Análise Probabilística de Resultados Experimentais

Probabilidade

O método científico e especialmente o uso de grandezas físicas têm modificado a vida do ser humano profundamente. Todo processo de industrialização depende do uso criterioso de avaliações e previsões quantitativas envolvendo grandezas físicas. Parte essencial do conceito de grandeza física contínua é o conceito de erro experimental. Sem admitir que os resultados experimentais possam ter uma margem de erro, praticamente todas as comparações de resultados experimentais com previsões teóricas levariam à condenação dos modelos teóricos por não concordarem com os resultados observados. A própria existência de grandezas físicas seria impossibilitada. Somente a modéstia de admitir a possibilidade de erro permite fazer previsões e descobrir regularidades e leis. Foi a modéstia que levou ao extraordinário desenvolvimento da tecnologia moderna. Mas muitas vezes esta modéstia ainda não é o suficiente.

Nas primeiras experiências feitas nesta disciplina vimos que a avaliação de erro experimental é difícil e incerta. Por causa disso é prudente formular as previsões de forma ainda mais cautelosa. Em vez de afirmar que mediremos um valor que coincidirá com um determinado valor teórico dentro da precisão experimental é melhor afirmar apenas que *provavelmente* esta coincidência acontecerá. Se ocasionalmente a esperada coincidência não for verificada, não é preciso jogar fora uma teoria que tal vez tenha servido milhares de vezes de forma perfeita. Mas se amolecermos nossas afirmações desta forma não chegaríamos a um "vale tudo"? Para que isto não aconteça é preciso definir regras para afirmações probabilísticas.

Uma probabilidade é um número $P_M[A]$ entre 0 e 1 que é associado a um possível acontecimento[1] A e que faz parte de um modelo teórico M. O significado deste número é determinado por um conjunto de regras. A idéia é de generalizar as antigas afirmações definitivas: em vez de afirmar "*A* acontecerá" vamos dizer "$P_M[A]=1$" e em vez de afirmar "*A* não acontecerá" vamos dizer "$P_M[A]=0$". A generalização destas afirmações definitivas reside na possibilidade de termos valores de $P_M[A]$ entre estes extremos; $0<P_M[A]<1$. A probabilidade é intimamente entrelaçada com o difícil conceito de tempo: da mesma forma que a afirmação "*A* acontecerá" perde seu sentido no instante que observamos o acontecimento *A*, a probabilidade $P_M[A]$ também perde o seu sentido. Valem as seguintes regras:

> **Regra da negação:** Se $\neg A$ é a negação lógica de A vale $P_M[\neg A] = 1-P_M[A]$.

> **Regra da soma**: Se A e B são possíveis acontecimentos excludentes, isto é se for impossível acontecer A e B, vale que $P_M[A\vee B]=P_M[A]+P_M[B]$, onde $A\vee B$ é o acontecimento "*A* ou *B*".

[1] Na literatura sobre probabilidade usa-se a palavra "evento". Mas como físicos definimos evento como um acontecimento momentâneo que ocorre num ponto. Para não utilizar a mesma palavra para duas noções diferentes usaremos aqui a palavra "acontecimento".

As probabilidades servem para formular afirmações sobre possíveis futuros acontecimentos e fornecer critérios para decisões que temos que tomar numa situação de incerteza. Estas afirmações são naturalmente baseadas em experiência passada. A seguinte regra determina como experiência passada influencia os modelos aplicáveis a situações semelhantes no futuro:

> **Regra da condenação de modelo**: Podemos escolher um acontecimento A cujo valor de probabilidade $P_M[A]$ é muito perto de zero e tomar a decisão de abandonar o modelo M caso A seja observado.

É muito fácil aplicar esta regra de forma errada. Esta regra é muito sutil e profunda, ela está relacionada com a imersão da nossa existência no tempo. Devemos escolher o acontecimento A e tomar a decisão de abandonar o modelo M caso A seja observado **antes de fazer o experimento** (ou pelo menos antes de ver os resultados)! O seguinte exemplo ilustra a importância da escolha anterior ao experimento: Imagine uma loteria com um número gigantesco de possíveis resultados. Por exemplo, podemos sortear 12 vezes cifras entre 0 e 9 para compor um número decimal com 12 casas. Os 12 processos de escolha de cifras devem ser completamente independentes e cifras já usadas podem ser sorteadas de novo. Existem 10^{12} possíveis resultados deste jogo. Um modelo bastante razoável seria um que atribui probabilidades iguais a todos os possíveis resultados. Vamos chamar este modelo o *modelo do jogo honesto*. Então cada número teria a probabilidade minúscula de 10^{-12}. Se aplicarmos a regra da condenação de modelo erradamente, poderíamos chegar sempre à conclusão que este modelo está errado, pois qualquer que seja o número sorteado o seu aparecimento como resultado seria sempre um acontecimento muito improvável. Veremos agora o que seria um procedimento correto: imagine que este jogo ofereça um prêmio milionário ao apostador que acertou o número e que ficamos sabendo que o senhor X, dono de 40 apartamentos de luxo, 50 carros de luxo, amigo de 3 juizes e grande organizador de festas populares, apostou no número 482363927406 como única aposta. Frente a esta situação vamos decidir abandonar o modelo do jogo honesto caso o número 482363927406 seja sorteado. Devemos tomar esta decisão antes de ver o resultado do sorteio. A escolha do acontecimento A que levaria à condenação do modelo é geralmente motivada por uma hipótese alternativa. No caso do exemplo da loteria, a alternativa seria "a probabilidade que a loteria forneça um número igual à aposta do senhor X é 1".

Que significado têm os valores de probabilidades entre 0 e 1 que não ficam perto de 0 e 1? Eles não servem para condenar um modelo. Mas eles podem ser usados em regras de decisões. Imagine que temos que atravessar um rio e temos duas pontes improvisadas na nossa frente. Imagine que temos um modelo teórico, baseado em estudos de estática e testado anteriormente, que prevê que a ponte A agüenta nosso peso com probabilidade 0,8 e a ponte B com probabilidade 0,6. Neste caso vamos escolher a ponte A. Se atravessarmos a ponte A e ela quebrar não podemos condenar o modelo usado. Mas probabilidades que não ficam perto de zero ou 1 podem ser usadas indiretamente em processos para julgar a validade de um modelo teórico. Veremos como isto funciona na seção seguinte:

Ensembles

A pesquisa científica procura regularidades ou leis simples num mundo extremamente complicado e complexo. A extração da regularidade simples requer necessariamente a abstração de fatores que julgamos de pouca ou nenhuma importância para nossa experiência. Por exemplo, no estudo do movimento de um carrinho no trilho de ar não deve fazer nenhuma diferença se comi frango ou bife antes de entrar no laboratório, o

carrinho deve-se mover da mesma forma independente do meu almoço. Então podemos desconsiderar tal fato irrelevante. Imagine que faremos duas experiências, uma pode ser a repetição da outra ou não. Se julgarmos que os resultados de uma são irrelevantes para a outra, vamos chamar as experiências de *estatisticamente independentes*. Um crítico poderia dizer que esta definição é muito vaga: – julgamos isto e aquilo como irrelevante, mas com que critérios? Bem, primeiramente com o bom senso formado nas nossas atividades não científicas e depois com um bom senso apurado pela experiência científica. De toda forma, nada impede que o nosso julgamento possa estar errado. Isto apareceria eventualmente em forma de inconsistências de resultados experimentais e nos dará motivo para autocrítica e conseqüente melhoria dos julgamentos. Com a noção de experiências independentes podemos acrescentar mais uma regra sobre probabilidades:

> **Regra do produto de probabilidades de acontecimentos independentes:** Sejam E_a e E_b duas experiências estatisticamente independentes e A um possível resultado da experiência E_a e B um possível resultado da experiência E_b. Se um modelo M prevê as probabilidades $P_M[A]$ e $P_M[B]$ para a observação dos respectivos resultados A e B, o mesmo modelo deve prever para a probabilidade de observar A e B o produto de $P_M[A]$ e $P_M[B]$. Isto é, $P_M[A \wedge B] = P_M[A] \cdot P_M[B]$. [2]

É especialmente interessante aplicar esta regra para repetições de experiências independentes. Vamos chamar um conjunto de repetições de experiências independentes de *ensemble de experiências*. Seja E uma experiência e A um possível resultado da mesma. Vamos supor que a probabilidade de observar A seja diferente de 0 e 1, por exemplo, $P_M[A] = 0,3$. Então A não poderia ser usado diretamente para condenar o modelo M. Mas a partir de A podemos definir um outro acontecimento com probabilidade muito perto de 0. Vamos repetir a experiência N vezes de forma independente. Com a regra do produto das probabilidades sabemos que a probabilidade de observar na primeira experiência A, na segunda $\neg A$ (não-A), na terceira de novo $\neg A$ na quarta .....e na última A é

$$P_M[A, \neg A, \neg A, ..., A] = P_M[A](1 - P_M[A])(1 - P_M[A]).....P_M[A] \qquad (1.6.1)$$

