

ANÁLISE ESTATÍSTICA DE MEDIDAS EM CIÊNCIAS EXATAS

Vito R. Vanin, Philippe Gouffon, Otaviano Helene

Março 2023

Capítulo 9

Propriedades do estimador de Máxima Verossimilhança

As propriedades de não-tendenciosidade e variância mínima do estimador de mínimos quadrados estão ligadas à linearidade das estatísticas que se obtêm por meio dele. Assim, quando a função não é linear nos parâmetros de ajuste, essas propriedades não se verificam mais.

No caso não linear, porém, é possível mostrar que o método da máxima verossimilhança fornece estimadores consistentes e assintoticamente normais, não tendenciosos e de variância mínima. O termo *assintoticamente* precisa ser levado a sério, de modo que se deve evitar situações onde há poucos graus de liberdade porque as propriedades ótimas só se verificam quando o número de dados é grande.

Mostraremos por que as propriedades valem apenas no caso em que há um único parâmetro desconhecido e a f.d.p. é bem comportada. A generalização para qualquer f.d.p. e vários parâmetros não é simples e, além disso, foge do nosso objetivo, que é explicar o significado da propriedade, em particular seu caráter assintótico.

Uma aplicação importante e comum deste método corresponde ao ajuste dos parâmetros do modelo adotado a dados com f.d.p. normal, quando as estatísticas são as mesmas do método dos mínimos quadrados. Entretanto, a não-linearidade faz com que a técnica de solução das equações fique mais complicada e será detalhada em uma das seções. Escolhemos incluir a discussão deste tópico aqui e não no capítulo anterior porque as propriedades ótimas das estimativas estão ligadas ao estimador de máxima verossimilhança, que é o tema deste capítulo.

9.1 A função verossimilhança no limite assintótico – um exemplo

Antes de apresentarmos as demonstrações da consistência e normalidade assintótica do estimador de máxima verossimilhança, vamos verificar essas propriedades no caso particular de uma grandeza que obedece à função de probabilidade de Poisson. Nesse caso, quando o valor médio é a_0 , a probabilidade de que a i -ésima observação dê o valor n_i é

$$P_{a_0}(n_i) = \frac{a_0^{n_i} e^{-a_0}}{n_i!} .$$

Vamos examinar como a estimativa de a_0 , \hat{a} , a partir de um conjunto de N dados $\{n_i, i = 1, \dots, N\}$, comporta-se quando o número de dados for muito grande.

A função verossimilhança e seu logaritmo são dadas por

$$\mathcal{L}(a \mid n_1, n_2, \dots, n_N) = \prod_{i=1}^N \frac{a^{n_i} e^{-a}}{n_i!} \quad \text{e}$$

$$\ln[\mathcal{L}(a \mid \{n_i\})] = -Na + \sum_{i=1}^N n_i \ln(a) - \sum_{i=1}^N \ln(n_i!) \quad , \quad (9.1)$$

respectivamente. Note que a na função verossimilhança é uma variável e não mais o valor verdadeiro a_0 , que aparece na função de probabilidade. A figura 9.1 mostra $\ln[\mathcal{L}(a \mid \{n_i\})]$, calculada para uma medição com 10 dados, no gráfico em linha cheia.

A estimativa de máxima verossimilhança é obtida impondo que a derivada da expressão acima seja nula,

$$\left. \frac{\partial \ln \mathcal{L}}{\partial a} \right|_{\hat{a}} = 0 \quad ,$$

que dá

$$\hat{a} = \frac{1}{N} \sum_{i=1}^N n_i \quad . \quad (9.2)$$

Q9.1 A partir do resultado (9.2), mostre que a estimativa de máxima verossimilhança não é tendenciosa e tem variância

$$\text{var}(\hat{a}) = \frac{a_0}{N} \quad . \quad (9.3)$$

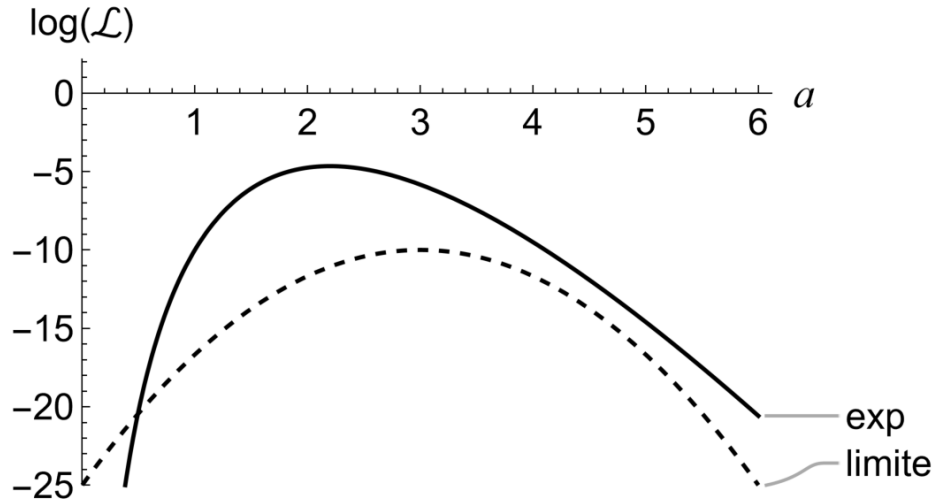


Figura 9.1: Em linha contínua, o logaritmo da função verossimilhança na medida cujos dados são $\{0, 3, 2, 3, 1, 1, 6, 2, 1, 3\}$, distribuídos conforme a f.p. de Poisson. Em linha tracejada, o logaritmo de uma f.d.p normal de média a_0 e desvio-padrão igual ao limite mínimo da variância para $N = 10$. Os termos independentes de a foram ignorados e o gráfico da função limite foi deslocado na ordenada de modo a facilitar a comparação com $\ln(\mathcal{L})$.

Já temos todos os elementos para entender que a estimativa de máxima verossimilhança tende ao valor verdadeiro assintoticamente. Como a diferença entre a_0 e \hat{a} é da ordem do desvio padrão de \hat{a} , que vai a zero quando $N \rightarrow \infty$ porque é inversamente proporcional a \sqrt{N} conforme a equação (9.3), essa diferença tende a zero quando o número de dados vai a infinito, ou seja, a estimativa de máxima verossimilhança é *consistente*.¹

A fim de comprovar que o comportamento assintótico da função verossimilhança em torno do valor verdadeiro a_0 é normal, vamos expandir o logaritmo da função verossimilhança da equação (9.1) em torno de \hat{a} . Usando que

$$\left. \frac{\partial^2 \ln \mathcal{L}}{\partial a^2} \right|_{\hat{a}} = -\frac{N}{\hat{a}} \quad \text{e} \quad \left. \frac{\partial^3 \ln \mathcal{L}}{\partial a^3} \right|_{\hat{a}} = 2 \frac{N}{\hat{a}^2}$$

¹A desigualdade de Chebyshev (seção 6.17) garante que a probabilidade da diferença entre a estimativa e o valor médio ser muito maior que o desvio-padrão é muito pequena. Note também que podíamos ter aproveitado que a fórmula (9.2) é linear em n_i e aplicar o Teorema Central do Limite para provar a consistência da estimativa, mas aqui estamos buscando entender como a função verossimilhança evolui com o número de dados.

a expansão em série de Taylor até ordem 3 é

$$\ln[\mathcal{L}(a | \{n_i\})] \approx C - \frac{N}{2\hat{a}}(a - \hat{a})^2 + \frac{N}{3\hat{a}^2}(a - \hat{a})^3 \quad (9.4)$$

onde usamos que a primeira derivada é nula em \hat{a} e os termos que não envolvem a estão reunidos em C .

A fim de interpretar esse resultado, define-se a grandeza α por meio da relação

$$\alpha = \frac{a - \hat{a}}{\sqrt{\text{var}(\hat{a})}} \quad (9.5)$$

que é uma medida normalizada da diferença de a em relação à estimativa. De acordo com a fórmula (9.3), a variância de \hat{a} pode ser estimada como $\text{var}(\hat{a}) = \frac{\hat{a}}{N}$, que, substituída na expressão acima, dá

$$a - \hat{a} = \alpha \sqrt{\text{var}(\hat{a})} = \alpha \sqrt{\frac{\hat{a}}{N}} \quad ,$$

que já foi rearranjada para facilitar sua substituição na expressão (9.4), o que resulta em

$$\ln \mathcal{L} = C - \frac{\alpha^2}{2} + \frac{\alpha^3}{3\sqrt{\hat{a}N}} \quad .$$

Como a probabilidade de α ser muito grande é muito pequena (por que α é o número de desvios-padrões entre a e a estimativa, conforme (9.5)), o último termo da expressão acima tende a zero quando $N \rightarrow \infty$ e, portanto,

$$\ln \mathcal{L} \xrightarrow{N \rightarrow \infty} C - \frac{\alpha^2}{2} \quad . \quad (9.6)$$

Substituindo α por sua definição, relação (9.5), e \hat{a} por a_0 , uma vez que $\hat{a} \xrightarrow{N \rightarrow \infty} a_0$, obtém-se

$$\mathcal{L} \xrightarrow{N \rightarrow \infty} A \exp\left(-\frac{(a - a_0)^2}{2 \text{var}(\hat{a})}\right) \quad (9.7)$$

Interpreta-se que, nesse limite em que $N \rightarrow \infty$, a função (9.7) representa a f.d.p. da estimativa de máxima verossimilhança. Assim, a região de interesse no gráfico de $\ln \mathcal{L}$ se restringe à vizinhança do máximo, com uma largura de alguns desvios-padrões. Por exemplo, ao fixar a região de interesse em

$[\hat{a} - n\sigma, \hat{a} + n\sigma]$ com σ o desvio-padrão de \hat{a} , a forma da função só importa entre o máximo e uma ordenada igual a

$$\ln[\mathcal{L}(a | \{n_i\})] = \ln[\mathcal{L}(\hat{a} | \{n_i\})] - \frac{n^2}{2} .$$

Na prática do ajuste de parâmetros por máxima verossimilhança, sempre se verifica que a função $\ln \mathcal{L}$ seja parabólica na região de interesse, de modo a certificar-se que a estimativa obtida seja não tendenciosa e normal.

