

Quantas Medidas são Necessárias para o Conhecimento de uma Grandeza Física?

(How many measurements one needs in order to know a physical quantity?)

V.P. Likhachev, M.T.F. da Cruz e J. Mesa

Instituto de Física da Universidade de São Paulo

Laboratório do Acelerador Linear

Rua do Matão, Travessa R, 187

05508-900 São Paulo, SP

Recebido em 28 de Fevereiro, 2000. Aceito 11 de Julho, 2000.

Discutimos aqui como expressar o resultado de um conjunto de medidas segundo as prescrições da estatística e estendemos essa discussão até o significado estatístico do intervalo $\bar{Y} \pm s_m$, variando o tamanho da amostra (número de medidas, N) desde $N = 2$ até $N = 100$. Fazemos um estudo com o método de Monte Carlo e examinamos a dispersão do nível de confiança (probabilidade) do intervalo acima, mostrando que esta é elevada quando o número de medidas é pequeno ($N < 5$). Concluimos afirmando que atingimos o “conhecimento de uma grandeza” quando a dispersão da probabilidade do resultado obtido não é importante quando comparada a essa mesma probabilidade.

We discuss how to express the result of a set of measurements according to the prescriptions of the statistics and extend the discussion to the statistical meaning of the $\bar{Y} \pm s_m$ interval, when changing the size of the data set, N , from $N = 2$ up to $N = 100$. With the help of the Monte Carlo method, we have surveyed the dispersion of the confidence level (probability) of the interval above, showing that such dispersion is important when the data set is small ($N < 5$). We conclude by stating that we have reached “knowledge of a quantity” when the dispersion of the probability contents of its resulting interval is not important when compared to the probability itself.

I Introdução

O objetivo inicial deste trabalho era apresentar, de forma clara, conceitos estatísticos essenciais para a representação correta de resultados de experimentos, para estudantes de áreas de cunho experimental. A discussão do assunto durante o preparo do texto, juntamente com os resultados das simulações efetuadas, conduziu à revisão de temas sedimentados, que receberam nova interpretação.

A urgência em se utilizar ferramentas que possibilitem um tratamento mais sofisticado e completo de um assunto causa às vezes inversões, pois se é levado a utilizar a ferramenta antes mesmo de estudá-la em um curso adequado sobre o tema. Essa é uma situação comum, sucedendo tanto em cursos teóricos como nos laboratórios didáticos. No caso particular dos laboratórios didáticos, muitas vezes os estudantes começam a utilizar diversas ferramentas estatísticas antes sequer de se familiarizarem com o seu significado preciso e com o aparato matemático que as envolve.

Esse procedimento pode deixar deficiências na formação do estudante. Dentre as deficiências encontradas estão a falta do entendimento e do uso corretos da linguagem estatística. A motivação didática deste trabalho é esclarecer o significado de ferramentas estatísticas, como o desvio padrão da amostra e aquele da média.

A simulação de medidas e o estudo das propriedades dos diferentes intervalos estatísticos a elas associados permitem observar o efeito de se fazer um número reduzido N de medidas (e.g. $N \leq 5$) e também o que sucede em casos de N mais elevado. Muito embora a linguagem estatística não possua qualquer ambigüidade, os resultados dessas simulações têm aspectos essenciais, que permitem ilustrar que, quando se expressa o nível de confiança (probabilidade) associado a um determinado intervalo, esse nível de confiança é um valor médio, tendo certa dispersão, associada à dispersão da média e à dispersão do desvio padrão do intervalo apresentado. A dispersão do nível de confiança será tão maior quanto menos medidas forem feitas. Assim, o termo “conhe-

cimento de uma grandeza física” representará aqui a obtenção de um intervalo (dado pelo melhor valor \pm desvio padrão) com um conteúdo de probabilidade que tenha uma dispersão pouco importante.

II A linguagem estatística

Nessa linguagem, os resultados obtidos em um experimento onde foram feitas N medidas da grandeza chamada de Y , representam uma *amostra* dessa grandeza. Representa-se o conjunto desses resultados por $\{Y_i\}_N$. O conjunto de todos os resultados possíveis de serem obtidos para Y é chamado de população. Denomina-se Y uma *variável aleatória* e diz-se que esta é regida por uma *função densidade de probabilidade*, $F(Y)$, cuja forma funcional pode ser ou não conhecida previamente [2].

