
Métodos Estatísticos em Física Experimental

Vitor Oguri

Rio de Janeiro, 24 de agosto de 2011

Sumário

1	O acaso e a necessidade	1
1.1	A aleatoriedade dos fenômenos	1
1.2	Experimentos em altas energias	5
2	As distribuições de probabilidades	8
2.1	Probabilidades a priori	8
2.1.1	Eventos equiprováveis	9
2.1.2	As regras básicas	10
2.1.3	Eventos compostos	14
2.2	Probabilidade condicional e a fórmula de Bayes	18
2.3	Probabilidades a posteriori	24
2.4	Distribuições de probabilidades contínuas	30

2.5	Os axiomas de Kolmogorov	36
2.6	Distribuições limites de frequências	40
2.7	A lei dos grandes números	45
2.8	A distribuição uniforme	48
2.9	A distribuição binomial	53
2.10	A distribuição de Poisson	63
2.11	A distribuição de Gauss	65
2.12	Exercícios	78
3	Métodos de simulação a Monte Carlo	85
3.1	O experimento de Buffon	86
3.2	Geração de eventos a Monte Carlo	93
3.2.1	Método de inversão	93
3.2.2	Método de rejeição simples	97
3.3	Cálculo de integrais e médias	100
3.3.1	Método de rejeição simples	100
3.3.2	Método direto	101
3.4	Redução de incertezas	104
3.4.1	Transformação e inversão	105
3.4.2	Métodos de rejeição	107
3.5	O método de Metropolis	110
3.6	Exercícios	117

4	Estimadores de máxima verossimilhança	124
4.1	As origens dos métodos estatísticos	124
4.2	O método da máxima verossimilhança de Fisher	126
4.3	Limite da função de verossimilhança	132
4.4	Estimadores para a distribuição de Poisson	135
4.5	Estimadores para a distribuição de Gauss	136
4.6	Estimadores para parâmetros de outras distribuições	141
4.7	Estimadores para distribuições multiparamétricas	143
4.8	Propriedades dos estimadores de máxima verossimilhança	148
4.9	Ajuste de funções a histogramas	151
5	Testes estatísticos paramétricos	154
5.1	O teste de significância de Fisher	156
5.2	Os testes de hipóteses de Neyman-Pearson	158
5.3	Intervalos de confiança	161
5.4	O teste de χ^2 de Pearson	165
5.5	Lançamento de dados	172
5.6	Exercícios	179
A	Tópicos sobre probabilidades	180
A.1	Eventos equivalentes	180
A.1.1	Distribuições de probabilidades de eventos equivalentes	181
A.1.2	Valor médio de função de uma variável aleatória	183

A.2	Variáveis aleatórias multidimensionais	184
A.2.1	Variáveis independentes	185
A.2.2	Funções de várias variáveis aleatórias	185
A.3	A desigualdade de Markov	189
A.4	A desigualdade de Chebyshev	189
A.5	A lei dos grandes números	190
A.6	Funções caracteríssticas	192
A.7	O teorema do limite central	195
A.8	A distribuição gama	197
A.9	A distribuição de χ^2	200
A.10	A distribuição de Student	203
B	Simulação Direta	209
	O Experimento de Rutherford-Geiger	210
	Referências Bibliográficas	212

Capítulo 1

O acaso e a necessidade

A incerteza é inevitável no cálculo de resultados observacionais. Em geral, a teoria nos permite calcular somente a probabilidade de obtermos um resultado particular.

P.A.M. Dirac (1930)

1.1. A aleatoriedade dos fenômenos

Os fenômenos naturais podem ser classificados como determinísticos ou aleatórios. Se os efeitos associados a um fenômeno, devido a determinadas influências (causas), são inequivocamente previsíveis, diz-se que os processos envolvidos são determinísticos. Por outro lado, se os efeitos associados a um fenômeno não são

exatamente previsíveis, mas podem ser associados a certas expectativas relativas de ocorrência, os processos envolvidos são ditos aleatórios.

Segundo a visão clássica, a não previsibilidade dos efeitos associados a um fenômeno estava associada a processos complexos que envolviam a interação de um grande número de sistemas simples. Assim, o conceito de aleatoriedade estava vinculado ao comportamento coletivo das moléculas de um gás, ou à enorme quantidade de núcleos que participam do fenômeno da radioatividade.

Uma vez que a teoria fundamental da Física Clássica – a Mecânica de Newton – descreve a evolução temporal de um sistema composto de um pequeno número de partículas por equações diferenciais ordinárias, pressupunha-se que seu comportamento seria completamente determinado por sua condição inicial.¹ A aleatoriedade e o acaso em um fenômeno, ou em um experimento, eram atribuídos à incapacidade do observador em determinar as condições iniciais, ou à complexidade dos arranjos experimentais.

Em princípio, uma teoria fundamental deveria ser determinística, tal que uma dada condição inicial estaria associada a um único resultado para a evolução de um sistema físico. Assim, teorias probabilísticas não seriam fundamentais, uma vez que poderiam admitir vários possíveis resultados para a evolução de um sistema, ao associar expectativas distintas a cada um desses possíveis resultados.²

¹Caracterizada pelas posições e velocidades iniciais de suas partículas constituintes.

²Sabe-se hoje que, mesmo para sistemas com poucos graus de liberdade, descritos por teorias causais, em princípio, determinísticas, pequenas perturbações iniciais podem dar origem a fenômenos caóticos não previsíveis.