Se chamarmos $P_M[A]$ de p podemos escrever a equação (1.6.1) da seguinte forma:

$$P_M[A, \neg A, \neg A, ..., A] = p^k (1-p)^{(N-k)} \qquad (1.6.2)$$

onde k é o número de vezes que o resultado A aparece no colchete e (N-k) é o número de vezes que $\neg A$ aparece. Todas as seqüências de resultados que têm o mesmo número k, e portanto diferem apenas na ordem, têm a mesma probabilidade. Existem exatamente

$$\frac{N!}{k!(N-k)!} \qquad (1.6.3)$$

seqüências deste tipo. Se queremos saber qual é a probabilidade de observar k vezes o resultado A temos que multiplicar a probabilidade (1.6.2) pelo número (1.6.3) (regra da soma):

---

[2] $A \wedge B$ significa "A e B".

$$P_M\left[\text{o resultado } A \text{ aparece } k \text{ vezes}\right] \;=\; \frac{N!}{k!(N-k)!}\,p^k\,(1-p)^{(N-k)} \qquad (1.6.4)$$

É praxe definir uma grandeza chamada *freqüência relativa* como o quociente do número de vezes que A é observado e o número total de observações:

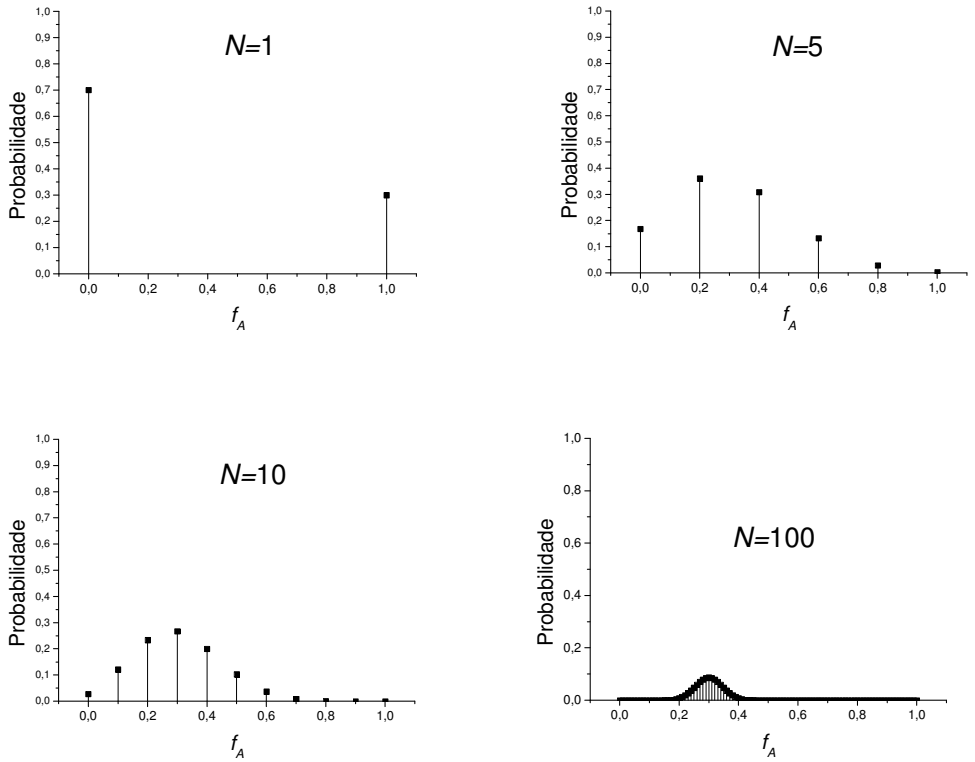
$$f_A = \frac{\text{número de vezes que } A \text{ é observado em } N \text{ experimentos}}{N} \qquad (1.6.5)$$

Com esta grandeza podemos escrever a equação (1.6.4) da seguinte forma:

$$P_M\left[f_A = k/N\right] \;=\; \frac{N!}{k!(N-k)!}\,p^k\,(1-p)^{(N-k)} \qquad (1.6.6)$$

$P_M\left[f_A = k/N\right]$ significa a probabilidade de encontrar o valor da freqüência relativa igual ao número k/N. As figuras 1-4 dão exemplos para as probabilidades $P_M\left[f_A = k/N\right]$ no caso de $p = 0{,}3$ e para $N = 1$, $N = 5$, $N = 10$ e $N = 100$.

Fig. 1-4 Probabilidades para encontrar os possíveis valores da freqüência relativa f_A para uma probabilidade $P_M\left[A\right] = 0,3$ com $N = 1$, 5, 10 e 100 medidas. Para melhorar a visibilidade cada valor é marcado não apenas com um ponto, mas também com uma barra vertical.



Quanto maior o número de experiências maior é o número de possíveis valores da freqüência relativa e as probabilidades dos valores individuais ficam menores, pois a soma de todas as probabilidades deve ser sempre 1. Na medida em que N cresce, a probabilidade se concentra cada vez mais em torno do valor $f_A = 0,3$. Por exemplo, com 100 medidas a probabilidade de achar um valor da freqüência relativa que se afasta por mais de 0,1 do valor 0,3 é muito perto de zero. Então este tipo de acontecimento

pode ser usado para refutar uma hipótese do tipo $P_M[A] = 0,3$. Podemos decidir medir 100 vezes e descartar a hipótese $P_M[A] = 0,3$ se encontrarmos f_A fora do intervalo $[\,0,2 \quad , \quad 0,4\,]$.

O que vimos com este exemplo vale de forma geral: se a probabilidade de um acontecimento A vale p, a probabilidade de encontrar a freqüência relativa f_A em um grande número de medidas longe do valor p é muito pequena. Uma probabilidade não é uma grandeza física, ela é um instrumento da nossa linguagem para formular afirmações sobre os futuros acontecimentos no laboratório. Freqüências relativas são grandezas físicas que podemos medir. Mas, se o nosso modelo teórico for razoável as freqüências relativas não ficarão longe das probabilidades. Se imaginarmos um número infinito de repetições de uma experiência e se o nosso modelo for adequado teríamos

$$P_M[A] = \lim_{N \to \infty} f_A \tag{1.6.7}$$

Esta equação não pode ser usada como definição de probabilidade, ela vale apenas para um modelo muito especial. Claro, este modelo seria o melhor que poderíamos usar, mas em geral não temos tal modelo ideal à nossa disposição. Além disso, a equação (1.6.7) pode ser formulada somente para acontecimentos que podem fazer parte de um ensemble.

Probabilidades Condicionais

Na primeira seção falamos que uma probabilidade $P_M[A]$ perde seu sentido na hora de observar o acontecimento A. Além disso, valores de probabilidades de outros acontecimentos B têm que ser alterados na hora que se registra um acontecimento A. Veremos um exemplo: imagine que estejamos jogando dados e o nosso modelo do dado prevê probabilidades iguais para todas as seis faces do dado. Então a probabilidade de eu obter cinco pontos na próxima jogada é 1/6:

$$B = \text{obter cinco pontos} \quad ; \qquad P_M[B] = 1/6 \tag{1.6.8}$$

Imagine que eu joguei o dado e antes de poder olhar, outro jogador que já viu o resultado me informa que obtive um número ímpar de pontos. Com esta constatação a probabilidade que minha jogada tenha dado cinco pontos deixa de ser 1/6 e passa para 1/3, pois existem apenas três possíveis resultados ímpares. A constatação A que o resultado é ímpar nos obriga a substituir o antigo modelo M por um outro M'. Mas este novo modelo difere do antigo de forma mínima, apenas os resultados pares foram excluídos, a idéia básica da igualdade dos possíveis resultados foi mantida. Por isso vamos escrever a probabilidade não com um índice M' , mas com o mesmo M e anotar que se trata de uma probabilidade alterada pela observação de um acontecimento A escrevendo o A no colchete separado do B por uma barra vertical:

$$P_M[B \mid A] \;=\; \begin{array}{l}\text{probabilidade de } B \text{ de acordo com o modelo } M, \\ \qquad\qquad \text{minimamente modificado pela observação de } A.\end{array} \tag{1.6.9}$$

Este tipo de probabilidade se chama *probabilidade condicional*. Vimos o exemplo do dado perfeitamente simétrico. No caso geral a modificação mínima do modelo consiste simplesmente na renormalização das probabilidades de acontecimentos compatíveis com a observação A:

$$P_M\left[B \mid A\right] = \frac{P_M\left[B \wedge A\right]}{P_M\left[A\right]} \tag{1.6.10}$$

Então em palavras: a probabilidade condicional de observar B dado que A foi observado é a probabilidade de observar o acontecimento (A e B) dividido pela probabilidade do acontecimento A. (**Exercício**: que podemos afirmar sobre $P_M\left[B \mid A\right]$ se A e B forem estatisticamente independentes?).

