ANÁLISE ESTATÍSTICA DE MEDIDAS EM CIÊNCIAS EXATAS

Vito R. Vanin, Philippe Gouffon, Otaviano Helene

Março 2023

Capítulo 7

Probabilidade e estimação

Neste capítulo, vamos construir métodos de análise aplicáveis aos dados obtidos em experimentos de laboratório com base em critérios interpretados no quadro da Teoria da Probabilidade, que foi apresentada no capítulo anterior. Assim, como nenhuma função dos dados (estatística) permite obter o valor verdadeiro da grandeza observada, apenas estimá-lo, os critérios de escolha da estimativa *ótima* estarão associados à maneira como ela (estimativa) se distribui ao longo de uma sequência de medidas.

Começaremos por atribuir um significado ao termo *probabilidade* que se aplica aos experimentos comuns e é tão concreto quanto possível. Nos deteremos um pouco na interpretação bayesiana do Teorema de Bayes, tanto para permitir seu uso nas situações em que ela cabe, quanto para evitar usá-la involuntária ou desnecessariamente.

Com o conceito de flutuação estatística, podemos dizer que um método é bom quando retorna valores em torno do valor verdadeiro, às vezes maior, às vezes menor. Essa é a idéia básica contida no critério da *não-tendenciosidade*, que discutiremos na seção 7.8. Infelizmente, nem sempre é possível encontrar um método com essa propriedade, de modo que discutiremos primeiro o critério da *consistência* (seção 7.6): o estimador permitiria calcular o valor verdadeiro se usado em uma medida que se efetuasse com esforço ilimitado, com a obtenção de infinitos dados. Essa é uma característica muito mais do estimador, enquanto maneira de estimar, que das estimativas que obteremos na realidade, onde o número de dados será finito.

Os dois critérios acima podem ser preenchidos por dois ou mais estimadores, de modo que é preciso ordenar as várias possibilidades com base em algum índice. Assim, o melhor método é aquele que, para uma medida com um certo número de dados, resulte na estimativa de *menor incerteza*. Esse é o critério da *eficiência*, discutido na seção 7.9.

Um quarto critério, a *robustez*, corresponde à independência do estimador em relação ao comportamento estatístico dos dados, mas não o discutiremos aqui. O uso da estatística não-paramétrica, da qual falamos um pouco na seção 1.11, é recomendado quando não se pode ou não se deseja conhecer a forma da função de probabilidade das observações.

Neste capítulo, portanto, estabeleceremos critérios para qualificar os estimadores e discutiremos a existência de métodos *ótimos*, em um sentido aceitável, enquanto nos seguintes mostraremos as condições nas quais os estimadores habituais têm essas propriedades.

7.1 Algumas definições de probabilidade

A ligação entre o universo dos fenômenos mensuráveis com a Teoria que vimos no capítulo precedente pode, em princípio, ser estabelecida por qualquer método que siga as regras lá desenvolvidas. Nesta seção, apresentaremos as abordagens mais usadas: eventos equiprováveis, probabilidade como frequência relativa e probabilidade como grau de confiança relativa, todas compatíveis com a Teoria da Probabilidade.

7.1.1 Eventos Equiprováveis

Uma hipótese válida em certos casos é a de todos os eventos de um espaço amostral serem equiprováveis. Por exemplo, se um dado, desses usados em jogos, não apresenta defeitos — tem massa distribuída uniformemente, faces planas, arestas iguais e todas as demais características que o façam ser simétrico em relação a todas as faces — e é lançado de forma totalmente aleatória, então a probabilidade de qualquer uma das faces cair "para cima" é igual às outras. Nessas condições, todos os seis eventos possíveis são equiprováveis.

Outro exemplo é o sistema de três detetores de radiação do capítulo anterior, com um arranjo simétrico como ilustra a figura 7.1, em que cada detetor cobre um gomo de 120º de uma esfera, no centro da qual está uma fonte radioativa, que emite radiação de forma totalmente isotrópica. Essas hipóteses garantem que os eventos: {o detetor **a** foi ativado}, {o detetor **b** foi ativado} e {o detetor **c** foi ativado} são igualmente prováveis.

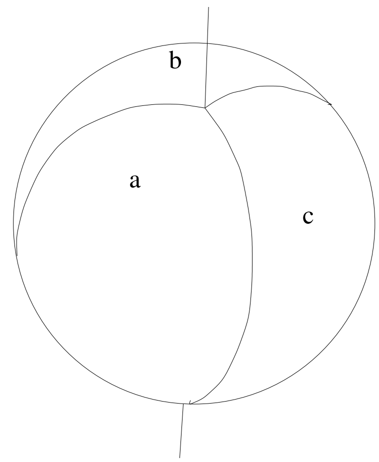

Figura 7.1: Sistema de três detetores na forma de gomos idênticos que formam uma esfera. Uma fonte radioativa, que emita radiação de forma isotrópica, vai gerar eventos de deteção com mesma probabilidade em cada um dos gomos.

Uma vez identificados os eventos equiprováveis, podemos determinar a probabilidade de outros eventos relacionados com eles. Por exemplo, a partir do conhecimento de que as diferentes faces de um dado podem cair para cima com chances iguais, podemos deduzir as probabilidades dos valores possíveis da soma das faces quando jogamos um dado duas vezes. Noutro exemplo, considerando o sistema dos três detetores, concluímos que a probabilidade da radiação disparar o detetor **b** ou **c** é duas vezes maior que a do detetor **a** ser acionado.

Em ambos os exemplos, o exame das condições permite concluir que os eventos considerados são equiprováveis, o que tanto fornece uma definição de probabilidade como permite o cálculo das probabilidades de certos eventos compostos. Em geral, porém, há limitações práticas nessa abordagem, pois nem sempre o sistema em análise é simétrico sob todos os aspectos relevantes e, caso o espaço amostral seja infinito, ela não se adapta, mesmo que os eventos sejam enumeráveis.

Vamos elaborar um pouco mais o exemplo do detetor da figura 7.1, na condição do capítulo precedente, em que a fonte de radiação emitia duas radiações simultaneamente, o que dá origem ao espaço amostral da Tabela 6.1, que numera os eventos para facilitar sua identificação. Nesse caso, a hipótese de eventos equiprováveis significa a atribuição de uma probabilidade $\delta = 1/N = 1/9$ a cada um dos $N = 9$ eventos que formam o espaço amostral. Vamos estabelecer, a partir de princípios físicos, a equiprobabilidade dos eventos.

O sistema de detetores é simétrico por rotação de 120º em torno do eixo. Assim, uma rotação de 120º em torno do eixo no sentido $\mathbf{b} \to \mathbf{a}$ transforma o evento 4 no evento 8, o evento 2 no evento 1 e assim por diante. Uma rotação de 240º transforma o evento 4 no 6, o 2 no 3, ... Não há, porém, nenhuma operação que transforme os eventos 1, 2 ou 3 em qualquer dos eventos 4 a 9. Concluímos, então, que os eventos 1 a 3 são equiprováveis e os eventos 4 a 9 também são, mas não há uma razão fundamental para que **todos** sejam equiprováveis. Na Tabela 6.3 do capítulo precedente, chamamos de χ a probabilidade dos eventos 1 a 3 e de ϕ, a dos eventos 4 a 9 — essa era a *construção* II. A *construção* I correspondia à situação $\chi = \phi$. Na seção 6.5, vimos que, na construção I em que todos eventos eram equiprováveis, o evento "deteção de G no detetor **a**" era independente do evento "deteção de g no detetor **b**", enquanto que, na construção II, $\chi \neq \phi$, esses eventos eram *dependentes*. Assim, a hipótese de eventos equiprováveis é aceitável se a emissão de um gama numa certa direção não influi na direção de emissão do outro. Uma experiência que meça a frequência relativa dos eventos da tabela 6.3 permite determinar a existência dessa propriedade. Aliás, as leis físicas determinam que os dois gamas são emitidos independentemente em certas situações *particulares*, mas $\chi \neq \phi$ é mais comum.