Essa visão dos fenômenos, há milênios arraigada no homem, foi também compartilhada por grandes expoentes do pensamento ocidental:

“Nada acontece aleatoriamente; tudo acontece por alguma razão e por necessidade.”

Leucipo

“Todos os eventos, mesmo aqueles que por sua irrelevância parecem não se relacionar às grandes leis da natureza, delas constituem uma série tão necessária quanto as revoluções do Sol”.

P. S. Laplace

“Deus não joga dados.”

A. Einstein

No entanto, com o surgimento da Mecânica Quântica, em 1925, a aleatoriedade passa a ser uma característica intrínseca da evolução dos fenômenos e sistemas físicos, mesmo aqueles com poucos graus de liberdade. A Mecânica Quântica estabelece que, para cada problema, há de se calcular uma distribuição de probabilidades que conterá as informações necessárias à descrição do fenômeno estudado.

Segundo a Mecânica Quântica[8], as medidas das grandezas físicas estão associadas a certas distribuições de probabilidades de ocorrência. Ou seja, mesmo que as medidas associadas a uma grandeza não sejam inteiramente previsíveis quando efetuadas sob as mesmas condições experimentais, elas não se manifestam de

modo totalmente imprevisível, uma vez que, em geral, seus valores, além de limitados a um intervalo definido, estão associados a certas expectativas de ocorrência.

Assim, uma teoria probabilística para o estudo dos fenômenos naturais ou dos sistemas físicos deve prover regras que permitam determinar:

- ▷ os valores (medidas) possíveis para as grandezas associadas ao fenômeno ou ao sistema físico;
- ▷ as respectivas probabilidades de ocorrência das medidas das grandezas, ou a distribuição dessas probabilidades.

Em geral, uma distribuição de probabilidades associadas às medidas de uma grandeza é caracterizada por parâmetros que indicam o valor esperado de cada grandeza. Desse modo, os objetivos de qualquer experimento em Física que envolva processos aleatórios podem ser expressos como:

A partir de amostras de medidas de grandezas associadas a um fenômeno, e de um modelo ou de uma teoria física, os objetivos de um experimento em Física são:

- ▷ *determinação de valores esperados;*
- ▷ *observação de um resultado previsto pela teoria.*

Na linguagem da Estatística, os objetivos dos experimentos do primeiro grupo estão associados aos chamados problemas de estimação de parâmetros e, os do segundo, aos chamados testes de hipóteses.

1.2. Experimentos em altas energias

Todo experimento em Física de Partículas envolve processos aleatórios e, portanto, seus resultados devem ser confrontados com as distribuições de probabilidades que se obtêm a partir de modelos baseados em teorias quânticas como a Eletrodinâmica (QED), a Eletrofraca e a Cromodinâmica (QCD).

Por exemplo, as Fig. 1.1 e 1.2 mostram possíveis distribuições de *momentum* transverso (p_t) e de pseudorapidez (η) de múons resultantes de colisões próton-próton.

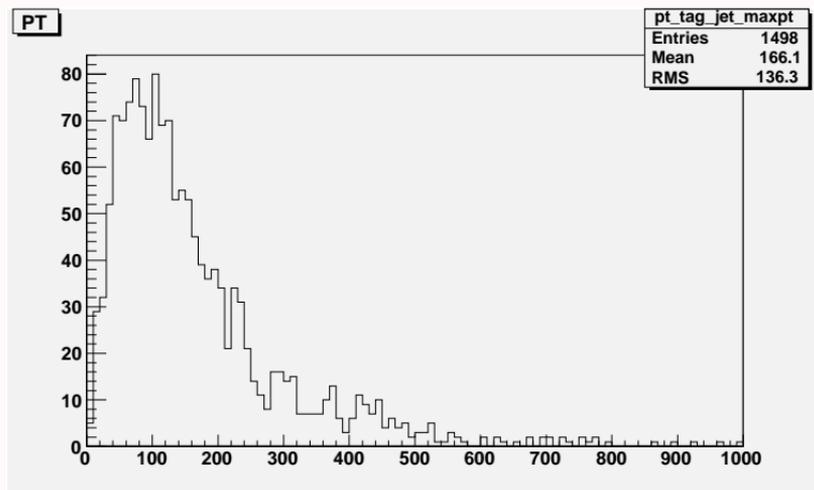


Figura 1.1: Distribuição de p_t de múons.

Segundo as teorias quânticas, alguns desses múons resultam de vários processos físicos aleatórios (como a hadronização e o decaimento de hádrons) que ocorrem após uma colisão (*quark-quark*) entre os constituintes dos prótons. Mesmo que esses processos elementares sejam bem descritos por modelos teóricos, as distribuições das grandezas físicas associadas a esses múons são modificadas por interações com os materiais que compõem o sistema de detecção de um experimento de colisões de partículas em altas energias.

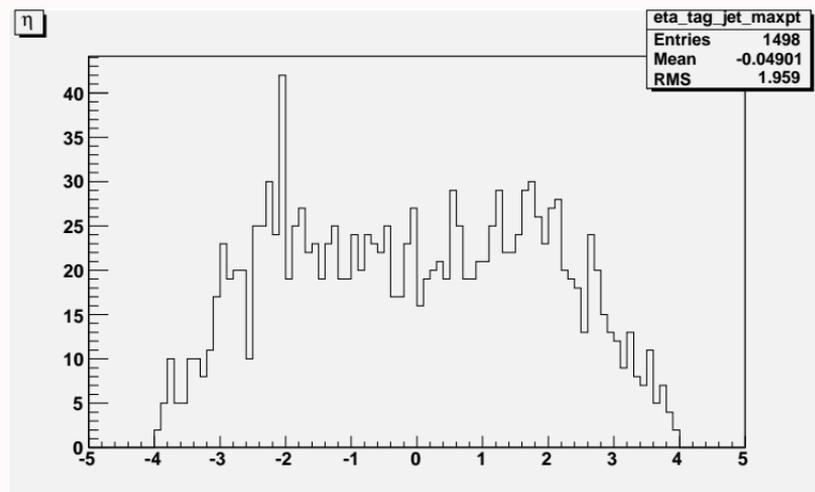


Figura 1.2: Distribuição de η de múons.

