Departamento de Física Experimental

Teste de Significância

03-04 de junho de 2014

Paulo R. Pascholati

Provinha 27 de maio de 2014

Item a)

Considere o conjunto de valores e incertezas $\{x_i, s_i\}$ resultantes de N medições de uma grandeza X. O valor médio da grandeza X é expressado como:

$$\overline{x} = \frac{\sum_{i=1}^{N} \frac{x_i}{s_i^2}}{\sum_{i=1}^{N} \frac{1}{s_i^2}}$$

Obtenha por propagação de incertezas a incerteza em $\overline{x}$.

A expressão de propagação de incerteza de uma função $f(x_1, x_2, \ldots x_N)$ pode ser escrita como

$$s_f^2 = \sum_{i=1}^{N} \left(\frac{\partial f}{\partial x_i} \right)^2 s_{x_i}^2$$

Provinha 27 de maio de 2014

Item a)

$$s_{\overline{x}}^2 = \sum_{i=1}^{N} \left\{ \left(\frac{\partial \overline{x}}{\partial x_i} \right)^2 s_{x_i}^2 \right\} = \sum_{i=1}^{N} \left\{ \left(\frac{\partial}{\partial x_i} \frac{\sum_{k=1}^{N} \frac{x_k}{s_k^2}}{\sum_{k=1}^{N} \frac{1}{s_k^2}} \right)^2 s_{x_i}^2 \right\}$$

$$= \sum_{i=1}^{N} \left\{ \left(\frac{1}{\sum_{k=1}^{N} \frac{1}{s_k^2}} \frac{\partial}{\partial x_i} \sum_{k=1}^{N} \frac{x_k}{s_k^2} \right)^2 s_{x_i}^2 \right\}$$

$$= \frac{1}{\left(\sum_{k=1}^{N} \frac{1}{s_k^2} \right)^2} \sum_{i=1}^{N} \left\{ \left(\frac{\partial}{\partial x_i} \sum_{k=1}^{N} \frac{x_k}{s_k^2} \right)^2 s_{x_i}^2 \right\}$$

$$= \frac{1}{\left(\sum_{k=1}^{N} \frac{1}{s_k^2} \right)^2} \sum_{i=1}^{N} \left\{ \left(\sum_{k=1}^{N} \frac{\delta_{ik}}{s_k^2} \right)^2 s_{x_i}^2 \right\} = \frac{1}{\left(\sum_{k=1}^{N} \frac{1}{s_k^2} \right)^2} \sum_{i=1}^{N} \left\{ \left(\frac{1}{s_i^2} \right)^2 s_{x_i}^2 \right\}$$

$$= \frac{1}{\left(\sum_{k=1}^{N} \frac{1}{s_k^2} \right)^2} \sum_{i=1}^{N} \left\{ \frac{1}{s_i^2} \right\} = \frac{1}{\sum_{k=1}^{N} \frac{1}{s_k^2}}$$

Provinha 27 de maio de 2014

Item b)

A probabilidade de ocorrer um dígito n como dígito mais significativo em um número que segue a lei de Newcomb-Benford pode ser escrita como:

$$P(n) = \ln\left(1 + \frac{1}{n}\right)$$

Verifique se essa probabilidade é normalizada. Coloque em detalhe os cálculos utilizados.
A propriedade de normalização é expressa como:

$$\sum_{i=1}^{N} P(i) = 1$$

Provinha 27 de maio de 2014

Item b)

A probabilidade de ocorrer um dígito *n* como dígito mais significativo em um número que segue a lei de Newcomb-Benford pode ser escrita como:

$$\sum_{i=1}^{9} P(i) = \sum_{i=1}^{9} \ln\left(1 + \frac{1}{i}\right) = \sum_{i=1}^{9} \ln\left(\frac{i+1}{i}\right)$$

$$= \ln\left(\frac{2}{1}\right) + \ln\left(\frac{3}{2}\right) + \cdots + \ln\left(\frac{9}{8}\right) + \ln\left(\frac{10}{9}\right)$$

$$= \ln(2) + \ln(3) - \ln(2) + \cdots + \ln(9) - \ln(8) + \ln(10) - \ln(9)$$

$$= \ln(10) = \log_e(10) \neq 1$$

Provinha 27 de maio de 2014
Item b)