Um ponto importante é como representar a amostra $\{Y_i\}_N$. Intuitivamente, o aluno/experimentador quer apresentar um melhor valor juntamente com alguma característica que indique a reprodutibilidade desses dados. A maneira adotada é apresentar-se o *valor médio*, \bar{Y} , e o *desvio padrão da média*, s_m . Este último é calculado a partir do desvio padrão da amostra, s , ambos definidos adiante. O que se está informando, dentro de uma linguagem *padronizada*, são os intervalos $[\bar{Y} - s, \bar{Y} + s]$ e $[\bar{Y} - s_m, \bar{Y} + s_m]$.

Costuma-se afirmar que os intervalos, centrados em \bar{Y} , estipulados pelo desvio padrão e pelo desvio padrão da média contém três tipos de informação, ou têm três conteúdos estatísticos:

1. $[\bar{Y} - s, \bar{Y} + s]$ encerra uma probabilidade conhecida de se obter uma particular medida Y_j em seu interior, *no caso de se repetir uma medida em condições idênticas*;
2. $[\bar{Y} - s_m, \bar{Y} + s_m]$ encerra uma probabilidade conhecida de se obter uma nova média Y' dentro dele, *no caso de se repetir todo o experimento em condições idênticas*. Isso significa que, tendo feito N medidas no primeiro experimento, a chance de que numa repetição de N medidas nas mesmas condições experimentais a nova média Y' esteja dentro do intervalo apresentado é conhecida;

3. $[\bar{Y} - s_m, \bar{Y} + s_m]$ encerra também uma probabilidade conhecida de se achar o *valor verdadeiro* Y_0 de Y dentro do intervalo obtido, quando Y_0 existir. Aqui, a existência ou não do valor verdadeiro depende do experimento (*e.g.* pode-se calcular o mesmo intervalo para as velocidades das moléculas de um gás à temperatura ambiente, mesmo não existindo o valor verdadeiro dessa velocidade, pois as variações obtidas durante as medidas não se devem apenas a flutuações estatísticas da grandeza em si ou àquelas introduzidas pelo processo de medida. Aqui, a principal origem das variações re-

side na distribuição Maxwelliana de velocidades, *i.e.*, há *de fato* moléculas movendo-se mais lentamente e outras mais rápidas).

O primeiro e segundo conteúdos representam como flutuam resultados obtidos dentro das condições do experimento. O terceiro não, pois ele é baseado na existência de um valor limite, Y_0 , para o qual convergirá a média de um número de medidas muito grande, com o concomitante estreitamento do intervalo $[\bar{Y} - s_m, \bar{Y} + s_m]$ ao redor dela. Este valor limite jamais será conhecido a não ser que a população seja finita e conhecida.

Estes três conteúdos de fato coincidem no caso de um grande número de medidas, N , mas os resultados são bem diferentes se poucas medidas são efetuadas e, além disso, o primeiro e o segundo conteúdos probabilísticos deixam de fazer sentido e mesmo ter utilidade prática para N pequeno, o que será demonstrado adiante, com simulações de Monte Carlo.

III Definições e métodos de simulação

A Fig.1 ilustra a diferença entre os parâmetros da população e aqueles da amostra.

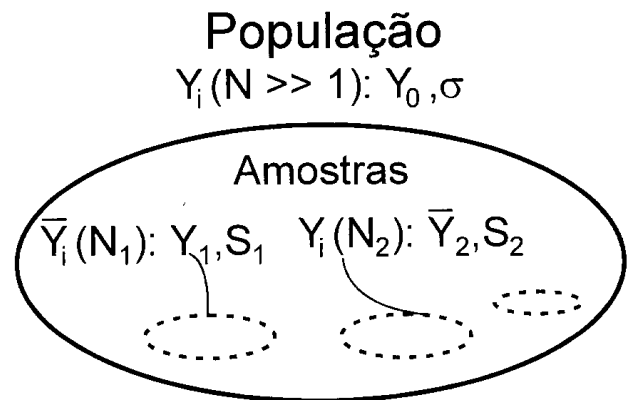


Figura 1. Definição geral.