Desse modo, para a comparação das distribuições experimentais e teóricas, em geral, utilizam-se modelos aleatórios dos diversos processos que modificam as grandezas (*momentum*, energia e pseudorapidez)

associadas aos múons (partículas observadas), quando esses atravessam os vários subsistemas de detecção do experimento.

Assim, todo experimento em altas energias envolve tanto a elaboração de programas de simulação de amostras aleatórias (geração de eventos) segundo distribuições teóricas, como a de programas de simulação de interação de partículas com diversos materiais que compõem os detectores (simuladores de detectores).

Capítulo 2

As distribuições de probabilidades

Os observáveis na ciência são as distribuições estatísticas que descrevem as probabilidades associadas às observações.

K. Pearson (1892)

2.1. Probabilidades a priori

A teoria da probabilidade teve sua origem na análise dos jogos de azar (Bennet, David), ao se quantificar a expectativa de ocorrência do resultado associado a um fenômeno não aleatório, como um jogo de cartas ou o lançamento de dados. Essa quantificação iniciou-se com a correspondência entre os matemáticos franceses

Blaise Pascal (1623-1662) e Pierre Fermat (1601-1665), em 1654, e foi sintetizada pelo também francês Pierre Simon Laplace (1749-1827), em sua clássica obra *Théorie analytique des probabilités*, em 1812.

Devido a essa gênese, associada aos jogos, os conceitos e métodos probabilísticos elementares são usualmente apresentados a partir de exemplos que envolvem processos combinatoriais.

2.1.1. Eventos equiprováveis

O lançamento de dados proporciona um meio simples e acessível para se estudar e estabelecer as regras básicas dos fenômenos aleatórios ou do “acaso”.

Os únicos resultados possíveis para os lançamentos de um dado são os números naturais $\{1, 2, 3, 4, 5, 6\}$. Para um dado não viciado, devido à simetria do problema, a probabilidade *a priori* atribuída a ocorrência de um determinado número possível é igual a $1/6$. Assim, a definição de probabilidade, segundo Laplace, é dada pela razão entre o número de casos favoráveis possíveis para a ocorrência de um evento e o número total de alternativas igualmente possíveis, ou equiprováveis.

$$P(i) = \frac{1}{6} \quad (i = 1, 2, 3, 4, 5, 6)$$

Nesse sentido, o cálculo de probabilidades consiste na contagem das possibilidades de ocorrência de um evento. Esse tipo de cálculo ou estimativa, porém, é possível somente em situações simples, nas quais os possíveis resultados, em número finito, são conhecidos *a priori*, e por hipótese têm as mesmas chances de ocorrência, ou seja, os eventos são igualmente prováveis.

$$P(\text{evento}) = \frac{\text{número de casos favoráveis}}{\text{número de casos possíveis}} \quad (\text{probabilidade para eventos equiprováveis})$$

Um exemplo ainda mais simples de eventos equiprováveis são os lançamentos de uma moeda. Nesse caso, a probabilidade associada a cada um dos dois únicos resultados possíveis e equiprováveis, {cara, coroa}, é igual a $1/2$, ou seja,

$$P(\text{cara}) = P(\text{coroa}) = \frac{1}{2}$$

A definição clássica de Laplace torna-se impraticável quando o número de casos possíveis for muito grande, como no caso dos estados acessíveis às moléculas de um gás.

2.1.2. As regras básicas

A partir da análise da ocorrência de eventos simples, no entanto, pode-se estabelecer as regras básicas para o cálculo de probabilidades.

Por exemplo, no lançamento de dados, a probabilidade de ocorrência de um resultado (i) maior que 4 é igual a probabilidade associada à ocorrência dos dois resultados, 5 ou 6, entre os seis igualmente possíveis e mutuamente excludentes,

$$P(i > 4) = P(5 \text{ ou } 6) = \frac{2}{6} = \frac{1}{3}$$

ou seja, igual à soma das respectivas probabilidades,

$$P(5 \text{ ou } 6) = P(5) + P(6) = \frac{1}{6} + \frac{1}{6} = \frac{1}{3}$$

Por outro lado, a probabilidade de um resultado par, {2, 4, 6}, ou maior que 4, {5, 6}, é dada por

$$P(i = \text{par} \text{ ou } i > 4) = \frac{4}{6} = \frac{2}{3}$$

pois, entre os cinco resultados igualmente possíveis, mas não mutuamente excludentes, aqueles distintos são $\{2, 4, 5, 6\}$.