A curva em linha tracejada na figura 9.1 corresponde à da relação (9.6), em que C foi escolhido arbitrariamente de modo a facilitar a comparação dos dois gráficos, \hat{a} foi substituído por a_0 e $\text{var}(\hat{a}) = \sqrt{\frac{a_0}{N}}$. A fim de examinarmos o intervalo de 4 desvios-padrões em torno do valor que maximiza a verossimilhança, é preciso verificar se a curva cheia da figura 9.1 é parabólica no intervalo de a para o qual $\ln(\mathcal{L})$ não difere do seu valor seu máximo (perto de $\ln(\mathcal{L}) \approx -5$) por mais que $4^2/2$ unidades², que corresponde ao intervalo $[-13, -5]$, aproximadamente. Vê-se que nessa região a semelhança com a forma assintótica é razoável, embora esteja um pouco mais côncava nos valores pequenos de a . Conclui-se que, com $N = 10$ observações de uma grandeza distribuída conforme a f.p. de Poisson, a estimativa de verossimilhança é aproximadamente normal nos intervalos de alguns desvios-padrões em torno dela, mas não, para regiões distantes da estimativa.

Nesta seção, em um exemplo baseado na distribuição de Poisson, vimos que a função verossimilhança tende assintoticamente a uma gaussiana centrada no valor verdadeiro. As seções seguintes mostram que essas propriedades valem em geral: a função verossimilhança, no limite assintótico, é sempre uma normal, centrada no valor verdadeiro do parâmetro, com variância igual ao limite mínimo da variância.

9.2 Consistência do estimador de Máxima Verossimilhança

Vamos chamar de \mathbf{A} à medida $\{\theta_i; i = 1, \dots, N\}$, com N observações da grandeza θ , cujo valor verdadeiro é θ_0 . A função verossimilhança é obtida a partir

²O fator 2 no denominador vem da expressão (9.6).

da f.d.p. f dos dados θ_i da medida,

$$\mathcal{L}(\theta | \mathbf{A}) = \prod_i^N f(\theta_i | \theta) \quad ,$$

em que θ é um parâmetro, que substitui θ_0 na f.d.p. A estimativa de máxima verossimilhança é o valor $\hat{\theta}$ que maximiza \mathcal{L} , ou seja,

$$\ln \mathcal{L}(\hat{\theta} | \mathbf{A}) \geq \ln \mathcal{L}(\theta | \mathbf{A}) \quad , \quad \forall \theta \quad . \quad (9.8)$$

Na demonstração das propriedades do estimador, vamos nos limitar a este caso da estimação de um único parâmetro; a extensão do resultado para o caso multiparamétrico é possível.

Pode-se demonstrar com grande generalidade que o estimador de máxima verossimilhança é consistente ([Kendall, seção 18-10] ou [Eadie, seção 7.2.3]). Aqui, nos restringiremos em mostrar que, se a função verossimilhança $\mathcal{L}(\theta | \mathbf{A})$ é *regular*, o estimador é consistente. *Regular* tem o mesmo significado usado na seção 7.10: a função \mathcal{L} é contínua e derivável e a ordem das operações “derivada em relação a θ ” e “integração em $d\theta_1 \dots d\theta_N$ ” pode ser invertida. Essas condições são usualmente satisfeitas quando os limites de integração não dependem dos parâmetros, o que ocorre na maior parte das situações de medição em ciências exatas.

Usaremos uma propriedade das funções que é básica no cálculo diferencial, expressa pelo *Teorema do Valor Médio*, que garante que, se $f(x)$ é contínua no intervalo $[a, b]$ e derivável no intervalo $]a, b[$, então

$$f(b) - f(a) = \left. \frac{df}{dx} \right|_{x^*} (b - a) \quad ,$$

onde x^* é algum ponto do intervalo $]a, b[$.

Como \mathcal{L} é regular por hipótese, temos que

$$E \left(\frac{\partial \ln \mathcal{L}}{\partial \theta} \right) = 0 \quad , \quad (7.11)$$

conforme demonstrado na seção 7.10.

Ao aplicar o Teorema do Valor Médio, como enunciado acima, para

$$f = \frac{\partial \ln \mathcal{L}}{\partial \theta} \quad , \quad a = \hat{\theta} \quad \text{e} \quad b = \theta_0 \quad ,$$

obtém-se que

$$\left(\frac{\partial \ln \mathcal{L}}{\partial \theta}\right)_{\hat{\theta}} = \left(\frac{\partial \ln \mathcal{L}}{\partial \theta}\right)_{\theta_0} + (\hat{\theta} - \theta_0) \left(\frac{\partial^2 \ln \mathcal{L}}{\partial \theta^2}\right)_{\theta^*}, \quad (9.9)$$

onde θ^* é um valor no intervalo definido por $\hat{\theta}$ e θ_0 , que pode ser tanto $[\hat{\theta}, \theta_0]$ quanto $[\hat{\theta}_0, \theta]$.

Agora, a condição (9.8), de máxima verossimilhança da estimativa $\hat{\theta}$, corresponde, na situação de regularidade suposta, a

$$\left(\frac{\partial \ln \mathcal{L}}{\partial \theta}\right)_{\hat{\theta}} = 0 \quad \text{com} \quad \left(\frac{\partial^2 \ln \mathcal{L}}{\partial \theta^2}\right)_{\hat{\theta}} < 0. \quad (9.10)$$

A primeira destas relações implica que é nulo o membro esquerdo da equação (9.9), cujo primeiro termo do membro direito pode ser reescrito como

$$\left(\frac{\partial \ln \mathcal{L}}{\partial \theta}\right)_{\theta_0} = \sum_{i=1}^N \left(\frac{\partial [\ln f(\theta_i | \theta)]}{\partial \theta}\right)_{\theta_0}, \quad (9.11)$$

que, quando $N \rightarrow \infty$, tende ao valor médio da fórmula (7.11) pela Lei dos Grandes Números, seção 6.18. Esse termo da fórmula (9.9), portanto, também é nulo.

Como ilustrado na figura 9.1, a derivada parcial de segunda ordem de $\ln \mathcal{L}$, que comparece no segundo termo do membro direito de (9.9), não é nula.³ Assim, como esse fator não é nulo, deduz-se que, quando N tende a infinito,

$$(\hat{\theta} - \theta_0) = 0 \quad \Leftrightarrow \quad \hat{\theta} = \theta_0.$$

Portanto, quando $N \rightarrow \infty$, a estimativa de Máxima Verossimilhança é o valor verdadeiro, ou seja, o estimador é consistente. Uma outra maneira de dizer isso é que o estimador é *assintoticamente* não-tendencioso.

Há um outro resultado que será necessário à próxima seção. Como θ^* está entre θ_0 e $\hat{\theta}$, ele também tende a θ_0 com infinitos dados, de modo que

$$\left(\frac{\partial^2 \ln \mathcal{L}}{\partial \theta^2}\right)_{\theta^*} \xrightarrow{N \rightarrow \infty} \left(\frac{\partial^2 \ln \mathcal{L}}{\partial \theta^2}\right)_{\theta_0} = E \left(\frac{\partial^2 \ln \mathcal{L}}{\partial \theta^2}\right), \quad (9.12)$$

³Somente seria nula se a primeira derivada fosse constante e, portanto, a verossimilhança teria a forma $\exp k\theta$, o que não acontece com as distribuições usuais. Mesmo uma f.d.p. exponencial, como a do exemplo 9.1, não origina uma função verossimilhança dessa forma.

onde a última igualdade deve-se à lei dos grandes números. Usando o resultado da seção 7.10

$$E \left\{ \left(\frac{\partial \ln \mathcal{L}}{\partial \theta} \right)^2 \right\} = -E \left(\frac{\partial^2 \ln \mathcal{L}}{\partial \theta^2} \right) . \quad (7.12)$$

e notando que $\frac{\partial \ln \mathcal{L}}{\partial \theta}$ só depende de θ e dos dados e *não* é nula no espaço todo, verifica-se que $E \left\{ \left(\frac{\partial \ln \mathcal{L}}{\partial \theta} \right)^2 \right\}$ dá um número positivo, que é a variância de $\frac{\partial \ln \mathcal{L}}{\partial \theta}$, uma vez que seu valor esperado é nulo. É costume definir uma grandeza $R(\theta)$ por meio de

$$[R(\theta)]^2 = -E \left(\frac{\partial^2 \ln \mathcal{L}}{\partial \theta^2} \right) > 0 , \quad (9.13)$$

a fim de simplificar a notação, ao mesmo tempo que recordará a independência em relação aos dados (que foram integrados ao calcular o valor esperado) e seu caráter de número (estritamente) *positivo*, uma vez que é o valor esperado do quadrado de uma grandeza. Assim, verifica-se que também a segunda das condições de (9.10) é obedecida.

