

ANÁLISE ESTATÍSTICA DE MEDIDAS EM CIÊNCIAS EXATAS

Vito R. Vanin, Philippe Gouffon, Otaviano Helene

Março 2023

Capítulo 2

Funções de probabilidade comuns e transformações

Embora nem sempre conheçamos a *verdadeira* função de probabilidade que governa uma medida, existem algumas funções que fornecem aproximações adequadas a muitos experimentos. Este capítulo apresenta as funções de probabilidade **Binomial** e de **Poisson**, de variável discreta, e as funções densidade de probabilidade (f.d.p.) **Normal**, **Multinormal**, de χ^2 e **Uniforme**, de variável contínua. Aproveitamos para introduzir propriedades e métodos gerais de lidar com essas funções, em particular definiremos a *função característica* de uma f.d.p. e mostraremos uma forma de lidar com transformações de variáveis aleatórias. Esses recursos serão aplicados na determinação das f.d.p.s das estimativas da média e da variância.

2.1 A função de probabilidade binomial

A probabilidade de observar-se n eventos de um certo tipo, em N tentativas, é dada pela f.p. Binomial,

$$P_{N,p}(n) = \binom{N}{n} p^n (1-p)^{N-n} \quad , \quad (2.1)$$

onde p é a probabilidade de observação do evento do tipo em que estamos interessados. O símbolo

$$\binom{N}{n} = \frac{N!}{n!(N-n)!} = \frac{N \cdot (N-1) \cdots (N-n+1)}{1 \cdot 2 \cdots n} \quad ,$$

corresponde ao número de combinações de N objetos n a n . Na expansão do binômio de Newton, $(p + q)^N$, $\binom{N}{n}$ é o coeficiente do termo $p^n q^{N-n}$, de onde vem o nome “binomial”. As duas situações já estudadas neste texto em que esta f.p. se aplica correspondem à da evaporação de metal sobre plástico (seção 1.1) e à do cálculo de certas probabilidades em intervalos construídos por métodos não paramétricos (seção 1.11).

As **hipóteses** necessárias para a aplicabilidade da f.p. binomial são:

- i) Há apenas dois tipos de eventos, mutuamente exclusivos e exaustivos.
- ii) A probabilidade de cada um dos tipos de evento é fixa, independente das demais observações.
- iii) O número total de observações é fixo e conhecido *a priori*, ou seja, ele não é uma variável aleatória.

Um experimento em que o resultado obedece às hipóteses **i)** e **ii)** é chamado *Ensaio de Bernoulli* e o resultado de N repetições é descrito pela f.p. binomial, cujas propriedades exploramos nesta seção.

Quando q representa a probabilidade do evento que *não* é do tipo que interessa, a hipótese **i)** acima corresponde a

$$p + q = 1 \quad . \quad (2.2)$$

A partir da função de probabilidade da fórmula (2.1), calcula-se (veja Exemplo 2.1) o valor médio da variável aleatória,

$$n_0 = \langle n \rangle = \sum_{n=0}^N n \cdot P_{N,p}(n) = Np \quad . \quad (2.3)$$

Note que $\langle n \rangle$, nos casos reais, raramente será um número inteiro. Com um pouco de reflexão é possível convencer-se que não há qualquer motivo para $\langle n \rangle$ pertencer ao espaço amostral. Aliás, se lembrarmos que ele é a média ponderada de números inteiros com pesos fracionários, dados pela função de probabilidade, concluiremos que $\langle n \rangle$ só será inteiro acidentalmente.

A variância também pode ser calculada a partir da f.p.,

$$\sigma_0^2 = \langle (n - n_0)^2 \rangle = \sum_{n=0}^N (n - n_0)^2 \cdot P_{N,p}(n) = Npq \quad , \quad (2.4)$$

donde se deduz o desvio padrão

$$\sigma = \sqrt{Npq} = \sqrt{Np(1 - p)} \quad . \quad (2.5)$$

Exemplo 2.1

A propriedade de normalização da f.p. binomial é demonstrada com a aplicação do binômio de Newton e da identidade (2.2):

$$\sum_{n=0}^N P_{N,p}(n) = \sum_{n=0}^N \binom{N}{n} p^n q^{N-n} = (p+q)^N = 1^N = 1 \quad .$$

O valor médio de n é calculado como

$$\begin{aligned} n_0 &= \sum_{n=0}^N n \cdot P_{N,p}(n) = \sum_{n=0}^N \frac{nN!}{n!(N-n)!} p^n q^{N-n} \\ n_0 &= \sum_{n=1}^N \frac{N!}{(n-1)!(N-n)!} p^n q^{N-n} \quad , \end{aligned}$$

onde, na última expressão, explicitou-se que a somatória começa em $n = 1$ porque o fator n anula o termo $n = 0$ da soma.

Com a mudança de variável $n' = n - 1$ e a definição $N' = N - 1$, chega-se a

$$n_0 = \sum_{n'=0}^{N'} \frac{N(N'!)}{n'!(N'-n')!} p p^{n'} q^{N'-n'} \quad .$$

Na somatória acima, podemos fatorar N e p , porque independem de n' . Com essa fatoração, obtém-se

$$n_0 = Np \sum_{n'=0}^{N'} \frac{(N'!)}{n'!(N'-n')!} p^{n'} q^{N'-n'} = Np(p+q)^{N'} = Np \quad .$$

A fórmula (2.4) pode ser deduzida de maneira semelhante, com um pouco mais de trabalho, e fica para o exercício 2.1.

Vamos detalhar um exemplo, que ocorre frequentemente na análise de histogramas de dados experimentais. Considere que observamos uma grandeza x , de valor verdadeiro x_0 , e $f(x)$ é a f.d.p. das observações. A partir da medida $\{x_i, i = 1, \dots, N\}$, construímos o histograma dos dados. Dessas informações, podemos deduzir quantos dados esperamos obter no intervalo $[x_a, x_b]$. A probabilidade de *um* dado cair no intervalo $[x_a, x_b]$, p , é determinada pela f.d.p. $f(x)$,

$$p = \int_{x_a}^{x_b} f(x)dx$$

A partir daí, estimamos que, dentre N dados, cerca de Np cairão nesse canal $]x_a, x_b]$.

Embora o número de dados nesse canal, n , possa ser o inteiro mais próximo a Np , o valor obtido pode ser diferente e existe uma barra de incerteza associada a ele, cuja determinação requer a f.p. dessa variável aleatória, n . Para calcular essa barra de incerteza, então, começaremos por provar que a f.p. de n é a binomial, verificando se as hipóteses i), ii) e iii) acima são satisfeitas por n .

A hipótese i) é verificada, desde que cada dado ou pertence a $]x_a, x_b]$ ou está fora deste intervalo, havendo portanto apenas dois eventos possíveis que se excluem mutuamente — a ocorrência de um deles garante que o outro não ocorre. Além disso, o dado é obrigatoriamente um número real e a união dos dois intervalos corresponde ao espaço amostral — são, portanto, eventos exaustivos, no sentido que qualquer evento pode ser encaixado em um dos dois eventos definidos, não havendo necessidade de definir um terceiro para encaixar qualquer evento particular. Costumamos chamar de p a probabilidade do evento que nos interessa (por isto, habitualmente chamado de *sucesso*) e de q a probabilidade do evento que não nos interessa, isto é, $x \leq x_a$ ou $x > x_b$.

A condição ii) depende apenas da constância de $f(x)$ ao longo da medida, hipótese que sempre fazemos ao analisar dados desse tipo. É essencial que usemos instrumentos e processos que não variem durante a medida.

Como o número total de dados no histograma corresponde ao número de dados obtidos, o item iii) também é verificado. Note que o número de dados deve ser fixado *antes* de começarmos a medida. Se escolhermos como critério: *interrompe-se a medida quando o número de dados no intervalo $]x_b, x_c]$ for igual a M* , a f.p. do número de dados no intervalo $]x_a, x_b]$ não será mais binomial¹!

Assim, o número médio de dados em $]x_a, x_b]$ é, simplesmente, o valor médio de n , Np , e a barra de incerteza, que usualmente corresponde ao desvio padrão σ , é dada pela expressão (2.5), $\sigma = [Np(1 - p)]^{1/2}$.

Q2.1 Considere um histograma de uma grande quantidade de dados que se espalham por muitos canais. Em um determinado canal, o número de dados obtidos é n .

¹Esta questão será tratada detalhadamente na seção 6.14, quando disporemos de mais recursos para resolvê-la.

a) Mostre que a barra de incerteza em n é dada, aproximadamente, por \sqrt{n} , e que a aproximação tende a melhorar com o aumento no número de canais do histograma e do valor de n .

b) A aproximação do item a) falha quando a maioria dos dados cai em um mesmo canal e o número total de eventos no histograma é um número N conhecido a priori. Encontre uma aproximação melhor para o desvio-padrão de n desse canal que concentra os eventos.

Exemplo 2.2

EXEMPLO NUMÉRICO

Vamos considerar o histograma da figura 1.3, seção 1.1. Queremos avaliar a barra de incerteza associada ao *número de casas* com área igual a b^2 que foram observadas recobertas por 2 átomos. Em cada canal do histograma, a abscissa identifica o número dos átomos, m , que recobrem uma casa da folha (um inteiro), enquanto a ordenada dá o número de casas observadas com esse específico valor de m . Se interpretamos p como a probabilidade de observar uma casa com $m = 2$, $p = P(m = 2)$ (probabilidade de ocorrer o evento $m = 2$), a discussão acima se aplica. O valor de p estimado a partir do número de casas com $m = 2$ no histograma é $p = 5/15 = 0,33$, mas na próxima seção encontraremos uma estimativa melhor, $p = P(m = 2) = 0,27$, que será obtida a partir da f.p. da variável aleatória a que se refere o histograma².

Então, quando n é o número de casas observadas com $m = 2$, a função de probabilidade de n é a binomial com $N = 15$ (número total de casas/eventos fixado pelo arranjo experimental) e $p = 0,27$. O número médio esperado é $15 \cdot 0,27 = 4,1$, com desvio padrão $(15 \cdot 0,27 \cdot 0,73)^{1/2} = 1,7$ ou, em notação resumida, escrevendo entre parênteses o desvio padrão em unidades do último algarismo da média,

$$\bar{n}_{m=2} = 4,1(17) \quad .$$

A interpretação desse resultado é a seguinte. Quando se repete o experimento e $N = 15$ eventos (=casas) são observados, o número de casas com

²A próxima seção traz a f.p. que rege esse histograma, a Poisson. Este exemplo mistura as duas f.p.s, de Poisson e Binomial, de modo que os dois aspectos com que se pode ler o histograma são indissociáveis e fica impossível evitar este zig-zag – a discussão de uma das f.p.s tem que vir antes e usará a da seção seguinte.

$m = 2$ será provavelmente 3, 4 ou 5. Já o número de casas observadas com $m = 2$ será $n = 1, 2, 6$ e 7 vai ocorrer em muitas das repetições, embora com menor frequência que 3, 4, ou 5; $n = 0, 8$ e 9 poderão acontecer, mas raramente, e $n = 10$ ou mais será muito pouco provável. Essa descrição qualitativa, baseada na ideia de que eventos diferentes da média por menos de um desvio padrão são os mais comuns, entre 1 e 2 desvios acontecem, entre 2 e 3 são raros e mais que 3 desvios-padrão são muito raros e praticamente não acontecem quando o número total de observações não é grande, pode ser substituída por uma descrição precisa, porque conhece-se a f.p. da grandeza e as probabilidades podem ser estimadas,

$$\begin{aligned} P(n = 0) &= (1 - 0,27)^{15} = 0,009 = 0,9\% \\ P(n = 1) &= 15 \cdot 0,27 \cdot (1 - 0,27)^{14} = 0,049 = 4,9\% \\ P(n = 4) &= \frac{15 \cdot 14 \cdot 13 \cdot 12}{1 \cdot 2 \cdot 3 \cdot 4} \cdot 0,27^4 \cdot (1 - 0,27)^{11} = 0,228 = 22,8\% \end{aligned}$$

e assim por diante.