Para ser normalizada a probabilidade deve ser expressa como

$$P(n) = \frac{\log_e\left(1+\frac{1}{n}\right)}{\log_e(10)}$$

$$= \frac{\log_e(10)\cdot\log_{10}\left(1+\frac{1}{n}\right)}{\log_e(10)}$$

$$P(n) = \log_{10}\left(1+\frac{1}{n}\right)$$

Transformação de Distribuição de Probabilidade
Helene e Vanin, pág. 47

Pode-se escrever que:

$$\Delta N(g, g + \Delta g) = N(x, x + \Delta x) \tag{1}$$

Como $F(x)$ é a densidade de probabilidade de x tem-se ainda:

$$N(x, x + \Delta x) \cong N \cdot F(x)|\Delta x| \tag{2}$$

$$\cong N \cdot F(x) \cdot \frac{\Delta g}{|\frac{dg}{dx}|} \tag{3}$$

$$\Delta N(g, g + \Delta g) \cong N \cdot F(x) \cdot \frac{\Delta g}{|\frac{dg}{dx}|} \tag{4}$$

Portanto

$$H(g) = \frac{F(x)}{|\frac{dg}{dx}|} \tag{5}$$

Transformação de Distribuição de Probabilidade

Aplicação no Caso do Volume de Esfera

Tomando o raio de uma esfera r_0 com distribuição de probabilidade normal, qual será a distribuição de probabilidade do volume V da esfera?

O volume da esfera é dado por:

$$V = \frac{4}{3}\pi r_0^3 \tag{6}$$

logo

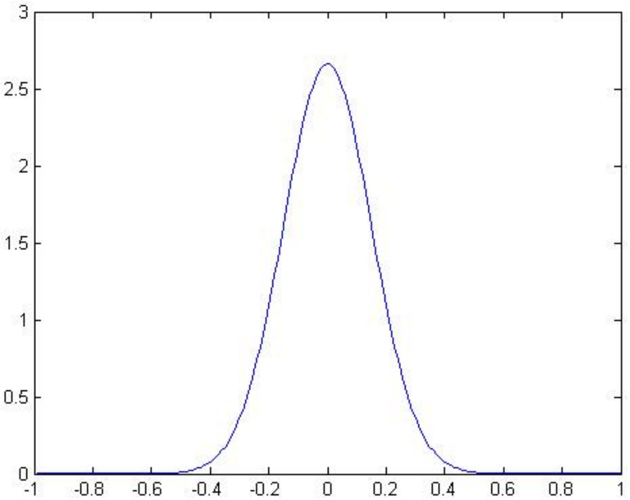
$$g(V) = \frac{4}{3}\pi r_0^3 \tag{7}$$

e

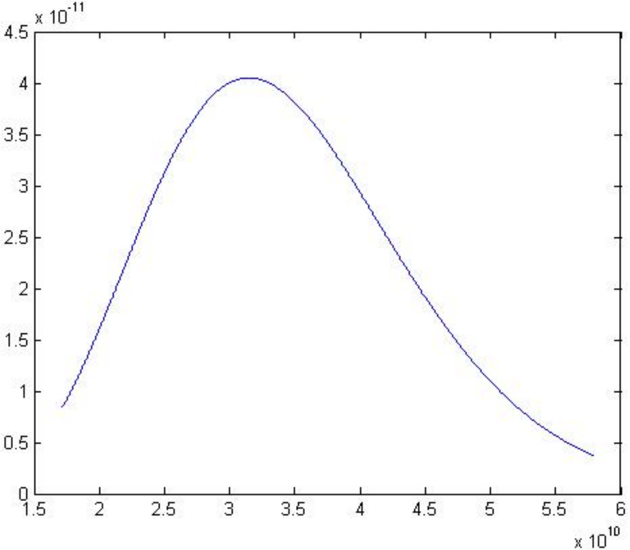
$$\left|\frac{dg}{dr_0}\right| = 4\pi r_0^2 \tag{8}$$

$$H(g) = \frac{F(r_0)}{4\pi r_0^2} = \frac{1}{4\pi r_0^2}\frac{1}{\sqrt{2\pi}\sigma_i} e^{-\left(\frac{r_i - r_0}{2\sigma_i}\right)^2} \tag{9}$$