9.3 Tendência à normalidade da estimativa de Máxima Verossimilhança

No capítulo 4, o método da máxima verossimilhança foi aplicado para deduzir as estatísticas com que se estimam o valor da grandeza e a variância de um conjunto de dados gaussianos. Verificou-se que a média tem f.d.p. gaussiana, mas a estimativa da variância tem f.d.p. de χ^2 ; uma vez que o estimador não costuma ser linear nos parâmetros, raramente as estimativas têm distribuição normal. Entretanto, quando o número de dados vai a infinito, o estimador de máxima verossimilhança é *normal*, sob as mesmas condições de regularidade que adotadas ao demonstrar sua consistência e não tendenciosidade assintótica, como será mostrado aqui.

A relação (9.9), quando se substitui a primeira das condições de máxima verossimilhança (9.10), com um ligeiro rearranjo, fica

$$(\hat{\theta} - \theta_0) = \frac{\left(\frac{\partial \ln \mathcal{L}}{\partial \theta} \right)_{\theta_0}}{- \left(\frac{\partial^2 \ln \mathcal{L}}{\partial \theta^2} \right)_{\theta^*}} . \quad (9.14)$$

9.3. TENDÊNCIA À NORMALIDADE DA ESTIMATIVA DE MÁXIMA VEROSSIMILHANÇA

Na seção 9.2, mostrou-se que (fórmulas 9.12 e 9.13)

$$\left(\frac{\partial^2 \ln \mathcal{L}}{\partial \theta^2} \right)_{\theta^*} \xrightarrow{N \rightarrow \infty} -[R(\theta_0)]^2 \quad ,$$

o que faz o denominador de (9.14) tender a uma constante, quando $N \rightarrow \infty$. No numerador, a derivada de $\ln \mathcal{L}$ vai a zero por causa de (7.11), com variância igual a $[R(\theta_0)]^2$ por causa de (7.12) e da identidade (9.13). Assim, o membro direito tende, quando $N \rightarrow \infty$, a uma variável aleatória de média nula e desvio padrão $[R(\theta_0)]^2$. Como o numerador pode ser encarado como uma soma de variáveis aleatórias, veja relação (9.11), pelo Teorema Central do Limite sua f.d.p. tende à normal. Assim, o membro esquerdo de (9.14), $(\hat{\theta} - \theta_0)R(\theta_0)$, também tende à normal de média nula e desvio padrão $[R(\theta_0)]^2$. Portanto, $\hat{\theta}$ tende à normal com média igual a θ_0 e variância dada por

$$\text{var}(\hat{\theta}, N \rightarrow \infty) = \frac{1}{R(\theta_0)^2} = - \left[E \left(\frac{\partial^2 \ln \mathcal{L}}{\partial \theta^2} \right) \right]^{-1} \quad . \quad (9.15)$$

Exemplo 9.1

A fim de ilustrar o comportamento assintótico da função verossimilhança, considere o caso de um decaimento exponencial, no qual a distribuição de eventos ao longo do tempo é dada por $\lambda e^{-\lambda t}$. Isso poderia corresponder, por exemplo, à probabilidade do decaimento radiativo de átomos ou núcleos ou ao tempo decorrido até que um determinado tipo de equipamento apresente defeito; neste exemplo, porém, não vamos nos preocupar com o realismo e usar tempos da ordem de segundos, por simplicidade.

Suponha que, a fim de estimar o parâmetro λ , meçamos os instantes de ocorrência dos eventos. No caso em que apenas um evento é observado, em $t = 28,6$ s, a função verossimilhança está representada pela curva tracejada da figura 9.2, normalizada de modo que o valor máximo é igual a 1, que é atingido próximo ao valor $0,04/s$ da abscissa, mas não tem uma forma que se pareça com uma gaussiana. Para uma outra medida, em que são observadas três ocorrências do evento em questão, nos instantes (em segundos): 13,4; 14,7 e 20,1, a função verossimilhança (curva pontilhada da figura 9.2), que é o produto da função probabilidade calculada nos instantes acima, é menos assimétrica que no caso anterior. Em um terceiro caso, em que são observados dez eventos nos instantes

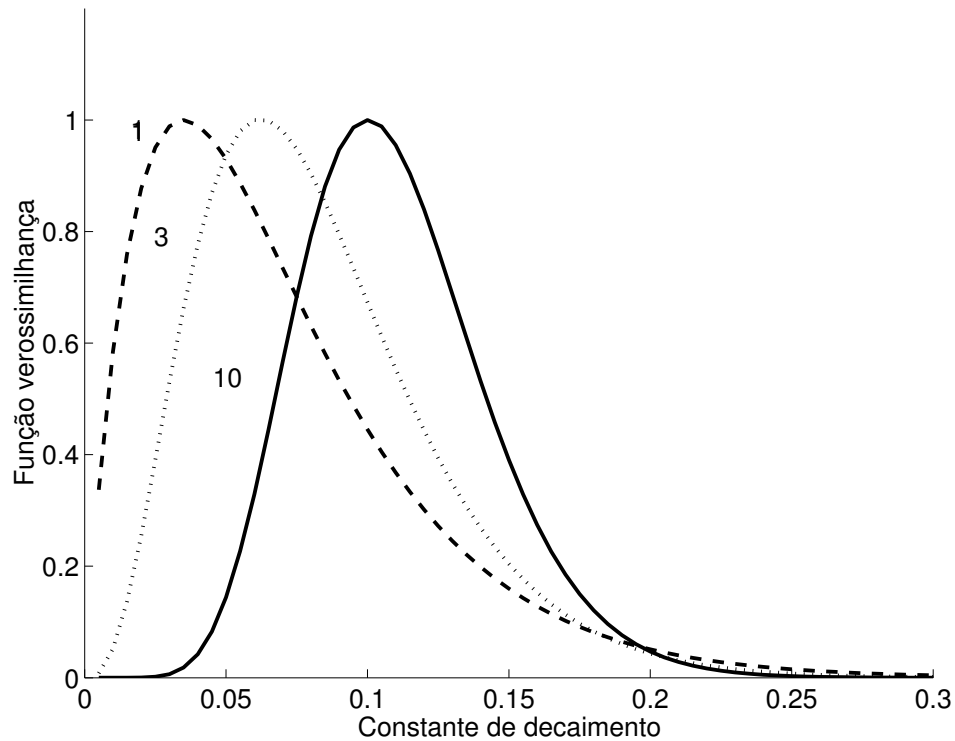


Figura 9.2: Função verossimilhança para conjuntos de N dados que têm f.d.p. exponencial, com $N = 1, 3, 10$, correspondentes às linhas tracejada, pontilhada e contínua, respectivamente.

(em segundos) $\{1, 1; 1, 4; 1, 9; 2, 5; 6, 0; 7, 0; 7, 4; 9, 9; 15, 7; 47, 0\}$, a função verossimilhança corresponde à curva contínua da figura, que tem uma forma mais próxima de uma gaussiana.

9.4 Eficiência assintótica

Comparando a expressão (9.15) da seção anterior com a fórmula (7.17), verifica-se que a variância da estimativa de máxima verossimilhança é igual ao Limite Mínimo de Variância que qualquer estimador t de θ poderá alcançar quando o número de dados for *infinito*. Assim, o estimador de Máxima Verossimilhança é *assintoticamente eficiente* (veja seção 7.14 para a definição de

eficiência), ou seja, tem a menor variância que se pode obter com estimadores não tendenciosos.

Com um raciocínio análogo ao da seção 9.2, a lei dos grandes números permite concluir que

$$\left(\frac{\partial^2 \ln \mathcal{L}}{\partial \theta^2} \right)_{\hat{\theta}} \xrightarrow{N \rightarrow \infty} E \left(\frac{\partial^2 \ln \mathcal{L}}{\partial \theta^2} \right)_{\theta_0} .$$

Ao substituir esse resultado na expressão (9.15) acima, encontra-se uma maneira de estimar a variância de $\hat{\theta}$ quando o número de dados é grande, por meio da relação

$$\text{var}(\hat{\theta}) \cong - \left[\left(\frac{\partial^2 \ln \mathcal{L}}{\partial \theta^2} \right)_{\hat{\theta}} \right]^{-1} . \quad (9.16)$$

9.5 Ajuste simultâneo de vários parâmetros

O Método da Máxima Verossimilhança foi aplicado na seção 4.2 para deduzir as estatísticas usadas para estimar o valor da grandeza e a variância de um conjunto de dados com f.d.p. normal. Nesta seção, ele será aplicado para ajustar múltiplos parâmetros quando os dados **não** têm distribuição normal; o caso gaussiano será retomado na próxima seção.