2.2 A função de probabilidade de Poisson

2.2.1 A Poisson como limite da Binomial

Quando tende ao infinito o número N de ensaios de Bernoulli em que a probabilidade de um sucesso, p , tende a zero, mas o número médio de sucessos, Np , tende a uma constante, a , a f.p. resultante é a de Poisson:

$$P_a(n) = \frac{a^n e^{-a}}{n!} . \quad (2.6)$$

Já vimos uma situação em que essa condição limite ocorre, que é o caso da metalização explorado na seção 1.1. Naquele caso, o número de átomos evaporados é muito grande e a probabilidade de cada átomo evaporado grudar numa certa casa em particular é muito pequena, mas o número médio de átomos grudados numa casa é finito. Este é o exemplo que será detalhado adiante, nesta seção.

Uma maneira de deduzir essa f.p. é passar a expressão (2.1), f.p. binomial, ao limite $p \rightarrow 0$ e $N \rightarrow \infty$ com $Np = a$, o que fica como exercício, questão 2.2.

As expressões da média e da variância da binomial, fórmulas (2.3) e (2.4), também podem ser passadas ao mesmo limite, e deduz-se que, para uma variável aleatória com f.p. de Poisson, a média do número de observações n é

$$n_0 = \langle n \rangle = a \quad (2.7)$$

e a variância de n ,

$$\sigma^2 = a \quad (2.8)$$

Perceba que a f.p. de Poisson tem a propriedade notável de depender de um único parâmetro, a média a .

As condições de validade da f.p. de Poisson são semelhantes às da binomial, exceto pelo número de eventos observados:

- i) Os eventos são independentes.
- ii) Todos eventos têm a mesma probabilidade de sucesso.
- iii) O número total de eventos é sempre muito maior que o número de sucessos em qualquer medição.

Um fenômeno que dá origem a eventos que se distribuem conforme a f.p. de Poisson é a *desintegração radioativa* de fontes de meia-vida muito maior que o intervalo de tempo considerado, porque: o número de átomos radioativos numa fonte é, na maioria dos casos, muito grande; a probabilidade de um átomo desintegrar-se em um certo intervalo de tempo é constante no tempo, não depende do átomo particular e o número de átomos que se desintegram nesse intervalo é muito menor que o total; e a desintegração de um átomo não altera a probabilidade de desintegração dos demais.

Q2.2 Mostre que quando $N \rightarrow \infty$ e $p \rightarrow 0$ com $Np = a$, a função de probabilidade binomial tende à função de probabilidade de Poisson. Você precisará substituir $p = a/N$ e aproximar $N \cdot (N-1) \cdots (N-n+1) \cong N^n$ na equação 2.1.

Q2.3 Mostre que a função de probabilidade da expressão (2.6) está corretamente normalizada. Deduza o resultado (2.7) a partir do cálculo da média do número de eventos observados por meio de

$$\langle n \rangle = \sum_{n=0}^{\infty} n \frac{a^n e^{-a}}{n!} = \sum_{n=1}^{\infty} n \frac{a^n e^{-a}}{n!} \quad .$$

Exemplo 2.3

Para determinar a variância a partir da f.p. da expressão (2.6) e da definição, usamos a identidade

$$\sigma^2 = \langle (n - n_0)^2 \rangle = \langle n^2 \rangle - n_0^2 \quad ,$$

que representa para variáveis discretas a relação equivalente à (1.23) da seção 1.6, que você deduziu na questão Q1.1 para variáveis aleatórias contínuas. Calculamos

$$\langle n^2 \rangle = \sum_{n=0}^{\infty} n^2 \frac{a^n e^{-a}}{n!} = \sum_{n'=0}^{\infty} (n' + 1) \frac{aa^{n'} e^{-a}}{n'!} \quad ,$$

onde $n' = n - 1$. Agora, fatorando a e distribuindo o termo $n' + 1$, encontramos

$$\langle n^2 \rangle = a \sum_{n'=0}^{\infty} n' \frac{a^{n'} e^{-a}}{n'!} + ae^{-a} \sum_{n'=0}^{\infty} \frac{a^{n'}}{n'!} \quad ,$$

onde o primeiro somatório é $\langle n' \rangle = a$, conforme expressão (2.7) acima, e o segundo é a expansão em série de potências de e^a . Então

$$\langle n^2 \rangle = a^2 + a \quad .$$

Com esse resultado substituído na identidade deduzida no início deste exemplo, calculamos finalmente

$$\sigma^2 = \langle n^2 \rangle - n_0^2 = a^2 + a - a^2 = a \quad .$$

No caso do exemplo da metalização da seção 1.1, por menor que seja a massa de material evaporado, o número de Avogadro é tão grande — $6 \cdot 10^{23}$ átomos por mol — que o número total de átomos evaporados é muito grande, o que satisfaz à condição **iii)** acima. A probabilidade do átomo grudar numa casa de área b^2 específica é muito pequena dada a pequenez de b , da ordem de uma fração de nanômetro, comparada com a dimensão da placa: uma placa de alguns mm^2 conterá da ordem de 10^{14} casas, garantindo uma probabilidade da ordem de 10^{-14} de um átomo grudar numa casa determinada. Assim, as condições **i)** e **ii)** acima também serão satisfeitas, desde que um átomo depositado não altere a probabilidade de outro átomo se depositar nas vizinhanças, o que é uma hipótese que precisa ser verificada caso a caso, mas que vamos aceitar aqui, a fim de completar o exemplo.

Para estimar a f.p. que governa o histograma da figura 1.3, precisamos estimar um único parâmetro, a média, que simbolizamos por a . Somando o número total de átomos grudados, \mathcal{N} , representados naquele histograma, obtemos

$$\mathcal{N} = 1 \cdot 0 + 5 \cdot 1 + 5 \cdot 2 + 3 \cdot 3 + 1 \cdot 4 = 28 \text{ átomos}$$

e o número de casas consideradas, N , é

$$N = 1 + 5 + 5 + 3 + 1 = 15 \text{ casas.}$$

O número médio de átomos ocupando uma casa é, simplesmente,

$$\hat{a} = \mathcal{N}/N = 28/15 = 1,87.$$

Adicionamos um chapéu sobre o símbolo do valor verdadeiro da média, a , para marcar a diferença entre ele e sua *estimativa* \hat{a} — estimativa não é valor verdadeiro.

A figura 2.1 apresenta, em linha tracejada, a f.p. de Poisson de média 1,87, multiplicada por N , comparada aos dados experimentais da figura 1.3, em linha cheia. A multiplicação por N é necessária para a comparação porque o histograma corresponde ao *número* de casas onde foi observado um número particular de átomos grudados e não à *probabilidade* de ocupação; como o *número* de casas com um dado m tem uma f.p. binomial, conforme discutimos na seção anterior, usamos a relação (2.4) $n_0 = Np$.

O histograma calculado, dado pela função $F(n) = 15P_{\hat{a}}(n)$, não é idêntico ao histograma experimental, e nem poderia, devido à flutuação estatística. Em princípio, tanto \hat{a} como o número de casas ocupadas por um número definido de átomos estão afetados por incertezas. Porém, a estimativa \hat{a} é bastante mais precisa que o número de casas que foram observadas ocupadas por um número definido de átomos, por basear-se no número total de casas (e de átomos), que é bem maior. Assim, você pode usar a discussão da seção 2.1 acima e supor que a curva tracejada seja mais próxima da verdadeira que a cheia (ao menos, ela tem uma imprecisão menor), para verificar que a diferença entre a curva cheia e a tracejada está quase sempre dentro de uma barra de incerteza do valor experimental. Note que a discussão efetuada na seção 2.1 vale para qualquer histograma e, portanto, também para este em particular.

2.2.2 A Poisson a partir de princípios básicos

A f.p. de eventos independentes que têm uma probabilidade constante de acontecer em um *intervalo* de alguma grandeza contínua (seja de tempo, distância,

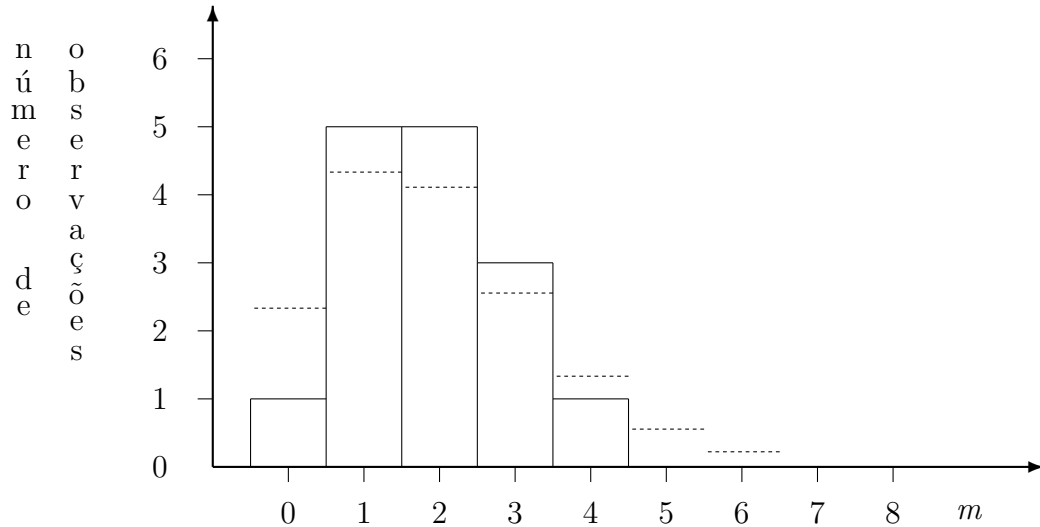


Figura 2.1: Histograma do número de átomos empilhados sobre uma área b^2 da folha, de acordo com a discussão da seção 1.1. Em linha cheia, o resultado experimental de $N = 15$ observações, correspondendo ao que se vê na figura 1.2 da seção 1.1. Em linha tracejada, o produto do número de observações pela f.p. de Poisson (fórmula (2.6)) de média $\hat{a} = 1,87$, calculada de acordo com o texto.

etc.), independente da origem escolhida, é a de Poisson, fórmula (2.6), conforme será deduzido abaixo.

Vamos supor que a grandeza contínua seja tempo, para efeito de raciocínio, sem perda de generalidade. Assim, t é o instante de observação e δt um intervalo de tempo, que vamos supor muito pequeno — a estratégia desta demonstração é estabelecer uma relação diferencial e passá-la ao limite, assim começamos com intervalos pequenos desde o princípio. Nessa situação, a probabilidade de encontrar um evento entre t e $t + \delta t$ é proporcional a δt :

$$P_1 = \lambda \delta t \quad .$$

P_1 independe de t , característica do evento tipo Poisson. Vamos supor δt suficientemente pequeno para que a probabilidade de ocorrer mais de um evento durante o intervalo δt possa ser ignorada. A probabilidade de ocorrerem $n \geq 1$ eventos entre 0 e $t + \delta t$ é a soma de duas possibilidades:

A: n eventos entre 0 e t , nenhum entre t e $t + \delta t$

B: $n - 1$ eventos entre 0 e t , 1 entre t e $t + \delta t$,

o que pode ser escrito como:

$$P_A = P(n, t)(1 - \lambda\delta t)$$

$$P_B = P(n - 1, t)\lambda\delta t \quad ,$$

onde $P(n, t)$ indica a probabilidade de se observar n eventos entre 0 e t . A probabilidade de se observar n eventos entre 0 e $t + \delta t$ é portanto a soma de P_A e P_B :

$$P(n, t + \delta t) = P(n, t)(1 - \lambda\delta t) + P(n - 1, t)\lambda\delta t \quad ,$$

o que pode ser re-escrito como

$$\frac{P(n, t + \delta t) - P(n, t)}{\delta t} = \lambda (P(n - 1, t) - P(n, t)) \quad .$$

No limite $\delta t \rightarrow 0$, o primeiro termo torna-se uma derivada:

$$\frac{dP(n, t)}{dt} = \lambda (P(n - 1, t) - P(n, t)) \quad ,$$

o que leva à equação diferencial

$$\frac{dP(n, t)}{dt} + \lambda P(n, t) = \lambda P(n - 1, t) \quad . \quad (2.9)$$