Transformação de Distribuição de Probabilidade
Aplicação no Caso do Volume de Esfera



Transformação de Distribuição de Probabilidade
Aplicação no Caso do Volume de Esfera



Transformação de Distribuição de Probabilidade

Aplicação no Caso da Lei de Newcomb-Benford

It has been observed that the pages of a much used table of common logarithms show evidences of a selective use of the natural numbers. The pages containing the logarithms of the low numbers 1 and 2 are apt to be more stained and frayed by use than those of the higher numbers 8 and 9. Of course, no one could be expected to be greatly interested in the condition of a table of logarithms, but the matter may be considered more worthy of study when we recall that the table is used in the building up of our scientific, engineering, and general factual literature. There may be, in the relative cleanliness of the pages of a logarithm table, data on how we think and how we react when dealing with things that can be described by means of numbers.

F. Benford, *The Law of Anomalous Numbers*, Proc. Am. Phil. Soc., vol. 78, n. 4 (1938)551-572

Transformação de Distribuição de Probabilidade

Aplicação no Caso da Lei de Newcomb-Benford

The law of probability of the occurrence of numbers is such that all mantissæ of their logarithms are equally probable.
S. Newcomb, *Note on the Frequency of Use of the Different Digits in Natural Numbers.*, Am. J. of Math., vol. 4 (1881)39-40

Seja s a mantissa de um número qualquer y em uma base i — $y = i^{s+s'}$, s' um inteiro — , então

$$0 \leq s \leq 1 \tag{10}$$

A função densidade de probabilidade de s é uniforme ou $F(s) =$ constante. Então a função densidade de probabilidade de y, $F(y)$ é dada por

$$F(y) = \frac{F(s)}{|\frac{dy}{ds}|} = \frac{cte}{i^{s+s'}} = \frac{cte}{y} \tag{11}$$

Transformação de Distribuição de Probabilidade

Aplicação no Caso da Lei de Newcomb-Benford

A probabilidade de encontrar um dígito d_n em um número com n digítos $d_1, d_2, \cdots, d_n$ é expresso por:

$$P(d_1\ d_2 \cdots d_n) = \int_{d_1\ d_2 \cdots d_n}^{d_1\ d_2 \cdots d_n + 1} G(y)dy \tag{12}$$

$$= \int_{d_1\ d_2 \cdots d_n}^{d_1\ d_2 \cdots d_n + 1} \frac{1}{y}dy = ln\left(\frac{d_1\ d_2 \cdots d_n + 1}{d_1\ d_2 \cdots d_n}\right) \tag{13}$$

$$P(d_1\ d_2 \cdots d_n) = ln\left(1 + \frac{1}{d_1\ d_2 \cdots d_n}\right) \tag{14}$$

Se é somente o primeiro dígito do número que é de interesse, pode-se escrever:

$$P(d_1) = ln\left(1 + \frac{1}{d_1}\right) \tag{15}$$

Distribuição de χ^2 - Helene e Vanin *Tratamento Estatísticos de Dados ... - pág. 77*

Aqui é apresentada a função de distribuição de probabilidade de χ^2. Supondo que existência da função $y(x)$, conhecida, e que é a verdadeira função que relaciona as variáveis x e y. Supondo ainda o conjunto de N dados experimentais x_i, y_i com desvio padrão da média si na variável y_i. Uma medida da dispersão de quanto os pontos estão distribuídos em relação a curva verdadeira é dado por:

$$\chi^2 = \sum_{i=1}^{N} \frac{(yi - y(x_i))^2}{s_i^2} \tag{16}$$

A partir da função densidade de probabilidade de cada ponto y_i

$$F(y_i) = \frac{e^{-(y_i - y(x_i))^2/2s_i^2}}{\sqrt{2\pi} s_i} \tag{17}$$

Distribuição de χ^2

Fazendo uma séria de mudanças de variáveis na última equação do quadro anterior e somando para todos os i's obtem-se a função densidade de probabilidade de χ^2 como[1]

