

André Fabiano Steklain

APOSTILA DE ERROS

Instrumentação para ensino - F809

Primeiro Semestre de 2002
ORIENTADOR: Prof. Mario Antonio Gneri
Instituto de Física Gleb Wataghin - UNICAMP

Sumário

Prefácio	5
1 Erros (versão simplificada)	7
1.1 Estimando incertezas a partir das medidas	7
1.1.1 Instrumentos Eletrônicos Digitais	8
1.2 Informando e usando incertezas	8
1.3 Tratamento Estatístico dos Erros	9
1.4 Propagação de Erros	10
1.4.1 Método da Cota Máxima	10
1.4.2 Método Estatístico	11
2 Erros (versão avançada)	13
2.1 Modelo de medição	13
2.2 Estimador da média de X (\bar{X})	14
2.3 Estimador da variância de X (S^2)	14
2.4 Desvio padrão da média	15
2.5 A Distribuição Normal	15
2.6 Propagação de Erros	16
A Histogramas	19
B Independência de Variáveis aleatórias	21
C Propriedades da Esperança e da Variância	25
D Conceitos Gerais de Estimação	27
E Teorema Central do Limite	31
E.1 História	31

Prefácio

Na análise de um problema, as conclusões são baseadas apenas parcialmente nos “dados”. Outro suporte tão importante quanto os “dados” é constituído por hipóteses ou suposições a respeito da situação sob análise. Dado que na prática é impossível considerar todos os aspectos de uma realidade, o uso destas hipóteses se torna necessário para permitir a modelagem ou abordagem matemática do problema. Estas suposições são racionalizações matemáticas ou simplificações de conhecimentos imprecisos ou até de palpites, não sendo, portanto, estritamente verdadeiras. A arte na escolha de tais hipóteses reside em que as restrições matemáticas por elas impostas não sejam conflitantes com aspectos relevantes da realidade.

A matemática tem um papel fundamental na resolução de qualquer tipo de problema, mas ela tem uma limitação: suas conclusões são sempre baseadas em hipóteses estritas, como ocorre em qualquer teorema.

Mesmo assumindo que as hipóteses não são satisfeitas à risca, algumas pessoas argumentam da seguinte forma: dado que as hipóteses são uma boa aproximação da realidade, então as conclusões tiradas a partir delas serão aproximadamente verdadeiras.

Este vago “princípio da continuidade- pequenos desvios nos supostos garantem pequenos desvios nas conclusões - não funciona necessariamente em qualquer situação, e muitos procedimentos científicos são muito sensíveis a pequenos desvios nas hipóteses em que se baseiam.

Capítulo 1

Erros (versão simplificada)

1.1 Estimando incertezas a partir das medidas

Uma questão-chave para a determinação de erros em medidas experimentais é: *Até que ponto podemos confiar nos valores fornecidos pelos instrumentos de medida?* Esta questão é bastante subjetiva, uma vez que o erro pode ter inúmeras fontes, e portanto depende basicamente da apreciação de quem faz a medida (afinal, os erros não podem ser nada mais do que estimados; não haveria sentido em se saber *exatamente* qual o erro de determinada medida). Existem, no entanto, fontes de erro que têm grande influência sobre os resultados e que podem ser facilmente identificadas. Estas fontes de erro são bastante diferentes dependendo do tipo do instrumento usado

Instrumentos analógicos

Os instrumentos analógicos possuem uma escala a partir da qual a medida é feita, seja diretamente (régua) ou através de um indicador (como a agulha de um voltímetro ou o nível do mercúrio em um termômetro). As principais incertezas estão relacionadas às escalas destes instrumentos.

O primeiro passo para se medir com estes instrumentos é definir o zero da escala. A posição do zero, contudo, pode não ser perfeitamente regular, o que influenciará deste modo todas as medidas. Existe também o fator visual ao se ajustar o zero, que depende do sistema de visão de quem está medindo (fator de erro aleatório) e da grossura das linhas das divisões, que pode mascarar um deslocamento da posição correta (veja figura 1.1).

Ao se fazer uma medida não ocorre sempre uma sobreposição do indicador com uma das divisões da escala. Desta forma, deve-se fazer uma estimativa da leitura, sendo esta estimativa o fator mais afetado pelo erro. Suponha uma leitura como a da figura 1.2. A agulha está apontando entre duas divisões da escala, de forma que é preciso estimar o valor medido. neste caso, pode-se fazer uma “divisão mental” do espaço entre as duas divisões em três espaços iguais, como na figura (???), e então, verificando em qual dos espaços está localizada a agulha, estimar a medida. Neste caso, o erro da medida deveria ser $1/3$ da menor escala. Juntando-se a este os erros na fabricação da escala, chega-se então na regra conhecida de que o erro neste caso é dado pela “metade da menor escala”. De fato, estes instrumentos são fabricados de modo a atender

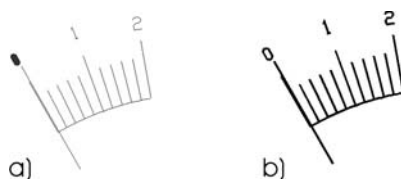


Figura 1.1: Mascaramento da posição do zero pela grossura das linhas. Em a), a leitura parece estar no zero. Em b), vê-se que o zero está deslocado.