O membro esquerdo dessa última equação pode ser reescrito usando o fator integrante $e^{\lambda t}$, uma vez que

$$\frac{d}{dt} (e^{\lambda t} P(n, t)) = e^{\lambda t} \left(\frac{dP(n, t)}{dt} + \lambda P(n, t) \right) \quad ,$$

que, substituído no membro esquerdo da relação 2.9, dá

$$e^{-\lambda t} \frac{d}{dt} (e^{\lambda t} P(n, t)) = \lambda P(n - 1, t) \quad ,$$

o que leva ao resultado

$$P(n, t) = e^{-\lambda t} \int \lambda e^{\lambda t'} P(n - 1, t') dt' \quad , \quad (2.10)$$

que é uma relação de recorrência. Assim, se encontrarmos $P(0, t)$, poderemos determinar $P(n, t)$, pela aplicação repetida dessa fórmula. Como há uma única

maneira de não ocorrer nenhum sucesso no intervalo $0 \leq t \leq t + dt$, deduzimos que

$$P(0, t + \delta t) = P(0, t)(1 - \lambda \delta t) \quad .$$

Essa equação pode ser rearranjada como

$$\frac{P(0, t + \delta t) - P(0, t)}{\delta t} = -\lambda P(0, t) \quad .$$

Passando a expressão acima ao limite $\delta t \rightarrow 0$, obtém-se uma equação diferencial para $P(0, t)$, cuja solução é

$$P(0, t) = Ae^{-\lambda t} = e^{-\lambda t} \quad ,$$

onde escolhemos a constante multiplicativa $A = 1$ porque a probabilidade de 0 sucessos em um intervalo de tempo nulo é 1. Iterando a relação de recorrência (2.10) n vezes a partir de $P(0, t)$, chega-se à solução:

$$P(n, t) = \frac{(\lambda t)^n e^{-\lambda t}}{n!} \quad .$$

Se lembrarmos que λ é a probabilidade de se observar um evento por unidade de tempo, então λt é o número médio de eventos no intervalo de 0 a t . Se escrevermos $a = \lambda t$, temos a fórmula da f.p. de Poisson (2.6).

Variáveis aleatórias que se distribuem de acordo com a f.p de Poisson têm uma série de propriedades importantes, que vão aparecer ao longo do texto e dos exercícios. Aqui, destacamos a condição da independência da probabilidade de se observar um evento em relação à origem escolhida — se depender, o evento não é do tipo Poisson.

2.2.3 Soma de eventos com f.p. de Poisson

Uma propriedade interessante das variáveis aleatórias que seguem a f.p. de Poisson é que sua soma se distribuirá da mesma forma. Assim, quando n e n' são variáveis aleatórias com f.p.s de Poisson de médias a e a' , $P_a(n)$ e $P_{a'}(n')$, respectivamente, a variável aleatória

$$m = n + n' \tag{2.11}$$

tem a f.p.

$$P(m) = \sum_{n=0}^m \sum_{n'=0}^m P_a(n) \cdot P_{a'}(n') \quad .$$

Na expressão acima, o símbolo " após a segunda somatória lembra que deve-se somar apenas os termos que obedecem à restrição (2.11). Esta restrição pode ser explicitada com um delta de Kronecker,

$$P(m) = \sum_{n=0}^m \sum_{n'=0}^m P_a(n) \cdot P_{a'}(n') \cdot \delta_{m,n+n'} \quad ,$$

o que conduz finalmente à expressão com uma única somatória.

$$P(m) = \sum_{n=0}^m P_a(n) \cdot P_{a'}(m-n) \quad .$$

Ao desenvolvermos as f.p.s, obtemos

$$P(m) = \sum_{n=0}^m \frac{a^n e^{-a}}{n!} \frac{a'^{(m-n)} e^{-a'}}{(m-n)!} = e^{-(a+a')} \sum_{n=0}^m \frac{1}{n!(m-n)!} a^n a'^{(m-n)} \quad .$$

Multiplicando por $m!/m!$, definindo

$$A = a + a'$$

e agrupando os termos de maneira conveniente, obtém-se

$$P(m) = \frac{e^{-A}}{m!} \sum_{n=0}^m \frac{m!}{n!(m-n)!} a^n a'^{(m-n)} = \frac{e^{-A}}{m!} (a + a')^m = \frac{e^{-A}}{m!} A^m \quad ,$$

que é a f.p. de Poisson de média igual à soma das médias das duas variáveis aleatórias.

Vimos na seção anterior que a desintegração de uma fonte radioativa é um fenômeno descrito pela f.p. de Poisson. A adição de outra fonte radioativa altera a quantidade de desintegrações, mas o número total continua seguindo a f.p. de Poisson, porque as desintegrações de cada uma das fontes obedece essa mesma f.p. e a origem do processo faz com que uma não interfira na outra.

2.3 A função densidade de probabilidade normal ou gaussiana

É muito comum em um experimento que os dados obtidos sigam uma f.d.p. gaussiana,

$$N(x; x_0, \sigma_0) = \frac{1}{\sqrt{2\pi}\sigma_0} \exp\left\{-\frac{(x-x_0)^2}{2\sigma_0^2}\right\} . \quad (2.12)$$

onde o parâmetro x_0 é o valor esperado de x , $\langle x \rangle$,

$$\langle x \rangle = \int_{-\infty}^{\infty} xN(x; x_0, \sigma_0)dx = x_0 ,$$

que aqui coincide com a *moda* e a *mediana*, e o parâmetro σ_0^2 corresponde à variância, uma medida da largura quadrática média da função,

$$\langle (x-x_0)^2 \rangle = \int_{-\infty}^{\infty} (x-x_0)^2 N(x; x_0, \sigma_0)dx = \sigma_0^2 .$$

Adiante, daremos o motivo pelo qual essa f.d.p. é muito comum nas ciências experimentais. Entretanto, o fato de ser possível determinar as f.d.p.s das estatísticas associadas à média e à variância de dados gaussianos ajudou a difundir sua aplicação. A frase jocosa a respeito da aparente universalidade da gaussiana foi elaborada ainda no século XIX pelo astrônomo Lippman [Benzécri]:

“... lei em que todos creêm. Os experimentais pensam que é um teorema matemático e os matemáticos, que é um fato experimental.”

Veremos no curso que nem a gaussiana é tão universal nem podemos dispensar um pouco de reflexão antes de utilizá-la, mas mostraremos que há alguns teoremas que justificam sua aplicação em muitos casos.

2.4 A função densidade de probabilidade multinormal

Em muitas situações concretas, é preciso trabalhar com funções densidade de probabilidade multidimensionais, que foram definidas na seção 1.7. Quando

as n variáveis aleatórias x_1, x_2, \dots, x_n são distribuídas como gaussianas, pode-se escrever

$$f(x_1, x_2, \dots, x_n) = \frac{1}{(2\pi)^{n/2} \|\mathbf{V}\|^{1/2}} \exp\left(-\frac{1}{2} \Delta \vec{x}^t \mathbf{V}^{-1} \Delta \vec{x}\right) \quad , \quad (2.13)$$

onde o conjunto de variáveis foi identificado como um vetor, $(x_1, x_2, \dots, x_n) = \vec{x}$, o símbolo “ t ” denota transposição, o vetor $\Delta \vec{x}$ corresponde a

$$\Delta \vec{x} = \begin{pmatrix} x_1 - x_{1_0} \\ x_2 - x_{2_0} \\ \vdots \\ x_n - x_{n_0} \end{pmatrix} \quad (2.14)$$

e \mathbf{V} é a matriz de covariâncias³ das variáveis aleatórias,

$$\mathbf{V} = \begin{pmatrix} \sigma_{1_0}^2 & \text{cov}(x_1 x_2) & \dots & \text{cov}(x_1 x_n) \\ \text{cov}(x_2 x_1) & \sigma_{2_0}^2 & \dots & \text{cov}(x_2 x_n) \\ \vdots & \vdots & \ddots & \vdots \\ \text{cov}(x_n x_1) & \text{cov}(x_n x_2) & \dots & \sigma_{n_0}^2 \end{pmatrix} \quad , \quad (2.15)$$

em que $\sigma_{i_0}^2$ e $\text{cov}(x_i x_j)$ são definidas como em (1.31) e (1.36), substituindo $f(x, y)$ por $f(x_1, \dots, x_i, \dots, x_j, \dots, x_n)$ e integrando em todas as n variáveis aleatórias; essa matriz é simétrica, pois $\text{cov}(x_i x_j) = \text{cov}(x_j x_i)$, e semi-definida positiva, portanto, com determinante $\|\mathbf{V}\|$ positivo ou nulo em geral, mas que neste caso não pode ser nulo para que f seja uma função densidade de probabilidade. O coeficiente de correlação é definido como a grandeza adimensional

$$\rho(x_i x_j) = \frac{\text{cov}(x_i x_j)}{\sigma_{i_0} \sigma_{j_0}} \quad , \quad (2.16)$$

com a propriedade $|\rho(x_i x_j)| \leq 1$ para todos os pares $i \neq j$.

Vamos obter a forma explícita da f.d.p. normal bivariada, com valores esperados x_{1_0} e x_{2_0} e matriz de covariância

$$\mathbf{V} = \begin{pmatrix} \sigma_{1_0}^2 & \rho \sigma_{1_0} \sigma_{2_0} \\ \rho \sigma_{1_0} \sigma_{2_0} & \sigma_{2_0}^2 \end{pmatrix} \quad .$$

³Essa matriz \mathbf{V} tem o nome de matriz de variância-covariância ou, mais curto e mais usado, matriz de covariâncias ou matriz de variâncias. Usaremos *matriz de covariâncias* ao longo do texto, para lembrar da importância dos termos não diagonais, embora muitos autores prefiram o nome matriz de variâncias por entender que ela é simplesmente a generalização da variância, quando há mais do que uma variável.

O expoente da equação (2.13),

$$\begin{pmatrix} x_1 - x_{1_0} \\ x_2 - x_{2_0} \end{pmatrix}^t \mathbf{V}^{-1} \begin{pmatrix} x_1 - x_{1_0} \\ x_2 - x_{2_0} \end{pmatrix} ,$$

quando expandido, fica igual a

$$\frac{(x_1 - x_{1_0})^2 \sigma_{2_0}^2 - 2(x_1 - x_{1_0})(x_2 - x_{2_0})\rho\sigma_{1_0}\sigma_{2_0} + (x_2 - x_{2_0})^2 \sigma_{1_0}^2}{\sigma_{1_0}^2 \sigma_{2_0}^2 (1 - \rho^2)} .$$

Assim, a equação (2.13), em função dos desvios-padrões e da covariância, tem a forma

$$f(x_1, x_2) = \frac{1}{2\pi\sigma_{1_0}\sigma_{2_0}\sqrt{1-\rho^2}} \cdot \exp \left[\frac{-1}{2(1-\rho^2)} \left(\frac{(x_1 - x_{1_0})^2}{\sigma_{1_0}^2} - \frac{2\rho(x_1 - x_{1_0})(x_2 - x_{2_0})}{\sigma_{1_0}\sigma_{2_0}} + \frac{(x_2 - x_{2_0})^2}{\sigma_{2_0}^2} \right) \right] .$$

Nesse caso bidimensional, vamos calcular a densidade de probabilidade de x_1 independentemente do valor da variável x_2 , $f_{x_1}(x_1)$, usando o procedimento da seção 1.7:

$$f_{x_1}(x_1) = \int f(x_1, x_2) dx_2 , \quad (2.17)$$

onde a integral é em todo o espaço amostral de x_2 , no caso $]-\infty, +\infty[$. Para simplificar as expressões sem perder a generalidade, vamos supor $x_{1_0} = 0$ e $x_{2_0} = 0$. Usando a expressão de $f(x_1, x_2)$ do exemplo acima temos

$$f_{x_1}(x_1) = \frac{1}{2\pi\sigma_{1_0}\sigma_{2_0}\sqrt{1-\rho^2}} \int_{-\infty}^{+\infty} \exp \left[\frac{-1}{2(1-\rho^2)} \left(\frac{x_1^2}{\sigma_{1_0}^2} - \frac{2\rho x_1 x_2}{\sigma_{1_0}\sigma_{2_0}} + \frac{x_2^2}{\sigma_{2_0}^2} \right) \right] dx_2 .$$