Thomas Bayes (1702 – 1761) observou um fato muito simples, mas sumamente importante: Caso ambas as probabilidades $P_M\left[A\right]$ e $P_M\left[B\right]$ sejam diferentes de zero, podemos relacionar as duas probabilidades condicionais $P_M\left[B \mid A\right]$ e $P_M\left[A \mid B\right]$. Pois a probabilidade do acontecimento (A e B) pode ser escrita de duas formas:

$$P_M\left[B \wedge A\right] \quad = \quad P_M\left[B \mid A\right] \cdot P_M\left[A\right] \tag{1.6.11}$$

e

$$P_M\left[B \wedge A\right] \quad = \quad P_M\left[A \mid B\right] \cdot P_M\left[B\right] \tag{1.6.12}$$

Então segue

$$\boxed{P_M\left[A \mid B\right] \cdot P_M\left[B\right] \quad = \quad P_M\left[B \mid A\right] \cdot P_M\left[A\right]} \tag{1.6.13}$$

ou

$$\boxed{P_M\left[A \mid B\right] \quad = \quad P_M\left[B \mid A\right] \frac{P_M\left[A\right]}{P_M\left[B\right]}} \tag{1.6.14}$$

Usaremos este resultado na análise de dados experimentais. Mas antes vamos citar um exemplo tomado do livro de B. W. Lindgren "Statistical Theory" que ilustra o poder deste resultado de Bayes de forma excelente:

Imagine que certa doença costuma ocorrer na população com freqüência de uma pessoa doente em cada 1000 habitantes. Imagine também que exista certo tipo de exame de sangue para diagnosticar esta doença. Experimentos de laboratório mostram que este exame de sangue fornece um resultado positivo para uma pessoa com a doença com probabilidade 0,99. Mas existe também a possibilidade de falso positivo: quando o exame é aplicado numa pessoa sadia, o exame de sangue dá um resultado positivo com probabilidade 0,05. Agora um paciente fez o exame de sangue e o resultado foi positivo. Qual é a probabilidade que este paciente tem a doença?

Vamos escrever os acontecimentos de obter um resultado positivo (negativo) como + (-) e de ter um paciente sadio (doente) com S (D). Temos então

$$P\left[D\right] = 0{,}001 \qquad P\left[+ \mid D\right] = 0{,}99 \qquad P\left[+ \mid S\right] = 0{,}05$$

Queremos calcular $P\left[D \mid +\right]$. Com a equação (1.6.14) temos

$$P\left[D \mid +\right] \quad = \quad P\left[+ \mid D\right] \frac{P\left[D\right]}{P\left[+\right]} \tag{1.6.15}$$

Falta apenas o dado $P\left[+\right]$. Podemos calcular esta probabilidade da seguinte forma:

$$
\begin{aligned}
P[+] &= P[+\mid D]\cdot P[D]+P[+\mid S]\cdot P[S]= \\
&= P[+\mid D]\cdot P[D]+P[+\mid S]\cdot\left(1-P[D]\right)= \\
&= 0,99\cdot 0,001 \;+\; 0,05\cdot 0,999 \;=\; 0,05094
\end{aligned}
\tag{1.6.16}
$$

Inserindo isto em (1.6.15) obtemos:

$$
P\left[D\mid +\right] \;=\; 0.99\cdot\frac{0,001}{0,05094}=0,019.... \tag{1.6.17}
$$

Então apesar de termos um exame de sangue aparentemente bastante confiável, a probabilidade de um paciente com resultado positivo estar doente é apenas 1,9%.

Variáveis estocásticas e o Teorema Central de Limite[3]

Uma variável estocástica é uma grandeza física G para a qual temos algum modelo teórico que determina probabilidades para o aparecimento dos valores de G num experimento.

Para começar com os casos mais simples, vamos considerar primeiramente uma grandeza não contínua. Por exemplo, G poderia ser o número de pontos na face de um dado. Chamar G de variável estocástica significa que estamos supondo que tenhamos alguma opinião formada sobre as probabilidades com as quais os valores de G aparecem numa experiência. Então podemos associar a cada um dos possíveis valores de G uma probabilidade. Vamos escrever os valores de G com a letra minúscula g para distingui-los da própria grandeza. A probabilidade $P_M\left[\text{observaremos o valor } g\right]$ será escrita como p_g. Vale naturalmente que a soma de todas estas probabilidades é 1:

$$
\sum_g p_g = 1 \tag{1.6.18}
$$

onde a soma se estende sobre os possíveis valores de g. Os seguintes valores resultam ser úteis para a caracterização do modelo: o valor esperado de G:

$$
\langle G\rangle_M \;\underset{def.}{=}\; \sum_g g\, p_g \tag{1.6.19}
$$

e a variância de G:

$$
\mathrm{var}_M\left(G\right) \;\underset{def.}{=}\; \sum_g \left(g-\langle G\rangle_M\right)^2 p_g \tag{1.6.20}
$$

O valor esperado $\langle G\rangle_M$ não é necessariamente um dos possíveis valores da grandeza G. Por exemplo, no caso do número de pontos na face de um dado e com o modelo que atribui probabilidades iguais a todas as faces do dado, o valor esperado vale 3,5. Este valor nunca aparecerá num jogo de dados. O papel do valor esperado é de caracterizar, de forma bastante incompleta, onde estão as probabilidades. Se o modelo atribui maiores probabilidades para os valores grandes de G o valor esperado estaria na região dos valores grandes. O papel da variância é de caracterizar o quanto as probabilidades estão espalhadas. Quanto mais concentrada for a probabilidade numa região pequena,

[3] Este teorema foi descoberto em 1733 pelo matemático Abraham de Moivre e divulgado pelo matemático Laplace no livro "Théorie Analytique de Probabilités". Em francês o nome do teorema é "Théorème Central Limite" e em inglês "Central Limit Theorem", porque o teorema tem um papel central e muito importante no ramo da estatística. Alguns autores Brasileiros chamam este teorema erroneamente de "teorema do limite central". O que é central é o teorema, não o limite!

menor será a variância. Por exemplo, vamos considerar a freqüência relativa como grandeza. As figuras 1-4 mostram quatro atribuições diferentes de probabilidade. Todas as quatro têm o mesmo valor esperado e este vale 0,3. As variâncias são diferentes, a figura 1 mostra a maior variância e a 4 a menor.

Exercício: mostre que $\mathrm{var}_M(G) = \langle G^2 \rangle_M - \langle G \rangle_M{}^2$.

Se a grandeza G for contínua, a probabilidade de observar um exato valor de G em geral é zero. Então não faz muito sentido fazer afirmações sobre probabilidades de determinados valores. Mas, o que faz sentido é falar de probabilidades de observar um valor de G dentro de um intervalo (se G for unidimensional). Por exemplo, imagine que pegamos um grão de feijão de um saco de feijão e colocamos este grão numa balança. Então nesta situação podemos fazer afirmações do seguinte tipo: "a probabilidade de encontrar um valor de massa entre 0,15g e 0,16 g vale 0,02" . Claro, o valor 0,02 é apenas um exemplo; qual valor que deve aparecer depende do modelo teórico que vamos adotar. Então podemos usar probabilidades do seguinte tipo:
$P_M\left[\text{o valor observado estará no intervalo } [g_1, g_2]\right]$. O modelo probabilístico consistiria de todas as probabilidades deste tipo. Como há uma infinidade de diferentes intervalos, esta maneira de descrever o modelo não parece ser muito prática. Os físicos preferem uma outra forma de descrever a atribuição de probabilidades, que se assemelha mais aos p_g que usamos no caso de variáveis discretas. Podemos definir uma função que vamos chamar de *densidade de probabilidade da variável estocástica G*:

$$\rho_{G,M}(g) = \lim_{\tilde{g} \to g} \frac{P_M\left[\text{o valor observado estará no intervalo } [g, \tilde{g}]\right]}{\tilde{g} - g} \qquad (1.6.21)$$

A densidade de probabilidade é obtida a partir da probabilidade através de uma divisão – uma divisão sofisticada que envolve um limite e que corresponde exatamente à noção de derivada. Inversamente, podemos recuperar a probabilidade através de uma multiplicação. Mas tem que ser também uma multiplicação sofisticada. Esta multiplicação sofisticada é a integração:

$$P_M\left[\text{o valor observado estará no intervalo } [g_1, g_2]\right] = \int_{g_1}^{g_2} \rho_{G,M}(g)\, dg \qquad (1.6.22)$$

A probabilidade de observar algum valor qualquer de G é naturalmente 1 e consequentemente temos

$$\int_{-\infty}^{\infty} \rho_{G,M}(g)\, dg = 1 \qquad (1.6.23)$$

Esta condição é análoga da (1.6.18) . Apenas a soma transformou-se numa integral. Analogamente vamos definir o valor esperado e a variância

$$\langle G \rangle_M \underset{Def.}{=} \int_{-\infty}^{\infty} g\, \rho_{G,M}(g)\, dg \qquad (1.6.24)$$

$$\mathrm{var}_M(G) \underset{Def.}{=} \int_{-\infty}^{\infty} \left(g - \langle G \rangle_M\right)^2 \rho_{G,M}(g)\, dg \qquad (1.6.25)$$

Estas definições serão de grande utilidade na formulação do teorema central de limite.