Figura 1.2: Posição de medida desfavorável.

a esta regra. Contudo, este erro vale apenas para o aspecto visual da medida. Muitos outros erros podem estar presentes, de forma que *em geral o erro de uma medida feita com um instrumento analógico é maior do que a metade da menor divisão.*

Outro aspecto a se levar em conta é o erro devido à *paralaxe*. Este erro ocorre pelo fato da agulha estar acima do plano da escala. Deste modo, dependendo da posição com que o observador olha o instrumento, verá a agulha marcando valores diferentes. Este problema, contudo, é resolvido colocando-se um pequeno espelho na escala, de maneira que para se fazer a medida deve-se observar o instrumento de forma que a agulha esteja alinhada com a sua imagem no espelho. Deste modo, obriga-se que o observador tenha sempre uma visão frontal do aparelho.

1.1.1 Instrumentos Eletrônicos Digitais

Os medidores eletrônicos são diferentes dos elétricos, uma vez que incorporam elementos ativos, necessitando assim de uma fonte de energia. São mais instáveis e estão sujeitos a mudanças dos materiais devido ao uso, clima ou mesmo à ação do tempo. Medidores deste tipo, como o multímetro digital costumam ser fabricados de forma que o erro do instrumento coincida com o valor correspondente ao dígito que não é mostrado, utilizando-se de um processo de arredondamento. Apesar destes instrumentos terem amplificadores e portanto possuírem mais sensibilidade e precisão, estes aparelhos eletrônicos possuem tendência a mudar as suas características por envelhecimento.

1.2 Informando e usando incertezas

O resultado de qualquer medida de uma quantidade x pode ser expresso como

$$x_M \pm \delta x \quad (1.1)$$

Sendo x_M a melhor estimativa para a quantidade x , e δx a incerteza, ou erro da medida. Isto significa dizer que o valor real de x é provavelmente um valor entre $x_M - \delta x$ e $x_M + \delta x$. Existem algumas convenções e outros resultados estatísticos que servem como regras, e que não podem ser perdidas de vista ao se inscrever as incertezas.

- O valor da incerteza δx é sempre *positivo*. Desta forma, $x_M + \delta x$ é sempre o maior valor possível para x e $x_M - \delta x$ é o menor possível para x .
- O erro é sempre escrito *com um único algarismo significativo*. Por exemplo, caso se obtenha $x = 0,87 \pm 0,023$, deve-se escrever $x = 0,87 \pm 0,02$. Para $x = 0,87 \pm 0,027$, deve-se escrever $x = 0,87 \pm 0,03$.
- O último algarismo significativo da melhor estimativa deve ser da mesma ordem de grandeza (na mesma posição decimal) do que a incerteza. Por exemplo, não há sentido em se escrever $0,8735 \pm 0,02$. Neste caso, escreve-se $0,87 \pm 0,02$.
- Para medidas com unidades deve-se utilizar parênteses. Por exemplo, pode-se medir a aceleração da gravidade, g , como $g = (9,87 \pm 0,05)m/s^2$
- Diversas constantes utilizadas nos cálculos são medidas, e por isso também apresentam erro. Contudo, freqüentemente estes valores apresentam erros tão baixos que não chegam a afetar as medidas, podendo, então, ser tratados como valores exatos. Por exemplo, o valor da constante universal dos gases R , é conhecido como $R = (8,31451 \pm 0,00007)J/(mol \cdot K)$.

Ao se comparar medidas com outras, ou com o valor aceito, pode-se verificar a *discrepância* entre elas. A discrepância é definida como a diferença entre dois valores que se deseja comparar. Ou seja:

$$x_{1M} - x_{2M} \quad (1.2)$$

A discrepância pode ser significativa ou não. Suponha que dois valores obtidos para a aceleração da gravidade sejam $g = (9,35 \pm 0,04)m/s^2$ e $g = (9,76 \pm 0,05)m/s^2$. A discrepância entre estas medidas é, então, $9,76 - 9,35 = 0,41m/s^2$. Esta discrepância é significativa, pois não existe nenhum valor de g que seja compatível com as duas medidas. Isto significa que alguma coisa está errada. Para o caso de uma discrepância não significativa devem existir, portanto, valores que se ajustem às duas medidas (veja a figura XX). De modo geral, para duas medidas $x_{1M} \pm \delta x_1$ e $x_{2M} \pm \delta x_2$ tem-se:

- Se $|x_{1M} - x_{2M}| > \delta x_1 + \delta x_2$, a discrepância é significativa.
- Se $|x_{1M} - x_{2M}| < \delta x_1 + \delta x_2$, a discrepância é não significativa.

1.3 Tratamento Estatístico dos Erros

Nem sempre é possível estimar o erro diretamente a partir das medidas. Ao se medir o tempo com um cronômetro, por exemplo, muitos fatores devem ser levados em conta: o tempo de reação de quem está medindo, a eletrônica do aparelho usado, etc. Contudo, ao se fazer várias medidas da mesma grandeza o

seu erro pode ser determinado estatisticamente.

Suponha que uma grandeza x seja medida N vezes, e sejam obtidos os valores x_1, \dots, x_N . Em geral, a melhor estimativa para o valor correto da grandeza é a média amostral (\bar{x}), definida por

$$\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i \quad (1.3)$$

Define-se também a dispersão do conjunto de medidas como

$$\Delta x = S = \sqrt{\frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})^2} \quad (1.4)$$

Esta dispersão fornece uma indicação sobre a *precisão* das medidas: quanto menor, mais precisas as medidas são.

O erro padrão da média \bar{x} é definido por

$$\Delta \bar{x} = S_m = \frac{S}{\sqrt{N}} \quad (1.5)$$

Desta forma, o valor da grandeza x pode ser estimado como

$$\bar{x} \pm \Delta \bar{x} \quad (1.6)$$

Observe que, quanto maior o número N de medidas, menor é o erro padrão da média, e, portanto, mais exata é a estimativa de x .