Suponha que as observações y_i sejam variáveis aleatórias que obedecem a uma função de probabilidade P que depende de uma grandeza z_i , dada por $P[y_i | z_i]$. Agora, z_i é uma função não linear nos parâmetros $\mathbf{p}_0 = (p_1, \dots, p_\mu)$ e depende de um conjunto de variáveis independentes \mathbf{x}_i , de modo que $z_i = z(\mathbf{x}_i, \mathbf{p}_0)$. Assim, a função de probabilidade (ou densidade de probabilidade) dos dados fica

$$P[y_i | z(\mathbf{x}_i, \mathbf{p}_0)] = P[y_i | \mathbf{p}_0] . \quad (9.17)$$

Construindo a função verossimilhança e calculando seu logaritmo, chega-se a

$$\ln \mathcal{L}(\mathbf{p} | \{y_i\}) = \sum_i^N \ln P(y_i | \mathbf{p}) .$$

A estimativa $\hat{\mathbf{p}}$ de máxima Verossimilhança é obtida pela equação (9.10),

$$\left(\frac{\partial \ln \mathcal{L}}{\partial p_\nu} \right)_{\hat{\mathbf{p}}} = 0, \quad \nu = 1, \dots, \mu , \quad (9.18)$$

que constitui um sistema de μ equações não lineares de solução usualmente difícil.

Exemplo 9.2

As estimativas de máxima verossimilhança podem ser encontradas por procedimentos analíticos, numéricos ou gráficos, e a forma usada para encontrá-las não altera suas propriedades.

Considere, por exemplo, um determinado tipo de evento cuja frequência obedece a uma distribuição de Poisson com parâmetro

$$a(t; \lambda_1, \lambda_2) = a(t) = 3000 e^{-\lambda_1 t} + 500 e^{-\lambda_2 t} \quad , \quad (9.19)$$

com t em segundos. Foram medidas as quantidades de eventos de 0,5 em 0,5 s, a partir do instante $t = 0$ s, obtendo-se os valores: 3502, 2108, 1396, 909, 640, 432, 305, 195 e 143 eventos, nessa sequência. A função verossimilhança é

$$\mathcal{L}(\lambda_1, \lambda_2) = \prod_{i=1}^9 \frac{e^{-a(t_i)} a(t_i)^{n_i}}{n_i!}$$

Uma possível maneira de encontrar as estimativas $\hat{\lambda}_1$ e $\hat{\lambda}_2$ é fazer um gráfico do logaritmo da função verossimilhança, $\ln \mathcal{L}$, para diferentes valores de λ_1 e λ_2 , e localizar o máximo dessa função. A figura 9.3, com as curvas de nível de $\ln \mathcal{L}$ com os dados acima, ilustra esse procedimento. Um exame das curvas de nível indica que os valores que maximizam $\ln \mathcal{L}$ e, portanto, também \mathcal{L} , são $\hat{\lambda}_1 = 1,03/\text{s}$ e $\hat{\lambda}_2 = 0,39/\text{s}$.

A opção pelo uso de um procedimento gráfico ou numérico para localizar a solução de máxima verossimilhança depende da situação concreta encontrada.

As equações (9.18) podem ser resolvidas por um procedimento iterativo. Começa-se pela expansão de (9.18) em série de Taylor em torno de uma aproximação \mathbf{p}' da solução, retendo somente o termo de primeira ordem:

$$\left. \frac{\partial \ln \mathcal{L}}{\partial p_\nu} \right|_{\hat{\mathbf{p}}} \cong \left. \frac{\partial \ln \mathcal{L}}{\partial p_\nu} \right|_{\mathbf{p}'} + \sum_{\eta=1}^{\mu} \left. \frac{\partial^2 \ln \mathcal{L}}{\partial p_\eta \partial p_\nu} \right|_{\mathbf{p}'} (\hat{p}_\eta - p'_\eta) \quad .$$

Substituindo-se (9.18) na expressão acima, obtém-se um sistema linear cuja solução é

$$(\hat{\mathbf{p}} - \mathbf{p}') = -\mathbf{M}^{-1} \mathbf{D} \quad , \quad (9.20)$$

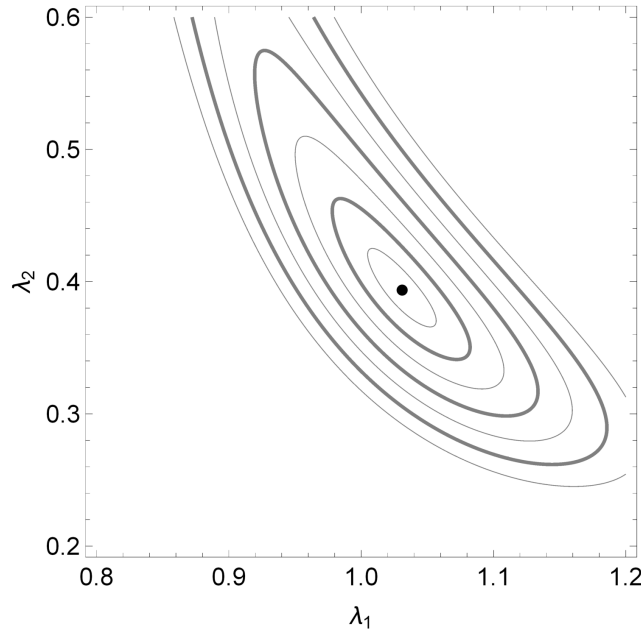


Figura 9.3: Curvas de nível da função verossimilhança em função de λ_1 e λ_2 da relação (9.19). As cotas das curvas correspondem a $\ln \mathcal{L}(\hat{\lambda}_1, \hat{\lambda}_2 | \mathbf{A}) - \frac{n^2}{2}$, com $n = 1, 2, \dots, 7$ e o ponto marca o máximo da função.

equivalente ao procedimento apresentado na seção 4.8 e no capítulo 8, com

$$M_{\eta\nu} = \frac{\partial^2 \ln \mathcal{L}}{\partial p_\eta \partial p_\nu} \Big|_{\mathbf{p}'} \quad \text{e} \quad (9.21)$$

$$D_\nu = \frac{\partial \ln \mathcal{L}}{\partial p_\nu} \Big|_{\mathbf{p}'} . \quad (9.22)$$

Como a expansão em série terá um resto significativo se a aproximação \mathbf{p}' não for muito próxima de $\hat{\mathbf{p}}$, tudo o que se obtém na primeira iteração é uma melhor aproximação \mathbf{p}'' de $\hat{\mathbf{p}}$. Deve-se retornar esta nova aproximação no lugar de \mathbf{p}' em (9.20), (9.21) e (9.22) e reiterar ao longo de todo o processo até convergir. As questões de convergência e de cálculo numérico envolvidas são equivalentes às que serão discutidas adiante, em particular o método de Gauss-Marquardt (seção 9.6.2 abaixo) adapta-se a este problema, de maneira que não entraremos nestes detalhes agora.

As estimativas das incertezas nos parâmetros são obtidas por meio de um generalização da expressão (9.16) acima, de maneira que se estima a matriz de covariâncias dos parâmetros como

$$\mathbf{V}(\hat{\mathbf{p}}) \cong -\mathbf{M}^{-1} \quad . \quad (9.23)$$

A matriz \mathbf{M} da fórmula (9.21) é conhecida como Hessiano da função $\ln \mathcal{L}$.

O Método da Máxima Verossimilhança é a escolha adequada *sempre que a f.d.p. dos dados não for gaussiana e a função a ajustar não for linear nos parâmetros*. É bom recordar que as propriedades ótimas desse estimador só valem para um número grande de dados, de maneira que se deve olhar com reserva as estimativas dos parâmetros e suas respectivas variâncias e covariâncias quando o número de graus de liberdade do problema for pequeno. Em termos práticos, dois ou três graus de liberdade podem originar desastres; uma dezena, deve dar bons resultados, mas sem garantia e desconfie em particular das variâncias (9.23), enquanto que uma centena de graus de liberdade funciona muito bem. Em todos os casos, fazer uma simulação para conhecer o comportamento das estimativas em seu caso específico é sempre conveniente.

9.6 Dados com distribuição normal

Mostramos na seção 4.6 que, na situação onde a f.d.p. dos dados é gaussiana, os métodos dos mínimos quadrados e da máxima verossimilhança são idênticos, uma vez que a função verossimilhança \mathcal{L} fica

$$\mathcal{L}(\mathbf{a} \mid \{y_i\}) \propto \exp(-Q/2) \quad , \quad (9.24)$$

onde Q representa a soma quadrática da relação (8.5) do capítulo anterior, em que \mathbf{a} é o vetor de parâmetros, desde que a relação entre os dados \mathbf{y} e os parâmetros seja linear; esta é uma versão com notação simplificada da fórmula (4.19), e verificamos, agora, que se aplica também a funções que não são lineares nos parâmetros. No caso não-linear que será tratado nesta seção, define-se, além do vetor \mathbf{y} com N linhas que contém os dados experimentais, um vetor \mathbf{g} que corresponde aos valores calculados pela função quando os parâmetros assumem o valor \mathbf{a} , ou seja, $\mathbf{g} = \mathbf{g}(\mathbf{a})$ com $g_i = g(\mathbf{x}_i; \mathbf{a})$. A função Q se escreve, então,

$$Q = [\mathbf{y} - \mathbf{g}(\mathbf{a})]^t \mathbf{V}^{-1} [\mathbf{y} - \mathbf{g}(\mathbf{a})] \quad , \quad (9.25)$$

de maneira a incluir o caso em que as covariâncias entre os dados não são nulas. O estimador representado pelo mínimo desta função Q também é chamado de Mínimos Quadrados e se aplica a funções não lineares nos parâmetros, embora possa apresentar propriedades muito diferentes do caso em que a função é linear nos parâmetros.