Como a integral é apenas em x_2 , os termos em x_1 podem ser extraídos da integração. Para integrar em x_2 , os termos do expoente que dependem dessa variável devem ser reescritos na forma de um quadrado perfeito em x_2 ,

$$-\frac{2\rho x_1 x_2}{\sigma_{1_0}\sigma_{2_0}} + \frac{x_2^2}{\sigma_{2_0}^2} = \frac{(x_2 - \rho \frac{\sigma_{2_0}}{\sigma_{1_0}} x_1)^2}{\sigma_{2_0}^2} - \frac{\rho^2 x_1^2}{\sigma_{1_0}^2} . \quad (2.18)$$

A equação 2.18 pode então ser escrita

$$f_{x_1}(x_1) = \frac{1}{2\pi\sigma_{1_0}\sigma_{2_0}\sqrt{1-\rho^2}} \exp\left[\frac{-1}{2(1-\rho^2)}\left(\frac{x_1^2}{\sigma_{1_0}^2} - \frac{\rho^2 x_1^2}{\sigma_{1_0}^2}\right)\right] \cdot \int_{-\infty}^{+\infty} \exp\left[-\frac{\left(x_2 - \rho\frac{\sigma_{2_0}}{\sigma_{1_0}}x_1\right)^2}{2(1-\rho^2)\sigma_{2_0}^2}\right] dx_2 \quad (2.19)$$

Como

$$\frac{1}{\sqrt{2\pi}\sigma_{2_0}\sqrt{1-\rho^2}} \int_{-\infty}^{+\infty} \exp\left[-\frac{\left(x_2 - \rho\frac{\sigma_{2_0}}{\sigma_{1_0}}x_1\right)^2}{2(1-\rho^2)\sigma_{2_0}^2}\right] dx_2 = 1 \quad , \quad (2.20)$$

por causa da normalização da f.d.p., temos, finalmente,

$$f_{x_1}(x_1) = \frac{1}{\sqrt{2\pi}\sigma_{1_0}} \exp\left(-\frac{x_1^2}{2\sigma_{1_0}^2}\right) \quad . \quad (2.21)$$

Este resultado nos permite concluir que, se x_1 e x_2 são duas variáveis com f.d.p. multinormal e matriz de covariância dada pela equação (2.15), então a f.d.p. da variável x_i , independente do valor assumido pela outra variável, é uma gaussiana com desvio padrão σ_{i_0} .

Q2.4 Além da função de probabilidade marginal, é possível construir funções condicionais. Mostre que, com x_2 fixo, a f.d.p. condicional de x_1 , $h(x_1 | x_2)$, é gaussiana de média $x_{1_0} + \rho(x_2 - x_{2_0})\sigma_1/\sigma_2$ e desvio padrão $\sigma_1\sqrt{1-\rho^2}$.

É mais fácil representar graficamente a função (2.17) se normalizamos as variáveis,

$$\zeta_1 = \frac{x_1 - x_{1_0}}{\sigma_{1_0}} \quad \text{e} \quad \zeta_2 = \frac{x_2 - x_{2_0}}{\sigma_{2_0}} \quad .$$

Nessas variáveis, tem-se

$$f(\zeta_1, \zeta_2) = \frac{1}{2\pi\sqrt{(1-\rho^2)}} \exp\left(-\frac{\zeta_1^2 - 2\rho\zeta_1\zeta_2 + \zeta_2^2}{2(1-\rho^2)}\right) \quad . \quad (2.22)$$

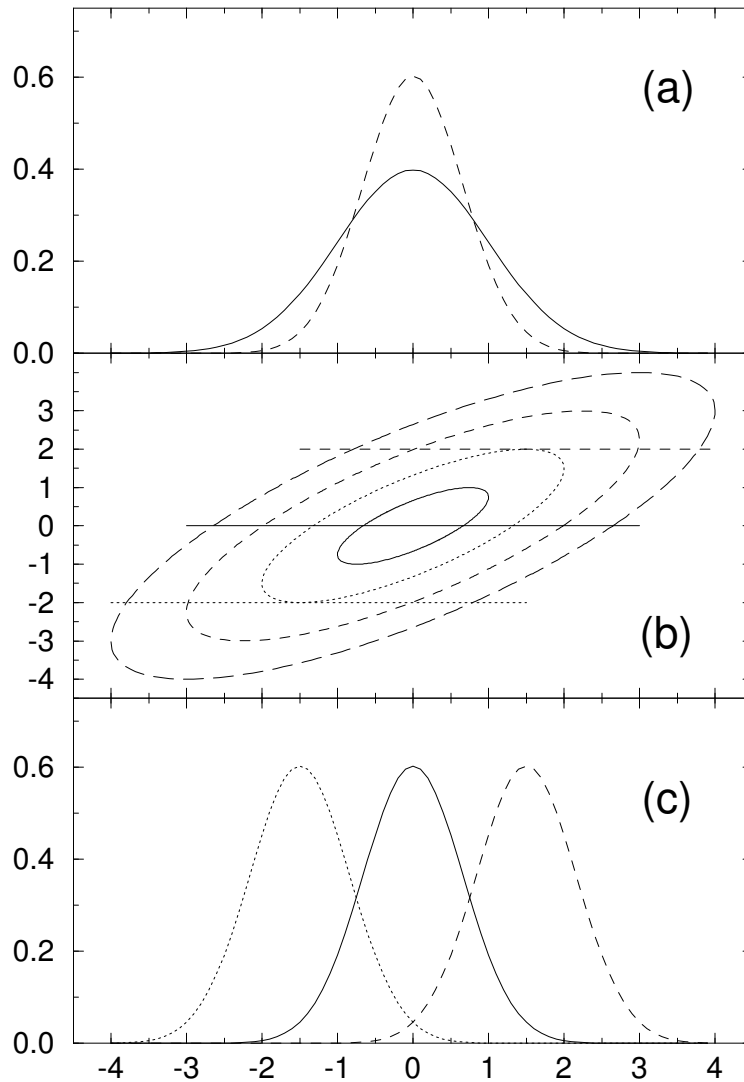


Figura 2.2: Gaussiana bi-dimensional para $\rho = 0,75$. **(a)** A curva em linha cheia representa a *distribuição marginal* e a *curva tracejada*, a *distribuição condicional para $\zeta_2 = 0$* . **(b)** Curvas de nível para os valores do expoente da equação (2.22) iguais a $n^2/2$, com $n = 1, 2, 3$ e 4 , em linha cheia, pontilhada, tracejada e tracejada longa, respectivamente. **(c)** Curvas para $f(\zeta_1)$ quando $\zeta_2 = -2; 0$ e 2 , respectivamente em linha pontilhada, cheia e tracejada. As linhas de corte aparecem na parte (b).

cuja curvas de nível e funções densidade de probabilidade marginal e condicionais estão representadas na figura 2.2, para uma correlação $\rho = 0,75$. Em geral, as curvas de nível são elipses e, quando ρ é positivo, a elipse tem seu eixo maior ao longo da bissetriz dos primeiro e terceiro quadrantes do sistema $O\zeta_1\zeta_2$, e para $\rho < 0$, o eixo maior da elipse está ao longo da bissetriz dos segundo e quarto quadrantes. O coeficiente de correlação ρ está relacionado com a razão entre os eixos maior e menor, mas, quando $\rho = 0$, os eixos tornam-se iguais e as curvas de nível tornam-se círculos; se $\rho = \pm 1$, a elipse transforma-se numa reta. Na parte (a) dessa figura, pode-se ver o efeito de estreitamento (veja questão Q2.4) da função de probabilidade condicional: a linha contínua mostra a f.d.p. marginal (com desvio-padrão do parâmetro ζ_1 igual a 1, por causa da normalização) enquanto que a linha tracejada mostra a f.d.p. de ζ_1 quando o valor de ζ_2 é fixado, mais estreita. Como aqui ρ é positivo, os valores mais prováveis de ζ_1 crescem quando cresce o valor fixado para a outra variável, ζ_2 , o que está ilustrado na parte (c), para $\zeta_2 = -2, 0$ e 2 , que correspondem às curvas pontilhada, contínua e tracejada, respectivamente, cujas linhas de corte aparecem na parte (b) com o mesmo tipo de linha. Note que o desvio padrão não muda com a mudança no valor fixado de ζ_2 .

2.5 Função característica

No estudo teórico das f.d.p.s, as suas transformadas de Fourier desempenham um papel importante. Chama-se *função característica* da f.d.p. $f(x)$ a transformada de Fourier $\phi(t)$,

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

A f.d.p. pode ser recuperada calculando a anti-transformada de Fourier,

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

Q2.5 Mostre que a função característica da gaussiana é

$$\phi_{\text{gaussiana}}(t; x_0, \sigma_0) = \exp \left\{ ix_0t - \frac{\sigma_0^2 t^2}{2} \right\} \quad . \quad (2.25)$$

O exemplo de aplicação desse resultado é importante e está destacado como a seção seguinte.

2.6 A função densidade de probabilidade da média de dados gaussianos independentes

A distribuição da média de um conjunto de dados $\{x_i, i = 1, \dots, N\}$, todos obtidos nas mesmas condições de modo que têm a mesma f.d.p., depende dessa função. Quando ela é a normal de média x_0 e variância σ_0^2 , $N(x_i; x_0, \sigma_0)$ para qualquer um dos dados x_i , a f.d.p. da média pode ser calculada de várias maneiras. Aqui, escolhemos o método baseado na função característica pela sua simplicidade.

Primeiro, construímos a f.d.p. da medida, $h(x_1, x_2, \dots, x_N)$, que é simplesmente o produto das f.d.p.s de cada dado, uma vez que supõe-se que os dados sejam estatisticamente independentes, ou seja,

$$h(x_1 \dots x_N) = \frac{1}{(\sqrt{2\pi}\sigma_0)^N} \exp\left\{-\frac{(x_1 - x_0)^2}{2\sigma_0^2}\right\} \dots \exp\left\{-\frac{(x_N - x_0)^2}{2\sigma_0^2}\right\} \quad (2.26)$$

A f.d.p. da média \bar{x} , com

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_N}{N},$$

corresponde à integral de h , em todas as suas N dimensões, mas sujeita ao vínculo acima. Vamos nos referir à variável \bar{x} como z , para nos concentrarmos no papel de vínculo desempenhado pela relação e facilitar a notação. Com esse mudança, obtemos

$$g(z) = C \int \dots \int \exp\left\{-\frac{(x_1 - x_0)^2}{2\sigma_0^2}\right\} \dots \exp\left\{-\frac{(x_N - x_0)^2}{2\sigma_0^2}\right\} \times \\ \delta\left(\frac{x_1 + x_2 + \dots + x_N}{N} - z\right) dx_1 \dots dx_N \quad .$$

onde C representa a constante multiplicativa que aparece em h . As regiões de integração cobrem todo o espaço e utilizamos a função δ para impor a restrição $z = \frac{x_1 + x_2 + \dots + x_N}{N} = \bar{x}$. Ao invés de efetuarmos essa integral múltipla

em $dx_1 \dots dx_N$, calculamos a função característica de $g(z)$, que definimos em (2.23) como

$$\phi(t) = \int \exp(izt)g(z)dz \quad ,$$

que, por causa da função δ , fica

$$\begin{aligned} \phi(t) = & C \int \dots \int \exp \left\{ -\frac{(x_1 - x_0)^2}{2\sigma_0^2} \right\} \dots \exp \left\{ -\frac{(x_N - x_0)^2}{2\sigma_0^2} \right\} \\ & \cdot \exp \left(i \frac{x_1 + \dots + x_N}{N} t \right) dx_1 \dots dx_N \end{aligned}$$

Essa expressão pode ser reorganizada como

$$\begin{aligned} \phi(t) = & C \int \exp \left\{ -\frac{(x_1 - x_0)^2}{2\sigma_0^2} \right\} \exp \left(\frac{ix_1 t}{N} \right) dx_1 \\ & \dots \int \exp \left\{ -\frac{(x_N - x_0)^2}{2\sigma_0^2} \right\} \exp \left(\frac{ix_N t}{N} \right) dx_N \quad . \end{aligned}$$

A única diferença entre cada uma das integrais acima e aquela utilizada na dedução da fórmula (2.25) é a troca de t por t/N . Assim, adaptando a função da fórmula (2.25), calculamos a função característica da média como

$$\phi(t) = \left[\exp \left\{ \frac{ix_0 t}{N} - \frac{\sigma_0^2 t^2}{2N^2} \right\} \right]^N = \exp \left\{ ix_0 t - \frac{\sigma_0^2 t^2}{2N} \right\} \quad ,$$

que pode ser reconhecida, novamente por meio da fórmula (2.25), como a função característica de uma gaussiana de média x_0 , mas com desvio padrão σ_0^2/N , ou seja,

$$g(\bar{x}) = \frac{\sqrt{N}}{\sqrt{2\pi}\sigma_0} \exp \left\{ -\frac{N(\bar{x} - x_0)^2}{2\sigma_0^2} \right\} \quad , \quad (2.27)$$

onde σ_0 é o desvio padrão *dos dados*.