7.1.2 Probabilidade como frequência relativa

Interpretar probabilidade como o limite da frequência relativa do evento numa série infinita de observações, como fizemos na seção 1.1,

$$p = \lim_{N\to\infty} \frac{n}{N} \quad , \tag{1.1}$$

é compatível com os axiomas básicos da Teoria da Probabilidade. Esta interpretação é bastante intuitiva depois que o(a) experimentador(a) se habitua com o fenômeno da flutuação estatística dos resultados das suas medições e é bastante adequada à análise de experimentos em Laboratório, na maior parte das vezes.

7.1.3 Probabilidade como grau de confiança ou crença relativa

Nesta interpretação, há duas possibilidades, uma dita *lógica* ou *objetiva*, outra *pessoal* ou *subjetiva*. Não vamos detalhar aqui as diferenças entre essas possibilidades.

Para fenômenos que **não se repetem**, a probabilidade só pode representar um grau de crença. Neste caso, todas as possibilidades são condicionais, dado o conhecimento (uma hipótese H) e um evento observado (o conjunto de dados que constitui a medida, $\{x_i\}$), que designaremos por A. Assim, a probabilidade

$$P(A \mid H)$$

descreve a extensão em que a hipótese implica no evento A. Essa é uma afirmação lógica ($H \Rightarrow A$) que é verdadeira ou falsa, a probabilidade sendo o **grau de crença** na sua veracidade.

Embora não nos alonguemos na discussão, a interpretação do resultado final de uma medida pode ser efetuada dentro deste quadro. Como acreditamos que a grandeza física tenha um valor verdadeiro, o intervalo apresentado ou contém ou não contém esse valor verdadeiro, mas, nem por isso, atribuímos a esses intervalos uma probabilidade igual a 100% ou igual a 0%, simplesmente porque não há maneira de saber o que acontece. Afirmamos que um dado intervalo contém o valor verdadeiro da grandeza com uma certa *probabilidade* α, que é o *nível de confiança* que definimos no capítulo 5. Na interpretação frequencista, este nível de probabilidade corresponde à fração de vezes que *acertaremos* o resultado, na repetição (infinita) da medida. A idéia de repetição da medida é natural nas ciências da natureza — ninguém acredita em um experimento que não possa ser repetido — mas, em uma situação **impossível de repetir** — a erupção do vulcão Etna no ano que vem ou o lançamento de uma nave tripulada rumo a Alfa Centauro no século XXX — a probabilidade α atribuída a intervalos de confiança das grandezas medidas corresponde ao grau em que **acreditamos** que o intervalo contenha o valor verdadeiro, numa escala de 0 a 1. Nos exemplos mencionados, grandezas como a quantidade de matéria expelida pelo vulcão Etna e o número de seres humanos que alcançariam Alfa Centauro são variáveis aleatórias que podem ser previstas, mas que não serão observadas uma segunda vez.

7.2 Teorema de Bayes

Consideremos um espaço amostral Ω e um conjunto $B_1, B_2, \ldots, B_N$ de eventos definidos sobre Ω, exclusivos e exaustivos (os B_i formam uma partição de Ω, veja seção 6.3), e tomemos dois eventos X e Y quaisquer. Usando os resultados e definições do capítulo anterior, podemos deduzir que

$$P(X \cap Y) = P(X \mid Y)P(Y)$$

$$\text{e} \qquad \sum_i P(X \cap B_i) = P(X) \quad .$$

Vamos agora realizar uma medida e supor que ela resulte no evento A e definir uma hipótese H, que também corresponda a um evento. Para um determinado evento B_r da partição definida acima, podemos verificar que

$$P[B_r \cap (A \cap H)] = P(B_r \mid A \cap H) \cdot P(A \cap H) \quad . \tag{7.1}$$

Podemos trocar A e B_r de posição – é a mesma intersecção de conjuntos – e reescrever a probabilidade em termos de outra probabilidade condicional,

$$P[A \cap (B_r \cap H)] = P(A \mid B_r \cap H)P(B_r \cap H) \quad . \tag{7.2}$$

Como as probabilidades dos membros esquerdos das expressões (7.1) e (7.2) são idênticas, deduz-se que

$$P(B_r \mid A \cap H) \cdot P(A \cap H) = P(A \mid B_r \cap H) \cdot P(B_r \cap H) \quad .$$

O segundo fator de cada membro da expressão acima também pode ser reescrito como uma probabilidade condicional,

$$P(A \cap H) = P(A \mid H)\, P(H) \qquad \text{(6.8) reformulada,}$$

e, como $P(H)$ aparece em ambos os membros e não é nula, chega-se a

$$P(B_r \mid A \cap H) \cdot P(A \mid H) = P(A \mid B_r \cap H) \cdot P(B_r \mid H) \quad ,$$

que pode ser reescrita como

$$P(B_r \mid A \cap H) = \frac{P(A \mid B_r \cap H)P(B_r \mid H)}{P(A \mid H)} \quad . \tag{7.3}$$

Esse é o teorema de Bayes.

7.3 Teorema de Bayes: interpretação bayesiana

Na relação (7.3), se B corresponde ao parâmetro que pretendemos medir, B_r corresponde a um valor possível desse parâmetro. O evento A é a medida e

H é a hipótese, que engloba todas as características da medida que a experimentadora acredita verdadeiras, por exemplo, a f.d.p. dos dados é gaussiana. Assim, $P(B_r \mid H)$ é a probabilidade *a priori* de B_r ser o valor verdadeiro da grandeza dentro da hipótese considerada; $P(B_r \mid A \cap H)$ é a probabilidade *a posteriori* e $P(A \mid B_r \cap H)$ é a função Verossimilhança, que já conhecemos (seção 4.1). Se definimos um "espaço amostral de hipóteses" acerca do valor da grandeza física e a cada valor possível da grandeza medida atribuímos uma certa probabilidade de observação por meio da medida efetuada, o teorema expresso pela relação (7.3) permite recalcular essas probabilidades de maneira que incluam a informação obtida pela medida efetuada. Assim, $P(B_r \mid A \cap H)$ é o grau de confiança do experimentador quanto a B_r ser o valor verdadeiro da grandeza.

Quando os parâmetros são grandezas contínuas, devemos substituir as probabilidades por f.d.p.s. Se θ for usado como parâmetro contínuo e valor verdadeiro da grandeza, $\mathcal{L}$ representa a função verossimilhança, e a medida A é o conjunto de dados $\{x_i\}$, o Teorema de Bayes fica

$$g(\theta \mid \{x_i\}, H) \propto f(\theta \mid H) \cdot \mathcal{L}(\{x_i\} \mid \theta, H) \quad , \tag{7.4}$$

onde $g(\theta)$ e $f(\theta \mid H)$ são funções densidade de probabilidade *a posteriori* e *a priori*, respectivamente.