Como $g(\mathbf{x}_i, \mathbf{a})$ não é linear em \mathbf{a} , as equações

$$\left. \frac{\partial Q}{\partial a_\nu} \right|_{\hat{\mathbf{a}}} = 0 \quad (9.26)$$

não conduzem a um sistema linear em $\hat{\mathbf{a}}$. Ao contrário, estas equações são habitualmente difíceis de resolver, embora possam dar soluções relativamente simples em casos particulares.

Um procedimento possível consiste em calcular numericamente $Q = Q(\mathbf{a})$ para diversos valores \mathbf{a} e ir *cercando* o mínimo. Este método é prático e conveniente quando o número de parâmetros é igual a um ou dois, mas pode ser muito trabalhoso para um número maior de parâmetros.

9.6.1 O método de Gauss

Uma abordagem que pode ser usada sistematicamente consiste em expandir $g(\mathbf{x}_i; \mathbf{a})$ em série de Taylor em torno de um valor \mathbf{a}' , que se supõe próximo de $\hat{\mathbf{a}}$, e tomar apenas o termo de primeira ordem, ou seja

$$g(\mathbf{x}_i; \hat{\mathbf{a}}) \cong g(\mathbf{x}_i; \mathbf{a}') + \sum_{\nu=1}^{\mu} \left. \frac{\partial g(\mathbf{x}_i; \mathbf{a})}{\partial a_\nu} \right|_{\mathbf{a}'} (a_\nu - a'_\nu) \quad . \quad (9.27)$$

Com a definição

$$\Delta g_i = g(\mathbf{x}_i; \hat{\mathbf{a}}) - g(\mathbf{x}_i; \mathbf{a}') \quad , \quad (9.28)$$

$$X_{i\nu} = \left. \frac{\partial g(\mathbf{x}_i; \mathbf{a})}{\partial a_\nu} \right|_{\mathbf{a}'} \quad \text{e} \quad (9.29)$$

$$\Delta a_\nu = \hat{a}_\nu - a'_\nu \quad , \quad (9.30)$$

pode-se reescrever (9.27) como

$$\Delta g_i = (\mathbf{X}\Delta\vec{a})_i \quad . \quad (9.31)$$

Se a expansão em primeira ordem for boa (isto é, se o resto da expansão (9.27) em primeira ordem for pequeno), Q da fórmula (9.25) pode ser aproximada por

$$Q_{\text{apr}} = [\mathbf{y} - \mathbf{g}(\mathbf{a}') - \mathbf{X}\Delta\vec{a}]^t \mathbf{V}^{-1} [\mathbf{y} - \mathbf{g}(\mathbf{a}') - \mathbf{X}\Delta\vec{a}] \quad .$$

Para tornar a expressão acima mais parecida com a que se aplica no caso de funções lineares, define-se

$$\mathbf{y} - \mathbf{g}(\mathbf{a}') = \mathbf{y}' \quad , \quad (9.32)$$

e obtém-se, finalmente,

$$Q_{\text{apr}} = [\mathbf{y}' - \mathbf{X}\vec{\Delta a}]^t \mathbf{V}^{-1} [\mathbf{y}' - \mathbf{X}\vec{\Delta a}] \quad . \quad (9.33)$$

(Note que a matriz de covariância de \mathbf{y}' é igual à de \mathbf{y} , uma vez que $\mathbf{g}(\mathbf{a}')$ é um vetor cujos elementos são valores numéricos sem erro.) É fácil minimizar Q_{apr} da expressão acima, porque o conjunto das equações definidas por

$$\left. \frac{\partial Q_{\text{apr}}}{\partial \Delta a_\nu} \right|_{\hat{\Delta a}} = 0$$

é o sistema linear

$$\mathbf{X}^t \mathbf{V}^{-1} \mathbf{y}' = (\mathbf{X}^t \mathbf{V}^{-1} \mathbf{X}) \hat{\Delta a} \quad , \quad (9.34)$$

cuja solução permite determinar $\hat{\Delta a}$, analogamente à expressão (8.7) do MMQ no modelo linear. No entanto, a equação (9.34) usa a definição (9.30), que só fornece o vetor $\hat{\mathbf{a}}$ que minimiza $Q(\mathbf{a})$ se o resto da expansão (9.27) em série de Taylor for pequeno. Espera-se, porém, que os novos valores

$$\mathbf{a}'' = \mathbf{a}' + \vec{\Delta a} \quad (9.35)$$

sejam mais próximos de $\hat{\mathbf{a}}$ que as estimativas \mathbf{a}' anteriores. O método de Gauss baseia-se nessa ideia.

As etapas do algoritmo iterativo de procura de \hat{a} são:

- i)* inicia-se o procedimento com uma aproximação \mathbf{a}' para $\hat{\mathbf{a}}$ e calcula-se $Q(\mathbf{a}')$ pela expressão (9.25);
- ii)* calculam-se o vetor \mathbf{y}' e a matriz \mathbf{X} por meio das expressões (9.32) e (9.29), respectivamente;
- iii)* calcula-se $\vec{\Delta a}$ a partir do sistema linear da fórmula (9.34);
- iv)* calculam-se \mathbf{a}'' e $Q(\mathbf{a}'')$ pelas relações (9.35) e (9.25), respectivamente, e
- v)* se as variações na soma dos quadrados dos resíduos, $Q(\mathbf{a}'') - Q(\mathbf{a}')$, e nos μ parâmetros, Δa_ν , forem pequenas ao ponto de poderem ser ignoradas, termina-se o processo iterativo e adota-se $\hat{\mathbf{a}} = \mathbf{a}''$; caso contrário, repetem-se as etapas a partir de *ii*, substituindo \mathbf{a}' por \mathbf{a}'' bem como $Q(\mathbf{a}')$ por $Q(\mathbf{a}'')$.

Uma vez que o processo convirja, a matriz de covariância dos parâmetros pode ser estimada como

$$\mathbf{V}(\hat{\mathbf{a}}) \cong (\mathbf{X}^t \mathbf{V}^{-1} \mathbf{X})^{-1} \quad , \quad (9.36)$$

com a matriz \mathbf{X} calculada com o valor obtido para $\hat{\mathbf{a}}$. Este resultado corresponde à generalização da fórmula (9.16) e da identidade (9.13), sendo, portanto, a matriz de variâncias *assintótica*. Por outro lado, pode-se obter (9.36) pela fórmula aproximada da variância de uma função de variáveis aleatórias, fórmulas (3.6) e (3.7), onde ela é aproximada justamente pelo fato da função não ser linear nos parâmetros.

O critério de convergência sobre Δa_ν consiste em compará-lo com a incerteza estimada para o parâmetro a_ν por meio da fórmula (9.36), e o relativo a Q , em verificar se sua variação é muito menor que 1, por uma razão que discutiremos adiante.

Para a inversão de $\mathbf{X}^t \mathbf{V}^{-1} \mathbf{X}$, procede-se como descrito na seção 8.14, uma vez que as diferenças nas ordens de grandeza dos valores numéricos dos diferentes parâmetros também causa dificuldades nas contas quando as funções não são lineares nos parâmetros de ajuste.

9.6.2 O método de Gauss-Marquardt

O método de Gauss é caracterizado por convergir lentamente para $\hat{\mathbf{a}}$, o que significa que se realizam dezenas de iterações, tipicamente. Além disso, é necessário disparar o processo com uma boa estimativa inicial. O procedimento que descreveremos nesta seção, conhecido como *método de Gauss-Marquardt* [Marquardt], é mais adequado — converge mais rapidamente e é mais tolerante em relação à estimativa inicial. Além do artigo original, a descrição do algoritmo pode ser encontrada também, por exemplo, em [Bevington], e uma descrição completa das razões pelas quais ele converge melhor que o método de Gauss para $\hat{\mathbf{a}}$ no livro [Bard].