Não deixe de perceber o alcance do resultado acima, que vai além de demonstrar que o desvio padrão da média, σ_{m_0} , é

$$\sigma_{m_0} = \frac{\sigma_0}{\sqrt{N}} \quad .$$

Se os dados são gaussianos, a f.d.p. da *estatística* \bar{x} , com a qual normalmente estimamos o valor verdadeiro, é a gaussiana, *qualquer que seja o número de dados* da medida. Mostraremos adiante que a f.d.p. da média de um número *grande* de dados é essa mesma gaussiana, com a condição que a f.d.p. dos dados tenha variância finita.

2.7 Transformação de variável aleatória

Considere uma variável aleatória x com distribuição conhecida, $f(x)$, e que desejamos calcular a f.d.p. de uma função dessa variável aleatória, $v(x)$, ou seja, queremos a f.d.p. $g(v)$. Para facilitar a discussão, vamos considerar um exemplo, em que x é o raio de um círculo e $f(x)$ é a normal,

$$f(x; x_0, \sigma_0) = \frac{1}{\sqrt{2\pi}\sigma_0} \exp \left\{ -\frac{(x - x_0)^2}{2\sigma_0^2} \right\} .$$

O objetivo é calcular a f.d.p. da área do círculo, $v = \pi x^2$. Neste primeiro exemplo, vamos nos restringir ao caso em que $\sigma_0 \ll x_0$, de forma que valores negativos de x podem ser ignorados. Como $v(x)$ é uma função crescente de x , ao intervalo $[x_1, x_2]$ corresponde um intervalo $[v(x_1), v(x_2)] = [v_1, v_2]$. Assim, a probabilidade de obter um dado no intervalo $[x_1, x_2]$ é a mesma de obter um dado no intervalo $[v_1, v_2]$, ou seja

$$\int_{x_1}^{x_2} f(x) dx = \int_{v_1}^{v_2} g(v) dv .$$

Passando ao limite em que $x_2 \rightarrow x_1$, obtemos

$$|f(x_1)dx| = |g(v_1)dv| .$$

Trocando x_1 por x' e chamando de v' o ponto correspondente, $v(x')$, podemos resumir o resultado como

$$g(v') = \frac{f(x')}{\left| \frac{dv}{dx} \right|_{x=x'}} \quad \text{com } v' = v(x') . \quad (2.28)$$

Em geral, a derivada deve ser tomada em módulo para garantir que g seja definida positiva quando a transformação $v(x)$ é decrescente com x crescente. No exemplo do círculo, a variável aleatória x é definida positiva, o que torna a relação $v(x) = \pi x^2$ monótona, de modo que

$$g(v) = \frac{1}{2\sqrt{2v} \pi \sigma_0} \exp \left\{ -\frac{(\sqrt{v/\pi} - \sqrt{v_0/\pi})^2}{2\sigma_0^2} \right\} .$$

onde definimos

$$v_0 = \pi x_0^2 .$$

Note que g é uma função de forma muito diferente da gaussiana. Entretanto, se $v \cong v_0$, é possível aproximar g pela normal ⁴

$$g(v; v_0, \sigma_v) = N(v; \pi x_0^2, 2\pi x_0 \sigma_0) \quad \text{para } v \approx v_0 \quad , \quad (2.29)$$

na notação da fórmula (2.12).

Q2.6 *Demonstre que a função $g(v)$ pode ser aproximada pela fórmula (2.29), expandindo $v^{1/2}$ até primeira ordem em série de Taylor em torno de v_0 .*

2.7.1 Correspondência não biunívoca

Quando há mais de um valor x que conduz ao mesmo valor v , precisamos generalizar a fórmula (2.28) de maneira a somar sobre todos os valores de x que conduzem ao mesmo valor de v . Vamos dar como exemplo uma transformação que nos será útil logo mais.

Considere uma variável aleatória x com f.d.p. $f(x)$ gaussiana, de média $x_0 = 0$, e que se busca determinar a f.d.p. da variável $v = x^2$. Então, como a f.d.p. de x é simétrica em torno de $x = 0$, ou seja,

$$f(x') = f(-x') \quad \text{e}$$

$$\left| \frac{dv}{dx} \right|_{x'} = \left| \frac{dv}{dx} \right|_{-x'} \quad ,$$

os dois valores x' e $-x'$ que conduzem ao mesmo v contribuem igualmente para $g(v)$. Por isso, a f.d.p. de v , $g(v)$, pode ser obtida a partir da fórmula (2.28) com o membro direito multiplicado por 2,

$$g(v) = 2 \frac{1}{\sqrt{2\pi}\sigma_0} \exp\left(-\frac{x^2}{2\sigma_0^2}\right) \frac{1}{|2x|}$$

$$g(v) = \frac{1}{\sqrt{2\pi v}\sigma_0} \exp\left(-\frac{v}{2\sigma_0^2}\right) \quad , \quad v \geq 0 \quad . \quad (2.30)$$

⁴Essa aproximação *não* quer dizer que $g(v)$ é aproximadamente normal para qualquer v ; para isso, também é preciso que $\frac{\sigma_0}{x_0} \ll 1$. Caso o desvio-padrão seja grande, então $g(v)$ será muito diferente da normal longe de v_0 .

2.7.2 Muitas variáveis

As transformações de variáveis não se limitam a uma dimensão. Considere a f.d.p. $f(\vec{x})$, com m variáveis aleatórias $x_1, x_2, \dots, x_i, \dots, x_m$, onde as m variáveis aleatórias foram representadas pelo vetor \vec{x} . Queremos escrevê-la em função das variáveis aleatórias $z_1, z_2, \dots, z_i, \dots, z_m$, obtidas pelas regras de transformação

$$\begin{aligned} z_1 &= z_1(x_1, x_2, \dots, x_m) \\ z_2 &= z_2(x_1, x_2, \dots, x_m) \\ &\vdots \\ z_m &= z_m(x_1, x_2, \dots, x_m) \end{aligned} \quad (2.31)$$

A f.d.p. das variáveis z_1, \dots, z_m está relacionada a $f(\vec{x})$ por

$$g(\vec{z}) = \frac{f(\vec{x})}{\left| \frac{\partial(z_1, \dots, z_m)}{\partial(x_1, \dots, x_m)} \right|}, \quad (2.32)$$

onde representamos o conjunto das coordenadas z_1, \dots, z_m por \vec{z} . O símbolo $\left| \frac{\partial(z_1, \dots, z_m)}{\partial(x_1, \dots, x_m)} \right|$ representa o Jacobiano da transformação de variáveis (2.31). Relembrando, o Jacobiano de uma transformação corresponde à razão entre os elementos dos hipervolumes dos sistemas de coordenadas \vec{z} e \vec{x} . Ele é calculado como o módulo do determinante da matriz cujo elemento (i, j) é a derivada parcial,

$$\frac{\partial z_i}{\partial x_j} \Bigg|_{x_1, \dots, x_{j-1}, x_{j+1}, \dots, x_m}.$$

Note que a fórmula (2.28), de mudança de variável unidimensional, é um caso particular dessa última fórmula (2.32), em que o Jacobiano reduz-se a $\left| \frac{dv}{dx} \right|$.

Exemplo 2.4

Suponhamos uma medida de uma variável gaussiana de valor verdadeiro x_0 com apenas 2 dados estatisticamente independentes, $\{x_1, x_2\}$. Vamos escolher $x_0 = 0$, para simplificar as contas, sendo que o resultado pode ser facilmente estendido para $x_0 \neq 0$. A f.d.p. conjunta dos dois dados é

$$g(x_1, x_2) = \left(\frac{1}{\sqrt{2\pi}\sigma_0} \right)^2 \exp \left\{ -\frac{x_1^2}{2\sigma_0^2} \right\} \cdot \exp \left\{ -\frac{x_2^2}{2\sigma_0^2} \right\}.$$

Dentre as transformações possíveis das variáveis x_1 e x_2 nas variáveis u e v , vamos escolher a transformação

$$\begin{aligned} u &= \frac{1}{2}(x_1 + x_2) & \mathbf{e} \\ v &= \frac{1}{2}(x_1 - x_2) & , \end{aligned} \quad (2.33)$$

porque estamos interessados no resultado, como veremos na questão **Q2.7** abaixo. O Jacobiano dessa transformação é, simplesmente, $J = 1/2$. A f.d.p. de (u, v) é, então,

$$\eta(u, v) = \frac{h(x_1, x_2)}{J} = \frac{1}{\pi\sigma_0^2} \exp\left\{-\frac{u^2 + v^2}{\sigma_0^2}\right\} . \quad (2.34)$$

Note que as variáveis u e v são estatisticamente independentes, ou seja, sua f.d.p. conjunta pode ser separada em dois fatores, um que depende apenas de u e, outro, apenas de v .

Q2.7 Mostre que u é a média e $|v|$ o desvio padrão da média, estimados numa medida com dois dados. Deduza a variância da média examinando a fórmula (2.34).

2.8 A f.d.p. de χ^2 (qui-quadrado)

Ao longo do curso, vamos usar uma variável aleatória, que tem o nome de χ^2 (qui-quadrado), que se relaciona com os N dados de uma medição de acordo com a relação

$$\chi^2 = \sum_{i=1}^N \frac{(y_i - y_{0,i})^2}{\sigma_{i,0}^2} , \quad (2.35)$$

onde y_i representa uma observação de uma grandeza distribuída conforme a *normal* de média $y_{0,i}$ e variância $\sigma_{0,i}^2$ ⁵. Em seguida, vamos determinar a f.d.p. de χ^2 , que vai depender do número de termos na somatória, N .

A primeira etapa do cálculo consiste em normalizar as variáveis aleatórias, pelas transformações

$$z_i = \frac{y_i - y_{0,i}}{\sigma_{i,0}} . \quad (2.36)$$

⁵Nesta demonstração, estamos supondo dados estatisticamente independentes; mais adiante, generalizaremos a expressão (2.35) para incluir dados correlacionados

A partir da f.d.p. de y_i , fórmula (2.12), determina-se a da variável normalizada, que se simplifica para

$$f(z_i) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z_i^2}{2}\right) = N(z_i; 0, 1) \quad , \quad (2.37)$$

para qualquer uma das N variáveis z_i . Note que $f(z_i)$ é a *normal de média 0 e desvio padrão 1*.

Em termos das variáveis normalizadas, qui-quadrado fica

$$\chi^2 = X = z_1^2 + z_2^2 + \cdots + z_i^2 + \cdots + z_N^2 \quad . \quad (2.38)$$

Há várias formas de prosseguir neste cálculo da f.d.p. de χ^2 . O exercício 2.10 mostra como encontrá-la pelo método da função característica. Já no texto que segue, vamos usar o caminho da mudança de variáveis, que usa uma transformação interessante e uma ideia recorrente na determinação de f.d.p.s de funções de variáveis aleatórias. Nos cálculos intermediários, escreveremos X no lugar de χ^2 , a fim de simplificar a notação, apenas.

2.8.1 Caso $N = 1$

Quando $N = 1$, qui-quadrado reduz-se à variável $X = z^2$, onde a f.d.p. de z é normal, problema que já resolvemos. A f.d.p. corresponde à da fórmula (2.30) acima, que repetimos aqui,

$$F_1(\chi^2) = \frac{1}{\sqrt{2\pi\chi^2}} \exp\left(-\frac{\chi^2}{2}\right) \quad , \quad (2.39)$$

onde o sub-índice de F corresponde ao número de termos da somatória da fórmula (2.35), que define a variável χ^2 .