O Teorema de Bayes fornece f.d.p.s a posteriori legítimas toda vez que a f.d.p. a priori for legítima, ou seja, se f estiver de acordo com os axiomas da Teoria da Probabilidade, a relação (7.4) fornecerá g também de acordo, onde g incluirá a informação obtida por meio de $\{x_i\}$. A fórmula (7.4) é um resultado lógico dedutível na Teoria da Probabilidade e inquestionável, como deve ter ficado claro pela dedução. A dificuldade no seu uso está nos critérios de construção da f.d.p. a priori e é acerca da possibilidade ou não de obter-se uma f.d.p. a priori objetiva que deve centrar-se a preocupação do(a) experimentador(a).

Na maior parte dos experimentos em Física, supõe-se que nada se conhece previamente ao experimento. Nesses casos, o *Postulado de Bayes* corresponde a admitir que todas as hipóteses são equiprováveis, o que frequentemente corresponde a utilizar f=constante. Neste caso, passa-se a interpretar a função Verossimilhança como a f.d.p. do parâmetro que se pretende medir por meio de $\{x_i\}$. É pela dificuldade em arranjar-se uma justificativa lógica para o Postulado de Bayes ou em escolher-se uma f.d.p. a priori que não é comum utilizá-lo. Quem quiser verificar as possibilidades dessa aplicação *Bayesiana* do Teorema de Bayes na análise de experimentos em física, pode consultar os artigos [Helene 83], [Helene 84], [Helene 91b] e também [Escoubes]. Para uma discussão abrangente, veja o livro [Kendall, vol 1, cap.7 e 8].

7.4 Procedimento que adotaremos

O procedimento geral será usarmos o Método da Máxima Verossimilhança ou dos Mínimos Quadrados para descobrir uma *estatística* que estime o parâmetro de interesse. Em seguida, verificaremos se esse estimador obedece os seguintes critérios:

i) Consistência.

ii) Não tendenciosidade.

iii) Eficiência.

Procuraremos delinear completamente as situações onde esses métodos (Máxima Verossimilhança e Mínimos Quadrados) têm essas propriedades desejadas.

Veremos que a função Verossimilhança desempenha um papel importante no estudo das propriedades dos estimadores. Com frequência, poderemos obter respostas completas ao adotarmos uma hipótese *apenas* acerca da f.d.p. *dos dados*, seja da sua forma, seja de algum dos seus parâmetros. Poderemos estabelecer o comportamento estatístico das estimativas, ou seja, verificar como elas se distribuem, sem necessidade de ter qualquer informação a priori acerca *do valor verdadeiro* da grandeza medida. Muitas vezes, porém, não obteremos respostas tão completas quanto seriam obtidas se utilizássemos a interpretação da função Verossimilhança dada pelo Postulado de Bayes. A razão de não aderirmos sistematicamente ao postulado de Bayes é que, normalmente, sua utilização produz resultados que são considerados subjetivos.

7.5 Estimadores e estimativas

Em toda a discussão que segue, vamos supor a inexistência de erros de medida sistemáticos. Se existem interferências não aleatórias, pode-se deduzir muito pouco a respeito das grandezas que procuramos encontrar. Suporemos também que a medida se destina a determinar uma única grandeza, ou seja, limitaremos a discussão ao caso em que a f.d.p. dos dados é conhecida a menos de um parâmetro θ. A medida será um conjunto de dados $\{x_1, x_2, \ldots, x_N\}$.

Uma *estatística* t (ou t_N, como simbolizaremos frequentemente para deixar claro quantos dados entram em sua determinação) é uma função cujas únicas variáveis aleatórias são os dados $x_1, \ldots, x_N$, portanto $t = t(x_1, \ldots, x_N)$. Distinguiremos o *estimador*, que é o método de obter estatísticas para as grandezas, da *estatística*, que é uma função dos dados da medida, e da *estimativa*, que corresponde ao valor numérico obtido, associado a uma medida particular, exatamente como distinguimos a operação derivada d/dx da função $df(x)/dx$, para uma função $f(x)$ particular, e do valor da derivada em um ponto a, $df/dx\,|_{x=a}$.

Por exemplo, o quadro abaixo associa os nomes estimador, estatística e estimativa obtidos pelo método dos mínimos quadrados no ajuste de parâmetros de funções, que discutimos no capítulo 4, às grandezas ou idéias correspondentes. Incluímos no quadro, também, a analogia com a derivada de uma função.

$\dfrac{d}{dx}$	$\dfrac{df}{dx}$	$\left.\dfrac{df}{dx}\right\|_{x=a}$
$\updownarrow$	$\updownarrow$	$\updownarrow$
método dos mínimos quadrados	$\mathbf{M}^{-1}\mathbf{D}$	$\hat{\mathbf{p}}\|_{(x_i,y_i,\sigma_i)}$
$\updownarrow$	$\updownarrow$	$\updownarrow$
estimador (método)	estatística (fórmula)	estimativa (valor numérico)

Se usamos[1] t para estimar θ, nem sempre o resultado (a estimativa) será aproximadamente igual a θ, algumas vezes poderá ser bastante *diferente* de θ. Isso não significa que t seja uma má estatística — queremos que t dê bons resultados ao longo dos experimentos que realizamos, ou dê bons resultados em média ou com grande probabilidade. Se t é muito diferente de θ com muita frequência, aí sim, rejeitaremos a estatística t como ferramenta para obter as estimativas de θ. Enfim, não devemos julgar t por um resultado particular, mas sim considerá-la como uma variável aleatória e determinar a distribuição das estimativas geradas.

Um exemplo que discutimos exaustivamente no texto corresponde ao da estimativa do valor esperado θ de uma grandeza quando os dados obtidos se

[1]Estamos usando a letra t para identificar um estimador, a estatística ou a estimativa correspondente; t não deve ser confundido com t de Student. Neste capítulo, identificaremos o valor verdadeiro de uma grandeza por θ; posteriormente, voltaremos a adotar um subíndice "$_0$" para identificar valores verdadeiros (x_0, σ_0, a_0, b_0, ...), como já haviamos feito anteriormente.

distribuem como uma normal. Usamos como estimador a *média dos dados*, cuja f.d.p. tem as propriedades:

- *A estimativa tem valor esperado θ, independentemente do número de dados.*

- *Quando o número de dados cresce, as estimativas se espalham menos em torno de θ.*

Estas propriedades não são inerentes ao estimador, pois dependem da forma da f.d.p. dos dados. Por exemplo, se os dados se distribuem de acordo com a f.d.p. de Cauchy,

$$c(x;\theta) = \frac{1}{\pi[1+(x-\theta)^2]} \quad ,$$

pode-se demonstrar que a média se distribui em torno de θ de maneira independente do número de dados, sempre obedecendo à mesma f.d.p. $c(x;\theta)$, portanto a largura da f.d.p. da média é a *mesma*, qualquer que seja o número de dados. Ou seja, nesse caso, a lei dos grandes números não se aplica, pois a variância $E((x-\theta)^2)$ é infinita e a desigualdade de Chebyshev não tem qualquer utilidade prática. Em geral, o comportamento de um estimador depende da f.d.p. que governa a medida; é por isso que não existe um estimador universal, que resolva todos os problemas de estimação que encontramos.