1.4 Propagação de Erros

Muitas quantidades físicas não podem ser medidas de forma direta. Elas podem ser, contudo, funções de grandezas as quais podem ser medidas diretamente. Desta forma, deve-se saber qual a influência das incertezas destes valores medidos diretamente no valor da grandeza que se deseja medir.

1.4.1 Método da Cota Máxima

O método absoluto (ou da cota máxima) é uma das maneiras mais simples para se calcular a incerteza de uma função f de grandezas mensuráveis. Este método consiste em se calcular os valores máximo e o mínimo (f_{max} , f_{min}) que a função atinge ao se alternar os valores das medidas x_i dentro do intervalo $(x_i - \Delta x_i, x_i + \Delta x_i)$. A medida, então, é dada por

$$\bar{f} \pm \Delta f, \quad (1.7)$$

sendo \bar{f} e Δf dados por

$$\bar{f} = \frac{f_{max} + f_{min}}{2} \quad (1.8)$$

$$\Delta f = \frac{f_{max} - f_{min}}{2} \quad (1.9)$$

Este método fornece o valor do erro sem se utilizar nenhuma aproximação. Contudo, pode se tornar complicado encontrar os valores de f_{max} e f_{min} , dependendo da função utilizada. Por isso, é necessário uma fórmula geral válida para os diferentes tipos de funções, e que cujos cálculos sejam mais fácil de se efetuar. Esta fórmula é fornecida a partir da estatística, se utilizando de algumas aproximações. Este será o tema da próxima seção.

1.4.2 Método Estatístico

Seja f uma função de n variáveis: x, y, \dots as quais são medidas e dadas como $x_M \pm \delta x, y_M \pm \delta y, \dots$. Desta forma, a melhor estimativa do valor de f é dada por $f(x_M, y_M, \dots)$, e o seu erro é dado por:

$$\delta f = \sqrt{\left(\frac{\partial f}{\partial x} \delta x\right)^2 + \left(\frac{\partial f}{\partial y} \delta y\right)^2 + \dots} \quad (1.10)$$

Em qualquer caso, o valor δf nunca é maior do que a soma ordinária:

$$\delta f \leq \left|\frac{\partial f}{\partial x}\right| \delta x + \left|\frac{\partial f}{\partial y}\right| \delta y + \dots \quad (1.11)$$

Às vezes pode-se tornar bastante complicado calcular as derivadas envolvidas no processo. Contudo, existem casos simples que são fáceis de serem calculados. Por exemplo, para a soma ou subtração de grandezas ($f(x, y, \dots) = x + y + \dots$), tem-se

$$\delta f = \sqrt{(\delta x)^2 + (\delta y)^2 + \dots} \quad (1.12)$$

Já para a multiplicação ou divisão ($f(x, y, \dots) = xy \dots$) tem-se

$$\frac{\delta f}{f_M} = \sqrt{\left(\frac{\delta x}{x_M}\right)^2 + \left(\frac{\delta y}{y_M}\right)^2 + \dots} \quad (1.13)$$

Capítulo 2

Erros (versão avançada)

2.1 Modelo de medição

Toda a grandeza a ser medida representa, em termos estatísticos, uma *variável aleatória*. Isto significa que, por trás desta grandeza, existe uma distribuição probabilística que fornece a probabilidade de se obter um determinado valor ao se medir a grandeza. Neste contexto, define-se um *modelo de medição*, definido por

$$x = \alpha + e \quad (2.1)$$

Sendo a variável aleatória X o valor obtido a cada medida. Ela está associada a um valor constante, α , que representa o valor correto da grandeza, e a uma variável aleatória e , que representa o *erro* da medida. Caso tivéssemos a distribuição da variável X ou e , seria fácil obter o valor correto α . Contudo isso significaria ter que se fazer infinitas medidas, até obter todos os pontos da distribuição, o que, na prática, é impossível. Por isso, é preciso um método de *estimar* o valor de α por meio de uma amostra de X . Para fazer isto, vamos formular algumas hipóteses de trabalho as quais em geral são bastante razoáveis se aplicarmos às condições experimentais encontradas na maioria das medidas. Estas hipóteses são:

- $Média(e) = 0$
- $0 < Variância(e) = \sigma^2 < \infty$

Pode-se entender pela primeira hipótese que erros “para mais” devem ocorrer de tal forma que, em média, cancelem os erros “para menos” (note que esta hipótese é mais fraca do que a hipótese de que a distribuição seja simétrica). A segunda hipótese implica que erros muito grandes devem ser muito pouco prováveis. A maioria das situações experimentais se encaixam nestas hipóteses, de forma que as conclusões obtidas podem ser aplicadas para a maioria das vezes.

Tomando as hipóteses, vamos fazer N medições de X , X_1, \dots, X_N . Estas N variáveis aleatórias são i.i.d. (independentes e identicamente distribuídas). Tem-se desta forma

$$\text{Média}(X_i) = \alpha \quad (2.2)$$

$$\text{Variância}(X_i) = \sigma^2 \quad (2.3)$$

2.2 Estimador da média de X (\bar{X})

A média amostral, \bar{X} , definida por

$$\bar{X} = \frac{1}{N} \sum_{i=1}^N x_i \quad (2.4)$$

é um bom estimador de α , visto que

$$\text{Média}(\bar{X}) = \alpha \quad (2.5)$$

$$\text{Variância}(\bar{X}) = \frac{\sigma^2}{N} \quad (2.6)$$

Conforme se pode perceber por (3.6), a variância de \bar{X} diminui conforme o número de medições N aumenta, o que significa que o valor \bar{X} se torna mais preciso. Desta forma, para um grande número de medições, é possível estimar o valor de α com bastante precisão.