Nesse caso,

$$P(i = \text{par ou } i > 4) = P(i = \text{par}) + P(i > 4) - P(i = \text{par e } i > 4) = \frac{3}{6} + \frac{2}{6} - \frac{1}{6} = \frac{2}{3}$$

Assim, quando a ocorrência de um evento implica necessariamente a não ocorrência do outro, como a ocorrência dos números $A = 5$ ou $B = 6$ no lançamento de um dado, ou seja, quando os eventos são ditos mutuamente excludentes, a probabilidade associada à ocorrência de um ou de outro é dada pela soma das probabilidades:

$$\boxed{P(A \text{ ou } B) = P(A) + P(B)} \quad (A \text{ e } B \text{ mutuamente excludentes}) \quad (2.1)$$

Por outro lado, se a ocorrência de um evento $A = \{i = \text{par}\}$ não exclui a ocorrência de outro evento $B = \{i > 4\}$, a probabilidade associada à ocorrência de A ou B é dada por

$$\boxed{P(A \text{ ou } B) = P(A) + P(B) - P(A \text{ e } B)} \quad (A \text{ e } B \text{ não excludentes}) \quad (2.2)$$

onde $P(A \text{ e } B)$ é a probabilidade de ocorrência simultânea dos eventos A e B .

Se dois eventos A e B são independentes, no sentido de que a ocorrência de um não afeta a ocorrência do outro, isto é, não está condicionado ao outro, a probabilidade de que A e B ocorram simultaneamente é igual ao produto de suas probabilidades individuais, ou seja,

$$\boxed{P(A \text{ e } B) = P(A) \times P(B)} \quad (A \text{ e } B \text{ independentes}) \quad (2.3)$$

Uma vez que $P(A)$ e $P(B)$ são menores que a unidade, a probabilidade de que A e B ocorram simultaneamente nunca pode ser maior que as probabilidades individuais.

Um exemplo um pouco mais complicado consiste em um dos problemas propostos a Pascal pelo jogador francês Chevalier de Méré, que originou a correspondência entre Pascal e Fermat.

Méré queria uma explicação sobre o fato de que, em quatro lançamentos de um mesmo dado de pôquer, as chances de ocorrência de uma face marcada com um *ás* ($i = 1$) eram maiores do que, em vinte e quatro lançamentos de dois dados, a ocorrência de dois *ases* ($i = 2$).

Para ele, as chances seriam as mesmas, pois:

- ▷ se em uma jogada a chance de se obter um *ás* é de uma em seis, em 4 jogadas as chances de sair pelo menos um *ás* seria 4 vezes maior, ou seja, $4 \times \frac{1}{6} = \frac{2}{3}$;
- ▷ se as chances de se obter um par de *ases* ao se lançar um par de dados é de uma em 36, em 24 lançamentos seria 24 vezes maior, ou seja, $24 \times \frac{1}{36} = \frac{2}{3}$.

Aplicando-se esse conceito, quando uma moeda é lançada duas vezes, uma vez que as chances de sair cara em um lançamento é de uma em duas ($1/2$), as chances de sair pelo menos uma cara em duas jogadas seria $2 \times \frac{1}{2} = 1$, o que não é correto.

O erro nesse tipo de argumento está no fato de que, apesar de os eventos serem independentes, eles não são excludentes, ou seja, a ocorrência de um não exclui a do outro. Nesses casos, para o cálculo da probabilidade conjunta de eventos independentes, deve-se subtrair a probabilidade de interseção dos eventos.

O resultado pode ser obtido a partir da contagem de vezes da ocorrência de pelo menos um *ás* ($i = 1$) na listagem das 1296 ($6^{N=4}$) possíveis sequências dos resultados de eventos equiprováveis nos quatro lançamentos de um dado.

(1 1 1 1), (1 1 1 2), ... (1 1 1 6), (1 1 2 1), ... (1 1 2 6), ... (1 1 6 1), ... (1 6 6 6)
 (2 1 1 1), (2 1 1 2), ... (2 1 1 6), (2 1 2 1), ... (2 1 2 6), ... (2 1 6 1), ... (2 6 6 6)
 (3 1 1 1), (3 1 1 2), ... (3 1 1 6), (3 1 2 1), ... (3 1 2 6), ... (3 1 6 1), ... (3 6 6 6)
 (4 1 1 1), (4 1 1 2), ... (4 1 1 6), (4 1 2 1), ... (4 1 2 6), ... (4 1 6 1), ... (4 6 6 6)
 (5 1 1 1), (5 1 1 2), ... (5 1 1 6), (5 1 2 1), ... (5 1 2 6), ... (5 1 6 1), ... (5 6 6 6)
 (6 1 1 1), (6 1 1 2), ... (6 1 1 6), (6 1 2 1), ... (6 1 2 6), ... (6 1 6 1), ... (6 6 6 6)

Esse processo trabalhoso seria impraticável no caso de 24 lançamentos ($N = 24$) de dois dados. No entanto, sabendo-se que a probabilidade de ocorrência de dois *ases*, $P(i = 2, N = 1)$, em um único lançamento de dois dados é igual a

$$P(i = 2, N = 1) = \frac{1}{6} \times \frac{1}{6} = \frac{1}{36}$$

a probabilidade de não ocorrência de duplo *ás* é dada por

$$P(i \neq 2, N = 1) = 1 - \frac{1}{36} = \frac{35}{36}$$

Assim, em 24 lançamentos ($N = 24$) de dois dados, a probabilidade de ocorrência de pelo menos dois *ases* é dada por

$$P(i = 2, N = 24) = 1 - \left(\frac{35}{36}\right)^{24} \simeq 0,49$$

ou seja, ligeiramente inferior a 50%.

