inc. 1971
- [CRC] Handbook of Tables for Probability and Statistics, CRC
- [Eadie] Statistical Methods for Physicists, W.T.Eadie et al., North Holland Pub.Co. 1971
- [Escoubes] Experimental Signs Pointing to a Bayesian Instead of a Classical Approach for Experiments with Small Number of Events, B.Escoubes, S.De Unamuno e O. Helene, Nuclear Instruments and Methods A257(1987)346
- [Feller] Feller, An Introduction to Probability Theory and its Applications, John Wiley, 2ª Ed. (1957)
- [Feynman] Lectures on Physics Vol.I, Chap.6, Feynman Leighton & Sands

- [Firestone] Analysis of α , β , and γ ray emission probabilities, R.B. Firestone, Nuclear Instruments and Methods A286(1990)584
- [Forbes] Forbes, Eric G., *Gauss and the Discovery of Ceres*. Journal for the History of Astronomy. 2 (1971) 195-199.
- [Frieden] Fisher's Information as the basis for the Schrödinger wave equation, B. Roy Frieden, Am. J.Phys. 57(1989)11
- [Geraldo] L.P. Geraldo e D.L Smith, Nuclear Instruments and Methods A290(1990)499
- [Grosser] Morton Grosser, The Discovery of Neptune, Harvard University Press, Cambridge, Massachusetts (1962)
- [Gray] C.G.Gray, Am. J.Phys. 59(1991)282
- [Guimarães-Filho] Z.O. Guimarães-Filho e O. Helene, One Step Self-Calibration Procedure in Gamma-Ray Energy Measurements. Brazilian Journal of Physics, v. 33, n.2, (2003) 280-281.
- [Lyons] How to combine correlated estimates of a single physical quantity, L.Lyons, D.Gibaut e P. Clifford, Nuclear Instruments and Methods A270(1988)110
- [Helene] Tratamento Estatístico de dados em Física Experimental, O.Helene, V. R. Vanin, Ed. Edgard Blücher, 2ª Ed., 1991
- [Helene 83] Upper Limit of Peak Area, O.Helene, Nuclear Instruments and Methods 212(1983)319
- [Helene 84] Errors in Experiments with Small Number of Events, O.Helene, Nuclear Instruments and Methods 228(1984)120
- [Helene 91b] Determination of the Upper Limit of Peak Area, O.Helene, Nuclear Instruments and Methods A300(1991)132
- [Helene 91] O que é uma medida?, O. Helene, Shan.P.Tsai, R.P.Teixeira, preprint IFUSP/P-854 (1990) e Revista de Ensino de Física, Vol.13 p.12, SBF (1991).

- [Helene 93] O.Helene and V.R.Vanin, Nuclear Instruments and Methods A335(1993)227
- [Helene 2013] O. Helene, Método dos Mínimos Quadrados com formalismo matricial, Editora Livraria da Física, São Paulo, 2ª edição (2013).
- [James] A review of pseudorandom number generators, F.James, Computer Physics Communications 60(1990)329-344
- [Kendall] The Advanced Theory of Statistics, M.Kendall, A.Stuart & J.K.Ord, Charles Griffin & Company Limited, London
- [Magalhães] Noções de Probabilidade e Estatística, Marcos N. Magalhães e Antonio Carlos P. Lima, Editora da Universidade de São Paulo - EDUSP, 2011
- [Mannhart] A Small Guide to Generating Covariances of Experimental Data, Report PTB-FMRD 84, Berlin, 1981. ISSN 0341-6666
- [Marquardt] An Algorithm for Least-Squares Estimation of Nonlinear Parameters, D. Marquardt, SIAM J. Appl. Math. 11, 431-441, 1963
- [Merzbacher] Quantum Mechanics, E.Merzbacher, John Wiley & sons, New York 1961
- [Mises] Probability, Statistics and Truth, R.von Mises, Dover, 1955
- [Moralles] M.Morales, P.R.Pascholati, V.R.Vanin and O.Helene, Applied Radiation and Isotopes 46-2(1995)133
- [Mucciolo] E.R.Mucciolo and O.Helene, Nuclear Instruments and Methods A256(1987)153
- [Noether] Introdução à Estatística – Uma abordagem não paramétrica, G.E.Noether, Guanabara Dois, 1983
- [Smith] D.L. Smith, Nuclear Instruments and Methods A257(1987)361
- [Stigler] *Gauss and the Invention of Least Squares*. Stephen M. Stigler, Annals of Statistics, 9 (1981) 465-474 - doi:10.1214/aos/1176345451
- [Vanin 1989] V.R.Vanin e M.Aiche, Nuclear Instruments and Methods A284(1989)452

- [Vanin 1997] V.R.Vanin, G.Kenchian, M.Morales, O.Helene e P.R. Pascholati, Nuclear Instruments and Methods A391(1997)338
- [Vuolo] Fundamentos da Teoria de Erros, J.H.Vuolo, Ed. Edgard Blücher, 1992
- [Youden] Statistical Methods for Chemists, W.J.Youden, John Wiley 1951
- [Zar] J.H. Zar, Appl. Statist. 27(1978)n.3, 280-290