Em resumo, em uma medida que procura determinar uma certa grandeza com valor verdadeiro igual a θ, a flutuação estatística impede de construir uma estatística t que forneça sempre o valor exato θ. Em outras palavras, não existe nenhuma função algébrica $t = t(x_1, x_2, \ldots, x_N)$ que resulte no valor θ para qualquer medida $\{\, x_i, i = 1, \ldots, N\}$. Isso nos obriga a definir um critério onde θ represente um certo parâmetro de localização da f.d.p. das estimativas $\hat{\theta}$ obtidas com a estatística t, ou seja, θ é um parâmetro de *localização* da f.d.p. das estimativas obtidas por meio de t. As possibilidades sugeridas por esta discussão são:

i) A estimativa mais provável corresponde ao valor verdadeiro — θ é a *moda* da f.d.p. de $t(\{x_i\})$.

ii) A probabilidade de subestimar θ é a mesma de superestimá-lo — θ é a *mediana* da f.d.p. de t.

iii) θ é igual ao valor esperado das estimativas.

Todas as três opções acima são interessantes, mas é fácil verificar que raramente uma estatística satisfará todas. Por exemplo, a estimativa da variância obtida pelo estimador de Máxima Verossimilhança e corrigida da tendenciosidade (veja seções 4.2 e 4.3), para um conjunto de dados com f.d.p. gaussiana, tem uma f.d.p. assimétrica, como discutimos na seção 2.9[2]. Somos, então, obrigados a optar por uma delas (média, mediana ou moda) e a escolha, por razões de simplicidade algébrica, recai na propriedade *iii*), a qual designaremos por *não-tendenciosidade*. Provavelmente teríamos dado esse mesmo nome a qualquer uma dessas três propriedades que fosse a escolhida, mas a escolha é *sempre* a *iii*) e, daqui para frente, *tendenciosidade* se refere à propriedade *iii*) acima.

7.6 O critério da consistência

Uma das propriedades desejadas de um estimador é que a variância da distribuição das estimativas diminua com o aumento do número de dados, n, e que o valor central seja θ. Essa propriedade é chamada **consistência**.

Formalmente, se t_n é o estimador para uma medida com n dados, diz-se que ele é um **estimador consistente** quando vale a propriedade:

para quaisquer $\epsilon > 0$ e $0 < \eta < 1$, $\exists N$ tal que

$$P(| t_n - \theta |< \epsilon) > 1 - \eta \qquad \text{para } n > N \quad . \tag{7.5}$$

Na definição acima, P simboliza probabilidade.

A figura 7.2 procura esboçar funções de probabilidade associadas a estimadores consistentes e inconsistentes.

Exemplo 7.1

Na seção 6.18, vimos a *Lei dos Grandes Números***, a partir da qual se verifica que a média é uma estimativa** *consistente* **da grandeza, sujeita às mesmas condições em que ela (lei dos grandes números) pode ser aplicada.**

---

[2]Note, porém, que o mesmo estimador de Máxima Verossimilhança, aplicado à estimação do valor verdadeiro para dados gaussianos, seção 2.6, gera uma estatística que possui todas as três propriedades.

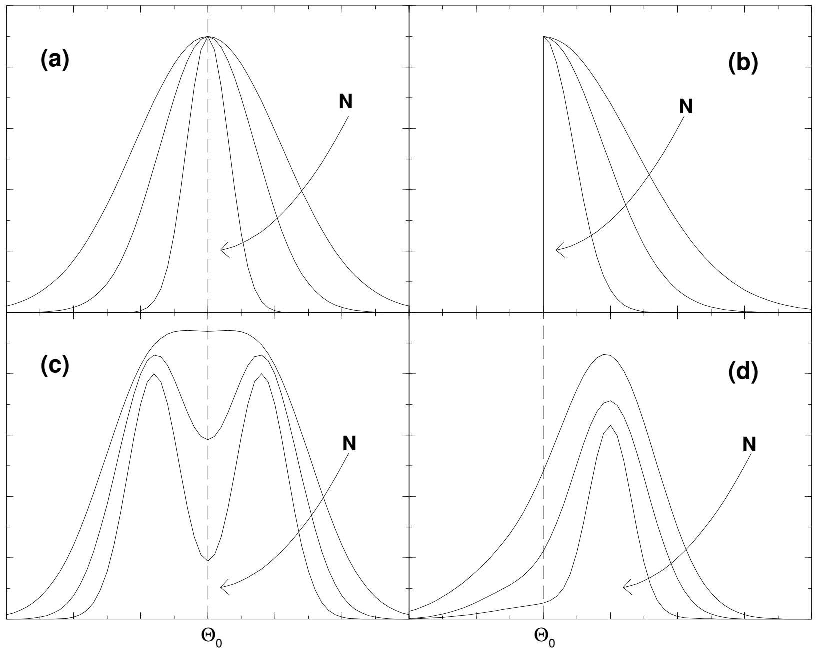

Figura 7.2: Esboço do possível comportamento com o número N de dados da f.d.p. das *estimativas* obtidas com um estimador: **a)** consistente e não-tendencioso, **b**) consistente e tendencioso, **c**) inconsistente e não-tendencioso e **d**) inconsistente e tendencioso. As f.d.p.s não estão normalizadas para ressaltar a dependência no número de dados, N. As setas identificam a direção em que N cresce.

7.7 O estimador consistente não é único

Consistência é uma propriedade assintótica, isto é, garante que o estimador fornece o valor correto de θ quando o número de dados é grande, mas não permite deduzir muito sobre o comportamento da estatística quando o número de dados é finito. Em consequência, se existe um estimador consistente, é possível construir uma infinidade de outros estimadores consistentes. Em particular, se t_n é um estimador consistente de θ em uma medida de n dados, então

$$\frac{n-a}{n-b}\, t_n \qquad \text{com } a \text{ e } b \text{ fixos}$$

também é consistente. Como exemplo, no caso de dados que obedecem a distribuições normais, ambas as estatísticas

$$\frac{\sum x_i}{n} \qquad e \qquad \frac{\sum x_i}{n-1}$$

são consistentes.

O raciocínio de limite infinito embutido nessa definição permite muita liberdade na escolha do estimador, de modo que vamos definir, na seção seguinte, um critério adicional para selecionar, dentre os possíveis estimadores consistentes, um—e apenas um—que seja adequado em algum sentido objetivo.

7.8 O critério da não-tendenciosidade

Chamamos de estimador *não-tendencioso* do parâmetro θ aquele para o qual vale a seguinte propriedade:

$$E(t) = \theta \quad , \quad \forall N \quad , \tag{7.6}$$

em que N é o número de dados, e não apenas para N grande. Lembre que $E(t)$ significa *valor esperado* de t. Assim, o nome "não-tendencioso" é dado a métodos de estimação que resultam em valores que flutuam estatisticamente em torno do valor verdadeiro, de maneira que seu valor esperado, ou seja, sua **média**, calculada a partir da sua f.d.p., corresponda ao valor verdadeiro.