2.8.2 Caso $N = 2$

Neste caso, as duas variáveis, z_1 e z_2 , podem ser transformadas em outras duas variáveis, uma das quais o qui-quadrado definido pela fórmula (2.38) e outra uma variável auxiliar θ , de acordo com as relações

$$\begin{aligned} z_1 &= X^{1/2} \operatorname{sen}(\theta) \\ z_2 &= X^{1/2} \operatorname{cos}(\theta) \quad , \end{aligned}$$

onde θ varia de 0 a 2π e X de 0 a ∞ , de modo que cada ponto (z_1, z_2) mapeia em um único ponto (X, θ) .

O Jacobiano da transformação é $J = 1/2$, de maneira que

$$\mathcal{F}(X, \theta) = \frac{1}{4\pi} \exp\left(-\frac{z_1^2}{2} - \frac{z_2^2}{2}\right) = \frac{1}{4\pi} \exp\left(-\frac{X}{2}\right) ,$$

independente de θ neste caso da soma de 2 termos. Para obter a f.d.p. apenas de X , basta integrar a expressão acima em θ , que resulta num fator 2π . O resultado final é

$$F_2(\chi^2) = \frac{1}{2} \exp\left(-\frac{\chi^2}{2}\right) . \quad (2.40)$$

2.8.3 Caso $N = 3$

Este é o menor valor de N para o qual pode-se compreender a estrutura particular da transformação de variáveis que permite deduzir a f.d.p. de χ^2 para qualquer N . Escolhendo um sistema de eixos tri-ortogonais Oz_1 , Oz_2 e Oz_3 , a expressão (2.38) define X como o quadrado da distância entre a origem O e um ponto P , de coordenadas (z_1, z_2, z_3) . O sistema de coordenadas esférico a que estamos acostumados descreve justamente as coordenadas de P em termos de uma distância e dois ângulos,

$$\begin{aligned} z_1 &= X^{1/2} \text{sen}(\theta_1) \cos(\theta_2) \\ z_2 &= X^{1/2} \text{sen}(\theta_1) \text{sen}(\theta_2) \\ z_3 &= X^{1/2} \cos(\theta_1) , \end{aligned}$$

onde os domínios de θ_1 e θ_2 são $[0, \pi]$ e $[0, 2\pi]$, respectivamente, sendo que estamos habituados a simbolizar o ângulo azimutal, θ_2 , por ϕ . Esta transformação é um pouco diferente da habitual transformação de coordenadas esféricas (r, θ, ϕ) para cartesianas (x, y, z) , porque a variável X corresponde a r^2 e não simplesmente a r . O Jacobiano resultante, portanto, é também um pouco diferente,

$$J = \frac{1}{2} X^{1/2} \text{sen}(\theta_1) .$$

Juntando todas as constantes de normalização numa mesma constante C , obtemos

$$\mathcal{F}(X, \theta_1, \theta_2) = C X^{1/2} \exp\left(-\frac{z_1^2}{2} - \frac{z_2^2}{2} - \frac{z_3^2}{2}\right) \text{sen}(\theta_1)$$

ou

$$\mathcal{F}(X, \theta_1, \theta_2) = CX^{1/2} \exp\left(-\frac{X}{2}\right) \text{sen}(\theta_1) \quad .$$

Para obter $F_3(X)$, basta integrar esta última expressão em θ_1 e θ_2 . A integral nas duas coordenadas angulares, porém, é independente de X , porque os limites de integração são independentes de X e a função \mathcal{F} fatora-se em um termo em X e um termo em θ_1 . Efetuando essa integral, obtém-se

$$F_3(\chi^2) \propto \sqrt{\chi^2} \exp\left(-\frac{\chi^2}{2}\right) \quad , \quad (2.41)$$

faltando só normalizar a f.d.p., o que faremos posteriormente.

2.8.4 Caso geral

A transformação de variáveis geral corresponde a escolher um sistema de eixos ortogonais Oz_1, Oz_2, \dots, Oz_N , onde X é o quadrado da distância à origem, O , de um ponto P de coordenadas (z_1, z_2, \dots, z_N) ,

$$\begin{aligned} z_1 &= X^{1/2} \text{sen}(\theta_1) \text{sen}(\theta_2) \dots \text{sen}(\theta_{N-2}) \text{sen}(\theta_{N-1}) \\ z_2 &= X^{1/2} \text{sen}(\theta_1) \text{sen}(\theta_2) \dots \text{sen}(\theta_{N-2}) \cos(\theta_{N-1}) \\ &\vdots \\ z_i &= X^{1/2} \text{sen}(\theta_1) \text{sen}(\theta_2) \dots \text{sen}(\theta_{N-i}) \cos(\theta_{N-i+1}) \\ &\vdots \\ z_{N-1} &= X^{1/2} \text{sen}(\theta_1) \cos(\theta_2) \\ z_N &= X^{1/2} \cos(\theta_1) \quad . \end{aligned}$$

O Jacobiano desta transformação depende de X pelo fator

$$\frac{1}{2}X^{-1/2}(X^{1/2})^{N-1} = \frac{1}{2}X^{N/2-1} \quad .$$

O fator $(X^{-1/2})/2$ provém das derivadas parciais em relação a X e o fator $(X^{1/2})^{N-1}$ vem das derivadas parciais em relação às $N - 1$ variáveis angulares θ_i , cada uma contribuindo com um fator $X^{1/2}$. O outro fator do Jacobiano corresponde a uma função nas variáveis angulares, que chamaremos $D(\theta_1, \dots, \theta_{N-1})$. Assim,

$$J = \frac{1}{2}X^{N/2-1}D(\theta_1, \dots, \theta_{N-1}) \quad .$$

Juntando todas as constantes de normalização numa constante C , dá

$$\mathcal{F}(X, \theta_1, \dots, \theta_{N-1}) = CX^{N/2-1} \exp\left(-\frac{z_1^2}{2} - \dots - \frac{z_{N-1}^2}{2}\right) D(\theta_1, \dots, \theta_{N-1})$$

ou

$$\mathcal{F}(X, \theta_1, \dots, \theta_{N-1}) = CX^{N/2-1} \exp\left(-\frac{X}{2}\right) D(\theta_1, \dots, \theta_{N-1}) \quad .$$

Para obter $F_N(X)$, basta integrar nas $N-1$ coordenadas angulares. A integral nos ângulos, porém, é independente de X , pelas mesmas razões apontadas no caso $N=3$, de maneira que podemos escrever

$$F_N(X) \propto X^{N/2-1} \exp\left(-\frac{X}{2}\right) \quad , \quad (2.42)$$

faltando somente normalizar a f.d.p.. Para isso, calcula-se

$$I_N = \int_0^\infty X^{N/2-1} \exp\left(-\frac{X}{2}\right) dX = 2^{N/2} \int_0^\infty t^{N/2-1} e^{-t} dt \quad ,$$

com a mudança de variável $t = X/2$. A última integral acima é a função gama de argumento $N/2$, de maneira que

$$I_N = 2^{N/2} \cdot \Gamma(N/2)$$

ou

$$I_N = \begin{cases} 2^{N/2}(N/2-1)! & , N \text{ par} \\ 2^{N/2}(N/2-1)(N/2-2)\cdots 1/2\sqrt{\pi} & , N \text{ ímpar.} \end{cases} \quad (2.43)$$

Juntando os resultados (2.42) e (2.43), chega-se à resposta final

$$F_N(X) = \frac{1}{2 \cdot \Gamma(N/2)} \left(\frac{X}{2}\right)^{N/2-1} \exp\left(-\frac{X}{2}\right) \quad (2.44)$$

ou, substituindo X por χ^2 ,

$$F_N(\chi^2) = \frac{1}{2 \cdot \Gamma(N/2)} \left(\frac{\chi^2}{2}\right)^{N/2-1} \exp\left(-\frac{\chi^2}{2}\right) \quad . \quad (2.45)$$

A figura 2.3 apresenta $F_N(\chi^2)$ para $N=1, 2, 3, 6$ e 10 . Note a assimetria pronunciada para N pequeno.

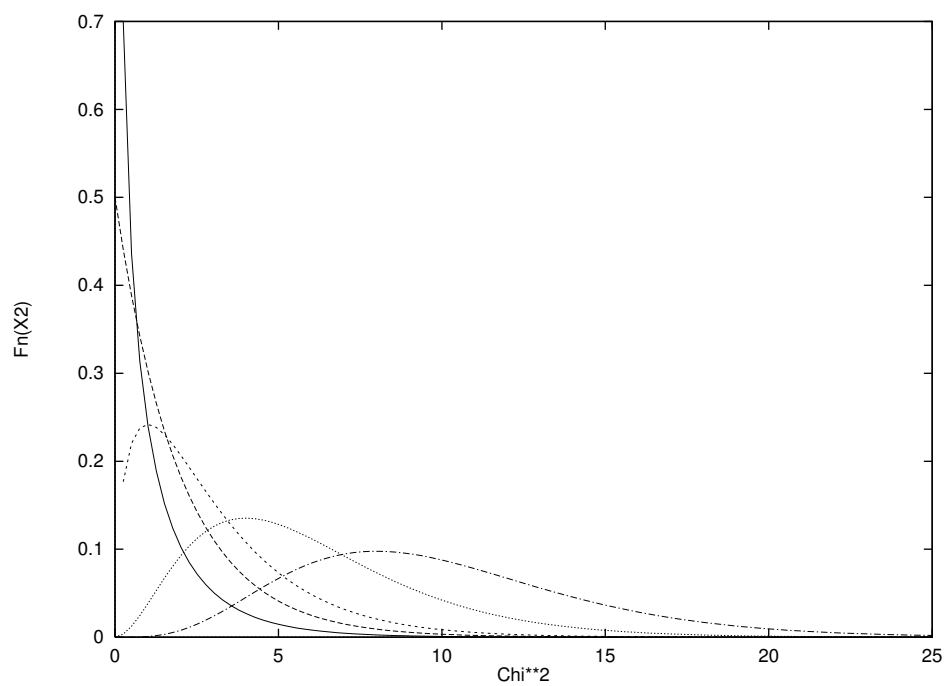


Figura 2.3: Funções densidade de probabilidade de qui-quadrado, $f_N(\chi^2)$, com diversos números de graus de liberdade, $N = 1, 2, 3, 6$ e 10 . Quando $N = 1$, a f.d.p. de qui-quadrado diverge na origem.

Tanto o valor médio como a variância de χ^2 podem ser obtidos a partir da equação (2.45) e as integrais necessárias recaem na função Γ . Obtém-se

$$\langle \chi^2 \rangle = N \quad (2.46)$$

$$\text{var}(\chi^2) = 2N \quad . \quad (2.47)$$

Q2.8 *Demonstre as relações (2.46) e (2.47) acima, a partir da f.d.p. de χ^2 . Mostre também que (2.46) pode ser obtida diretamente a partir da definição de χ^2 da fórmula (2.35) e não depende da f.d.p. dos dados ser gaussiana.*

2.9 O desvio-padrão do desvio-padrão

Na seção 2.6 determinamos a f.d.p. do valor médio da medida de uma grandeza no Modelo Normal⁶ e verificamos que depende da variância dos dados, isto é, sua f.d.p. é uma gaussiana de média centrada no valor verdadeiro e desvio padrão igual a σ_0/\sqrt{N} . Se conhecêssemos a variância verdadeira dos dados, poderíamos calcular a probabilidade de qualquer intervalo contendo o valor verdadeiro que escolhêssemos, porque a f.d.p. da média seria completamente determinada. Infelizmente, quase nunca conhecemos a variância verdadeira e o melhor que podemos fazer usualmente é estimá-la a partir da fórmula (1.10), copiada aqui para facilitar a leitura,

$$\sigma^2 = \frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})^2 \quad (1.10)$$

e determinar o comportamento estatístico dessa estimativa. O objetivo desta seção é apresentar a f.d.p. da estimativa da variância pela fórmula (1.10) no modelo Normal e explorar as consequências da flutuação estatística da estimativa.