É claro que, aumentando-se o número de tentativas de um evento, as chances de êxito também aumentam. Mas o mesmo acontece para todos os tipos de eventos, mesmo os indesejáveis:

- ▷ as chances de acidentes aumentam com o número de vezes que uma pessoa atravessa uma rua, viaja de automóvel ou de avião, ou com o número de dias de operação de uma usina nuclear, ou com a quantidade de vezes que uma pessoa se submete a uma cirurgia.¹

Por exemplo, cada vez que um avião sai para uma missão de guerra, as chances de ser abatido é de 2%. Aumentando-se o número de missões as chances de ser abatido também aumentam. Em 50 missões o avião será certamente abatido?

2.1.3. Eventos compostos

Para eventos compostos, mesmo que em número finito, a situação pode ser mais complicada. Galileu Galileu (1564-1642), em um texto escrito por volta de 1613,² identificou corretamente os 216 resultados equiprováveis no lançamento de três dados, sendo capaz de prever a ligeira diferença entre as probabilidades dos três dados somarem 9 e 10. Apesar de cada resultado estar associado a seis partições distintas, cada partição corresponde a diferentes multiplicidades.

¹Deve-se observar que esses eventos não são puramente aleatórios. Cuidados com a segurança certamente reduzem as chances dos acidentes. A maioria das pessoas morrem de outras causas.

²Sopra le scoperte dei dadi, tradução inglesa em David.

soma (S) de três dados	partições	partições equivalentes	multiplicidade
9	(126)	126 – 162 – 216 – 261 – 612 – 621	6
	(135)	135 – 153 – 315 – 351 – 513 – 531	6
	(144)	144 – 414 – 441	3
	(225)	225 – 252 – 522	3
	(234)	234 – 243 – 324 – 342 – 423 – 432	6
	(333)	333	1
10	(136)	136 – 163 – 316 – 361 – 613 – 631	6
	(145)	145 – 154 – 415 – 451 – 514 – 541	6
	(226)	226 – 262 – 622	3
	(235)	235 – 253 – 325 – 352 – 523 – 532	6
	(244)	244 – 424 – 442	3
	(334)	334 – 343 – 433	3

Uma vez que o número total de casos possíveis é 216, com um total de 25 resultados possíveis para a soma 9, e de 27 para a soma 10, as respectivas probabilidades são dadas por

$$P(S = 9) = 25/216 \quad \text{e} \quad P(S = 10) = 27/216$$

O fato de que muitos resultados, apesar de conhecidos pelos jogadores mais experientes da época, não serem por eles explicados, mostra que eventos compostos podem ser complexos demais para que suas alternativas ou resultados equiprováveis sejam diretamente determinados.

No entanto, utilizando-se as propriedades estabelecidas para a união e a intersecção de conjuntos de

eventos, pode-se fazer prever probabilidades associadas a eventos compostos. Por exemplo, a partir dos 36 pares ordenados de resultados possíveis e equiprováveis no lançamento de dois dados,

(1,1)	(1,2)	(1,3)	(1,4)	(1,5)	(1,6)
(2,1)	(2,2)	(2,3)	(2,4)	(2,5)	(2,6)
(3,1)	(3,2)	(3,3)	(3,4)	(3,5)	(3,6)
(4,1)	(4,2)	(4,3)	(4,4)	(4,5)	(4,6)
(5,1)	(5,2)	(5,3)	(5,4)	(5,5)	(5,6)
(6,1)	(6,2)	(6,3)	(6,4)	(6,5)	(6,6)

pode-se definir os seguintes conjuntos de eventos

$$\triangleright A = \{(i, j) \mid i + j = 7\} \quad (\text{pares de resultados com soma igual a 7})$$

$$\triangleright B = \{(i, j) \mid i = 1\} \quad (\text{pares nos quais o primeiro valor é igual a 1})$$

$$\triangleright C = \{(i, j) \mid i = 1 \text{ e } j \leq 4\} \quad (\text{pares nos quais o primeiro é 1 e o segundo menor ou igual a 4})$$

Desse modo, de acordo com as regras de cálculo de probabilidades, pode-se escrever para cada conjunto de eventos

$$\triangleright P(A) = \frac{6}{36} = \frac{1}{6}$$

$$\triangleright P(B) = \frac{6}{36} = \frac{1}{6}$$

$$\triangleright P(C) = \frac{4}{36} = \frac{1}{9}$$

$$\triangleright A \text{ e } B \text{ independentes} \implies \begin{cases} P(A \text{ e } B) = P(A) \times P(B) = \frac{1}{6} \times \frac{1}{6} = \frac{1}{36} \\ P(A \text{ ou } B) = P(A) + P(B) - P(A \text{ e } B) = \frac{1}{6} + \frac{1}{6} - \frac{1}{36} = \frac{11}{36} \end{cases}$$

$$\triangleright A \text{ e } C \text{ excludentes} \implies \begin{cases} P(A \text{ e } C) = 0 \\ P(A \text{ ou } C) = P(A) + P(C) = \frac{1}{6} + \frac{1}{9} = \frac{5}{18} \end{cases}$$

$$\triangleright C \subset B \implies \begin{cases} P(C \text{ e } B) = P(C) = \frac{4}{36} = \frac{1}{9} \\ P(C \text{ ou } B) = P(B) = \frac{1}{9} + \frac{1}{6} - \frac{1}{9} = \frac{1}{6} \end{cases}$$

$$\triangleright \bar{B} = \{(1, j) | i \neq 1\} \implies \begin{cases} P(\bar{B}) = \frac{30}{36} = \frac{5}{6} \\ P(A \text{ e } \bar{B}) = P(B) = \frac{5}{36} \end{cases}$$

2.2. Probabilidade condicional e a fórmula de Bayes

A probabilidade de ocorrência de B condicionada à ocorrência de A , denominada **probabilidade condicional** de B relativa a A , e denotada por $P(B|A)$, é igual a razão entre o número de ocorrências conjuntas de B e A e o de A , ou seja, de acordo com exemplo anterior (lançamento de dois dados),

$$P(B|A) = \frac{P(B \text{ e } A)}{P(A)} = \frac{1/36}{1/6} = \frac{1}{6}$$

Assim, a probabilidade condicional de um evento B com relação a um evento A , equivale a considerar o evento A como o universo dos resultados possíveis (população) para a ocorrência de B .