Neste método, procede-se como no método de Gauss descrito acima, mas os passos *iii*) e *iv*) são ligeiramente alterados. Define-se uma matriz \mathbf{M}^λ com elementos de matriz

$$M_{\eta\nu}^\lambda = (1 + \lambda \delta_{\eta\nu}) (\mathbf{X}^t \mathbf{V}^{-1} \mathbf{X})_{\eta\nu} \quad . \quad (9.37)$$

Inicialmente, escolhe-se λ com um valor moderadamente pequeno, digamos 10^{-3} . Com referência aos passos da seção anterior, as alterações são:

iii) calcula-se a matriz \mathbf{M}^λ pela expressão (9.37) e o vetor de incrementos resolvendo o sistema

$$\mathbf{X}^t \mathbf{V}^{-1} \mathbf{y}' = \mathbf{M}^\lambda \vec{\Delta a} \quad ;$$

iv) calculam-se \mathbf{a}'' e $Q(\mathbf{a}'')$ pelas relações (9.35) e (9.25), respectivamente, mas, se a variação na soma dos quadrados dos resíduos,

$$Q(\mathbf{a}'') - Q(\mathbf{a}'),$$

for **positiva**, aumenta-se λ , tipicamente por um fator 10, e repete-se os itens *iii)* e *iv)* até que se consiga obter um valor de $Q(\mathbf{a}'')$ **menor** que $Q(\mathbf{a}')$, quando se avança para o passo v e se reduz λ para a próxima iteração, tipicamente por um fator 10.

Ao final do processo, calcula-se a matriz de covariâncias pela expressão (9.36), evitando-se utilizar \mathbf{M}^λ para este fim, porque essa matriz pode embutir um λ significativamente diferente de zero. É também preciso escolher um valor máximo para λ — caso não se consiga reduzir Q com λ menor ou da ordem de 10, há pouca chance do processo convergir e é melhor reiniciá-lo com outra estimativa \mathbf{a}' . Às vezes, quando se está muito próximo ao mínimo, um algoritmo com critérios de convergência no passo v exageradamente exigentes pode tentar **aumentar** λ indefinidamente, desde que pode ser impossível reduzir Q por problemas de (im)precisão numérica.

Este método converge mais rapidamente que o de Gauss, tipicamente em cerca de 6 iterações, e é muito mais tolerante em relação à estimativa inicial dos parâmetros. Quando os parâmetros têm valores numéricos com ordens de grandeza muito distintas, é conveniente efetuar a transformação \mathbf{M} para \mathbf{M}' descrita na seção 8.13 antes de inverter a matriz ou usar rotinas de inversão de matrizes destinadas ao método dos mínimos quadrados. Caso se aplique a transformação \mathbf{T} descrita em 8.14, no passo *ii)* se calcula

$$\mathbf{M}' = \mathbf{T}(\mathbf{X}^t \mathbf{V}^{-1} \mathbf{X}) \mathbf{T}$$

e no passo *iii)* calcula-se

$$(\mathbf{M}^\lambda)^{-1} = \mathbf{T}[(\mathbf{M}' + \lambda \mathbf{I})^{-1}] \mathbf{T} \quad ,$$

de maneira a evitar o recálculo da matriz \mathbf{X} e do produto $\mathbf{X}^t \mathbf{V}^{-1} \mathbf{X}$ durante o ciclo de procura do λ necessário para reduzir o valor de Q (passos *iii)* e *iv)*). Isso importa especialmente quando $g(\mathbf{x}_i; \mathbf{a}')$ não é uma função analítica, por exemplo uma convolução calculada numericamente ou a solução numérica de uma equação diferencial, e/ou o número de dados é grande, talvez dezenas de milhares.

9.7 Simulação do ajuste de parâmetros não lineares

O método da máxima verossimilhança é usualmente *tendencioso* quando o número de dados é finito, de maneira que é uma boa ideia tomar muitos dados nos experimentos que serão analisados por esse método. No entanto, isso nem sempre é possível, pelas mais diversas razões, que vão desde custo até a raridade do fenômeno em observação. Nesses casos, é preciso verificar a qualidade do resultado, o que pode ser feito por simulação. Esta seção destina-se a apresentar um exemplo do procedimento de teste e as características gerais das situações que causam problemas no tratamento dos dados.

Vamos usar como exemplo a análise de um experimento que se destina a medir a amplitude, A , e a frequência, ω , de um movimento harmônico ao longo do eixo y com o passar do tempo t . O conjunto de dados que se obteve é $\{(t_i, y_i, \sigma_i), i = 1, \dots, N\}$, com

$$y_i = A \operatorname{sen}(\omega t_i) + \epsilon_i \quad , \quad (9.38)$$

onde o erro ϵ_i não é tendencioso, $\langle \epsilon_i \rangle = 0$, as variâncias são conhecidas, $\langle \epsilon_i^2 \rangle = \sigma_i^2$, e os dados são estatisticamente independentes, com f.d.p. normal. A unidade de distância é o μm , a de tempo é o milissegundo (ms) e a de frequência o rd/s; daqui para a frente, omitiremos essas unidades.

9.7.1 Uma situação confortável

A figura 9.4 ilustra um conjunto de dados experimentais com $N = 4$. Os parâmetros obtidos pelo MMQ foram $A = 1264(217)$ e $\omega = 60,4(14)$, mas esse resultado, que corresponde ao limite assintótico, não deve ser adotado sem verificar sua validade.

Sabemos que as propriedades do método dos mínimos quadrados, nesse caso de ajuste de parâmetros não lineares, decorrem de sua correspondência com o método da máxima verossimilhança, de modo que ele pode ser tendencioso, a distribuição estatística das estimativas pode não ser gaussiana e os desvios padrão calculados podem ser inadequados, uma vez que o número de dados é muito pequeno para que possamos aceitar como válidas as propriedades assintóticas de não tendenciosidade, eficiência e normalidade, sem testar.

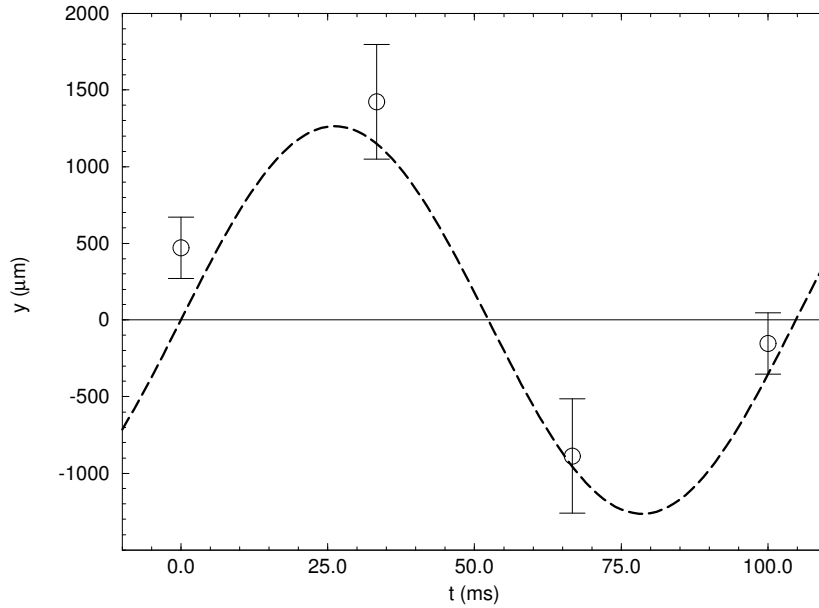


Figura 9.4: Posição em função do tempo do movimento harmônico descrito no texto. Os círculos representam os dados experimentais e as barras de incerteza cobrem o intervalo de um desvio-padrão para mais ou para menos do valor medido. A linha tracejada é a curva com os parâmetros ajustados.

Uma maneira de verificar a qualidade do ajuste é simular as medições, o que permite impor valores para as grandezas e avaliar a tendenciosidade ou não das estimativas pela comparação entre os resultados da análise com os valores verdadeiros impostos inicialmente. Neste exemplo, escolheu-se⁴ $A_0 = 1000$ e $\omega_0 = 20 \pi$ e a simulação foi repetida mil vezes, de acordo com as etapas dos itens *i*) e *ii*) a seguir:

- i*) Para cada valor da coordenada independente, calcula-se o valor y_i da coordenada dependente correspondente pela fórmula (9.38), com os valores A_0 e ω_0 acima, onde o pseudo aleatório gaussiano ϵ_i foi gerado de maneira a ter média zero e variância igual a

$$\sigma_i^2 = [f A_0 \text{sen}(\omega_0 t_i)]^2 + s^2 \quad , \quad (9.39)$$

com $f = 0,2$ e $s = 200$.

⁴Quaisquer valores A_0 e ω_0 próximos dos ajustados poderiam ter sido usados.

- ii) Com os valores simulados, calcula-se \hat{A} e $\hat{\omega}$ pelo procedimento descrito na seção (9.6.1) e os respectivos desvios padrão, σ'_A e σ'_ω , pela fórmula (9.36).

Em seguida, constrói-se o histograma dos mil valores da amplitude ajustados nas mil simulações, cujo valor médio calculado foi $\bar{A} = 1002$ com desvio padrão da série $\sigma_A = 217$, portanto não tendencioso. Um resultado análogo foi obtido para ω , com média $\bar{\omega} = 62,8$ e desvio padrão da série $\sigma_\omega = 2,2$. Os resultados para o desvio padrão de A em cada simulação flutuaram entre 216 e 218, com alguns valores menores até 210, de maneira que podemos adotar o valor calculado para σ_A . Já os resultados para o desvio padrão de $\hat{\omega}$ flutuaram entre 1,4 e 4, de maneira que σ'_ω é mal definido, e a melhor estimativa corresponde a usar o desvio padrão da simulação, $\sigma_\omega \approx 2,2$. Ou seja, o resultado a ser adotado é $A = 1264(217)$ e $\omega = 60,4(22)$.