Caracteriza-se a amostra, $\{Y_i\}_N$, de resultados de N medidas de uma grandeza através da sua média aritmética

$$\bar{Y} = \frac{1}{N} \sum_{i=1}^N Y_i \quad (1)$$

e da variância.

$$s^2 = \frac{\sum_{i=1}^N (Y_i - \bar{Y})^2}{N - 1}. \quad (2)$$

A raiz quadrada da variância é o desvio padrão, s , da amostra:

$$s = \sqrt{\frac{\sum_{i=1}^N (Y_i - \bar{Y})^2}{N - 1}} \quad (3)$$

este é uma *estimativa* do desvio padrão, σ , da população. O desvio padrão do valor médio, s_m , pode ser calculado através do desvio padrão da amostra:

$$s_m = \frac{s}{\sqrt{N}}. \quad (4)$$

Seja a hipótese típica: todos os dados são estatisticamente independentes e obedecem à distribuição normal, ou de Gauss, dada por

$$F(Y) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(Y - Y_0)^2}{2\sigma^2}\right), \quad (5)$$

É fácil verificar de (5) que o desvio padrão da população, σ , é a meia largura da distribuição $F(Y)$ na ordenada $e^{1/2} \cong 0,61$ do valor máximo da curva e que a área sob esta dentro do intervalo $[Y_0 - \sigma, Y_0 + \sigma]$ é $\approx 0,682$. Em outras palavras, a probabilidade de se achar um elemento da população dentro do intervalo de um desvio padrão é $\approx 68,2\%$.

A média introduzida no gerador de números aleatórios teve o papel do valor verdadeiro. A confiabilidade do gerador de números aleatórios é uma questão importante [3]. A consistência dos resultados foi verificada com o auxílio de dois geradores independentes, como explicado a seguir: (1) foi preparado um programa em linguagem FORTRAN, utilizando o algoritmo de Box-Muller [3], eq.(6), com o qual são gerados dois números aleatórios, X_1 e X_2 , distribuídos uniformemente dentro do intervalo $[0,1]$. A transformação na Eq.(6), abaixo, produz o número Y_i , pertencente ao intervalo aberto $]-\infty, +\infty[$, distribuído de acordo com a curva de Gauss de média Y_0 e desvio padrão σ :

$$Y_i = Y_0 + \sigma\sqrt{-2 \cdot \ln X_1} \cdot \cos(2\pi \cdot X_2) \quad (6)$$

aqui, Y_0 e σ têm os papéis de valor verdadeiro e de desvio padrão da população; (2) foram preparadas rotinas de simulação, na linguagem da plataforma MATLAB® [4].

A qualidade destes geradores de números aleatórios de gaussianos foi estudada e um teste pode ser visto na Fig.2, para o caso dos cálculos em FORTRAN, onde foi produzido um histograma a partir da geração de 10000 números com distribuição normal, com parâmetros $Y_0 = 2,0$ e $\sigma = 0,3$.

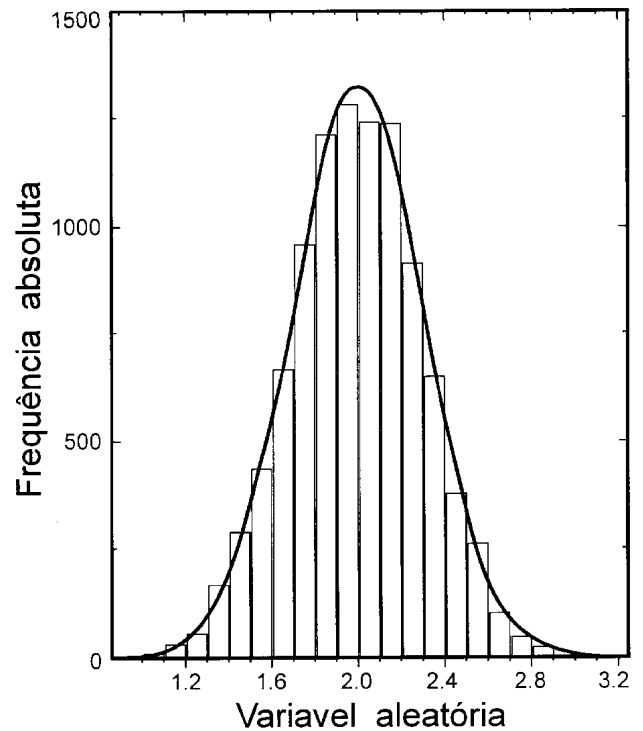


Figura 2. Histograma da distribuição de Y_i , produzida utilizando-se a linguagem FORTRAN.