A média amostral não é o único estimador do valor α . Contudo, é possível demonstrar que a média amostral é o estimador de variância mínima de α dentro da classe de todos os estimadores lineares sem viés (ver Apêndice 3). Desta forma, pode-se considerar a média amostral (\bar{X}) como a melhor estimativa do valor correto da grandeza α no *contexto destas hipóteses*.

2.3 Estimador da variância de X (S^2)

Uma vez estimado o valor de α , correspondente à média da distribuição de X , é preciso estimar a variância desta distribuição, σ^2 , a qual fornece uma idéia da precisão de α . Pode-se pensar que um bom estimador seja o desvio quadrático médio (ou desvio RMS, do inglês *Root Mean Square*), dado por

$$s = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - \bar{X})^2} \quad (2.7)$$

Contudo, calculando-se a média de s^2 , tem-se

$$\text{Média}(s^2) = \frac{N-1}{N} \sigma^2 \quad (2.8)$$

Ou seja, s^2 é um estimador viesado de σ^2 . De fato, para $N \gg 1$, tem-se $\text{Média}(s^2) \approx \sigma^2$, e pode-se desprezar esta diferença (ver apêndice 5). No entanto, a rigor utiliza-se como estimador de σ o desvio padrão S , definido por

$$S = \sqrt{\frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{X})^2} \quad (2.9)$$

O desvio padrão (e o seu quadrado) são funções de variáveis aleatórias. Desta forma S^2 é também uma variável aleatória, possuindo a sua própria distribuição. Isto significa que é possível estimar a variância $\sigma_{S^2}^2$, e saber deste modo com que precisão S^2 fornece o valor de σ^2 . Tem-se para σ_{S^2} :

$$\sigma_{S^2} = \sqrt{\text{Média}[(S^2 - \sigma^2)^2]} \quad (2.10)$$

Em distribuições de probabilidade arbitrárias esta quantidade não pode ser expressa em termos apenas da média e variância da distribuição de X . Contudo, para a distribuição *normal* tem-se

$$\sigma_{S^2} = \sqrt{\frac{2}{N-1}}\sigma^2 \quad (2.11)$$

2.4 Desvio padrão da média

O desvio padrão de \bar{X} ($\sigma_{\bar{X}}$), como já foi visto, é dado por

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{N}} \quad (2.12)$$

Este valor pode ser estimado por $\Delta\bar{X}$, definido por

$$\Delta\bar{X} = \frac{S}{\sqrt{N}} \quad (2.13)$$

Esta quantidade é chamada *desvio padrão na média*. Esta quantidade permite dizer que existe uma grande probabilidade de que o valor correto (α) esteja entre $\bar{X} + \Delta\bar{X}$ e $\bar{X} - \Delta\bar{X}$. Desta forma, tem-se

$$\alpha = \bar{X} \pm \Delta\bar{X} \quad (2.14)$$

2.5 A Distribuição Normal

As conclusões das seções anteriores são válidas para qualquer distribuição de probabilidade, conquanto sejam satisfeitas as hipóteses de trabalho propostas na seção 2.1. No entanto, existe uma distribuição que possui propriedades que a tornam melhor para se trabalhar do que as outras. Esta é a chamada *distribuição normal* ou *gaussiana*, sendo dada por

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \quad (2.15)$$

Dentre as propriedades da distribuição normal estão as seguintes:

- A função $p(x)$ depende de apenas dois parâmetros, μ e σ^2 , que são respectivamente a média e a variância da distribuição. Desta forma, o conhecimento destas duas grandezas fornece todas as informações necessárias a respeito da distribuição.
- A média amostral \bar{X} é o estimador de variância mínima entre *todos* os não viciados, (e não apenas o variância mínima entre os não viciados e lineares, como havia-se dito antes)

- \bar{X} é, além do estimador de quadrados mínimos, o estimador de máxima verossimilhança, e como tal, é ótimo em termos assintóticos, pois é assintoticamente consistente e eficiente.

Uma pequena digressão histórica

O nome “distribuição gaussiana” é inadequado a esta função. A descoberta desta fórmula não se deve a Gauss, como se poderia supor, mas a Abraham DeMoivre, que foi o primeiro a apresentar esta distribuição no ano de 1733 em um trabalho de distribuição privada, o qual foi incluído em 1738 na sua conhecida obra *Doctrine of Chances*. Foi Gauss, no entanto, quem tornou esta distribuição conhecida. Ele tentava determinar o valor mais provável de uma quantidade desconhecida, chegando à conclusão de que era impossível sem se saber a distribuição de probabilidade associada a ela. Gauss então supôs existente tal distribuição e supôs, como hipótese central, a *excelência da média amostral*. Desta forma ele deduziu qual deveria ser a forma da distribuição para que a solução de máxima verossimilhança fosse a média amostral. Neste ponto é importante ressaltar que o resultado de Gauss depende totalmente da hipótese de que a média amostral é o melhor estimador da medida, o que não é necessariamente verdadeiro.

Grandezas físicas e a distribuição normal

O físico francês Lippman, prêmio Nobel em 1908 certa vez disse o seguinte: “Todos acreditam na distribuição normal: os matemáticos porque pensam que é um fato experimental e os experimentadores porque pensam que é um teorema matemático...”