$$F_N(\chi^2) = \frac{1}{2\Gamma(\frac{N}{2})} \left(\frac{\chi^2}{2}\right)^{\frac{N}{2}-1} e^{-\chi^2/2} \tag{18}$$

onde Γ é a função gama que é definda como $\Gamma(z) = (z-1)!$ para z inteiro e $\Gamma(z) = (z-1)(z-2)...(1/2)\Gamma(1/2)$ para z ímpar com $\Gamma(1/2) = \sqrt{\pi}$.

Uma grandeza que tem interesse é a probalidade de encontrar um valor de χ^2 maior do que um determinado valor χ^2_α é

$$P_N(\chi^2_\alpha) = \frac{1}{2\Gamma(\frac{N}{2})} \int_{\chi^2_\alpha}^{\infty} \left(\frac{\chi^2}{2}\right)^{\frac{N}{2}-1} e^{-\chi^2/2}\, d\chi^2 \tag{19}$$

chamada de função cumulativa ou distribuição cumulativa de χ^2.

[1] A demonstração além de bastante trabalhosa está acima dos objetivos do programa da disciplina. Os interessados podem encontrar a dedução na referência citada.

Distribuição de χ^2

A última equação do quadro anterior não tem solução analítica e são utilizadas tabelas para obter o seu valor.

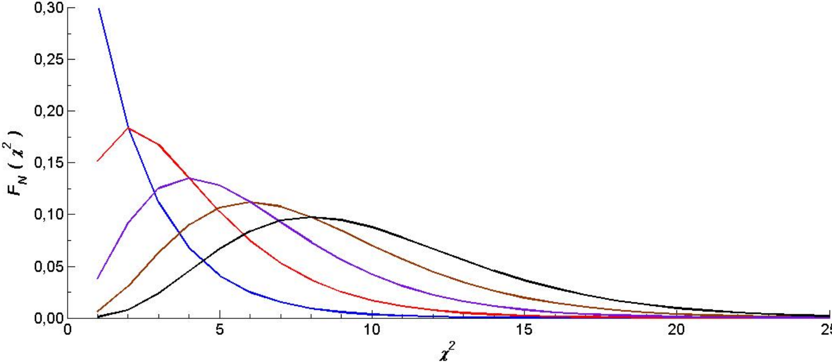


Figura : Função densidade de probabilidade de χ^2 para $N = 2$ (curva na cor azul), 4 (vermelha), 6 (lilás), 8 (marron) e 10 (preta).

Distribuição de χ^2

O valor médio de χ^2, N, pode ser obtido calculando a integral

$$\overline{\chi^2} = \int_0^\infty \chi^2 \ F_N(\chi^2) \ \mathrm{d}\chi^2$$

A rigor devemos usar o número de graus de liberdade $n_{gl} = N - p$ em lugar de N porque há uma quantidade de p de relações entre dos dados.
Pode-se encontrar a média do valor de χ^2 como o n_{gl}.

Distribuição de χ^2

Vai ser utilizada a equação que define o χ^2 para analisar a figura.

$$\chi^2 = \sum_{i=1}^{N} \frac{(yi - y(x_i))^2}{s_i^2} \tag{20}$$

Valores de χ^2 muito menores do que o n_{gl}, supondo que se está utilizando a função verdadeira, estão associados a uma superestimação das incertezas s_i. Estes estão presentes no denominador da equação de χ^2.
Caso contrário, valores de χ^2 maiores que n_{gl} podem estar associados a uma função que não corresponde a verdadeira de modo que as diferenças do numerador fiquem grandes ou a uma subestimação das incertezas s_i.