Os parâmetros da curva sólida foram obtidos através do ajuste dos dados da simulação e resultaram em $Y = 2,0031$ e $s = 0,3011$. Um teste similar foi feito com o gerador de números aleatórios gaussianos na plataforma MATLAB®, *randn*, produzindo $\bar{Y} = 1,9965$ e $\sigma = 0,3012$.

IV Estimativa dos parâmetros da população

Para o estudo do conteúdo probabilístico do desvio padrão da amostra, s , no caso de um pequeno número de medidas, $N = 2$, fixou-se $Y_0 = 2,0$ e $\sigma = 0,3$ e apenas dois dados aleatórios Y_1 e Y_2 foram gerados. O valor médio \bar{Y} e o desvio padrão do conjunto foram usados, determinando assim o intervalo $[\bar{Y} - s, \bar{Y} + s]$. Em seguida $N' = 1000$ dados, $\{Y_i\}_{1000}$, foram gerados e contadas as ocasiões em que Y_i está contido no intervalo definido acima. Repetindo este procedimento $N'' = 2000$ vezes e elaborando-se um histograma da distribuição de frequências relativas para a fração de sucessos, que é a probabilidade ou o nível de confiança daquele intervalo, produziu-se a Fig.3. Esta representa os resultados desse tipo de simulação para $N = 2, 3, 5, 10$ e 100 . Pode-se ver que no caso $N = 2$ a distribuição de probabilidade é monotônica, quase uniforme, com um pequeno aumento da probabilidade para valores baixos da fração de sucessos. A frequência de se achar 10% do conjunto dentro do intervalo é só ~ 2 vezes maior do que aquela de se achar 90%. Podemos

concluir que no caso $N = 2$ o intervalo $[\bar{Y} - s, \bar{Y} + s]$ não tem um nível de confiança bem definido.

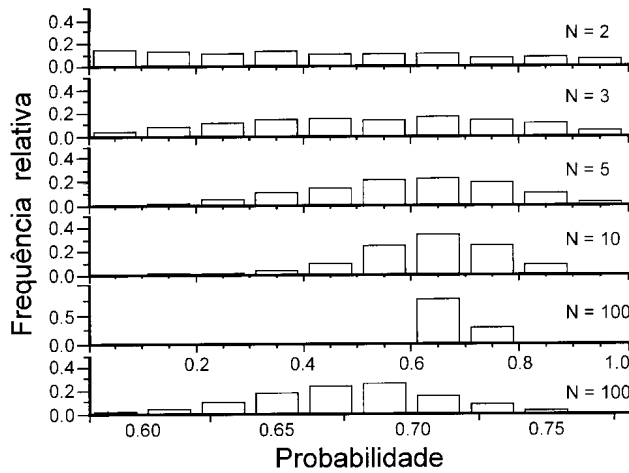


Figura 3. Histogramas da distribuição da probabilidade de se achar Y_i dentro do intervalo $[\bar{Y} - s, \bar{Y} + s]$ para vários números N de medidas. O último gráfico representa um detalhamento do resultado para $N = 100$.

No caso $N = 3$, a característica da distribuição mudou. A probabilidade de se achar a metade do conjunto dentro do intervalo é bem maior do que para frações muito pequenas e muito grandes. No caso $N = 5$, a distribuição torna-se ainda mais estreita e o valor máximo se desloca para a direita, para valores maiores da fração de sucessos. Todas as distribuições para $N > 3$ têm aspecto gaussiano. No caso $N > 5$ o máximo da distribuição corresponde à probabilidade igual a aproximadamente 68%.

V Estimativa da dispersão do valor médio

Examinou-se em seguida o conteúdo do intervalo $[\bar{Y} - s_m, \bar{Y} + s_m]$. Estudou-se a probabilidade de que a média \bar{Y}' de outras N medidas esteja dentro do intervalo $[\bar{Y} - s, \bar{Y} + s]$. Nesta situação, todo o experimento é repetido em condições idênticas. Foram gerados $N = 2$ dados aleatórios Y_1 e Y_2 , calculando-se o seu valor médio, \bar{Y} , e o desvio padrão da média, s_m . O intervalo $[\bar{Y} - s_m, \bar{Y} + s_m]$ foi determinado e gerou-se ainda $N' = 1000$ vezes conjuntos de $N = 2$ dados, $\{Y_i\}_2$, calculou-se a cada vez a média \bar{Y}' e foi feita a contagem do número de vezes em que \bar{Y}' caiu dentro do intervalo obtido no início. Este procedimento foi repetido $N'' = 2000$ vezes, produzindo-se um histograma da distribuição de probabilidade dos sucessos, ou seja, das frações do número de conjuntos $\{Y_i\}_N$, cujas médias caíram dentro do intervalo mencionado (Fig.4).