Ainda com relação ao exemplo anterior, os eventos B e A são independentes, portanto,

$$P(B|A) = P(B)$$

Analogamente,

$$P(A|B) = \frac{P(A \text{ e } B)}{P(B)} = \frac{1/36}{1/6} = \frac{1}{6} = P(A)$$

Por outro lado, no caso de eventos não independentes, como C e B ,

$$P(C|B) = \frac{P(C \text{ e } B)}{P(B)} = \frac{1/9}{1/6} = \frac{2}{3} \neq P(C) \quad \text{ou} \quad P(C|B) = \frac{4}{6} = \frac{2}{3}$$

e

$$P(B|C) = \frac{P(B \text{ e } C)}{P(C)} = 1 \quad \text{ou} \quad P(B|C) = \frac{4}{4} = 1$$

Assim, de um modo geral, a probabilidade de ocorrência simultânea de dois eventos A e B é dada por

$$\boxed{P(A \text{ e } B) = P(A|B) \times P(B)} \quad (2.4)$$

Como a probabilidade de ocorrência conjunta é simétrica,

$$P(A \text{ e } B) = P(B \text{ e } A)$$

a relação entre probabilidades condicionais entre eventos não excludentes pode ser expressa pela fórmula de Bayes,³

$$\boxed{P(B|A) = \frac{P(A|B) \times P(B)}{P(A)}} \quad (2.5)$$

De acordo com as propriedades dos conjuntos, a cada conjunto A de eventos, relativos a um universo de resultados possíveis, corresponde uma coleção \bar{A} (excludente à A) de eventos complementares tal que

$$P(A) + P(\bar{A}) = 1$$

pois, a condição (A ou \bar{A}) é igual ao Universo de resultados possíveis.

No exemplo dos dois dados, o complemento \bar{A} é o conjunto de pares, dentre os resultados possíveis, com soma diferente de 7,

$$\bar{A} = \{(i, j) \mid i + j \neq 7\}$$

³ Estabelecida pelo reverendo inglês Thomas Bayes (1702-1761), em *Essay towards solving a problem in the doctrine of chances*, publicado postumamente em 1763.

a probabilidade associada $P(\bar{A})$ é dada por

$$P(\bar{A}) = 1 - P(A) = 1 - \frac{1}{6} = \frac{5}{6}$$

e

$$P(A|\bar{B}) = \frac{5}{30} = \frac{1}{6} = P(B|\bar{A}) \quad P(B|\bar{C}) = \frac{2}{32} = \frac{1}{16} \quad P(C|\bar{B}) = 0$$

Desse modo, se \bar{A} e \bar{B} são complementares, respectivamente, a A e B , pode-se escrever

$$\begin{cases} P(A) = P(A|B) \times P(B) + P(A|\bar{B}) \times P(\bar{B}) \\ P(B) = P(B|A) \times P(A) + P(B|\bar{A}) \times P(\bar{A}) \end{cases}$$

Assim, alternativamente, a fórmula de Bayes pode ser expressa como

$$P(B|A) = \frac{P(A|B) \times P(B)}{P(A|B) \times P(B) + P(A|\bar{B}) \times P(\bar{B})} \quad (2.6)$$

Por exemplo, em uma dada cidade, estima-se que uma em mil pessoas estejam infectadas por um vírus. Um exame para a detecção do vírus tem 95% de eficiência e 6% de ineficiência.⁴ Qual a probabilidade de que uma pessoa com resultado positivo esteja realmente infectada?

Como a probabilidade de que uma pessoa qualquer esteja infectada (I) é dada por $P(I) = 0,001$ (uma

⁴ Enquanto a eficiência de um exame ou teste médico é determinada pela fração de casos (de resultados do teste médico) que classificam um paciente realmente doente como portador de um vírus ou de uma doença, a ineficiência, chamada também de taxa de resultados falsos-positivos, é fração do número de casos que classificam um paciente saudável como infectado.

em mil), a probabilidade do complemento $P(\bar{I})$, ou seja, de que uma pessoa não esteja infectada é dada por $P(\bar{I}) = 0,999$.

A probabilidade (condicional) de ocorrer um resultado positivo (P) quando uma pessoa está infectada é dada por $P(P, I) = 0,95$ (eficiência do exame), e a probabilidade de ocorrência de resultados falsos-positivos (ineficiência do exame), ou seja, a probabilidade de ocorrer um resultado positivo quando a pessoa não está infectada, é dada por $P(P, \bar{I}) = 0,06$.