9.7.2 Uma situação limite

Considere agora N pontos, tomados no intervalo $0 \leq t \leq \frac{T}{4}$, em que T é o período. Nessas condições e com incertezas equivalentes à do item anterior, é preciso que o conjunto de dados tenha dezenas de pontos para ser útil. A fim de preparar um exemplo das dificuldades quando N é pequeno, reduzimos os desvios-padrões por um fator 2, usando $f = 0,1$ e $s = 100$ na fórmula (9.39), mas com os mesmos valores de $A = 1000$ e $\omega_0 = 20\pi \text{ s}^{-1}$.

Adotando um procedimento análogo ao do item anterior, repetiu-se a simulação 10000 vezes para alguns valores de N , com os resultados listados na tabela 9.1. A última coluna da tabela dá o número de conjuntos de dados que deram estimativas aceitáveis para os parâmetros ($A > 0$ e ω da ordem de grandeza esperada para o fenômeno). Preliminarmente, note que $N = 4$ não dá resultados úteis — o desvio-padrão de \hat{A} é metade do seu valor, de modo que a estimativa é compatível com 0. $N = 6$ ainda dá resultados marginalmente utilizáveis e, mesmo para esse número de dados, cerca de 3 % das simulações dão parâmetros ajustados inaceitáveis; esses 3 % são uma estimativa da probabilidade de realizarmos essa medida nessas condições com 6 dados e terminar a análise com um resultado inconsistente ($A < 0$ ou ω com ordem de grandeza diferente da esperada para o fenômeno). Somente a partir de $N = 20$ a probabilidade de obter-se um conjunto de dados que não permitirá análise é pequena. Assim, caso essa simulação se destine a planejar o experimento, deve-se concluir que o número de dados a obter precisa ser maior que 20.

Verifica-se que o método é tendencioso, porém pouco – a diferença entre o valor médio das quase 10000 repetições que levaram a resultados aceitáveis (o número de sucessos no procedimento de ajuste está na última coluna da tabela 9.1) e o valor verdadeiro é muito menor que o desvio-padrão de um único ajuste, tanto para a amplitude quanto para a frequência. Por exemplo, quando $N = 20$, a tendenciosidade média em A é cerca de 38, a comparar com um desvio padrão de 186. Já as estimativas dos desvios padrões flutuam muito de caso a caso, podendo ser muito diferentes dos valores corretos, mesmo quando $N = 100$ dados.

N	simulação				ajuste		sucessos
	\bar{A}	$\bar{\omega}$	σ_A	σ_ω	σ'_A	σ'_ω	
4	1120	61,7	495	15,8	80-600	9-40	9494
6	1096	61,6	395	14,7	60-700	5-80	9683
10	1071	61,5	297	12,7	40-400	7-20	9887
20	1038	61,8	186	9,5	40-250	6-15	9958
30	1022	62,1	105	7,7	30-200	4-13	9988
100	1006	62,6	42	4,2	20-80	3-5	9998

Tabela 9.1: Valores médios dos resultados dos ajustes dos parâmetros A e ω aos conjuntos de dados simulados (2ª e 3ª colunas), e respectivos desvios padrão (4ª e 5ª colunas). A 6ª e 7ª colunas apresentam os desvios padrão de A e de ω estimados pelo método dos mínimos quadrados que, no caso de ω , é uma faixa de valores, porque esse resultado dependia muito de cada ajuste em particular. A primeira coluna corresponde ao número de dados experimentais. Os desvios-padrões de \bar{A} e $\bar{\omega}$ são aproximadamente iguais a 1 % dos respectivos desvios-padrões, uma vez que foram realizadas 10000 simulações. Foram simulados 10000 conjuntos de dados para cada N , mas nem todos convergiram para uma solução aceitável; o número de conjuntos em que houve convergência está apresentado na última coluna.

A tabela 9.1 sugere uma maneira de eliminar a tendenciosidade, que corresponde a utilizá-la para estimar a tendenciosidade para um certo N e adicionar ao resultado individual obtido. Da tabela, verifica-se que, para $N = 10$, a tendenciosidade é 71. Assim, quando se obtém um resultado para a amplitude igual a \hat{A} , devemos adotar como resultado⁵ $\hat{A}_{\text{corrigido}} = \hat{A} - 71$. A maior mudança no resultado final corresponderá, porém, a abandonar os valores dos desvios padrões da frequência obtido no ajuste e usar os valores das 4ª e 5ª colunas na tabela.

⁵Melhor seria efetuar simulações com A_0 em torno de $\hat{A} - 71$, a fim de encontrar que valor dá resultados iguais a \hat{A} , em média.

A outra maneira de eliminar a tendenciosidade exige mudanças no procedimento experimental, de modo que os desvios padrões de cada um dos dados seja menor. Reduzindo o desvio-padrão de cada dado por uma ordem de grandeza, mesmo $N = 4$ dá estimativas dos parâmetros com tendenciosidade muito menor que o desvio-padrão e das variâncias próximas do que se obtém por simulação, além da probabilidade de produzir um conjunto de dados que leve a estimativas divergentes em relação aos valores verdadeiros ser menor que 0,5 %.

9.7.3 Discussão

Do ponto de vista matemático, a origem da dificuldade observada no ajuste desses parâmetros decorre de que tanto a derivada como a curvatura da função seno, no quarto de período observado, não são muito pronunciadas e têm sempre mesmo sinal. Discriminar entre a solução verdadeira e outras com amplitude muito maior que 1000 e frequência bem menor que 20π exige precisão no conjunto dos dados, seja pela sua quantidade ou pelo desvio-padrão relativamente pequeno. Assim, com um número pequeno de pontos, fica até mesmo difícil falar em tendenciosidade, uma vez que a característica principal corresponde a obter soluções extremamente diferentes para medições semelhantes, sendo que apenas para $N > 10$ os resultados começam a ser independentes do particular conjunto de dados, mas ainda com algum risco de obter resultados inadequados quando $N \sim 10$.

Note que, no ajuste de parâmetros lineares, não há nenhum equivalente a esse comportamento inconveniente do ajuste de parâmetros não-lineares com o número de dados. Já a dependência com o intervalo escolhido para a variável independente existe também no modelo linear. Note, por exemplo, a dependência na precisão da inclinação de uma reta com a dispersão dos dados conforme a relação (8.23), embora nesse caso linear a dificuldade limita-se a obter um desvio-padrão grande quando a dispersão é pequena, mas não gera resultados tendenciosos.

Embora seja claramente inadequado medir um movimento harmônico por um intervalo de tempo muito inferior ao período, é possível que isso seja obrigatório, devido a circunstâncias ligadas ao sistema em estudo, tanto pela dificuldade em observá-lo demoradamente (o sistema pode aquecer, por exemplo, e mudar o comportamento) quanto pelo período muito longo (tal como na precessão dos equinócios). Nesses casos, não se pode esquecer que as propriedades das estimativas de máxima verossimilhança são assintóticas e o menor número

de observações que garantirá estimativas “seguras” — consistentes, não tendenciosas e com variância estimada corretamente — dependerá de detalhes do modelo e das incertezas experimentais.

Assim, concluímos que, em uma medida confortável, onde pode-se observar o fenômeno plenamente e com incerteza relativamente pequena, convém obter muitos pontos experimentais ao longo de toda a faixa de valores da variável independente, o que permitirá aplicar o estimador de máxima verossimilhança com confiança. Nas situações extremas, em que for obrigatório obter uma amostra limitada ou com incertezas grandes, faz-se necessário testar a aplicação do método da máxima verossimilhança por simulação, uma vez que não se pode esquecer de considerar as propriedades dentro das condições em que foram deduzidas.⁶

9.8 Estimativa dos intervalos de confiança

Na seção 9.4 foi mostrado que a função verossimilhança tende assintoticamente à f.d.p. normal, com variância

$$\sigma^2 = \text{var}(\hat{\theta})_{N \rightarrow \infty} = E \left[\left(\frac{\partial^2 \ln \mathcal{L}}{\partial \theta^2} \right) \right]^{-1}, \quad (9.15)$$

ou seja, nesse limite em que $N \rightarrow \infty$,

$$\mathcal{L} \propto \exp \left\{ -\frac{(\theta - \theta_0)^2}{2\sigma^2} \right\}.$$

ou, tirando o logaritmo da expressão acima,

$$\ln \mathcal{L} = -\frac{(\theta - \theta_0)^2}{2\sigma^2} + \text{constante}. \quad (9.40)$$

A figura 9.5 ilustra a função dada por (9.40), tomando-se a constante nula por conveniência. No limite assintótico, ela é exatamente uma parábola. Veja que, no exemplo da seção 9.1, a fórmula (9.6), que corresponde ao limite assintótico de $\ln \mathcal{L}$, tem exatamente a forma de (9.40).

É imediato que os valores q_1 e q'_1 para os quais $\ln \mathcal{L} = -1/2$ (equivalente a $\theta - \theta_0 = \sigma$) correspondem aos limites do intervalo de confiança de 68%. Os

⁶É a necessidade prática de entender as propriedades dos estimadores e suas condições de validade que nos levou a apresentar essas deduções ao longo do texto.