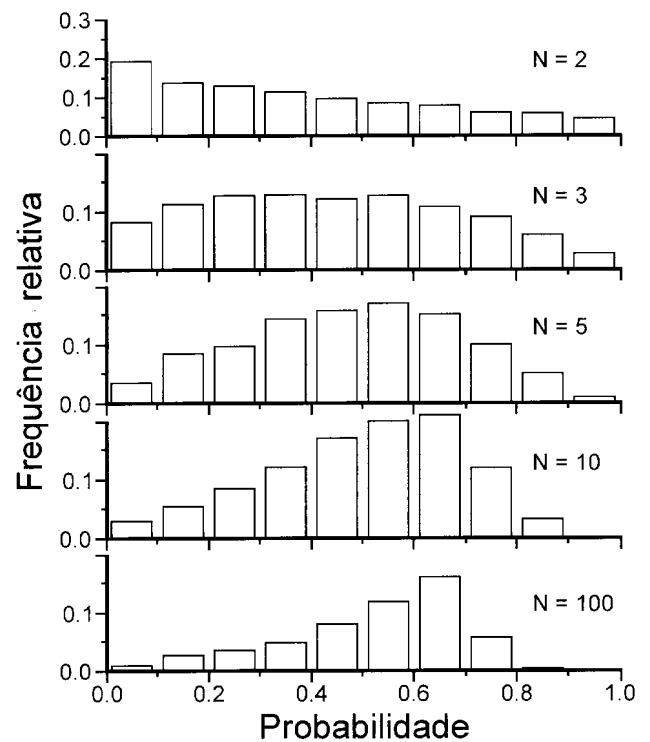


Figura 4. Histogramas da distribuição da probabilidade de se achar \bar{Y}' dentro do intervalo $[\bar{Y} - s_m, \bar{Y} + s_m]$ para vários números N de medidas.

A Fig.4 representa os resultados das simulações para $N = 2, 3, 5, 10$ e 100 . Vê-se que para $N = 2$ a distribuição de probabilidade é a mesma do caso anterior (Fig.3), ou seja, a distribuição é quase uniforme, com um aumento da probabilidade para frações pequenas. No caso $N = 3$, as características da distribuição também se alteram. Aqui, a probabilidade de que a metade do conjunto de \bar{Y}' esteja no interior do intervalo de teste é novamente maior que os valores na vizinhança. Nos casos $N = 5, 10$ e 100 , essa tendência é ainda mais forte. Porém, as características da distribuição de probabilidade já não são iguais às da distribuição de Gauss, como foi obtido na primeira simulação (Seção IV, Fig.3). Mesmo no caso $N = 100$, a probabilidade encontra-se relativamente mais dispersa. Isso é devido ao fato de que o centro do intervalo $[\bar{Y} - s_m, \bar{Y} + s_m]$ flutua com dispersão compatível com o tamanho do próprio intervalo, s_m , o desvio padrão da população das médias \bar{Y}' estudadas. Ao contrário do primeiro caso, em que o centro do intervalo $[\bar{Y} - s, \bar{Y} + s]$ flutua bem menos (s), quando comparada a flutuação com a semi-largura do intervalo, s .

VI Estimativa do valor verdadeiro

Finalmente, estudou-se o conteúdo probabilístico do desvio padrão do valor médio s_m com respeito à probabilidade do intervalo $[\bar{Y} - s_m, \bar{Y} + s_m]$ conter o valor verdadeiro, Y_0 . O estudo partiu da definição da probabilidade de se achar o valor verdadeiro dentro do intervalo obtido no experimento. O valor verdadeiro, Y_0 , não é uma grandeza aleatória, e por isso a probabilidade de achá-lo dentro de um intervalo experimental particular não tem sentido direto. Assim, a probabilidade de se achar o valor verdadeiro dentro intervalo $[\bar{Y} - s_m, \bar{Y} + s_m]$ é definida, por exemplo nas referências [5,6], como:

$$P(Y_0 \in [\bar{Y} - s_m, \bar{Y} + s_m]) = \lim_{N_{\text{total}} \rightarrow \infty} \frac{N_1}{N_{\text{total}}} \quad (7)$$

onde N_1 é o número de intervalos que contém Y_0 e N_{total} número total de intervalos do ensaio. A probabilidade de se encontrar o valor verdadeiro Y_0 , assim definida não é característica de um intervalo particular, obtido em um experimento, mas sim de *toda a população* de intervalos. Por essa razão ela é interpretada como uma constante para qualquer intervalo obtido no experimento. Neste ensaio gerou-se novamente $N = 2$ dados aleatórios, Y_1 e Y_2 , calculou-se o valor médio, \bar{Y} , e o desvio padrão da média, s_m . Determinou-se o intervalo $[\bar{Y} - s_m, \bar{Y} + s_m]$, e se analisou se o valor verdadeiro Y_0 estava em seu interior. Esse procedimento foi repetido $N' = 1000$ vezes, gerando-se outros conjuntos $\{Y_i\}_2$ e contando-se o número de intervalos $[\bar{Y} - s_m, \bar{Y} + s_m]$, gerados a partir deles, que continham Y_0 . Repetindo esse procedimento $N'' = 2000$ vezes, preparou-se o histograma da distribuição de frequência para a fração de sucessos, na Fig.5.

No caso $N = 2$ a distribuição de probabilidade é bastante diferente dos casos anteriores. Já nesse caso a distribuição é bem localizada e possui as características de uma gaussiana. A probabilidade é máxima para uma fração de cerca de 1/2. Com o aumento de N , a máxima probabilidade da distribuição se desloca para a direita e atinge o limite de 68% ($N = 100$). Pode-se corrigir s_m introduzindo-se um coeficiente k dependente de N [7], tal que (veja a Tab.1) $s'_m = k \times s_m$, para obter a máxima probabilidade em 68%, para qualquer N . A Fig.6 mostra os resultados da simulação com s'_m .

Tabela 1

n	2	3	4	5	10	20	∞
k	1,84	1,32	1,20	1,14	1,06	1,03	1

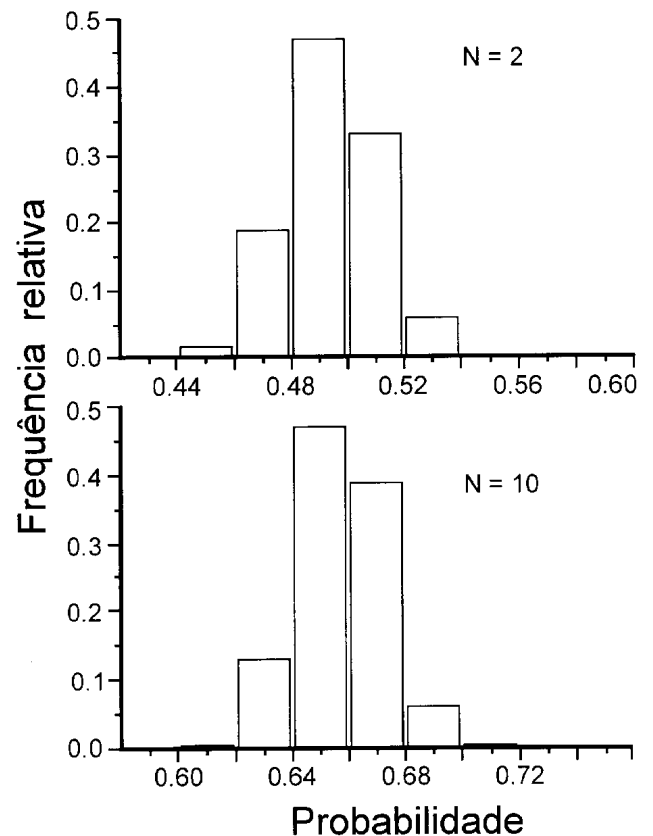


Figura 5. Histogramas da distribuição da probabilidade de se achar Y_0 dentro do intervalo $[\bar{Y} - s_m, \bar{Y} + s_m]$ para números de medidas $N=2$ e 10.

A largura dos histogramas diminui com o aumento do tamanho N' da amostra dos intervalos (veja a Fig.7). No caso de pequeno número $N' < 5$ o intervalo $[\bar{Y} - s'_m, \bar{Y} + s'_m]$ já não tem um nível de confiança bem definido em relação Y_0 .