Assim, de acordo com a fórmula de Bayes, a probabilidade de que uma pessoa esteja infectada, se o resultado for positivo, é dada por

$$\begin{aligned} P(I|P) &= \frac{P(P|I) \times P(I)}{P(P|I) \times P(I) + P(P|\bar{I}) \times P(\bar{I})} \\ &= \frac{1}{1 + \frac{P(P|\bar{I}) \times P(\bar{I})}{P(P|I) \times P(I)}} = \frac{95}{5995} \simeq 0,016 \end{aligned}$$

Esse resultado não deve surpreender pois, apesar do resultado positivo do exame, a probabilidade (*a priori*) de que uma pessoa qualquer esteja infectada é muito baixa (uma em mil).

Essa a probabilidade *a priori* é muito difícil de ser estimada, pois não se sabe o número exato de pessoas infectadas por um vírus mas, de qualquer modo, se essa estimativa for muito baixa, a fórmula de Bayes indica que qualquer que seja o resultado do teste a probabilidade de uma pessoa estar infectada também é muito baixa. Portanto, a aplicação da fórmula de Bayes nesse tipo de teste só é útil em regime de epidemia de um determinado vírus, quando a probabilidade *a priori* de infecção for relativamente alta.

- **Eventos excludentes**

Em geral, não existe, necessariamente, relação entre as probabilidades condicionais $P(A|B)$ e $P(A|\bar{B})$. No entanto, para eventos excludentes, como nos casos de testes de confiabilidade na identificação de acidentes (problema do táxi), essas probabilidades estão relacionadas por

$$P(A|B) = 1 - P(A|\bar{B})$$

De modo geral, se $\{B_i\}$ é um conjunto de eventos mutuamente excludentes, a fórmula de Bayes pode ser expressa também como

$$P(B_i|A) = \frac{P(A|B_i) \times P(B_i)}{\sum_j P(A|B_j) \times P(B_j)} \quad (2.7)$$

Ou seja, se $\{B_i\}$ é um conjunto de hipóteses mutuamente excludentes e A é um resultado que pode ser associado a cada uma das hipóteses, pode-se utilizar a fórmula de Bayes para se calcular a probabilidade *a posteriori*, $P(B_i|A)$, de qualquer uma das hipóteses condicionada à observação do resultado A , a partir das probabilidades *a priori* das hipóteses, $P(B_i)$, e das probabilidades de ocorrências de A devido a cada uma das hipóteses.

Atualmente, a grande aplicação da fórmula de Bayes talvez se dê no combate à disseminação de *spam* pela *internet*, utilizando os chamados filtros adaptativos bayesianos, propostos pela primeira vez em 1996 por Jason Rennie e, desde 2002, difundidos pela rede por Paul Graham.

Se a cada palavra suspeita A_i contida nas principais partes de um *e-mail* (*header, sender, subject, body, ...*) associa-se uma probabilidade $P(A_i, |S)$ de que essa palavra apareça em *e-mail* do tipo *spam*, e uma probabilidade $P(A_i, |\bar{S})$ de que apareça em *e-mail* que não seja *spam*, a partir da probabilidade (*a priori*)

$P(S)$ de que um *e-mail* qualquer seja um *spam*, a fórmula de Bayes permite avaliar a probabilidade $P(S|A_i)$ de que um futuro *e-mail* que contenha a palavra suspeita A_i seja um *spam*.

$$P(S|A_i) = \frac{P(A_i|S) \times P(S)}{P(A_i|S) \times P(S) + P(A_i|\bar{S}) \times P(\bar{S})}$$

onde $P(\bar{S}) = 1 - P(S)$.

Fazendo-se esse cálculo para cada palavra suspeita, se a probabilidade total exceder um certo nível (por exemplo, 95%), o filtro classifica o *e-mail* como *junk*, removendo-o para um *junk folder* ou apagando-o.

Além de eficiente, esse processo é adaptativo, pois a cada vez que o dono da conta classifica um *e-mail* como *junk*, aumenta-se a probabilidade condicional $P(A_i|S)$ de que uma palavra suspeita A_i apareça em *e-mail* do tipo *spam*.

A fórmula de Bayes é utilizada também como a principal ferramenta dos chamados métodos estatísticos bayesianos, que consideram a probabilidade como o grau de crença associado a um evento.

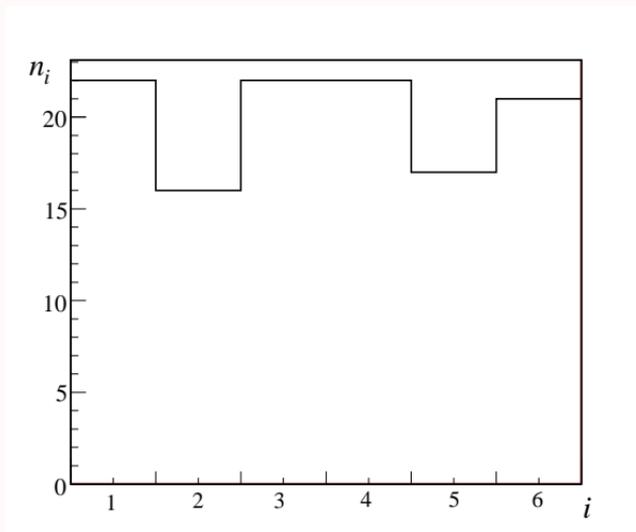
No exemplo da infecção por vírus, a probabilidade de que uma pessoa qualquer esteja infectada não sendo estimada a partir de observações estatísticas de casos, na concepção bayesiana, representa a expectativa de ocorrência de infecção antes de qualquer teste para detecção do vírus, ou seja, do grau (subjetivo) de crença de que uma certa parcela da população esteja infectada. De qualquer modo, se essa estimativa for muito baixa, a fórmula de Bayes indica que qualquer que seja o resultado do teste a probabilidade de uma pessoa estar infectada também é muito baixa. Portanto, a aplicação da fórmula de Bayes nesse tipo de teste só é útil em regime de epidemia de um determinado vírus, quando a probabilidade *a priori* de infecção for relativamente alta.