Distribuição de χ^2

	$P_N(\chi^2)$													
f	99,5	99	97,5	95	90	75	50	25	10	5	2,5	1	0,5	f
1	0,00004	0,00016	0,00098	0,00393	0,0158	0,102	0,455	1,32	2,71	3,84	5,02	6,63	7,88	1
2	0,01	0,0201	0,0506	0,103	0,211	0,575	1,39	2,77	4,61	5,99	7,38	9,21	10,6	2
3	0,0717	0,115	0,216	0,352	0,584	1,21	2,37	4,11	6,25	7,81	9,35	11,3	12,8	3
4	0,207	0,297	0,484	0,711	1,06	1,92	3,36	5,39	7,78	9,49	11,1	13,3	14,9	4
5	0,412	0,554	0,831	1,15	1,61	2,67	4,35	6,63	9,24	11,1	12,8	15,1	16,7	5
6	0,676	0,872	1,24	1,64	2,20	3,45	5,35	7,84	10,6	12,6	14,4	16,8	18,5	6
7	0,989	1,24	1,69	2,17	2,83	4,25	6,35	9,04	12,0	14,1	16,0	18,5	20,3	7
8	1,34	1,65	2,18	2,73	3,49	5,07	7,34	10,2	13,4	15,5	17,5	20,1	22,0	8
9	1,73	2,09	2,70	3,33	4,17	5,90	8,34	11,4	14,7	16,9	19,0	21,7	23,6	9
10	2,16	2,56	3,25	3,94	4,87	6,74	9,34	12,5	16,0	18,3	20,5	23,2	25,2	10
11	2,60	3,06	3,82	4,57	5,58	7,58	10,3	13,7	17,3	19,7	21,9	24,7	26,8	11
12	3,07	3,57	4,40	5,23	6,30	8,44	11,3	14,8	18,5	21,0	23,3	26,2	28,3	12
13	3,57	4,11	5,01	5,89	7,04	9,30	12,3	16,0	19,8	22,4	24,7	27,7	29,8	13
14	4,07	4,66	5,63	6,57	7,79	10,2	13,3	17,1	21,1	23,7	26,1	29,1	31,3	14
15	4,60	5,23	6,26	7,26	8,55	11,0	14,3	18,2	22,3	25,0	27,5	30,6	32,8	15
16	5,14	5,81	6,91	7,96	9,31	11,9	15,3	19,4	23,5	26,3	28,8	32.0	34,3	16
17	5,70	6,41	7,56	8,67	10,1	12,8	16,3	20,5	24,8	27,6	30,2	33,4	35,7	17
18	6,26	7,01	8,23	9,39	10,9	13,7	17,3	21,6	26,0	28,9	31,5	34,8	37,2	18
19	6,84	7,63	8,91	10,1	11,7	14,6	18,3	22,7	27,2	30,1	32,9	36,2	38,6	19
20	7,43	8,26	9,59	10,9	12,4	15,5	19,3	23,8	28,4	31,4	34,2	37,6	40,0	20
21	8,03	8,90	10,3	11,6	13,2	16,3	20,3	24,9	29,6	32,7	35,5	38,9	41,4	21
22	8,64	9,54	11,0	12,3	14,0	17,2	21,3	26,0	30,8	33,9	36,8	40,3	42,8	22
23	9,26	10,2	11,7	13,1	14,8	18,1	22,3	27,1	32,0	35,2	38,1	41,6	44,2	23
24	9,89	10,9	12,4	13,8	15,7	19,0	23,3	28,2	33,2	36,4	39,4	43,0	45,6	24
25	10,5	11,5	13,1	14,6	16,5	19,9	24,3	29,3	34,4	37,7	40,6	44,3	46,9	25
26	11,2	12,2	13,8	15,4	17,3	20,8	25,3	30,4	35,6	38,9	41,9	45,6	48,3	26
27	11,8	12,9	14,6	16,2	18,1	21,7	26,3	31,5	36,7	40,1	43,2	47,0	49,6	27
28	12,5	13,6	15,3	16,9	18,9	22,7	27,3	32,6	37,9	41,3	44,5	48,3	51,0	28
29	13,1	14,3	16,0	17,7	19,8	23,6	28,3	33,7	39,1	42,6	45,7	49,6	52,3	29
30	13,8	16,0	16,8	18,5	20,6	24,5	29,3	34,8	40,3	43,8	47,0	50,9	53,7	30