VII Conclusões

Os resultados das simulações efetuadas mostram que os três conteúdos discutidos acima coincidem, no limite de um grande número de medidas, N .

1. Situação 1: o máximo de frequência corresponde à probabilidade $\approx 68\%$ de encontrarmos o resultado individual Y_i dentro do intervalo $[\bar{Y} - s, \bar{Y} + s]$;

2. Situação 2: o máximo de frequência corresponde à probabilidade $\approx 68\%$ de encontrarmos uma nova média de N medidas, \bar{Y}' , dentro do intervalo $[\bar{Y} - s_m, \bar{Y} + s_m]$;

3. Situação 3: o máximo de frequência corresponde à probabilidade $\approx 68\%$ de encontrarmos o valor verdadeiro, Y_0 , dentro do intervalo $[\bar{Y} - s_m, \bar{Y} + s_m]$.

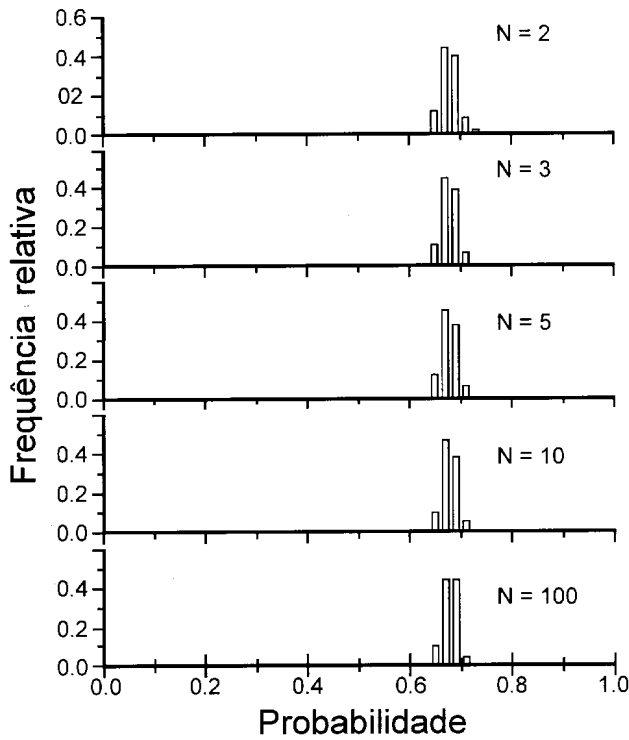


Figura 6. Histogramas da distribuição da probabilidade de se achar Y_0 dentro do intervalo corrigido $[\bar{Y} - s'_m, \bar{Y} + s'_m]$ para vários números de medidas, fazendo uso do coeficiente de correção, k .

Por outro lado, o desvio padrão da amostra, s , e o desvio padrão da média, s_m , podem não caracterizar a reprodutibilidade da medida, especialmente quando a amostra a partir da qual eles foram obtidos é pequena (N pequeno).

Os comportamentos apresentados estão ligados à flutuação estatística do próprio desvio, s , e da posição do centro do intervalo, ou seja, *o desvio padrão experimental e a posição do centro do intervalo são também grandezas aleatórias* e a dispersão destes afeta o conhecimento da probabilidade associada ao intervalo. A dispersão do desvio pode ser caracterizada através do desvio padrão do desvio padrão, dado por [2]

$$s_{s_m} = \frac{s_m}{\sqrt{2(N-1)}} \quad (8)$$

As distribuições de probabilidade obtidas são normais no primeiro e terceiro casos, e assimétrica no segundo. Quando diminui o número N de medidas, as larguras das distribuições de probabilidade aumentam no primeiro e segundo casos e praticamente não há alteração no terceiro caso.