Agora veremos de que trata este teorema e por que ele é tão central para o trabalho com probabilidades. As regras que descrevemos até agora fornecem relações entre

probabilidades de um dado modelo e um critério que permite descartar modelos. Mas não temos ainda nada para orientar o pesquisador na hora de inventar ou formular um modelo. Então estamos na situação de uma pessoa que aprendeu quando se deve jogar uma comida estragada fora, mas que não aprendeu onde se compra comida. Bem, em princípio podemos simplesmente dizer: o pesquisador pode inventar qualquer modelo e a regra da condenação de modelo vai eliminar aqueles modelos inapropriados de tal forma que sobram somente modelos bons. Mas este procedimento de tentativa e erro não levaria a resultado algum devido à enorme variabilidade de modelos. Por exemplo, para uma variável estocástica contínua temos uma infinidade de possíveis funções de densidade de probabilidade $\rho_{G,M}$. O teorema central de limite nos ensina a formar certas variáveis estocásticas para as quais a densidade de probabilidade não tem tanta variabilidade. Para estas variáveis estocásticas existe apenas um único tipo de função $\rho_{G,M}$, que depende de apenas dois parâmetros. Então para estas variáveis estocásticas é possível encontrar um modelo adequado mesmo usando apenas o método de tentativa e erro, pois neste caso a tarefa é apenas de encontrar os valores apropriados destes dois parâmetros. Fora do método de tentativa e erro existem também outros métodos para formar modelos, mas para estes também é útil ter que cuidar apenas de dois parâmetros.

Seja G uma variável estocástica contínua e $\rho_{G,M}$ a densidade de probabilidade para esta variável. No momento não vamos nos preocupar com a questão da determinação do $\rho_{G,M}$. Esta função pode ser qualquer densidade de probabilidade, a única hipótese que faremos sobre $\rho_{G,M}$ é que esta função caia para zero suficientemente rápido quando $g \to \pm\infty$ de tal forma que a variância de G exista. A partir da grandeza G vamos formar uma outra grandeza ${}^{N}\overline{G}$ que é definida para N repetições de medidas independentes de G. O valor de ${}^{N}\overline{G}$ é simplesmente a média dos N resultados das medidas de G:

$$ {}^{N}\overline{g} \underset{Def.}{=} \frac{1}{N}\{g_1 + g_2 + g_3 + ... + g_N\} \tag{1.6.26} $$

Com a regra do produto de probabilidades e a regra da soma, pode-se calcular a densidade de probabilidade da grandeza ${}^{N}\overline{G}$ a partir da densidade de probabilidade da grandeza G. O teorema central de limite afirma que, independente de qual era a função $\rho_{G,M}$, a densidade da variável ${}^{N}\overline{G}$ é essencialmente sempre a mesma coisa desde que N seja suficientemente grande. Apenas o valor esperado de G e a variância de G entram na função $\rho_{{}^{N}\overline{G},M}$. Então para grandes valores de N temos sempre

$$ \boxed{\rho_{{}^{N}\overline{G},M}\left({}^{N}\overline{g}\right) \approx \frac{1}{\sqrt{2\pi}\,\sigma}\exp\left\{-\frac{\left({}^{N}\overline{g}-\mu\right)^2}{2\sigma^2}\right\}} \tag{1.6.27} $$

onde

$$ \sigma = \frac{1}{\sqrt{N}}\sqrt{\mathrm{var}_M\left(G\right)} \tag{1.6.28} $$

e

$$ \mu = \langle G \rangle_M \tag{1.6.29} $$

A função $\exp\{x\}$ é a exponencial $e^x = \lim_{n\to\infty}\left(1+\frac{x}{n}\right)^n = \sum_{n=0}^{\infty}\frac{x^n}{n!}$. A notação com "exp" é frequentemente usada quando o x é uma expressão complicada que ficaria um tanto ilegível no expoente da exponencial. O "exp" não deve ser confundido com a tecla "EXP" das calculadoras! O "EXP" na calculadora serve para escrever números com fator de uma potência de 10 e não tem nada a ver com a função exponencial. A figura 5 mostra uma densidade de probabilidade do tipo da equação (1.6.27). Este tipo de curva é chamado de *distribuição normal* ou *Gaussiana*.

Aplicando as definições (1.6.24) e (1.6.25) na grandeza ${}^N\overline{G}$ pode-se mostrar (**exercício**) que o parâmetro μ é o valor esperado da grandeza ${}^N\overline{G}$ e o parâmetro σ é a raiz quadrada da variância de ${}^N\overline{G}$. O σ determina a largura do pico.

$$\left\langle {}^N\overline{G}\right\rangle = \mu \tag{1.6.30}$$

$$\mathrm{var}_M\left({}^N\overline{G}\right) = \sigma^2 \tag{1.6.31}$$

Como podemos ver pela equação(1.6.28) a variância de ${}^N\overline{G}$ é menor que a variância de G:

$$\boxed{\mathrm{var}_M\left({}^N\overline{G}\right) = \frac{1}{N}\mathrm{var}_M\left(G\right)} \tag{1.6.32}$$

Esta diminuição da variância é a razão pela qual costumamos medir grandezas muitas vezes e calcular as médias.

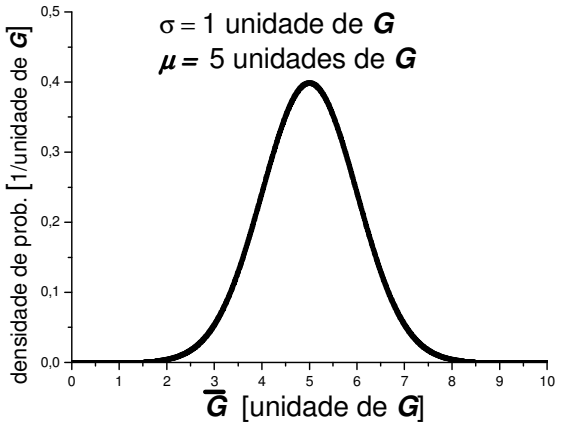


Fig. 5 Exemplo de uma densidade de probabilidade Gaussiana de uma grandeza qualquer $\overline{G}$ com valor esperado $\langle\overline{G}\rangle = 5$ unidades de G e variância $\mathrm{var}\left(\overline{G}\right) = 1\ \left(\text{unidade de } G\right)^2$.

É útil conhecer alguns valores de probabilidade calculados com uma densidade de probabilidade Gaussiana: A probabilidade de encontrar o valor da grandeza ${}^N\overline{G}$ dentro do intervalo $[\mu-\sigma,\mu+\sigma]$ vale aproximadamente 0,6826. A figura 6 mostra a correspondente área abaixo da curva. A probabilidade se aproxima rapidamente ao valor 1 quando aumentarmos o tamanho do intervalo simetricamente:

Tabela 1:

Intervalo ***I***	Probabilidade de encontrar o valor da grandeza no intervalo ***I***
$[\mu-\sigma,\mu+\sigma]$	0,6826...

$[\mu-2\sigma,\mu+2\sigma]$	0,9544...
$[\mu-3\sigma,\mu+3\sigma]$	0,9974...

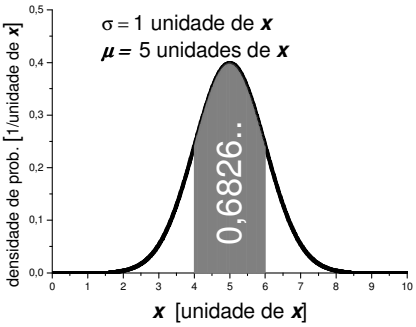


Fig. 6 Densidade de probabilidade Gaussiana de uma grandeza qualquer com marcação da área correspondente ao intervalo $[\mu-\sigma,\mu+\sigma]$.