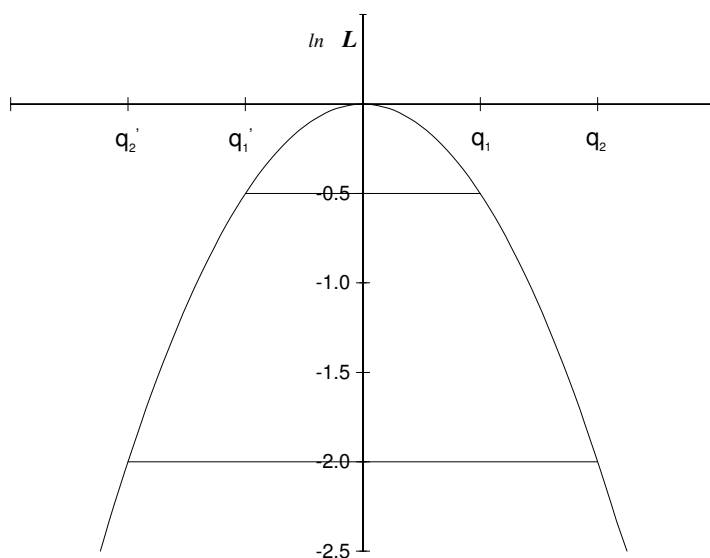


Figura 9.5: Limite assintótico da Função Verossimilhança (9.33), transladando-a de maneira a obter um máximo igual a 0.

valores q_2 e q'_2 tais que $\ln \mathcal{L}(q_2) = \ln \mathcal{L}(q'_2) = -2$ (equivalente a $\theta - \theta_0 = 2\sigma$), correspondem aos limites do intervalo de confiança de 95,5% e assim por diante, desde que neste limite de muitos dados \mathcal{L} é gaussiana e centrada no valor verdadeiro.

Essa discussão sugere dois procedimentos:

i) Verificar se a função $\ln \mathcal{L}$ é aproximadamente parabólica em torno do valor da estimativa $\hat{\theta}$, o que indica que o número de dados é suficientemente grande para valerem as propriedades assintóticas de não tendenciosidade e variância mínima, quando a estimativa do desvio-padrão é adequada. Essa prática é recomendada sempre, mas ainda mais fortemente se houver desconfiança acerca da validade do método de ajuste pela máxima verossimilhança (o que engloba mínimos quadrados quando os dados são gaussianos, porque $Q = -2 \ln \mathcal{L} + \text{Constante}$).

ii) Aproximar a f.d.p. da estimativa pela normal, de modo que o intervalo de confiança de 68% é representado por $[q_1, q_2]$ de modo que

$$\ln \mathcal{L}(q_1) - \ln \mathcal{L}(\hat{\theta}) = \ln \mathcal{L}(q_2) - \ln \mathcal{L}(\hat{\theta}) = -1/2 \quad (68\%) \quad ,$$

onde a subtração do termo $\ln \mathcal{L}(\hat{\theta})$ destina-se a deslocar a origem do eixo de modo que $\ln \mathcal{L}$ tenha o gráfico da figura 9.5. Assim, o intervalo de 95,4% de

confiança é $[q'_1, q'_2]$ em que

$$\ln \mathcal{L}(q'_1) - \ln \mathcal{L}(\hat{\theta}) = \ln \mathcal{L}(q'_2) - \ln \mathcal{L}(\hat{\theta}) = -2 \quad (95\%)$$

e assim por diante. A situação é a mesma para ajuste pelo método dos mínimos quadrados, onde o intervalo com nível de confiança equivalente a $[\hat{\theta} - n\sigma, \hat{\theta} + n\sigma]$, dada pela regra acima (transformada pela relação $\ln \mathcal{L} = -Q/2 + \text{constante}$ que faz o fator $-1/2$ desaparecer) é $[q_n, q'_n]$ tal que $Q(q_n) = Q(q'_n) = Q(\hat{\theta}) + n^2$. Esse procedimento é também o melhor quando se constatar, na etapa **i**), que a função $\ln \mathcal{L}$ é assimétrica, mas nesse caso os intervalos de confiança serão assimétricos, cujo uso no lugar de $\hat{\theta} \pm \sigma$ com σ calculado por meio da fórmula (9.16) revela que as condições de validade das propriedades de não tendenciosidade, eficiência e normalidade da estimativa não estão satisfeitas, o que é uma informação importante.

Toda a discussão pode ser estendida a várias variáveis. Considere o exemplo da seção 9.7, no ajuste dos parâmetros A e ω aos dados da figura 9.4, que formam um conjunto com apenas 4 pontos. A figura 9.6 mostra o gráfico de Q em função de ω quando o parâmetro $A = \hat{A}$. De um lado, note a parabolicidade em torno do mínimo absoluto, uma boa indicação a respeito da qualidade do ajuste. Porém, note que o mínimo não é suficientemente pronunciado. Como a relação entre $\ln \mathcal{L}$ e Q envolve um fator -2 , como discutido na seção precedente, os valores de ω a três desvios padrão são determinados pelos valores de ω para os quais $Q = Q_{\text{mínimo}} + 9$, onde Q já não tem mais a forma parabólica. As dificuldades encontradas neste ajuste estão relacionadas a este desvio da parabolicidade de Q nessa região. Note, também, que o desvio-padrão obtido desse gráfico, onde $A = \hat{A}$, um valor fixo e constante para o outro parâmetro, é o desvio-padrão condicional a um determinado valor de A , portanto, um pouco menor que o da tabela 9.1.

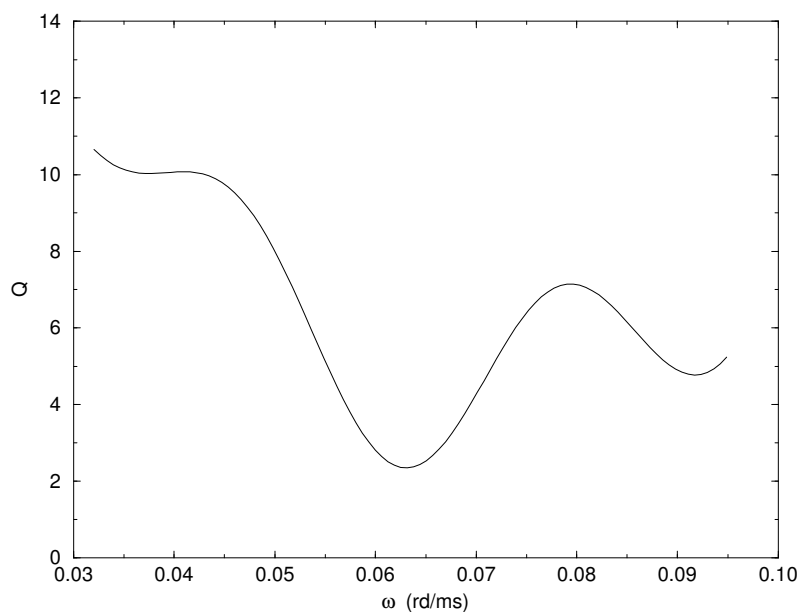


Figura 9.6: Soma dos quadrados dos resíduos ponderados pelo inverso das variâncias, no ajuste dos dados da figura 9.4 pela fórmula (9.38). A função está desenhada com $A = \hat{A}$, o valor de A que minimiza a soma dos quadrados dos resíduos.

EXERCÍCIOS

- 9.1. Deseja-se estimar a taxa de ocorrência λ de um determinado tipo de evento que obedece a uma distribuição de Poisson com parâmetro λt , sendo t o tempo de medida,

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

- Escreva a função verossimilhança $\mathcal{L}(\lambda | n)$.
- Determine o estimador de máxima verossimilhança de λ .
- Mostre que esse estimador é consistente, não tendencioso e de variância mínima.
- Mostre que no limite assintótico $\mathcal{L}(\lambda | n)$ é uma distribuição normal.

- 9.2. Considere a distribuição de Poisson

$$P(n) = \frac{e^{-t/\tau} (t/\tau)^n}{n!} ,$$

onde t é o tempo de medida e τ o parâmetro que se quer estimar. Mostre que a estimativa de máxima verossimilhança de τ é consistente, mas tendenciosa.

9.3. Considere uma distribuição binomial com parâmetros N e p .

(a) Mostre que a estimativa de máxima verossimilhança de $r = p^2$ é

$$\hat{r} = \frac{n^2}{N^2} \quad .$$

(b) Mostre que r é tendencioso e consistente.

(c) Determine a f.d.p. de \hat{r} quando $N \rightarrow \infty$, verificando que essa f.d.p. é normal.

9.4. A taxa de ocorrência de um determinado tipo de evento obedece a uma distribuição de Poisson com parâmetro $a_i = 100e^{-\lambda t_i}$. Considerando os dados abaixo (t_i, n_i) , ajuste λ e estime seu desvio padrão usando o método da máxima verossimilhança.

t_i	0	1	2	3	4	5	6	7	8	9	10	11	12
n_i	95	85	89	83	56	59	51	52	50	41	49	28	36

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