Exemplo 7.2

A estimativa da variância de dados gaussianos

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

é não tendenciosa (e consistente). A estimativa

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

deduzida na seção 4.3 com base no método da máxima verossimilhança, é tendenciosa (embora seja consistente).

Pode-se verificar que, em certos casos, não existe nenhum estimador não-tendencioso que seja consistente. E consistência, sem dúvida, é importante. Enfim, gostaríamos de usar sempre estimadores não-tendenciosos, mas isso nem sempre é possível.

Tabela 7.1: Razões entre os desvios padrão da mediana e da média, c_N, e entre os desvios padrão da média dos extremos e da média, d_N, em função do número de dados, N, para uma distribuição gaussiana. Note que a média dos extremos da relação (7.7) não é um estimador consistente. Também veja que a mediana é quase tão bom estimador de θ quanto a média, especialmente para um número pequeno de dados.

	mediana	média dos extremos
N	c_N	d_N
2	1,000	1,000
4	1,092	1,092
6	1,135	1,190
10	1,177	1,362
20	1,214	1,691
∞	1,253	∞

$$\sigma_{\text{mediana}} = c_N \frac{\sigma}{\sqrt{N}}$$

$$\sigma_{\text{media extremos}} = d_N \frac{\sigma}{\sqrt{N}}$$

Há algumas maneiras de corrigir a tendenciosidade, discutidas no [Kendall], em particular o método chamado "*jackknife*" [vol.2, seção 17.10], que pode fornecer uma saída em certos casos.

7.9 Eficiência

Em situações particulares, pode haver diversos estimadores consistentes e não-tendenciosos de um parâmetro. Usando novamente como exemplo o caso da medida de θ por meio de um conjunto de dados $\{x_i, i = 1, \ldots, N\}$ que segue uma f.d.p. normal, há uma infinidade de estimadores não-tendenciosos. Três deles são a *média* $\bar{x}$, a *mediana* x_m e a *média dos extremos*, ζ,

$$\zeta = \frac{x_{[1]} + x_{[N]}}{2} \quad , \tag{7.7}$$

onde $x_{[1]}$ é o menor dos valores x_i e $x_{[N]}$, o maior, na notação da seção 1.11. A tabela 7.1 apresenta, em função do número de dados em uma medida governada pela distribuição normal, as razões dos desvios padrão dos dois últimos estimadores com o da média,

$$\sigma_{\text{media}} = \frac{\sigma}{\sqrt{N}} \quad ,$$

em que σ^2 é a variância dos dados x_i.

A definição formal de *eficiência* está na seção 7.14, logo depois da discussão acerca da possibilidade de saber se a estatística que usamos é (ou não) a de menor variância que possa existir. Entretanto, podemos antecipar que eficiência é uma característica que reflete a rapidez com que o desvio-padrão se reduz com o aumento no número de dados. Por exemplo, os estimadores média, mediana e média dos extremos não são eficientes para todas as distribuições, como mostra a tabela 7.2, que compara suas eficiências para as distribuições normal, uniforme e de Cauchy. Os zeros dessa tabela indicam que as variâncias diminuem mais lentamente que $\frac{1}{N}$ com o crescimento de N ou, como no caso da média dos extremos em conjuntos de dados normais, não diminui. Esses zeros chamam a atenção para a importância da escolha do estimador correto.

Tabela 7.2: Eficiência assintótica para estimadores de localização (média, mediana e média dos extremos), para as distribuições normal, uniforme e de Cauchy.

Estatística	Mediana	Média	Média dos extremos
Distribuição			
Normal	0,64	1,00	0,00
Uniforme	0,00	0,00	1,00
Cauchy	0,81	0,00	0,00

Assim, o último critério que adotaremos se destina a selecionar, dentre os estimadores consistentes e não-tendenciosos, o de mínima variância.

7.10 Limite Mínimo de Variância

Se não é possível determinar o valor verdadeiro de uma grandeza a partir de uma medida com um número finito de dados, então a variância de um estimador deve ter um mínimo. Precisamos, portanto:

- Demonstrar a existência de um Limite Mínimo da Variância (LMV).
- Determinar o LMV.

A possibilidade de determinar o LMV dos estimadores de uma grandeza foi descoberta na década 1940 – 1950 por diversos pesquisadores (veja [Kendall, Vol.2, 17.15]). Como a determinação do limite mínimo implica na existência desse mínimo, obteremos a resposta às duas questões simultaneamente. A dedução do LMV é relativamente simples do ponto de vista do cálculo envolvido, embora não do conceitual, e a reproduziremos abaixo.

Primeiro, vamos deduzir uma propriedade da função verossimilhança válida sob certas condições de regularidade usualmente satisfeitas, que será útil também na determinação do LMV.

Como a função verossimilhança $\mathcal{L}$ é a f.d.p. conjunta das observações, então

$$\int \ldots \int \mathcal{L}\, dx_1 \ldots dx_N = 1 \quad , \tag{7.8}$$

em virtude da normalização das f.d.p.s. Suponha que existam as duas primeiras derivadas de $\mathcal{L}$ em relação a θ (esta é uma das condições para classificarmos $\mathcal{L}$ como regular). Derivando a expressão acima em relação a θ, obtem-se

$$\int \ldots \int \frac{\partial \mathcal{L}}{\partial \theta}\, dx_1 \ldots dx_N = 0 \quad , \tag{7.9}$$

que pode ser reescrita como

$$\int \ldots \int \left(\frac{1}{\mathcal{L}} \frac{\partial \mathcal{L}}{\partial \theta} \right) \mathcal{L}\, dx_1 \ldots dx_N = 0 \quad , \tag{7.10}$$

que é, simplesmente,

$$E\left(\frac{1}{\mathcal{L}} \frac{\partial \mathcal{L}}{\partial \theta} \right) = E\left(\frac{\partial \ln \mathcal{L}}{\partial \theta} \right) = 0 \quad . \tag{7.11}$$

Supos-se também que a ordem da integração (em x) e da derivação (em θ) podiam ser invertidas, outra condição de regularidade. Derivando a expressão (7.10) em relação a θ, chega-se a

$$\int \ldots \int \left\{ \left(\frac{1}{\mathcal{L}} \frac{\partial \mathcal{L}}{\partial \theta} \right) \frac{\partial \mathcal{L}}{\partial \theta} + \mathcal{L} \frac{\partial}{\partial \theta} \left(\frac{1}{\mathcal{L}} \frac{\partial \mathcal{L}}{\partial \theta} \right) \right\} dx_1 \ldots dx_N = 0 \quad ,$$

que, com o mesmo tipo de artifício usado para chegar na expressão (7.11), pode ser reescrita como

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

Reservemos a expressão (7.12) e passemos à dedução do LMV.

Faremos a dedução em toda sua generalidade, como é habitual na bibliografia, onde se considera a possibilidade de buscar estimar uma função de θ, $\tau(\theta)$, ao invés do próprio θ, de modo que para estimar o próprio θ pode-se usar $\tau(\theta) = \theta$. O fato de t ser um estimador não tendencioso de uma certa função de θ, $\tau(\theta)$, exprime-se matematicamente pela relação (7.6), neste caso

$$E(t) = \int \ldots \int t \, \mathcal{L} \, dx_1 \ldots dx_N = \tau(\theta) \quad , \tag{7.13}$$

onde, de novo, usou-se que $\mathcal{L}$ é a f.d.p. conjunta de todos os dados da medida.