A formulação exata do teorema com um limite $N \to \infty$ requer ainda alguns cuidados. Mas, estes são apenas detalhes técnicos. Aqui estamos mais interessados no caso prático que terá sempre um número N finito. Apesar da demonstração deste teorema não ser difícil, não nos preocuparemos com esta demonstração no primeiro semestre para não sobrecarregar o curso com matemática. No lugar de uma demonstração matemática veremos como este resultado funciona na prática.

Sabemos que as freqüências relativas de muitas medidas muito provavelmente coincidem com as probabilidades. Então temos boas chances de testar o teorema central de limite observando freqüências relativas. Para isto faremos as seguintes medidas:

Dois grupos de alunos fazem medições de massa de grãos de feijão e depois juntamos os resultados dos grupos.

Grupo A: Tire aleatoriamente 60 grãos de feijão de um saco e determine a massa de cada grão. (Podemos usar duas balanças em cada grupo para dar conta do número grande de medidas)

Grupo B: Tire aleatoriamente 60 grupos de 25 grãos de feijão do mesmo saco e determine a massa de cada grupo. Divida a massa de cada grupo por 25 para obter 60 valores de médias de massa ${}^{25}\overline{m}$.

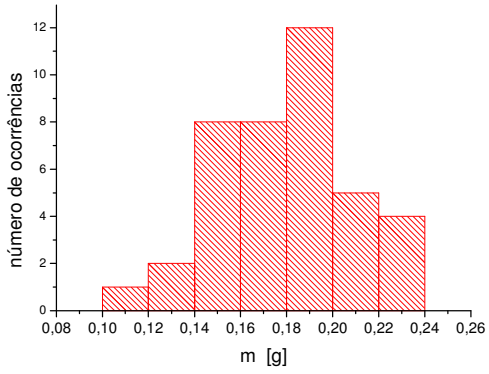


Fig. 7 Histograma de massas de 40 grãos de feijão.

Com os dois conjuntos de dados, isto é, 60 valores de massas individuais m_k e 60 valores de médias ${}^{25}\overline{m}_k$, $k=1,...,40$, faremos dois *histogramas*. Veremos com o exemplo das massas individuais o que é um histograma. Para elaborar um histograma das medidas das massas, dividimos o eixo das massas em intervalos de mesmo tamanho. Para o caso das massas individuais, uma boa divisão são intervalos do tamanho $0{,}02\text{g}$. Para cada intervalo contamos quantas medidas caíram dentro do intervalo e desenhamos este número num gráfico com uma coluna sobre do intervalo. A figura 7 dá um exemplo de um histograma feito a partir de 40 medidas de massa de grãos de feijão.

A tabela mostra os dados que deram origem a este histograma. Para dar um exemplo, os dados com valores no intervalo $[0,20\text{g}\,,\ 0,22\,\text{g})$ estão marcados. Temos 5 dados neste intervalo e a coluna montada sobre este intervalo tem a altura 5.

Tabela2: Dados correspondente ao histograma da figura 7:

	m [g]			m [g]			m [g]			m [g]
1	0,169		11	0,192		21	0,183		31	0,164
2	0,158		12	0,182		22	0,155		32	0,107
3	0,132		13	0,196		**23**	**0,208**		33	0,239
4	0,157		14	0,188		24	0,188		**34**	**0,200**
5	0,162		15	0,223		25	0,235		35	0,184
6	0,150		16	0,169		26	0,137		36	0,151
7	**0,203**		17	0,171		27	0,166		37	0,150
8	0,143		18	0,177		**28**	**0,215**		38	0,181
9	0,156		19	0,197		29	0,188		39	0,231
10	0,193		**20**	**0,202**		30	0,166		40	0,189

Se dividíssemos as alturas das colunas pelo número total de medidas (40), obteríamos um gráfico das freqüências relativas de ocorrências para cada intervalo. Estas freqüências relativas provavelmente não estariam longe das probabilidades de observar uma massa dentro do intervalo correspondente, de acordo com o modelo que descreve as massas dos grãos idealmente.

Vocês devem fazer um histograma para as 60 massas de grãos individuais e um para as 60 médias de massas ${}^{25}\overline{m}$. Use para os intervalos os seguintes tamanhos:

Tabela 3:

Grandeza	**T**amanho de **I**ntervalo recomendado (TI)
m	0,02g
${}^{25}\overline{m}$	0,004g

Se não tivermos muito azar, a forma do histograma para a grandeza ${}^{25}\overline{m}$ deve ter aproximadamente a forma de uma Gaussiana. Veremos se podemos realmente descrever este histograma por uma curva do tipo (1.6.27). Temos que fazer uma adaptação de escala já que a curva (1.6.27) é normalizada tal que a área abaixo dela vale 1 enquanto a área embaixo do histograma vale 60 vezes o tamanho do intervalo (0,004g). Então temos que multiplicar a função (1.6.27) por 60xTI (60 vezes 0,004g) para poder comparar a curva com o histograma no mesmo gráfico. A questão é: quais valores dos parâmetros μ e σ devemos usar? Poderíamos apelar para o método de tentativa e erro. Mas existem métodos que levam mais rapidamente a resultados satisfatórios. Os seguintes valores de μ e σ devem funcionar bem:

$$\mu = \frac{1}{60}\left({}^{25}\overline{m}_1 + {}^{25}\overline{m}_2 + {}^{25}\overline{m}_3 + \ldots\ldots\ldots\ldots + {}^{25}\overline{m}_{60}\right) = \frac{\sum_{k=1}^{60} {}^{25}\overline{m}_k}{60} = {}^{60}\overline{{}^{25}\overline{m}} \qquad (1.6.33)$$

e

$$\sigma = \sqrt{\frac{\sum_{k=1}^{60}\left({}^{25}\overline{m}_k - \mu\right)^2}{59}} \qquad (1.6.34).$$

Com estes valores dos parâmetros desenhe a curva

$$y(x) \;=\; 60\cdot(\mathrm{TI})\cdot\frac{1}{\sqrt{2\pi}\,\sigma}\exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\} \qquad (1.6.35)$$

na mesma folha do histograma da grandeza ${}^{25}\overline{m}$ e verifique visualmente se ela se adapta bem ao histograma. É recomendável fazer o histograma e esta curva com o computador.

Repare que a largura do histograma das massas individuais é muito maior que a largura do histograma das médias ${}^{25}\overline{m}$. A largura do histograma das médias podemos caracterizar pelo parâmetro σ. Analogamente podemos definir um parâmetro que caracteriza a largura do histograma das massas individuais:

$$\sigma^{\text{Individual}} = \sqrt{\frac{\sum_{k=1}^{60}\left(m_k - {}^{60}\overline{m}\right)^2}{59}} \qquad (1.6.36)$$

Vamos chamar esta grandeza de desvio padrão das 60 massas individuais. Verifique se o parâmetro σ da equação (1.6.34) é aproximadamente $\frac{1}{\sqrt{25}} = \frac{1}{5}$ do valor de $\sigma^{\text{Individual}}$. O desvio padrão (1.6.36) é uma raiz quadrada de uma média dos desvios quadráticos da média. O que naturalmente chama a atenção é o fato que nesta média dividimos por 59 e não por 60. Isto parece um tanto estranho. Mas repare que nas 60 parcelas da soma no numerador somente 59 termos são independentes por causa da presença da média.