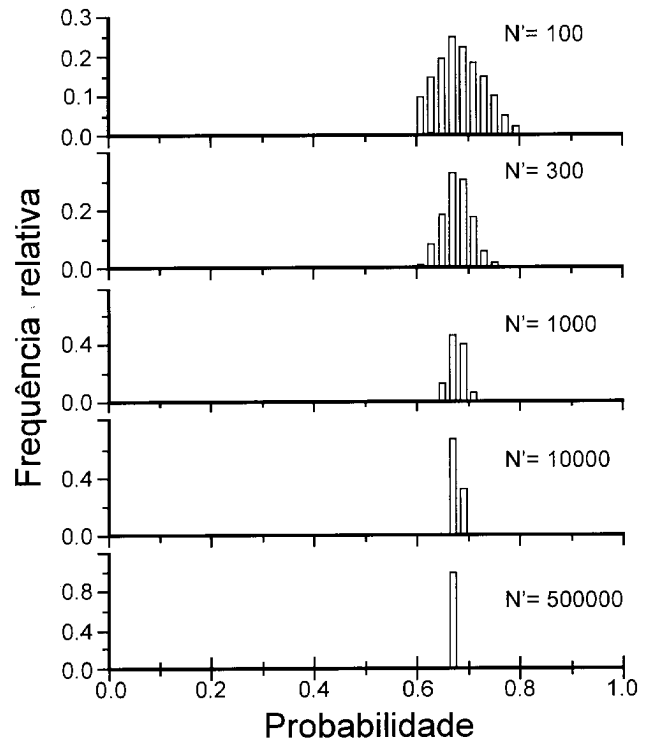


Figura 7. Histogramas da distribuição da probabilidade de achar Y_0 dentro do intervalo corrigido $[\bar{Y} - s'_m, \bar{Y} + s'_m]$ para vários números de simulações, N' (tamanho do conjunto dos intervalos).

Nos primeiros dois casos, para $N = 2$ o desvio padrão do conjunto s e da média s_m já não caracterizam dispersão. A probabilidade de se achar a grandeza testada dentro do intervalo pode ser qualquer e só para $N = 5$ tem-se um máximo de frequência correspondente a uma probabilidade em torno de $2/3$. A diferença de caráter da distribuição de probabilidade no primeiro e segundo casos em comparação com o terceiro está ligada à diferença entre os conteúdos desses intervalos. Nestes casos testa-se o intervalo particular, no terceiro testa-se o conjunto dos intervalos e só se os agrupa de maneiras diferentes, $N = 2, N = 10$ etc. Assim não há paradoxo no fato de que a probabilidade é a mesma para qualquer intervalo particular no terceiro caso.

A largura da distribuição de probabilidade para o primeiro e segundo casos praticamente não depende do tamanho N' do conjunto Y_i ou Y' e está ligada a flutuações estatísticas da posição e da largura do intervalo testado. No terceiro caso a largura não está ligada a flutuações estatísticas da posição e da largura do intervalo testado mas sim com as flutuações estatísticas de todo o conjunto dos intervalos. A largura diminui com o aumento do tamanho N' da amostra de intervalos. Para N' pequeno consegue-se reproduzir o mesmo efeito observado com N pequeno nos outros dois casos.

A probabilidade de se encontrar o valor verdadeiro Y_0 não é uma característica de um intervalo particular, obtido em um experimento, mas sim de toda a população de intervalos. Por essa razão, ela é inter-

pretada como uma constante para qualquer intervalo obtido no experimento e não tem sentido prático no caso de um pequeno número N de medidas, onde a largura e a posição do intervalo variam muito.

Agradecemos aos colegas, Prof. Otaviano Helene por apontar este problema e pelas valiosas discussões, Profs. H.R. Schelin, Ivan Evseev e Sergei Paschuk, e ao Sr. André Matias Larsen pela valiosas discussões.

References

- [1] J. H. Vuolo, *Fundamentos da Teoria de Erros*, Editora Edgard Blücher Ltda., São Paulo, 1992.
- [2] O. Helene e V.R. Vanin, *Tratamento Estatístico de Dados em Física Experimental*, Ed. Edgard Blücher Ltda., São Paulo, 1981.
- [3] W.H. Press, S.A. Teukolsky, W.A. Vetterling e B.P. Flannery, *Numerical Recipes in FORTRAN*, 2a. edição, Cambridge, 1992.
- [4] MATLAB®, *High-Performance Numeric Computation and Visualization Software*, The MathWorks Inc., versão 4.2.
- [5] P.A. Morettin e W.O. Bussab, *Métodos Quantitativos para Economistas e Administradores*, Atual Editora Ltda., São Paulo, 1981.
- [6] P.L.O. Costa Neto, *Estatística*, Editora Edgard Blücher Ltda, São Paulo, 1987.
- [7] O. Helene *et al.*, O que é uma medida?, *Revista de Ensino de Física*, **13**, 12 (1991).