Como ressaltado por Lippman, existe uma grande confusão em torno de quando considerar uma distribuição “normal”. O argumento mais comum aos cientistas vem do chamado *Teorema Central do Limite*. Segundo a opinião de muitos, o resultado da medida “depende de um número muito grande de processos aleatórios independentes microscópicos, que se somam no final dando um caráter aleatório a uma grandeza macroscópica”. O que geralmente se esquece é que o Teorema Central do limite exige também que outras hipóteses sejam satisfeitas por estes processos microscópicos. Contudo, estas hipóteses geralmente são de difícil, senão impossível verificação. Desta forma, torna-se mais fácil comprovar se a distribuição é normal diretamente através de testes apropriados do que fazer suposições *a priori* a respeito da validade do teorema. Para se comprovar a gaussianidade de determinada distribuição, pode-se utilizar de testes como o teste do χ^2 (qui-quadrado), ou, ainda, testes mais poderosos, como o de teste de Komogoroff-Smirnov.

2.6 Propagação de Erros

Antes de apresentar argumentos mais gerais, vamos discutir o caso mais simples. Suponha uma função de uma variável aleatória X , $f(X)$. Dados os valores estimados da média de X , \bar{X} e da sua variância, $(\Delta\bar{X})^2$, quais são os valores estimados para a média e variância de f ? Sendo o erro pequeno e sendo a média dos erros nula (conforme as hipóteses apresentadas), tem-se que as medidas de

X se distribuem em torno de \bar{X} , de forma que se pode fazer uma expansão da função em uma série de Taylor até primeira ordem, desprezando os demais termos:

$$f_i = f(x_i) \approx f(\bar{X}) + \left(\frac{\partial f}{\partial x} \right)_{x=\bar{X}} (x_i - \bar{X}) \quad (2.16)$$

Usando-se esta aproximação pode-se então calcular a média amostral de f , \bar{f} :

$$\bar{f} \approx \frac{1}{N} \sum_{i=1}^N f_i = f(\bar{X}) \quad (2.17)$$

Se devido ao fato de que $\sum_{i=1}^N (x_i - \bar{x}) = 0$. Para o desvio padrão de f , S_f , tem-se :

$$S_f^2 \approx \frac{1}{N-1} \sum_{i=1}^N (f_i - \bar{f})^2 = \left(\frac{\partial f}{\partial x} \right)^2 S^2 \quad (2.18)$$

Desta forma, o desvio na média ($\Delta \bar{f}$) é dado por:

$$\Delta \bar{f} \approx \left| \frac{\partial f}{\partial x} \right| \Delta \bar{X} \quad (2.19)$$

A interpretação desta formula pode ser vista graficamente na figura (???). Como pode-se ver, um incremento δx no argumento da função f causa um incremento δf nesta. Sendo δf e δx muito pequenos, então, pode-se considerar uma relação linear entre eles, sendo a coeficiente angular o mesmo da reta que passa tangente à curva, e, portanto, tem inclinação $\partial f / \partial x$. Desta forma, os dois incrementos guardam a relação (2.19)

O procedimento é análogo para mais de uma variável, lembrando-se da expansão de Taylor para o caso multidimensional. Para duas variáveis, x e y , tem-se:

$$\bar{f} \approx f(\bar{x}, \bar{y}) \quad (2.20)$$

$$S_f^2 \approx \left| \frac{\partial f}{\partial x} \right|^2 S_x^2 + \left| \frac{\partial f}{\partial y} \right|^2 S_y^2 + S_{xy} \quad (2.21)$$

Sendo S_{xy} dado por

$$S_{xy} = \frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{X})(y_i - \bar{Y}) \quad (2.22)$$

Sendo x e y independentes, é fácil verificar que, conforme o número de medidas aumenta, a quantidade S_{xy} tende a zero ¹, de forma que podemos desprezá-la, resultando:

$$S_f^2 \approx \left| \frac{\partial f}{\partial x} \right|^2 S_x^2 + \left| \frac{\partial f}{\partial y} \right|^2 S_y^2 \quad (2.23)$$

¹Isto se deve ao fato de que o S_{xy} amostral tende ao S_{xy} populacional (lei dos grandes números). Como as duas variáveis são independentes, este valor (a covariância) é zero.

Generalizando para N variáveis, tem-se então as fórmulas gerais:

$$\bar{f} \approx f(\bar{x}, \bar{y}, \dots) \quad (2.24)$$

$$\Delta \bar{f} \approx \sqrt{\left| \frac{\partial f}{\partial x} \right|^2 (\Delta \bar{x})^2 + \left| \frac{\partial f}{\partial y} \right|^2 (\Delta \bar{y})^2 + \dots} \quad (2.25)$$

Posto que o erro é dado por (2.25), tem-se que este nunca é igual à soma ordinária dos erros, isto é,

$$\Delta \bar{f} \leq \left| \frac{\partial f}{\partial x} \right| \Delta \bar{x} + \left| \frac{\partial f}{\partial y} \right| \Delta \bar{y} + \dots \quad (2.26)$$

Apêndice A

Histogramas

As freqüências relativas dos diferentes valores medidos podem ser dispostos graficamente em um *histograma*, o qual representa a freqüência relativa para cada intervalo (ou classe) de medidas (veja a figura A.1). Neste apêndice serão mostrados os passos para se construir um histograma.