Derivando ambos os membros de (7.13) em relação a θ, tem-se

$$\int \ldots \int t \, \frac{\partial \ln \mathcal{L}}{\partial \theta} \, \mathcal{L} \, dx_1 \ldots dx_N = \tau'(\theta) \quad ,$$

onde τ' é a derivada de τ em relação a θ e $\partial \mathcal{L}/\partial \theta$ foi escrita da mesma maneira artificiosa usada na passagem da fórmula (7.9) para a (7.10).

Multiplicando a expressão nula da fórmula (7.10) por $\tau(\theta)$ e subtraindo-a da expressão acima, obtém-se

$$\int \ldots \int [t - \tau(\theta)] \, \frac{\partial \ln \mathcal{L}}{\partial \theta} \, \mathcal{L} \, dx_1 \ldots dx_N = \tau'(\theta) \quad . \tag{7.14}$$

Usando a desigualdade de Schwarz, com os “vetores”

$$f = [t - \tau(\theta)] \cdot \sqrt{\mathcal{L}} \qquad , \qquad g = \frac{\partial \ln \mathcal{L}}{\partial \theta} \sqrt{\mathcal{L}} \quad , \tag{7.15}$$

obtem-se

$$\int \ldots \int [t - \tau(\theta)]^2 \, \mathcal{L} \, dx_1 \ldots dx_N \cdot \int \ldots \int \left(\frac{\partial \ln \mathcal{L}}{\partial \theta} \right)^2 \mathcal{L} dx_1 \ldots dx_N$$

$$\geq \left(\int \ldots \int [t - \tau(\theta)] \, \frac{\partial \ln \mathcal{L}}{\partial \theta} \, \mathcal{L} \, dx_1 \ldots dx_N \right)^2$$

(veja por exemplo [Merzbacher, p.169]). Identificando o membro direito com τ'^2 e os fatores do membro esquerdo com valores esperados, chega-se a

$$E\{[t - \tau(\theta)]^2\} \cdot E \left\{ \left(\frac{\partial \ln \mathcal{L}}{\partial \theta} \right)^2 \right\} \geq [\tau'(\theta)]^2 \quad ,$$

onde o primeiro fator corresponde exatamente à variância do estimador t, $\mathrm{var}(t)$. Assim,

$$E\{[t-\tau(\theta)]^2\} = \mathrm{var}(t) \geq [\tau'(\theta)]^2 E\left\{\left(\frac{\partial \ln \mathcal{L}}{\partial \theta}\right)^2\right\}^{-1} \quad , \tag{7.16}$$

ou ainda, usando a identidade (7.12) acima,

$$\mathrm{var}(t) \geq -[\tau'(\theta)]^2 \cdot \left\{E\left(\frac{\partial^2 \ln \mathcal{L}}{\partial \theta^2}\right)\right\}^{-1} \quad . \tag{7.16'}$$

A expressão do membro direito da desigualdade representa um limite abaixo do qual a variância de *nenhum* estimador não tendencioso da função $\tau(\theta)$ pode descer. Note que o lado direito desta última desigualdade independe do estimador t, o que dá sentido à frase anterior.

Apenas para completar a discussão das condições de regularidade necessárias para a dedução de (7.16) e (7.16'): normalmente as operações que efetuamos são sempre possíveis se os limites de integração são infinitos ou finitos, mas independentes de θ, se o integrando é contínuo em x e θ, e desde que a integral resultante seja uniformemente convergente para todo θ.

Em resumo, o valor mínimo da variância de qualquer estimador está ligado à forma da função densidade de probabilidade dos dados e é dada por

$$\mathrm{LMV} = \frac{[\tau'(\theta)]^2}{E\left\{\left(\frac{\partial \ln \mathcal{L}}{\partial \theta}\right)^2\right\}} = -\frac{[\tau'(\theta)]^2}{E\left(\frac{\partial^2 \ln \mathcal{L}}{\partial \theta^2}\right)} \quad ,$$

onde qualquer uma das formas pode ser usada.

7.11 Limite Mínimo de Variância de θ

Quando o parâmetro que se pretende estimar é o próprio θ, então $\tau(\theta) = \theta$, portanto

$$\tau'(\theta) = 1 \quad .$$

O LMV da fórmula (7.16') fica

$$\mathrm{var}(t) \geq -\left\{E\left(\frac{\partial^2 \ln \mathcal{L}}{\partial \theta^2}\right)\right\}^{-1} = \mathrm{LMV} \quad . \tag{7.17}$$

Denomina-se *Informação acerca de θ contida num conjunto de dados* ou, simplesmente, *Informação*, à grandeza I dada por

$$I = -E\left(\frac{\partial^2 \ln \mathcal{L}}{\partial \theta^2}\right) \quad , \tag{7.18}$$

de maneira que, quanto maior a informação, menor o LMV.

Exemplo 7.3

No capítulo 2, mostramos que a média de n dados estatisticamente independentes distribuídos de acordo com a normal de desvio-padrão conhecido e igual a σ_0 tem a f.d.p.

$$f(\bar{x} \mid \theta) = \frac{1}{\sqrt{2\pi}\sigma_m} \, \exp\left(-\frac{(x-\theta)^2}{2\sigma_m^2}\right) \quad , \tag{2.27}$$

com $\sigma_m = \sigma_0/\sqrt{n}$. Neste caso, a função verossimilhança coincide com a expressão de f acima e, portanto,

$$\frac{\partial^2 \ln \mathcal{L}}{\partial \theta^2} = -\frac{1}{\sigma_m^2} \quad ,$$

de modo que, de acordo com a relação (7.17),

$$\mathrm{LMV} = \sigma_m^2 \quad .$$

Como a variância da média é σ_m^2, concluímos que a média é o estimador de variância mínima do valor (verdadeiro) da grandeza, quanto os dados têm distribuição normal.

Podemos tentar localizar o estimador t cuja variância atinge o LMV por meio da Condição Necessária e Suficiente para a desigualdade de Schwarz transformar-se numa igualdade, que corresponde ao caso em que g e f das relações (7.15) são proporcionais,

$$g = Af \quad \Rightarrow$$

$$\frac{\partial \ln \mathcal{L}}{\partial \theta}\sqrt{\mathcal{L}} = A \cdot [t - \tau(\theta)]\sqrt{\mathcal{L}} \quad \Rightarrow$$

$$\frac{\partial \ln \mathcal{L}}{\partial \theta} = A(\theta) \cdot [t - \tau(\theta)] \quad , \tag{7.19}$$

onde ficou explícito que a constante de proporcionalidade A pode depender de θ, mas não de t, uma vez que nas integrais usadas na dedução da desigualdade de Schwarz θ é um parâmetro fixo.