Multiplicando-se ambos os membros da equação (1.10) por $(N-1)/\sigma_0^2$, ela pode ser reescrita

$$\frac{(N-1)\sigma^2}{\sigma_0^2} = \sum_{i=1}^N \frac{(x_i - \bar{x})^2}{\sigma_0^2} \quad . \quad (2.48)$$

⁶Modelo Normal corresponde à hipótese dos dados terem f.d.p. gaussiana.

O membro direito lembra bastante a definição de χ^2 da fórmula (2.35), e a única diferença é a troca de x_0 pela estimativa \bar{x} . Como há um vínculo entre os N termos da somatória, correspondente ao cálculo de \bar{x} , há apenas $N - 1$ termos independentes. Isso faz com que a variável aleatória definida pela equação (2.48) comporte-se como $F_{N-1}(\chi^2)$ [Kendall, vol. I, cap. 11]. Usando o resultado conhecido da variância de χ^2 , podemos deduzir

$$\text{desvio padrão de } \sigma^2 = \sqrt{\frac{2}{N-1}} \sigma_0^2 \cong \sqrt{\frac{2}{N-1}} \sigma^2 \quad . \quad (2.49)$$

É mais comum precisarmos da estimativa do desvio padrão do desvio padrão, de modo que precisamos da f.d.p. da raiz quadrada da variância e, por isso, aplicamos a fórmula (2.28) para a transformação de variável aleatória $\sigma_0 = \sqrt{\sigma_0^2}$ e deduzimos

$$\text{desvio padrão de } \sigma = \frac{\sigma_0}{\sqrt{2(N-1)}} \cong \frac{\sigma}{\sqrt{2(N-1)}} \quad . \quad (2.50)$$

Como o desvio-padrão da média difere do desvio-padrão apenas por um fator constante, de (2.49) e (2.50) deduz-se também

$$\text{desvio padrão de } \sigma_m \cong \frac{\sigma_m}{\sqrt{2(N-1)}} \quad . \quad (2.51)$$

A função de probabilidade da estimativa da variância é assimétrica, particularmente quando o número de dados, N , é pequeno (veja a dependência de $F_\nu(\chi^2)$ com ν , figura 2.3). Em particular, para $N = 2$, F_1 tem um máximo em $\sigma^2 = 0$, o que significa que, de *todos* os valores possíveis de serem obtidos na estimativa, zero é o mais provável! Por outro lado, F_1 cai relativamente devagar com o aumento da variância, tornando provável também obter-se valores muito maiores que a variância verdadeira. No caso $N = 2$, a probabilidade de subestimar a variância é de cerca de 70%. A estimativa da variância com 2 dados é, portanto, pouco informativa. Já para N maior ou da ordem de 5, a f.d.p. da estimativa é mais simétrica e mais centrada em torno do valor verdadeiro. Quanto maior N , mais simétrica é a f.d.p. e mais próximas ficam as probabilidades de subestimar e de superestimar a variância. Assim, para estimar a variância, é melhor não tomar poucos dados.

Exemplo 2.5

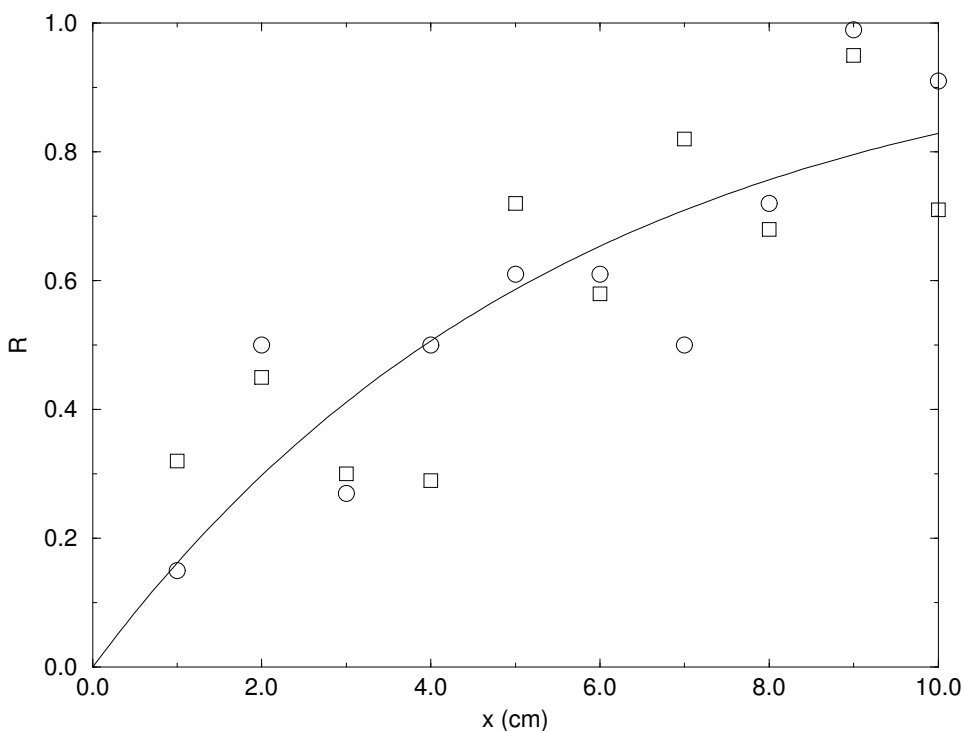


Figura 2.4: Gráfico dos dados da tabela (2.1) com a função R ajustada. Os círculos representam a primeira medida e os quadrados, a segunda.

Numa medida de atenuação da radiação pela matéria, um experimenter observou a razão entre a intensidade da radiação atenuada por placas de Pb e a intensidade incidente, R , que segue a lei

$$R = 1 - \exp(-\mu x) \quad ,$$

onde μ é uma constante que se pretende determinar e x é a espessura da placa, conhecida com desvio padrão tão pequeno que não precisa ser levado em conta.

A figura 2.4 apresenta os dados experimentais obtidos em 2 medições independentes realizadas sob as mesmas condições, onde os resultados da primeira delas estão representados pelos pontos marcados com o símbolo \circ e da segunda, com o quadrado vazio.

A tabela 2.1 dá os valores obtidos em cada uma dessas medidas, cujo desvio-padrão, $\sigma_R = 0,15$, é conhecido por outros meios. Ignorando por

Tabela 2.1: Dados de atenuação vs. espessura, para as duas medidas descritas no texto.

$x(cm)$	1ª medida	2ª medida	$x(cm)$	1ª medida	2ª medida
1	0,15	0,32	6	0,61	0,58
2	0,50	0,45	7	0,50	0,82
3	0,27	0,30	8	0,72	0,68
4	0,50	0,29	9	0,99	0,95
5	0,61	0,72	10	0,91	0,71

enquanto que o desvio padrão é conhecido, podemos estimá-lo a partir dos 2 dados obtidos para cada espessura. Neste caso, a estimativa da variância reduz-se a

$$\sigma^2 = \frac{1}{2}(y_1 - y_2)^2 \quad ,$$

onde y_1 e y_2 são os resultados da primeira e da segunda medida, respectivamente, para uma determinada espessura da placa. O desvio padrão de 2 dados é, simplesmente

$$\sigma = \frac{|y_1 - y_2|}{\sqrt{2}} \quad .$$

Calcule a estimativa de σ para cada par de dados obtido com um valor de x e verifique que a maioria é menor que 0,15, o valor conhecido para o instrumento, mostrando a tendência à subestimação da estimativa do desvio padrão com poucos dados.

Qual o risco dessa subestimação? Na figura 2.4, está representada a curva calculada com os parâmetros ajustados aos dados, usando uma mesma estimativa da variância para todos os pontos. Se o experimentador convencer-se erradamente que o desvio padrão é bem menor que 0,15, ele terá dificuldade em aceitar que o resultado da 2ª medida para $x = 4$ e o da primeira medida para $x = 7$ correspondem à mera flutuação estatística, o que poderia levá-lo a reexaminar todo o experimento para investigar a ocorrência de algum problema. Já se ele estiver convencido que $\sigma = 0,15$, esses pontos que acabamos de mencionar não têm nada de especial, estão entre mais que 1 e menos que 2 desvios padrão do valor médio estimado, o que é uma ocorrência muito provável.

É frequente efetuar-se medidas de curvas experimentais como a descrita neste exemplo. Muitas vezes, o desvio padrão da medida é ignorado

ou mal conhecido e sentimo-nos tentados a repetir os pontos experimentais para estimar o desvio padrão. Deve-se, porém, evitar repetir cada observação uma única vez. Se é impossível (caro, demorado, pouco prático ou pouco importante) repetir as observações para todos os valores da variável independente muitas vezes, é preferível repetir apenas um ou alguns pontos ao menos 5 vezes e, a partir dos desvios padrão obtidos para esses pontos e do conhecimento do instrumento, avaliar os desvios-padrão dos demais. Não se esqueça também de, ao substituir N pontos experimentais por sua média, usar o desvio padrão da *média* como barra de incerteza.

2.10 Quando as f.p.s e f.d.p.s. tendem à normal

Várias distribuições tendem à gaussiana ou outras distribuições para certos limites. A figura 2.5 resume o que é discutido nesta seção e inclui também as f.d.p.s das outras estatísticas que combinam variáveis aleatórias normais e que são conhecidas exatamente, t de Student (razão entre a média e o desvio-padrão de dados gaussianos) e F de Fisher (razão de variâncias de dados gaussianos), e que serão deduzidas nos capítulos 3 e 5, respectivamente, além da f.p. multinomial, que é uma generalização da binomial para mais de uma variável aleatória.

2.10.1 Binomial

Pode-se mostrar que a função de probabilidade binomial da fórmula (2.1) pode ser aproximada por uma gaussiana sob certas condições. Quando N é o número de dados, p a probabilidade de sucesso e q , a probabilidade complementar, a aproximação é boa para os valores da variável aleatória, n , que satisfaçam à condição [Feller]

$$\frac{1}{\sqrt{N}} \left(\frac{n - Np}{\sqrt{Npq}} \right)^3 \cong 0 \quad . \quad (2.52)$$

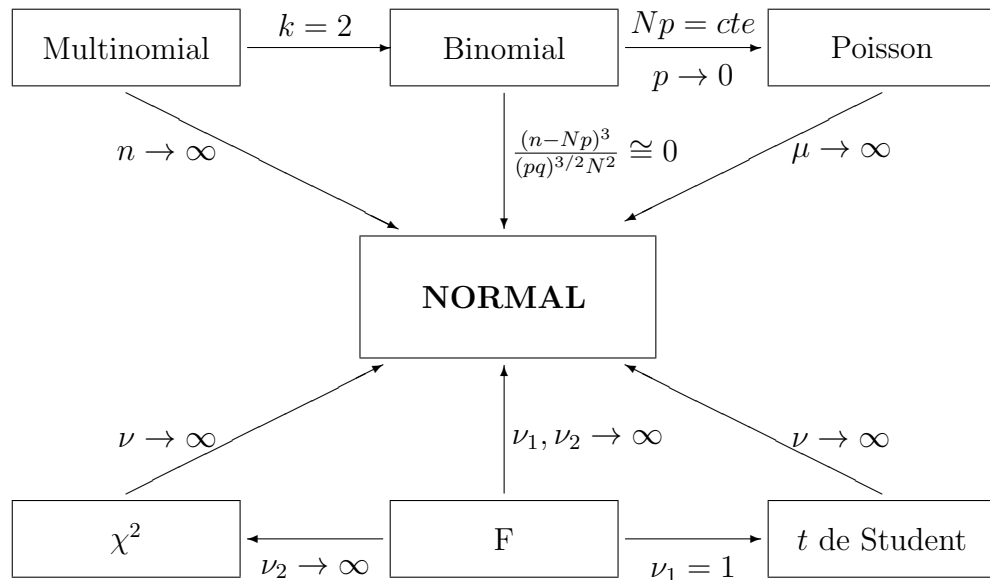


Figura 2.5: Relações entre as funções de probabilidade e densidade de probabilidade entre si e com a normal em diversas situações limites de seus parâmetros característicos, identificadas ao lado das setas que relacionam as funções. A f.p. multinomial é uma generalização da binomial para um conjunto de mais que dois eventos mutuamente exclusivos e exaustivos e as f.d.p.s de F e t de Student serão estudadas no próximo capítulo.