Aplicação no tratamento de erros experimentais

Imagine que temos uma balança e dois corpos metálicos a e b. Colocando estes corpos na balança observamos as massas $m_a = 2{,}034\,\text{g}$ e $m_b = 3{,}011\,\text{g}$. Colocando os dois corpos juntos na balança observamos um valor $M = 5{,}047\,\text{g}$. Notamos uma inconsistência, pois o valor de M difere da soma de m_a e m_b por 0,002 g. Então concluímos que existe erro experimental nestas medidas. Mas não sabemos qual das três medidas está errada. Agora repetimos as três medidas e obtemos não apenas mais uma inconsistência de soma, mas, além disso, três valores diferentes: $m_a' = 2{,}033\,\text{g}$, $m_b' = 3{,}012\,\text{g}$ e $M' = 5{,}046\,\text{g}$. Os corpos são de um metal que não oxida e nem gasta facilmente, digamos de aço inoxidável, e as massas certamente não mudaram. Então além de termos erros experimentais, estes erros não são sempre os mesmos. Esta situação descrita é típica para medidas. Mas ela é um pouco complicada porque ela envolve três medidas e não sabemos como os erros são distribuídos entre as três. Para analisar o erro experimental vamos imaginar uma situação mais simples, porém infelizmente menos comum na prática:

Imagine que temos um objeto O e dois aparatos A e R para medir alguma grandeza X que é propriedade deste objeto. R é um instrumento de referência que mede com uma precisão muito melhor que o aparato A. Vamos supor que os erros de R são tão pequenos que podemos considerar os resultados obtidos com este instrumento como valores exatos. Claro, na verdade eles não são exatos, mas enquanto tratarmos da análise dos dados do aparato A os resultados de R podem ser contados como exatos. Agora imagine que medimos o valor de X com o aparato de referência R obtendo um valor x_V, que vamos chamar o verdadeiro valor de X. Depois medimos ainda a grandeza X do objeto O com o aparato A um número imenso de vezes. Vamos chamar o k-esimo valor medido com o aparato A de x_k. A diferença $\delta x_k = x_k - x_V$ seria o erro da k-esima medida. Imagine que dividirmos o eixo x em intervalos muito pequenos e que o número de medidas é tão grande que mesmo com intervalos muito pequenos o número de vezes que os resultados x_k caem em cada um dos intervalos é ainda grande. Desta forma podemos construir um bom histograma que, depois de dividir pelo número total de medidas e pelo tamanho dos intervalos, fornece aproximadamente uma densidade de probabilidade $\rho_{X\,\infty;x_V}(x)$ que serve para calcular probabilidades de obter certos resultados quando se mede X do objeto O com o aparato A:

$$P_{\infty;x_V}\left[\text{a próxima medida de } X \text{ estará em } [x_1, x_2]\right] = \int_{x_1}^{x_2} \rho_{X\,\infty;x_V}(x)\,dx \qquad (1.6.37)$$

Chamamos o modelo desta previsão de "$\infty; x_V$" porque ele resultou da nossa experiência passada com um número imenso (quase infinito) de medidas e além disso ele se refere a medidas com um objeto cujo verdadeiro valor e X é x_V. A figura 8 mostra um exemplo de como esta distribuição de probabilidades poderia ficar numa situação real.

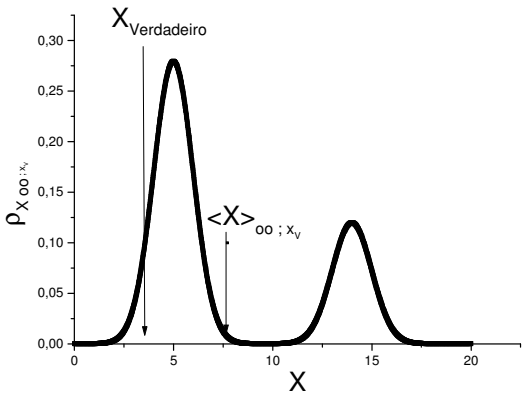


Fig. 8 Exemplo de uma distribuição de probabilidade que resultou de um grande número de medidas com o aparato A.

Na figura 8, que é nada mais que um exemplo hipotético, estão indicados também o valor verdadeiro de X (medido com o aparato R) e o valor esperado das medidas com A de acordo com o modelo "$\infty; x_V$". O valor verdadeiro em geral não coincide com o valor esperado de X. É prudente escrever o erro $\delta x_k = x_k - x_V$ de uma medida com o aparato A como soma de duas parcelas:

$$\delta x_k = x_k - x_V \quad = \quad \left[\langle x\rangle_\infty - x_V\right] + \left[x_k - \langle x\rangle_\infty\right] \qquad (1.6.38)$$

A primeira parcela, $\left[\langle x\rangle_\infty - x_V\right]$, é igual para todas as medidas do ensemble. Ela é chamada de *erro sistemático*. A segunda parcela, $\left[x_k - \langle x\rangle_\infty\right]$, é chamada de *erro estatístico*. Esta parcela tem o valor esperado zero. Esta separação do erro em duas partes se justifica porque na tarefa de combater os erros, na tentativa de minimizá-los,

cada uma destas parcelas requer um tratamento diferente. Ambas as parcelas podem ser reduzidas fazendo melhorias no aparato de medida. A parcela do erro estatístico pode ser diminuída ainda de uma outra forma, utilizando o teorema central de limite. Para fazer isto basta medir muitas vezes com o aparato A e calcular a média dos valores. A figura 9 mostra como ficaria a densidade de probabilidade da média de 64 medidas daquela grandeza X cuja densidade de probabilidade está mostrada na figura 8. Dos dois picos separados não sobrou nada. Temos apenas um único pico Gaussiano muito estreito na posição do valor esperado de X.

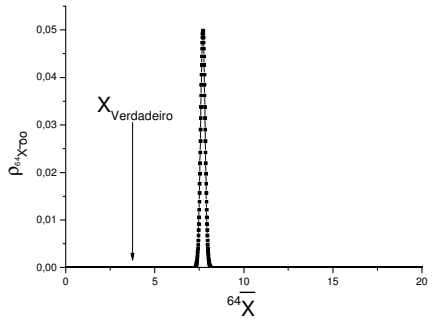


Fig.9 Densidade de probabilidade da grandeza $^{64}\overline{X}$

A grandeza $^{N}\overline{X}$ tem o mesmo erro sistemático que a grandeza X, pois o valor esperado de $^{N}\overline{X}$ é igual ao valor esperado de X. Mas o erro estatístico numa determinação de $^{N}\overline{X}$ é muito provavelmente muito pequeno. Se X, tinha no modelo "$\infty; x_V$", uma variância $\mathrm{var}_{\infty;x_V}(X)$, a probabilidade de achar um valor da média $^{N}\overline{X}$ fora do intervalo

$$\left[\langle X\rangle_{\infty;x_V} - \frac{3}{\sqrt{N}}\mathrm{var}_{\infty;x_V}(X) \quad , \quad \langle X\rangle_{\infty;x_V} + \frac{3}{\sqrt{N}}\mathrm{var}_{\infty;x_V}(X)\right]$$

seria apenas $0,0026 = 1 - 0,9974$.

É conveniente resumirmos neste ponto o que fizemos com o nosso exemplo dos dois aparatos A e R até agora: introduzimos uma densidade de probabilidade $\rho_{X\,\infty;x_V}$ que serve perfeitamente bem para descrever um ensemble infinito de medidas com o aparato A. Com esta densidade de probabilidade podemos calcular as probabilidades de observar certos valores numa próxima medida com o aparato A. Usamos ainda esta densidade de probabilidade para definir o erro sistemático e o erro estatístico. Finalmente consideramos a média de N medidas com o aparato A e vimos que esta grandeza tem o mesmo erro sistemático que X, mas muito provavelmente um erro estatístico bem pequeno.

Infelizmente esta situação não é aquela que enfrentamos normalmente num laboratório. Vários detalhes são diferentes na prática:

1) Normalmente temos somente o aparato A à nossa disposição e não o aparato de referência R.

2) Não temos tempo para um número gigantesco de medidas que permitiriam determinar a densidade $\rho_{X\,\infty;x_V}$.

3) As densidades $\rho_{X\,\infty;x_V}$ e $\rho_{^{N}\overline{X}\,\infty;x_V}$ permitem falar das probabilidades da próxima observação de X com o aparato A e da próxima média de N medidas com o aparato A. Mas o que nos interessa não é isto, o próximo valor não importa. <u>O que queremos realmente saber é qual é o verdadeiro valor.</u>

Então a situação real é tipicamente a seguinte: medimos uma vez ou N vezes o valor de X com o aparato A e baseado neste resultado ou nestes resultados gostaríamos de fazer alguma previsão sobre qual seria o valor obtido se alguém entrasse de repente no nosso

laboratório com aquele aparato maravilhoso R e medisse o verdadeiro valor. Claro, qualquer previsão sobre um resultado deste tipo tem que ser incerta, já que ela seria baseada nos resultados de medidas feitas com o aparato A que é de baixa qualidade. Então a linguagem adequada para estas previsões é naturalmente uma de probabilidades. No lugar das densidades $\rho_{X\,\infty;x_V}$ e $\rho_{N\,\overline{X}\,\infty;x_V}$, temos que tratar de uma densidade de probabilidade para o verdadeiro valor: $\rho_{X_V\,M}$. O modelo M seria baseado na medida ou nas N medidas feitas com o aparato A. É importante notar que as probabilidades calculadas com $\rho_{X_V\,M}$ **não** correspondem a freqüências relativas! A medida de X com o aparato de referência R, que eventualmente entrou com muita sorte no nosso laboratório, não faz parte de um ensemble! Não porque o aparato R é tão caro que o dono queira emprestá-lo somente uma vez. Mesmo que o aparato pudesse ser usado muitas vezes, a medida de X com R daria numa repetição simplesmente o mesmo valor e nada parecido com a função $\rho_{X_V\,M}$. Se quiséssemos considerar o valor verdadeiro como membro de um ensemble, teríamos que considerar um ensemble de objetos O', O'',... semelhantes ao objeto O e um ensemble de aparatos A', A'', ... semelhantes ao aparato A. Geralmente isto não é factível de uma forma bem definida. De fato, na prática usamos aqui o conceito de probabilidade sem relação com uma freqüência relativa.