Suponha que foram obtidas N medidas de uma grandeza X , x_1, \dots, x_N . Sendo x_{max} e x_{min} respectivamente o maior e o menor valor obtidos, define-se a amplitude amostral $(\Delta_h x)$ ¹ por:

$$\Delta_h x = x_{max} - x_{min} \quad (\text{A.1})$$

Deve-se saber qual o número de intervalos (ou classes) adequados ao histograma, de forma que este faça sentido (um histograma com um número muito grande ou pequeno de classes não traz nenhuma informação sobre a distribuição de X). Um valor razoável do número de classes (n) é dado pela *fórmula de Sturges*:

$$n \approx 1 + 3,3 \log N \quad (\text{A.2})$$

O próximo passo, então, é calcular a amplitude de cada intervalo ($\delta_h x$). Este valor é dado por:

$$\delta_h x = \frac{\Delta_h x}{n} \quad (\text{A.3})$$

Obtém-se, assim, a estrutura do histograma. Deve-se então fazer o levantamento de quantas medidas x_i estão em cada intervalo $(x_{min} + k\delta_h x, x_{min} + (k+1)\delta_h x)$, $k = 0, \dots, n$. Sendo N_k o número de medidas neste intervalo, a freqüência é dada por

$$f_k = \frac{N_k}{N} \quad (\text{A.4})$$

O tamanho das barras do histograma, então, são proporcionais às freqüências relativas. Tem-se, naturalmente, $\sum_{i=1}^n n = 1$. À medida que N cresce, f_k tende à probabilidade de se medir x no intervalo $\delta_h x$. Para $N \rightarrow \infty$, $f_k \rightarrow p(x)\delta_h x$, sendo $\delta_h x$ infinitesimal, e o histograma representa uma excelente aproximação da função densidade de probabilidade, $p(x)$.

¹Não confunda $\Delta_h x$ com Δx , o erro de cada medida

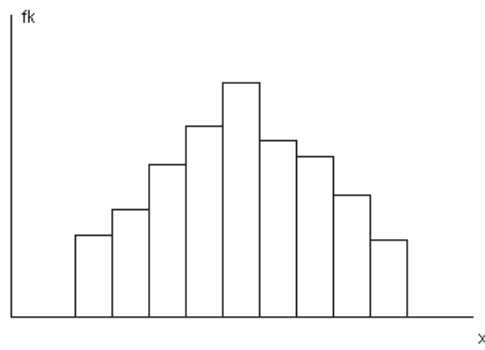


Figura A.1: *Exemplo de um histograma*

Apêndice B

Independência de Variáveis aleatórias

Definição B.1 *As variáveis aleatórias X e Y são ditas independentes se e somente se elas satisfazem qualquer uma das seguintes condições equivalentes:*

- (i) $Prob(X \in I \text{ e } Y \in J) = Prob(X \in I) \cdot Prob(Y \in J)$ para todo par de intervalos I e J ;
- (ii) $Prob(X \in I | Y \in J) = Prob(X \in I)$ para todo par de intervalos I e J onde $Prob(Y \in J) \neq 0$;
- (iii) $F_{XY}(x, y) = F_X(x) \cdot F_Y(y)$, para todos x e y reais, onde F_{XY} , F_X e F_Y denotam as funções de distribuição acumulada da conjunta, da variável X e da variável Y , respectivamente;
- (iv) O enunciado (iii) assume a seguinte forma se a variável for do tipo discreto: $Prob(X = x \text{ e } Y = y) = Prob(X = x) \cdot Prob(Y = y)$, para todos x, y reais;
- (v) O enunciado (iii) assume a seguinte forma se a variável for de tipo contínuo: $f_{XY}(x, y) = f_X(x) \cdot f_Y(y)$ para todos x, y reais, onde f_{XY} , f_X e f_Y denotam as funções de densidade conjunta, a densidade de X e a densidade de Y , respectivamente;
- (vi) No caso de variáveis qualitativas, a definição de independência assume as seguintes formas:
 - $Prob(X \in I \text{ e } Y \in J) = Prob(X \in I) \cdot Prob(Y \in J)$, para todo par de subconjuntos I e J da imagem de X ;
 - $Prob(X \in I | Y \in J) = Prob(X \in I)$, para todo par de subconjuntos I e J da imagem de X onde $Prob(Y \in J) \neq 0$.

Observações:

- (a) A palavra intervalo é usada aqui no sentido mais geral possível, desde um ponto até intervalos com algum ou ambos extremos infinitos, os intervalos contendo nenhum, um ou ambos extremos;

- (b) Sendo que as condições acima são todas equivalentes sempre pode ser escolhida aquela que resultar mais conveniente segundo seja o problema. A condições (ii) e (vi.b) dão uma idéia intuitiva do sentido da palavra independência; já as condições (iii), (iv) e (v), embora não tão transparentes, são de grande utilidade nas demonstrações. Também a condições (i) e (vi.a) têm seu apelo intuitivo, segundo podemos ver no seguinte exemplo.

Exemplo 1

Uma população é classificada segundo duas características: sexo (Feminino (F) ou Masculino (M)) e cor dos olhos (Claro (C) ou Escuro (E)). Suponhamos que:

$$\begin{aligned} Prob(F) &= 0.55 & Prob(M) &= 0.45 \\ Prob(C) &= 0.40 & Prob(E) &= 0.60; \end{aligned}$$

Se as características sexo e cor dos olhos fossem independentes teríamos:

$$\begin{aligned} Prob(F \cap C) &= 0,55 \times 0,40; & Prob(F \cap E) &= 0.55 \times 0,60; \\ Prob(M \cap C) &= 0,45 \times 0,40; & Prob(M \cap E) &= 0,45 \times 0,60; \end{aligned}$$

Isto é, a probabilidade das interseções seria diretamente proporcional à dos conjuntos intersectados. Observando este exemplo desde a condição (ii), podemos ilustrar desta forma o conceito de independência:

$$Prob(C/F) = Prob(C/M) = Prob(C) \quad (B.1)$$

$$Prob(E/F) = Prob(E/M) = Prob(E) \quad (B.2)$$

Ou seja, que a proporção de pessoas de olho claro entre as mulheres, entre os homens e no total da população é a mesma; também ocorrem aqui outras proporcionalidades:

$$Prob(F/C) = Prob(F/E) = Prob(F) \quad (B.3)$$

$$Prob(M/E) = Prob(M/C) = Prob(M) \quad (B.4)$$

Exemplo 2

Um conjunto de pessoas é classificado segundo peso e altura. Cada variável divide a população em duas classes disjuntas, segundo que o valor observado seja maior ou menor que a respectiva mediana. Temos então as duas dicotomias:

$$\begin{aligned} \{Leves(L), & \quad Pesados(P)\} \\ \{Baixos(B), & \quad Altos(A)\} \end{aligned}$$