Nessa forma, reconhecemos o desvio padrão $\sigma = \sqrt{Npq}$. Chamando o número médio de sucessos de n_0 , onde $n_0 = Np$, a aproximação gaussiana é

$$P_{N,p}(n) \cong \frac{1}{\sqrt{2\pi}\sigma} \exp \left\{ -\frac{(n - n_0)^2}{2\sigma^2} \right\} . \quad (2.53)$$

Essa aproximação é boa em duas situações. De um lado, quando o valor da variável aleatória, n , difere do valor médio, Np , por uma fração do desvio padrão σ , mesmo para um número de observações, N , pequeno. Por outro lado, quando a diferença entre n e n_0 é grande em relação ao desvio-padrão, é preciso que o número de observações N seja suficientemente grande para compensar a diferença entre n e n_0 , na forma explicitada em (2.52).

Q2.9 Compare as probabilidades calculadas pela f.d.p. binomial da fórmula (2.1) com a obtida pela fórmula aproximada, (2.53), quando $N = 25$ e $p = 1/2$, para os intervalos $[10, 15]$; $[8, 17]$ e $[5, 20]$.

2.10.2 Poisson

A f.p. de Poisson pode ser aproximada pela normal quando o valor médio, a , é suficientemente grande,

$$P_a(n) \xrightarrow{a \rightarrow \infty} \frac{1}{\sqrt{2\pi a}} \exp \left\{ -\frac{(n - a)^2}{2a} \right\} . \quad (2.54)$$

Para $a \sim 30$, a aproximação gaussiana já é muito boa e desde $a > 5$ já é possível utilizá-la. Analogamente ao caso da binomial, cabe uma verificação da qualidade da aproximação caso seja necessário conhecer precisamente a probabilidade.

Q2.10 Compare as probabilidades calculadas pela f.p. de Poisson da fórmula (2.6) com as calculadas pela aproximação (2.54) nos casos:

i) $a = 5$ e $x \in [3, 7]$, $[1, 9]$ e $x \geq 10$ e

ii) $a = 25$ e $x \leq 5$, $x \leq 15$ e $x \in [20, 30]$.

2.10.3 Qui-quadrado

A f.d.p. de χ^2 tende à normal à medida que N cresce, sendo que para $N > 30$ a aproximação

$$F_N(\sqrt{2\chi^2}) \cong \frac{1}{\sqrt{2\pi}} \exp \left\{ -\frac{(\sqrt{2\chi^2} - \sqrt{2N-1})^2}{2} \right\} \quad (2.55)$$

é bastante boa. Pode-se utilizar, então, tabelas da integral da gaussiana para determinar a probabilidade de um certo qui-quadrado ser excedido. A referência [Zar] discute a precisão de diversas aproximações conhecidas.

2.11 A f.d.p. uniforme

A figura 2.6 abaixo ilustra a f.d.p. uniforme. A distribuição uniforme é definida como:

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{para } a \leq x \leq b \\ 0 & \text{para } x < a \text{ e } x > b \end{cases} .$$

Essa f.p. é muito comum, em particular serve para modelar o erro de arredondamento ou truncamento de números e é a f.d.p. que modela o resultado de muitos geradores de números aleatórios.

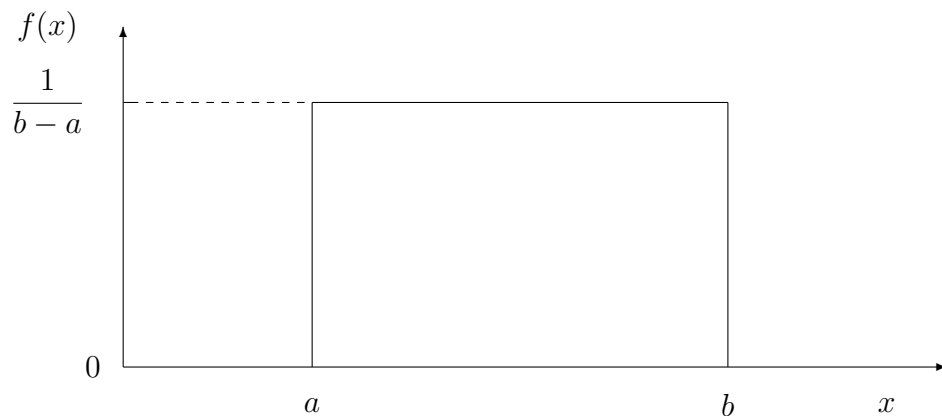


Figura 2.6: F.d.p. uniforme quando o espaço amostral é o intervalo de números reais $[a, b]$.

Q2.11 Calcule a média, x_0 , e o desvio padrão, σ , da f.d.p. de um gerador de números aleatórios, que procura gerar uma variável aleatória real, cuja distribuição seja uniforme no intervalo $[0,1[$.

O erro do arredondamento numa soma de muitas parcelas

Quando um número é arredondado pelos critérios habituais, o valor descartado é uma grandeza aleatória com f.d.p. uniforme. Vamos discutir a questão do erro de arredondamento numa soma para um caso particular, da soma de 12 parcelas p_i , onde p_i são dados de uma medição representados por um número decimal com casas após a vírgula que são arredondadas de maneira que, se a fração é maior ou igual a $0,50\dots$ arredonda-se p_i para o menor inteiro que supere p_i e se ela é menor que $0,50\dots$, desconsidera-se a fração.

A fração desconsiderada é uma variável aleatória se os dados são aleatórios e, se a variância dos dados p_i é muito maior que 1, podemos aproximar a f.d.p. da fração pela f.d.p. uniforme, com $a = -0,5$ e $b = 0,5$. A variância da fração desconsiderada é, então, $1/12$ (veja a questão Q2.12). Como a variância de uma soma de variáveis independentes é a soma das variâncias, a variância da soma de 12 frações é $12 \cdot (1/12) = 1$ e, portanto, o desvio padrão também é igual a 1.

A última etapa corresponde a determinar qual a probabilidade do erro cometido na soma dos números arredondados superar 1 desvio padrão, 2 desvios padrão, etc. Para isto, usaremos o fato da soma de muitos números aleatórios tender a gaussiana, um resultado que provaremos adiante. Assim, um erro de módulo menor que 1 (igual a um desvio padrão) deve ter uma probabilidade de 68%, menor que 2, 95% e etc. Olhando de outra maneira, o erro supera 1 provavelmente em $100\% - 68\% = 32\%$ dos casos, supera 2 em provavelmente 5% dos casos, etc.

Q2.12 Calcule a média, x_0 , e o desvio padrão, σ , da f.d.p. uniforme no intervalo $[a, b]$ (ou seja, a probabilidade de obter um valor fora desse intervalo é nula). Qual a probabilidade de obter um dado no intervalo $[x_0 - \sigma, x_0 + \sigma]$? e no intervalo $[x_0 - 2\sigma, x_0 + 2\sigma]$?

EXERCÍCIOS

- 2.1. Mostre que a variância de uma variável aleatória com uma f.p. binomial é $\sigma^2 = Npq = Np(1 - p)$. Sugestão: siga um procedimento semelhante ao que usou para calcular a variância da Poisson, seção 2.2.
- 2.2. Mostre que se a variável aleatória n segue a distribuição binomial $P_{N,p}(n)$, e a variável n' segue $P_{N',p'}(n')$, a variável $m = n + n'$ será uma distribuição

binomial se $p = p'$, mas não se $p \neq p'$. Este resultado permite, em certas condições, somar histogramas sem perder nenhuma informação estatística contida nos dados.

- 2.3. Indique quais das situações abaixo correspondem a variáveis aleatórias com distribuição binomial:
- (a) o número de vezes que um dado jogado N vezes não apresenta a face 5 para cima.
 - (b) o número de vezes que um dado jogado N vezes não apresenta as faces 1 ou 2 para cima.
 - (c) o número de vezes que um dado apresenta a face 5 para cima, sendo que se para de jogar o dado quando a face 6 aparece para cima N vezes.

Nos casos em que você supõe que a f.p. binomial seja adequada, procure determinar seus parâmetros característicos.

- 2.4. No exemplo da seção 1.11, estime a probabilidade do intervalo $(x_{[3]}, x_{[7]})$ conter a mediana. Estime, também, um intervalo de confiança para a mediana com probabilidade $P \approx 68\%$; se necessário, faça interpolações para encontrar os limites adequados. *Obs.: Para este exercício você deverá usar a distribuição binomial.*
- 2.5. Em um experimento, a probabilidade de um dado pertencer ao intervalo $\omega =]x_1, x_2]$ é $p = 0,15$. Em um experimento com 20 dados, determine:
- (a) a probabilidade de não obter nenhum dado em ω .
 - (b) a probabilidade de obter três dados em ω .
 - (c) a probabilidade de todos os dados pertencerem a ω .
 - (d) quantos dados são esperados no intervalo ω , em média.
 - (e) o desvio padrão da média calculada no item anterior.
- 2.6. Mostre que, para uma variável aleatória com f.p. de Poisson de média igual a a , a moda é única e igual ao maior inteiro contido em a se $a \notin \mathbf{N}$ (a não é um número natural), e que a e $a - 1$ são os dois valores mais prováveis quando $a \in \mathbf{N}$.

- 2.7. Mostre que a soma de duas variáveis n e n' , que obedecem a duas f.p. de Poisson com parâmetros a e a' , respectivamente, obedece a uma distribuição de Poisson com parâmetro $a + a'$.
- 2.8. Uma máquina produz capacitores, cuja capacidade nominal é 47 pF. A capacidade real desses componentes, porém, varia aleatoriamente de uma peça para outra, de acordo com uma f.d.p. gaussiana de média 47,0 pF e desvio padrão $\sigma=2,0$ pF.
- (a) Qual a fração de capacitores com capacitância entre 43,0 e 51,0 pF?
 - (b) O fabricante tem um equipamento de altíssima precisão para a medida de capacitância ($\sigma_{\text{ponte}} \ll 0,1$ pF). De um lote de 1000 capacitores, quantos ele selecionará, em média, com capacidade entre 46,5 e 47,5 pF?
 - (c) Qual deveria ser o desvio padrão para que 80% dos capacitores tivessem capacidade entre 45,0 e 49,0 pF?

Obs: Para responder os itens b) e c), você precisará da integral da Gaussiana (se estiver sem computador, veja as tabelas ?? e seguintes).

- 2.9. Em um experimento de medida do calor específico da água à temperatura ambiente, nove grupos de uma turma de estudantes adotaram o mesmo procedimento e usaram aparelhos de mesma precisão. Os resultados relatados foram (em $\text{J}/(\text{g}^\circ\text{C})$): 3,90(10); 4,34(14); 4,25(22); 4,31(5); 4,52(20); 4,20(40); 3,87(18); 4,09(4); 4,16(5). Cada turma obteve seus resultados a partir de medidas com 10 dados.
- (a) *Ignore* as estimativas do desvio padrão relatadas e suponha que todas as medidas sejam igualmente precisas e que os dados sigam uma f.d.p. normal. Nessas condições, determine, a partir dos resultados dos estudantes: o valor (e respectiva incerteza) do calor específico da água; o intervalo que contém o valor verdadeiro com 95% de probabilidade; a precisão do arranjo experimental usado por cada grupo.
 - (b) Os nove grupos da turma relataram resultados com desvios padrão muito diferentes. Indique quais estimativas dos desvios padrão possivelmente não sejam corretas. (Note que é impossível afirmar que uma determinada estimativa está errada, mas pode-se concluir que seria muito pouco provável.)

- 2.10. Determine a f.d.p. de χ^2 pelo método da função característica, usado na determinação da f.d.p. da média de dados normais.
- (a) Como no cálculo da seção 2.8, transforme χ^2 em uma soma dos quadrados de uma variável normalizada, z_i^2 , fórmula (2.38).
 - (b) Determine a f.d.p. de $v_i = z_i^2$ (conforme a seção 2.7).
 - (c) Determine a função característica de v_i . Note que essa função é a mesma para todos os v_i , uma vez que todas têm desvio padrão igual.
 - (d) Determine a função característica de χ^2 . Não deixe de usar que todas as variáveis v_i têm desvio padrão igual a 1.
 - (e) Determine a f.d.p. de χ^2 a partir da sua função característica.

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