A grande questão é: qual função densidade $\rho_{X_V\,M}$ escolher para fazer nossa previsão? Vamos tratar deste problema por partes. Primeiramente vamos supor uma situação ainda pouco complicada: imagine que alguém teve acesso ao instrumento de referência e fez anteriormente inúmeras medidas com os aparatos R e A com vários objetos e nos forneceu funções densidade $\rho_{X\,\infty;x_V}$ para diversos valores de x_V. Agora não temos mais acesso ao aparato de referência R e fazemos uma medida da grandeza X com o aparato A de um único objeto O. Queremos uma previsão probabilística sobre o verdadeiro valor de X deste objeto. Neste momento é conveniente que abandonemos o conceito de densidade de probabilidade e usemos apenas probabilidades. Consideramos as medidas com R como exatas. Mas na verdade o aparato de referência também tem um pequeno erro ε. Por exemplo, se X for a grandeza massa, imagine que os erros de A estejam na faixa de 0,01g e os erros de R são, digamos, na ordem de $\varepsilon = 0{,}000001$ g. Podemos dividir a escala de valores da grandeza X em intervalos de tamanho ε. Desta forma podemos trabalhar com probabilidades $\varepsilon \cdot \rho_{X\,\infty;x_V}(x)$. Esta probabilidade é uma probabilidade condicional:

$$P_A\left[\text{resultado} = x \mid \text{valor verdadeiro} = x_V\right] = \varepsilon \cdot \rho_{X\infty;x_V}(x) \qquad (1.6.39)$$

É a probabilidade de encontrar com o aparato A um resultado no intervalo de tamanho ε em torno do valor x, dado que o verdadeiro valor de X está no intervalo de tamanho ε em torno do valor x_V. O que queremos é a probabilidade condicional de ter o verdadeiro valor de X num intervalo de tamanho ε em torno do valor x_V, dado que o valor observado com o aparato A está no intervalo de tamanho ε em torno do valor x. Este é o típico problema que requer o método de Bayes. A equação (1.6.14) fornece esta probabilidade:

$$P_A\left[\text{v. verd.} = x_V \mid \text{res.} = x\right] = \frac{P\left[\text{v. verd.} = x_V\right]}{P\left[\text{res.} = x\right]} P_A\left[\text{res.} = x \mid \text{v. verd.} = x_V\right] \qquad (1.6.40)$$

O quociente que aparece nesta equação contem as probabilidades de ter um verdadeiro valor $x_V \pm \varepsilon/2$ e de obter um resultado de medida com o aparato A igual a $x \pm \varepsilon/2$. Que probabilidades são estas? Bem, qualquer probabilidade é uma opinião. $P[\text{v. verd.} = x_V]$ expressa nossa opinião sobre o verdadeiro valor antes de medir e $P[\text{res.} = x]$ nossa opinião sobre qual valor mediremos. A questão é: que opinião devemos formar? Imagine que X é uma massa e o objeto é um lápis. Certamente a massa do lápis é positiva e menor que 500 g. Mas dentro do intervalo [0g, 500g] que opinião devemos formar? Não sabemos nada sobre o lápis. O que nos salva nesta situação é o fato que este quociente vem multiplicado por uma função $P_A[\text{res.} = x \mid \text{v. verd.} = x_V]$ (visto como uma função de x_V) que é diferente de zero num pequeno intervalo que tem apenas o tamanho da imprecisão do aparato A, digamos algo da ordem de 0,01 g. Num intervalo deste tamanho nossas opiniões sobre o verdadeiro valor é praticamente constante. Desta forma podemos substituir a equação (1.6.40) pela seguinte equação:

$$P_A[\text{v. verd.} = x_V \mid \text{res.} = x] = \text{CONST.} \times P_A[\text{res.} = x \mid \text{v. verd.} = x_V] \qquad (1.6.41)$$

O valor da constante é naturalmente determinado pela condição de normalização:

$$P_A[\text{v. verd.} = x_V \mid \text{res.} = x] = \frac{P_A[\text{res.} = x \mid \text{v. verd.} = x_V]}{\sum_{\xi} P_A[\text{res.} = x \mid \text{v. verd.} = \xi]} \qquad (1.6.42)$$

Agora podemos voltar para a linguagem das densidades de probabilidade e temos o resultado desejado:

$$\rho_{X_V M}(x_V) = \frac{\rho_{X\infty;x_V}(x)}{\int_{-\infty}^{\infty} \rho_{X\infty;\xi}(x)\, d\xi} \qquad (1.6.43)$$

Em geral esta densidade de probabilidade para o valor verdadeiro tem uma forma bem diferente da densidade de probabilidade dos valores medidas com o aparato A. Mas existe um caso no qual a relação entre $\rho_{X\infty;x_V}(x)$ e $\rho_{X_V M}(x_V)$ é relativamente simples: se as densidades $\rho_{X\infty;x_V}(x)$ para diferentes valores de x_V diferem simplesmente por uma translação sem deformação da curva. Neste caso podemos escrever $\rho_{X\infty;x_V}(x)$ como uma função da diferença $x - x_V$:

$$\text{(caso simples):} \qquad \rho_{X\infty;x_V}(x) = f(x - x_V) \qquad (1.6.44)$$

Este caso simples ocorre quando os erros do aparato não dependem do valor verdadeiro. Neste caso a densidade $\rho_{X_V M}(x_V)$ descreveria uma curva que é a espelhada da curva de $\rho_{X\infty;x_V}(x)$ como se o eixo x corresse no sentido oposto. A figura 10 mostra a densidade $\rho_{X_V M}(x_V)$ correspondente á curva da figura 8 supondo que tenhamos o caso simples de erros independentes do valor verdadeiro.

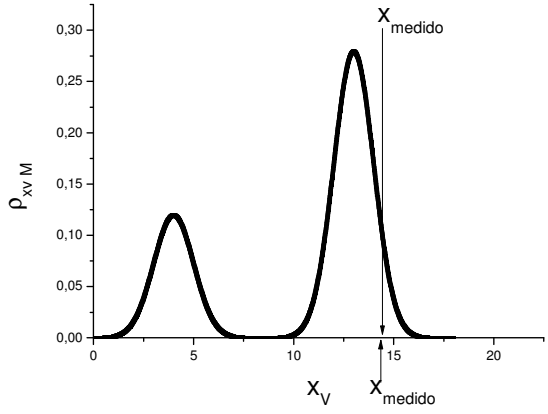


Fig. 10 Densidade de probabilidade para o verdadeiro valor baseada numa medida com o aparato A (compare a figura 8) e supondo erros independentes do valor verdadeiro.

Percebemos que o erro sistemático não apresenta nenhum problema, desde que ele for conhecido. O erro estatístico pode ser reduzido com o método de tomar médias. A mesma transformação de densidade de probabilidade da próxima medida para a densidade de probabilidade do verdadeiro valor pode ser feita com a densidade de probabilidade da média de N medidas: dado que a média de N medidas com o aparato A deram o valor ${}^{N}\overline{x}$ a densidade de probabilidade do verdadeiro valor é dado por

$$\rho_{X_V\,M}\left(x_V\right)=\frac{\rho_{{}^{N}\overline{X}\,\infty;x_V}\left({}^{N}\overline{x}\right)}{\int_{-\infty}^{\infty}\rho_{X\,\infty;\xi}\left({}^{N}\overline{x}\right)d\xi} \tag{1.6.45}$$

Se os erros não dependem do valor verdadeiro esta curve seria outra vez uma Gaussiana e podemos determinar o verdadeiro valor com boa precisão (compare a figura 11)

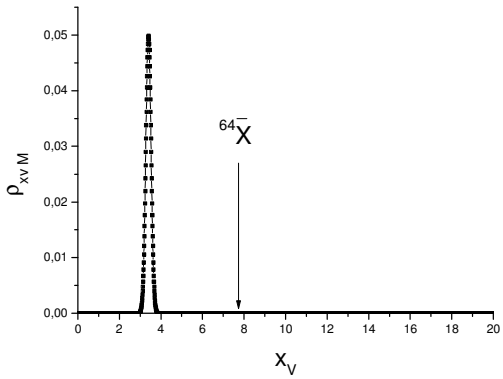


Fig. 11 Densidade de probabilidade para o verdadeiro valor baseada na medida de 64 medidas com o aparato A (supondo erros independentes do valor verdadeiro).