Sendo:

$$Prob(L) = Prob(P) = Prob(B) = Prob(A) = \frac{1}{2}.$$

Também sabe-se que:

$$\begin{aligned} Prob(L \cap B) &= 0,34 & Prob(L \cap A) &= 0,16 \\ Prob(P \cap B) &= 0,14 & Prob(P \cap A) &= 0,36 \end{aligned}$$

Assim, dadas as informações acima, observa-se que:

$$\begin{aligned} \text{Prob}(L \cap B) > \text{Prob}(L) \times \text{Prob}(B) & \quad \text{e} \quad \text{Prob}(P \cap A) > \text{Prob}(P) \times \text{Prob}(A) \\ \text{Prob}(L \cap A) < \text{Prob}(L) \times \text{Prob}(B) & \quad \text{e} \quad \text{Prob}(P \cap B) < \text{Prob}(P) \times \text{Prob}(A) \end{aligned}$$

O que mostra que os atributos peso e altura não são independentes, já que eles apresentam o seguinte tipo de associação: a maior altura, maior peso (sempre em termos gerais).

Apêndice C

Propriedades da Esperança e da Variância

A seguir apenas enunciaremos algumas das propriedades da média e da variância. As demonstrações podem ser encontradas em qualquer livro de probabilidade elementar.

Sejam X e Y variáveis aleatórias e c uma constante. Então:

$$\text{Média}(c) = c \quad (\text{C.1})$$

$$\text{Variância}(c) = 0 \quad (\text{C.2})$$

$$\text{Média}(cX) = c \cdot \text{Média}(X) \quad (\text{C.3})$$

$$\text{Média}(X + Y) = \text{Média}(X) + \text{Média}(Y) \quad (\text{C.4})$$

$$\text{Variância}(cX) = c^2 \cdot \text{Variância}(X) \quad (\text{C.5})$$

Se X e Y são independentes:

$$\text{Variância}(X + Y) = \text{Variância}(X) + \text{Variância}(Y) \quad (\text{C.6})$$

Apêndice D

Conceitos Gerais de Estimação

Conceitos Gerais de Estimação

Definição D.1 *No contexto de um determinado problema, a população ou universo é o conjunto de todos os elementos de interesse.*

Definição D.2 *Seja X uma variável aleatória definida na população. Dizer que a distribuição de X segue um modelo paramétrico significa assumir que a distribuição de X pertence a uma família de probabilidades*

$$\{P_\theta, \theta \in \Theta\}, \quad (\text{D.1})$$

Onde Θ é um subconjunto de \mathfrak{R} ou \mathfrak{R}^k e θ é um parâmetro desconhecido.

Definição D.3 *Uma amostra de tamanho n é uma família $\{X_1, X_2, \dots, X_n\}$ de variáveis aleatórias independentes e identicamente distribuídas (i.i.d.) definidas na população. Denotaremos por $\{x_1, x_2, \dots, x_n\}$ os valores observados da família $\{X_1, X_2, \dots, X_n\}$.*

Definição D.4 *Uma estatística é qualquer função da amostra, ou seja qualquer:*

$$g : (X_1, X_2, \dots, X_n) \rightarrow \mathfrak{R} \quad \text{ou} \quad \mathfrak{R}^k \quad (\text{D.2})$$

Uma estatística é uma variável aleatória (pois é função de variáveis aleatórias).

Exemplo. Seja $g : (X_1, X_2, \dots, X_n) \rightarrow \mathfrak{R}^2$ definida por:

$$g(X_1, X_2, \dots, X_n) = (\bar{X}, S^2) \quad (\text{D.3})$$

Na qual \bar{X} e S^2 são dados por:

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n x_i \quad (\text{D.4})$$

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{X})^2 \quad (\text{D.5})$$

Definição D.5 Um estimador do parâmetro $\theta \in \Theta$ é uma estatística utilizada para estimar o parâmetro θ . Uma estimativa é um valor observado do estimador.

Definição D.6 Seja $\hat{\theta} = \hat{\theta}(X_1, X_2, \dots, X_n)$ um estimador do parâmetro $\theta \in \Theta$. O viés de $\hat{\theta}$ é a função definida em Θ através da fórmula:

$$\text{Viés}(\hat{\theta}) = |E_{\theta}(\hat{\theta}) - \theta|, \quad (\text{D.6})$$

Sendo E_{θ} calculada a partir de P_{θ} .

Definição D.7 O erro quadrático médio do estimador do parâmetro θ é definido como:

$$EQM_{\theta} = E_{\theta}[(\hat{\theta} - \theta)^2] \quad (\text{D.7})$$

O EQM está definido em Θ e é utilizado para comparar o desempenho dos estimadores.