A expressão (7.19) corresponde à condição de existência de um estimador t para θ com variância mínima, ou seja, igual ao LMV. Nesse caso, a variância de t pode ser calculada multiplicando (7.19) por $t-\tau(\theta)$ e pela função densidade de probabilidade conjunta dos dados, $\mathcal{L}$, integrando-se a seguir. Obtém-se

$$\int \ldots \int [t-\tau(\theta)] \, \frac{\partial \ln \mathcal{L}}{\partial \theta} \, \mathcal{L} \, dx_1 \ldots dx_N =$$

$$A(\theta) \int \ldots \int [t-\tau(\theta)]^2 \mathcal{L} \, dx_1 \ldots dx_N = A(\theta) \, \mathrm{var}(t) \quad .$$

O membro esquerdo da equação acima pode ser identificado com τ', relação (7.14), o que resulta em

$$\tau'(\theta) = A(\theta) \, \mathrm{var}(t)$$

ou, o que é equivalente,

$$\mathrm{var}(t) = \frac{\tau'(\theta)}{A(\theta)} \quad . \tag{7.20}$$

Como exemplo, voltemos ao caso da estimativa do "valor verdadeiro" θ de uma grandeza por meio de uma medida, cujos dados têm f.d.p. gaussiana com variância σ_0^2 . Nesse caso,

$$\ln \mathcal{L} = -N \ln(\sigma_0 \sqrt{2\pi}) - \sum_{i=1}^{N} \frac{(x_i - \theta)^2}{2\sigma_0^2} \quad ,$$

em que usamos $\mathcal{L}$ da seção 4.2. A derivada em relação a θ resulta em

$$\frac{\partial \ln \mathcal{L}}{\partial \theta} = \sum_{i=1}^{N} \frac{(x_i - \theta)}{\sigma_0^2} = \frac{N}{\sigma_0^2} (\bar{x} - \theta) \quad . \tag{7.21}$$

A expressão acima é da forma da equação (7.19), donde pode-se identificar $A(\theta)$ e concluir, por meio de (7.20), que

$$\mathrm{var}(\bar{x}) = \sigma_0^2 / N \quad ,$$

e que $t = \bar{x}$ é o estimador de menor variância que pode ser construído para estimar θ numa medida onde os dados têm f.d.p. gaussiana.

Q7.1 *Mostre que existe um estimador para θ cuja variância atinge o LMV se $f(x \mid \theta) = \theta^x \exp(-\theta)/x!$ (Função de Probabilidade de Poisson). Qual é esse estimador e sua variância?*

7.12 Unicidade do estimador de variância mínima

Se um estimador de $\tau(\theta)$ tem variância igual ao LMV, pode-se mostrar que não existirá nenhum estimador de variância mínima para qualquer outra função de θ. Aqui, vamos somente exemplificar essa propriedade.

Retomemos o exemplo da seção anterior, só que, ao invés de estimar θ, procuremos estimar θ^2 de uma distribuição gaussiana. Toda a discussão acima vale, com $\tau(\theta) = \theta^2$. A condição de existência de um estimador que atinge o LMV está expressa pela relação (7.19) que, com o resultado da fórmula (7.21), fornece

$$\frac{N}{\sigma_0^2}(\bar{x} - \theta) = \tilde{A}(\theta) \cdot (\tilde{t} - \theta^2) \quad .$$

Vamos *tentar* resolver essa última equação. Determinemos primeiro $\tilde{A}(\theta)$. Para que a identidade acima possa valer para qualquer θ, é preciso que a dependência em θ nos dois membros seja a mesma, o que exige

$$\frac{N}{\sigma_0^2}\,\theta = \tilde{A}(\theta) \cdot \theta^2 \quad \Rightarrow$$

$$\tilde{A}(\theta) = \frac{N}{\sigma_0^2} \cdot \frac{1}{\theta} \quad .$$

Substituindo esse $\tilde{A}(\theta)$ na identidade, obtemos

$$\frac{\tilde{t}}{\theta} = \bar{x} \qquad\qquad \Rightarrow \qquad \tilde{t} = \bar{x} \cdot \theta \qquad .$$

A função $\tilde{t}$ definida pela expressão acima não constitui uma estatística válida para estimar θ^2 porque depende de θ, que é a grandeza a determinar. Assim, não existe nenhum estimador de variância igual ao LMV para o quadrado da média de dados gaussianos.

Exemplo 7.4

Suponha que se obtenham dados $\{x_i\}$ de uma grandeza com valor verdadeiro nulo e distribuição normal, com o propósito de determinar a variância σ^2 a partir dos dados. Um exemplo de grandeza física de interesse com essa propriedade é a projeção, em um eixo fixo e definido, das velocidades das moléculas de um gás, cuja *dispersão* se relaciona com a

temperatura. A f.d.p. de x é

$$f(x \mid \sigma) = \frac{1}{\sqrt{2\pi}\sigma}\ \exp\left(-\frac{x^2}{2\sigma^2}\right) \quad .$$

Calculando a função verossimilhança e derivando seu logaritmo, obtemos

$$\frac{\partial \ln \mathcal{L}}{\partial \sigma} = -\frac{N}{\sigma} + \sum_{i=1}^{N} \frac{x_i^2}{\sigma^3} = \frac{N}{\sigma^3}\left(\frac{1}{N}\sum_{i}^{N} x_i^2 - \sigma^2\right) \quad . \tag{7.22}$$

Concluímos que $\frac{1}{N}\sum_i^N x_i^2$ estima σ^2 com variância igual ao LMV.

Q7.2 *Usando o resultado do exemplo acima e as fórmulas (7.19) e (7.20), calcule a variância de $\frac{1}{N}\sum_i^N x_i^2$.*

Note a impossibilidade de obter uma expressão análoga à (7.19) ou (7.22) quando se insiste em ter $(t - \sigma)$ explicitado, ou $(\ t\ -\ h(\sigma)\ )$, para qualquer $h(\sigma) \neq \sigma^2$, confirmando outra vez a propriedade enunciada no começo desta seção.

7.13 F.d.p.s que permitem estimadores com variância igual ao LMV

A expressão (7.19) representa a condição para existência de um estimador com variância igual ao LMV. Integrando aquela equação, pode-se verificar que as f.d.p.s da forma

$$f(x \mid \theta) = \exp[F_1(\theta)F_2(t) + F_3(x) + F_4(\theta)] \quad , \tag{7.23}$$

onde F_1, F_2, F_3 e F_4 representam funções genéricas das variáveis indicadas, são as que permitem a obtenção de alguma estatística t de variância igual ao LMV. Para f.d.p.s fora dessa família exponencial, não contamos com nenhum apoio teórico na identificação do estimador ótimo nas situações onde o número de dados é pequeno.

Q7.3 *Verifique que as distribuições de Poisson e binomial e a f.d.p. gaussiana podem ser escritas na forma da eq. (7.23) para os parâmetros a (Poisson), p (binomial), x_0 e σ^2 (normal).*

7.14 Critério da eficiência

Mesmo quando a f.d.p. dos dados não permite a determinação de um estimador com variância igual ao LMV, podemos frequentemente encontrar algum que atinja o LMV quando $N \to \infty$. Nessa situação, é provável que o estimador seja também assintoticamente normal, em decorrência do Teorema Central do Limite, caso em que apenas dois parâmetros determinam a f.d.p. das estimativas, a média e a variância. A média será o valor esperado da grandeza estimada se o estimador é consistente, ou seja, se o estimador é também assintoticamente não tendencioso.