Até aqui temos uma solução do problema bem clara. A situação começa ficar complicada quando não temos as densidades de probabilidade $\rho_{X\,\infty;x_V}\left(x\right)$ fornecidas por alguém que teve acesso a um aparato de referência e que trabalhou com extrema paciência. Neste caso podemos usar as próprias medidas feitas com o aparato A para estimar o valor esperado e a variância da distribuição $\rho_{X\,\infty;x_V}\left(x\right)$.

Tendo feito N medidas com o aparato A é razoável estimar estes parâmetros da seguinte forma:

$$\text{estimativa do valor esperado} \;=\; \mu \;=\; {}^{N}\overline{x} \tag{1.6.46}$$

$$\text{estimativa da variância} \;=\sigma^2=\;\frac{\sum_{k=1}^{N}\left(x_k-{}^{N}\overline{x}\right)^2}{N-1} \tag{1.6.47}$$

Mas para poder concluir algo sobre os valores verdadeiros precisamos, sem falta, de alguma avaliação do erro sistemático do aparato A. De onde poderia vir esta avaliação?

Há essencialmente duas possibilidades: uma é realmente alguma comparação com um aparato de referência, mesmo sem ser uma determinação exaustiva das densidades $\rho_{X\infty;x_V}(x)$. A outra é que o próprio aparato A é um aparato de referência (um daqueles construídos nos grandes centros de medidas e padrões). Neste caso a avaliação de erro sistemático tem a forma de testes de autoconsistência. Nestes testes podemos também formar médias de grandes números de medidas de tal forma que a parcela estatística se torna desprezível e as inconsistências encontradas têm um caráter sistemático. Mas geralmente estas inconsistências permitem apenas uma estimativa do erro sistemático na forma de uma desigualdade do módulo do erro: $|\text{erro sistemático}| \le S$. Neste caso é prudente anotar no resultado de uma medida o erro sistemático separado do erro estatístico: resultado = valor principal $\pm S$ $\pm E$, onde E seria uma estimativa do erro estatístico. Tendo feito N medidas usaríamos como valor principal:

$$\text{valor principal} = \overset{N}{\overline{x}} \tag{1.6.48}$$

e como erro estatístico

$$E = K\sqrt{\frac{\sum_{k=1}^{N}\left(x_k - \overset{N}{\overline{x}}\right)^2}{N(N-1)}} \tag{1.6.49}$$

O fator K na frente da raiz depende do grau de segurança desejado. Com $K = 1$ e supondo que os erros são aproximadamente independentes do valor verdadeiro (então a $\rho_{X_V M}(x_V)$ da equação (1.6.45) seria uma gaussiana) o nível de segurança seria 68%, com $K = 2$ seria 95% e com $K = 3$ seria 99,7% (compare com os valores da tabela 1).

Repare no fator N no denominador embaixo da raiz. Este fator é o responsável para tornar o erro estatístico pequeno quando N assume valores grandes.

Tratamento estatístico da propagação de erros

Seja $F(A,B,C,.....,H)$ uma função que depende das grandezas físicas $A,B,C,.....H$. Imagine que medimos estas grandezas e obtivemos valores com estimativas de erro: $a \pm \delta_a$, $b \pm \delta_b$,.... $h \pm \delta_h$. Declararíamos como valor de F o valor $F(a,b,c,.....,h)$. Como estimativa de erro numa aproximação linear usamos até agora a fórmula

$$\delta_f \approx \left|\frac{\partial F}{\partial a}\right|\delta_a + \left|\frac{\partial F}{\partial b}\right|\delta_b + \left|\frac{\partial F}{\partial c}\right|\delta_c + \ldots\ldots + \left|\frac{\partial F}{\partial h}\right|\delta_h \tag{1.6.50}$$

Esta formula é de fato adequada quando se trata de poucas variáveis. Mas, se o numero de grandezas *A*, *B*, ... for grande esta fórmula estima um erro grande demais. Na equação (1.6.50) consideramos a possibilidade que os erros de todas as grandezas possam colaborar no mesmo sentido. Embora que isto possa realmente ocorrer esta situação seria muito pouco provável. Se adotarmos uma avaliação probabilística, o erro de F deve ser avaliado de forma diferente. Sejam os verdadeiros valores os seguintes:

$$a_V = a + \delta a\,,\quad b_V = b + \delta b\,,\quad c_V = c + \delta c\,,\ldots\quad .\ .\ .\ .\ ,\quad h_V = h + \delta h$$

com as estimativas

$$|\delta a| \le \delta_a\,,\quad |\delta b| \le \delta_b\,,\quad |\delta c| \le \delta_c\,,\quad .\ .\ .\ .\ ,\quad |\delta h| \le \delta_h$$

Vamos supor que não haja erros sistemáticos e as estimativas de erro $\delta_a, \delta_b, ...$ seriam baseadas nas variâncias das distribuições de probabilidade das respectivas grandezas:

$$\delta_a = K\sqrt{\mathrm{var}(A)}\ ,\quad \delta_b = K\sqrt{\mathrm{var}(B)}\ ,\ .....,\quad \delta_h = K\sqrt{\mathrm{var}(H)}$$

Usando a aproximação linear para a variação do valor de F temos:

$$F(a+\delta a, b+\delta b, ..., h+\delta h) - F(a,b,...,h) \approx \frac{\partial F}{\partial a}\delta a + \frac{\partial F}{\partial b}\delta b + ... + \frac{\partial F}{\partial h}\delta h \quad (1.6.51)$$

A variância desta grandeza é

$$\begin{aligned}
&\mathrm{var}\left(\frac{\partial F}{\partial a}\delta a + \frac{\partial F}{\partial b}\delta b + ... + \frac{\partial F}{\partial h}\delta h\right) = \\
&\quad \left(\frac{\partial F}{\partial a}\right)^2 \left\langle (\delta a)^2 \right\rangle + \frac{\partial F}{\partial a}\frac{\partial F}{\partial b}\left\langle \delta a \delta b \right\rangle + ... + \\
&\quad \frac{\partial F}{\partial b}\frac{\partial F}{\partial a}\left\langle \delta b \delta a \right\rangle + \left(\frac{\partial F}{\partial b}\right)^2 \left\langle (\delta b)^2 \right\rangle + ... + \\
&\quad . \\
&\quad . \\
&\quad . \\
&\quad + ... + \left(\frac{\partial F}{\partial h}\right)^2 \left\langle (\delta h)^2 \right\rangle
\end{aligned} \quad (1.6.52)$$

As medidas das grandezas A, B, ... H são estatisticamente independentes. Isto implica (**Exercício**: Mostre!) que os termos mistos $\langle \delta a \delta b \rangle$, $\langle \delta a \delta c \rangle$, $\langle \delta b \delta c \rangle$ ..... são todos zero. Então vale

$$\begin{aligned}
&\mathrm{var}\left(\frac{\partial F}{\partial a}\delta a + \frac{\partial F}{\partial b}\delta b + ... + \frac{\partial F}{\partial h}\delta h\right) = \\
&\quad \left(\frac{\partial F}{\partial a}\right)^2 \left\langle (\delta a)^2 \right\rangle + \left(\frac{\partial F}{\partial b}\right)^2 \left\langle (\delta b)^2 \right\rangle + ... + \left(\frac{\partial F}{\partial h}\right)^2 \left\langle (\delta h)^2 \right\rangle
\end{aligned} \quad (1.6.53)$$

Se avaliamos o erro de F também na forma $\delta_f = K\sqrt{\mathrm{var}(F)}$ segue da última equação para o erro de F:

$$\delta_f = \sqrt{\left(\frac{\partial F}{\partial a}\right)^2 (\delta_a)^2 + \left(\frac{\partial F}{\partial b}\right)^2 (\delta_b)^2 + ... + \left(\frac{\partial F}{\partial h}\right)^2 (\delta_h)^2} \quad (1.6.54)$$

Esta fórmula substitui a equação (1.6.50) que usamos até agora. Vale lembrar que isto vale apenas quando os erros foram pequenos por que usamos uma aproximação linear para a variação da função *F*. Aqui não vamos deduzir nenhuma fórmula para casos mais gerais por que o estudo da Física não é uma visita a um supermercado de fórmulas, mas a formação de um Físico deve ser o desenvolvimento da capacidade de criar as fórmulas adequadas para cada caso. **Exercício**: Deduza uma formula para os erros dos coeficientes linear e angular de uma reta ajustada com o método de mínimos quadrados.