Definição D.8 Um estimador é sem viés se e somente se:

$$E_{\theta}(\hat{\theta}) = \theta \quad \text{para todo} \quad \theta \in \Theta$$

Observação. No caso dos estimadores sem viés, $EQM_{\theta}(\hat{\theta})$ coincide com $Var_{\theta}(\hat{\theta})$:

$$Var_{\theta}(\hat{\theta}) = E_{\theta}[(\hat{\theta} - E_{\theta}(\hat{\theta}))^2] = E_{\theta}[(\hat{\theta} - \theta)^2] = EQM_{\theta}(\hat{\theta}) \quad (\text{D.8})$$

Quadrados Mínimos

Definição D.9 Seja X uma variável aleatória tal que $\mu = \text{Média}(X)$. O estimador de quadrados mínimos de μ baseado no conjunto de valores observados x_1, x_2, \dots, x_n da variável X é definido como o número real m que minimiza a expressão:

$$Q(m) = \sum_{i=1}^n (x_i - m)^2 \quad (\text{D.9})$$

Teorema D.1 \bar{X} é o estimador de quadrados mínimos $\mu = \text{Média}(X)$.

Teorema D.2 Sejam X uma variável aleatória com média μ e variância finita σ^2 e X_1, X_2, \dots, X_n i.i.d.. Então \bar{X} é o estimador de variância mínima de μ dentro da classe de todos os estimadores lineares e sem viés.

Observação. O teorema acima é a versão mais simples do denominado teorema de Gauss-Markov, que afirma que, sob condições bastante gerais no contexto dos modelos lineares, os estimadores de quadrados mínimos são os estimadores de variância mínima na família dos lineares e não viesados.

Máxima verossimilhança

Um pouco de história

Os estimadores de máxima verossimilhança foram introduzidos por Gauss (final do século XVIII) mas foi Fisher (1912) quem desenvolveu em forma plena a metodologia de máxima verossimilhança.

Cálculo de EMV no caso da distribuição Normal

Sejam X_1, X_2, \dots, X_n i.i.d. segundo uma distribuição Normal(μ, σ^2). Portanto, a função de densidade de probabilidade de X_i é dada por:

$$\phi(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \quad (\text{D.10})$$

O estimador de máxima verossimilhança de μ baseado nos valores observados x_1, x_2, \dots, x_n da variável aleatória X é o valor $\hat{\mu}$ que maximiza a função de verossimilhança:

$$\prod_{j=1}^n \phi(x_j, \hat{\mu}) = \prod_{j=1}^n \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x_j-\mu}{\sigma}\right)^2} = \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n e^{-\frac{1}{2}\sum_{i=1}^n \left(\frac{x_i-\mu}{\sigma}\right)^2} \quad (\text{D.11})$$

O que equivale a achar o valor de $\hat{\mu}$ que minimiza

$$\sum_{i=1}^n \left(\frac{x_i - \mu}{\sigma}\right)^2 \quad (\text{D.12})$$

Que é a mesma coisa que achar $\hat{\mu}$ que minimiza

$$\sum_{i=1}^n (x_i - \mu)^2 \quad (\text{D.13})$$

Ou seja, se X é normal o estimador de máxima verossimilhança coincide com o estimador de quadrados mínimos que, segundo já foi visto, é \bar{X} .

Apêndice E

Teorema Central do Limite

Teorema Central do Limite (versão i.i.d.)

Sejam X_1, X_2, \dots, X_n i.i.d. com $Média(X_1) = \mu$ e $0 < Variância(X_1) = \sigma^2 < \infty$. O Teorema Central do Limite afirma que a distribuição limite de

$$\frac{\bar{X} - Média(\bar{X})}{DesvioPadrão(\bar{X})} = \frac{(\bar{X} - \mu)}{\frac{\sigma}{\sqrt{n}}} = \sqrt{n} \frac{(\bar{X} - \mu)}{\sigma} \quad (\text{E.1})$$

É a $Normal(0,1)$.

E.1 História

O enunciado acima é a versão mais utilizada do Teorema Central do Limite. A primeira versão é de 1733 é devida a De Moivre, quem conjecturou que, se $X_n \sim Binomial(n, p)$, então a distribuição limite de

$$\frac{X_n - nE(X_n)}{DesvioPadrão(X_n)} = \frac{X_n - n \cdot p}{\sqrt{n \cdot p(1-p)}} \quad (\text{E.2})$$

é a $Normal(0,1)$.

De Moivre só provou o caso particular $p = 1/2$. Laplace em 1812 provou a conjectura de modo completo, isto é para $0 < p < 1$.

Observe-se que a $Binomial(n, p)$ é soma de n v.a. i.i.d. $Bernoulli(p)$. Tchebycheff e seus alunos Markov e Liapunov estenderam o resultado a somandos i.i.d. bastante gerais. Posteriormente, foi sendo relaxada a hipótese de que as variáveis tivessem a mesma distribuição (por exemplo, o teorema vale se a seqüência de v. a. for uniformemente limitada e ao menos uma das variáveis tiver $Variância > 0$).

No começo do século XX, Lindeberg e Feller estabeleceram que o teorema continua válido se as variáveis forem independentes e “nenhuma das variáveis for preponderante”, dando ao teorema sua forma mais geral. Existem também para o caso i.i.d. resultados sobre velocidade de convergência, dependendo da distribuição comum das X_j .

Apesar da importância do teorema, deve-se evitar o pensamento “toda v.a. é aproximadamente normal”, já que muitas vezes as hipóteses do TCL não ocorrem.

Referências Bibliográficas

- [1] Anders Hald. “*A history of probability and statistics and their applications before 1750*”. Wiley, 1990.
- [2] P. J. Huber. “Robust Statistics: a review”. *Ann. Math. Statist.*, 43:1041 – 1067, 1972.
- [3] John R. Taylor. “*An introduction to error analysis*”. University Science Books, 1997.
- [4] J. J. Lunazzi. “*Fundamentos básicos sobre erros - Com aplicações em eletricidade*”. IFGW - Unicamp, 2000.
- [5] Carlos Henrique Brito Cruz et. al. “*Física Experimental I - Apêndices e Complementos*”. IFGW- Unicamp, 1998.