O terceiro critério que estabelecemos na seção 7.4 para selecionar estimadores relaciona-se justamente com a variância do estimador em medidas com muitos dados, $N \to \infty$. Formalmente,

> **um estimador é chamado eficiente quando tem variância mínima em medidas com muitos dados.**

Nenhum estimador pode fornecer estimativas com variâncias abaixo do LMV. Assim, uma boa maneira de se comparar diferentes procedimentos para estimar parâmetros, ou seja, diferentes estimadores, é examinar quão próximas do LMV estão as variâncias que eles fornecem. Formalmente, a eficiência de um estimador, $\epsilon(t)$, é definida por

$$\epsilon(t) = \frac{1/I}{\mathrm{var}(t)} \quad ,$$

onde I é a informação, como definido pela equação (7.18). Por exemplo, no caso de dados que obedecem uma distribuição normal, como a variância da média é igual ao LMV, dado por I^{-1}, sua eficiência é igual a 1. Entretanto, no mesmo caso de dados independentes e gaussianos, a mediana (com exceção do caso $n = 2$, quando a mediana coincide com a média) tem eficiência menor do que um.

Q7.4 *Um experimentador, medindo uma determinada grandeza cujos dados são independentes e obedecem a uma fdp normal, decide usar como estimativa a média dos dados depois de excluir o maior e o menor deles, apesar de ter certeza que não houve nenhum erro no processo de medição que justificasse sua eliminação. Você acha que essa opção é melhor do que considerar todos os dados no cálculo da média?*

Por razões práticas, é comum usar-se a eficiência assintótica, que é simplesmente a eficiência de um estimador, t_n, quando o número de dados, n, tende a infinito:

$$\epsilon_a(t) = \lim_{n\to\infty} \frac{1/I_n}{\mathrm{var}(t_n)} \quad . \tag{7.24}$$

A tabela 7.2 mostra a eficiência assintótica para a média, a média dos extremos e a mediana nos casos das distribuições normal, uniforme e de Cauchy. Assim, no caso da distribuição de Cauchy, tanto a média como a média dos extremos tem eficiência assintótica nula e a mediana tem eficiência assintótica igual a 0,81. Portanto, quando se trabalha com dados que se distribuem segunda uma f.d.p. do tipo Cauchy, usar a média, como já foi dito, é inútil, sendo a mediana um estimador mais adequado.

7.15 Estatística suficiente

Há uma maneira mais fundamental de definir os critérios ótimos para seleção de um estimador, que está relacionada à característica de *suficiência* da estatística.

Se temos duas ou mais observações, considerando r estatísticas independentes $t, t_1, t_2, \ldots, t_{r-1}$, pode-se desmembrar a f.d.p. conjunta das r estatísticas numa f.d.p. de t e numa f.d.p. das demais estatísticas,

$$f_r(t, t_1, t_2, \ldots, t_{r-1} \mid \theta) = g(t \mid \theta) \cdot h_{r-1}(t_1, t_2, ..., t_{r-1} \mid t, \theta) \quad .$$

Se h_{r-1} é independente de θ, as estatísticas $t_1, \ldots, t_{r-1}$ não contribuem em nada ao conhecimento de θ. Assim, se

$$f_r(t, t_1, t_2, \ldots, t_{r-1} \mid \theta) = g(t \mid \theta) \cdot h_{r-1}(t_1, t_2, ..., t_{r-1} \mid t) \quad ,$$

então t é uma **estatística suficiente para** θ.

Essa é a definição formal, é a expressão matemática mais simples que podemos propor para o fato do estimador t sozinho carregar **toda** a informação contida na medida, mas não é prática. Ela pode também ser reduzida a uma condição sobre a função verossimilhança, $\mathcal{L}$. Se

$$\mathcal{L}(x_1, x_2, \ldots, x_N \mid \theta) = g(t \mid \theta) \cdot k(x_1, x_2, ..., x_N) \quad ,$$

com x_i representando os dados da medida, então t é uma estatística suficiente.

Verifica-se que o critério de suficiência sozinho é equivalente aos critérios que vínhamos usando e que há uma correspondência entre a existência de uma estatística suficiente e a existência de um estimador com variância igual ao LMV. Não vamos, porém, nos alongar nesta questão, que referimos apenas para completar a discussão do assunto.

EXERCÍCIOS

7.1. Usando a desigualdade de Chebyshev (6.56), determine quantos dados são necessários para que a média de eventos que obedecem a uma distribuição de Poisson com parâmetro a não difira de a, em módulo, mais do que $\sqrt{a}$, com no máximo 16% de probabilidade.

7.2. Use o teorema central do limite para determinar a f.d.p. da média de N dados que obedecem a uma f.p. de Poisson com parâmetro a, em que N tem o valor encontrado na solução do exercício 7.1. Com essa f.d.p., calcule a probabilidade do módulo da diferença entre a média dos N dados e a ser maior do que $\sqrt{a}$.

7.3. Faça uma tabela comparando os limites de probabilidade dados pela desigualdade de Chebyshev com as probabilidades exatas no caso de um dado que obedece a uma f.d.p. gaussiana centrada em x_0 e com desvio padrão σ. Faça essa tabela para $\epsilon = \sigma$, $\epsilon = 2\sigma$ e $\epsilon = 3\sigma$.

7.4. Deseja-se medir a taxa de ocorrência de um determinado evento, r, onde o número de eventos em um intervalo de tempo t obedece a uma distribuição de Poisson com parâmetro $r \cdot t$,

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

O valor estimado de r é

$$\hat{r} = \frac{n}{t} \quad .$$

(a) Mostre que o valor esperado de $\hat{r}$ é igual a r, para qualquer intervalo de tempo $t > 0$. (Isso significa que $\hat{r}$ é não-tendencioso).

(b) Usando a aproximação gaussiana para a distribuição de Poisson, eq. (2.54), mostre que quando $t \to \infty$ (e portanto $n \to \infty$, desde que r não seja nulo), então $\hat{r} \to r$ (ou seja, $\hat{r}$ é consistente).

7.5. A estimativa de máxima verossimilhança do parâmetro p de uma distribuição binomial com N tentativas é

$$\hat{p} = \frac{n}{N} \quad ,$$

onde n é o número de sucessos observados. Mostre que $\hat{p}$ é consistente e não-tendencioso.

7.6. No caso de uma distribuição binomial do número de sucessos na repetição de N ensaios com probabilidade p, podemos ficar tentados a estimar p^2 por

$$\frac{n^2}{N^2} \quad ,$$

em que n representa o número de sucessos observados. Mostre que

(a) essa estimativa é tendenciosa.

(b) a estimativa de p^2 dada por

$$\frac{n(n-1)}{N(N-1)}$$

não é tendenciosa.

7.7. Verifique que, para uma distribuição binomial com parâmetros N e p, a estimativa de p dada por n/N tem variância igual ao LMV.

7.8. Considere dois dados, x_1 e x_2, que obedecem a f.d.p.s gaussianas, independentes e com desvios padrões σ_1 e σ_2, respectivamente.

(a) Mostre que o LMV é

$$\text{LMV} = \frac{1}{\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2}} \quad .$$

(b) Mostre que a média ponderada dos dois dados, com pesos iguais aos inversos das respectivas variâncias, tem variância igual ao LMV.

Bibliografia

[Arfken] Mathematical Methods for Physicists, G.Arfken & H.Weber, Academic Press, 4^a 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, 2^a 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, 2^a 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.Moralles, 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.Moralles, 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