2.3. Probabilidades a posteriori

A probabilidade de $1/6$, que pode ser *a priori* atribuída à ocorrência de um determinado resultado no lançamento de um dado não viciado, pode ser obtida, também, *a posteriori*.

Tabela de distribuição de frequências (n_i) de cada uma das faces (i) de um dado, resultante da simulação de $N = 120$ lançamentos,⁵ e o histograma correspondente.

face (i)	freq. (n_i)
1	22
2	16
3	22
4	22
5	17
6	21



⁵A simulação de amostras de eventos ou fenômenos aleatórios é um procedimento usual no estudo desses fenômenos e será abordado e esclarecido em tópicos posteriores.

A tabela de frequências 2.1 mostra, por sua vez, a simulação de várias amostras de lançamentos de um dado.

face (i)	n_i	n_i/N	n_i	n_i/N	n_i	n_i/N
1	15	0.1250	197	0.1642	20108	0.1676
2	23	0.1917	206	0.1717	19854	0.1655
3	16	0.1333	196	0.1633	19859	0.1655
4	23	0.1917	187	0.1558	20149	0.1679
5	20	0.1667	202	0.1683	20036	0.1670
6	23	0.1917	212	0.1767	19994	0.1666
N	120	120	1200	1200	120000	120000

Tabela 2.1: Simulação de amostras de lançamentos de dados.

Desse modo, à medida que o número de lançamentos aumenta, a frequência relativa (f_i) de ocorrência de uma face i ,

$$f_i = \frac{n_i}{N} \quad (2.8)$$

aproxima-se cada vez mais de um número definido entre 0 e 1, nesse caso, igual a $1/6$.

Considerando que o número total N de medidas x_i de uma grandeza x em um experimento, ou que o número n_i de ocorrências de um evento sejam “suficientemente grande”, a probabilidade de ocorrência $P(x_i)$ da medida ou do evento em questão é definida como o limite da razão entre a frequência de observação do

evento e o número total de eventos observados, ou seja, como o limite da frequência relativa,

$$P(x_i) = \lim_{N \gg 1} f_i = \lim_{N \gg 1} \frac{n_i}{N} \quad (2.9)$$

A concepção frequentista ou estatística de probabilidade *a posteriori*, introduzida em 1718 por Abraham De Moivre (1667-1754), na obra *The doctrine of chances*, foi adotada por John Venn, em 1822, e sistematizada por Richard von Mises (1883-1953), em 1931, e pode ser aplicada mesmo quando o número total de eventos possíveis seja infinito, bastando que a frequência relativa aproxime-se de um limite experimental.

Assim as frequências relativas determinam uma distribuição de probabilidades associada às ocorrências de um determinado evento e, no caso dos lançamentos de um dado, em que a probabilidade de ocorrência de cada face é a mesma, essa distribuição é uniforme (figura 2.1).

Do mesmo modo que o valor médio (\bar{x}) de uma coleção $\{x_i\}$ de N valores distribuídos segundo um conjunto de frequências $\{n_i\}$ é caracterizado pela média dos valores ponderados pelas respectivas frequências relativas ($f_i = n_i/N$),

$$\bar{x} = \sum_{i=1}^N f_i x_i \quad (2.10)$$

o valor médio ou esperado, $E(x) = \mu = \langle x \rangle$, associado a uma coleção $\{x_i\}$ de N valores associados a uma distribuição de probabilidades, $P(x_i)$, é dado por

$$E(x) = \mu = \langle x \rangle = \sum_{i=1}^N x_i P(x_i) \quad (2.11)$$

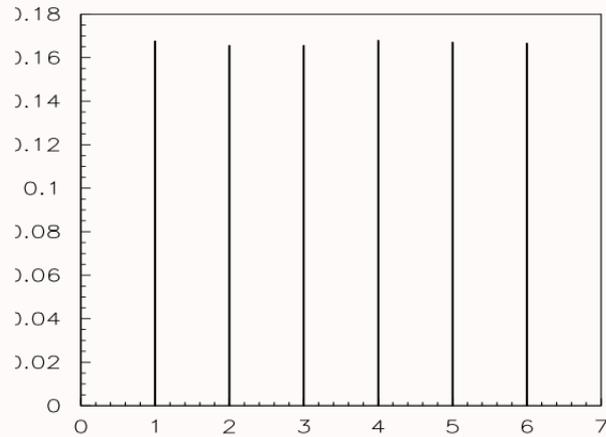


Figura 2.1: Distribuição de probabilidades para lançamentos de dados.

Analogamente, a dispersão desses valores pode ser caracterizada pela variância, $E[(x - \mu)^2] = V(x) = \sigma_x^2$, definida por

$$E[(x - \mu)^2] = V(x) = \sigma_x^2 = \sum_{i=1}^N (x_i - \langle x \rangle)^2 P(x_i) \quad (2.12)$$

ou pelo desvio-padrão,

$$\sigma_x = \sqrt{E[(x - \mu)^2]} = \sqrt{\sum_{i=1}^N (x_i - \langle x \rangle)^2 P(x_i)} \quad (2.13)$$

Do ponto de vista prático, a variância pode ser calculada por⁶

$$\sigma_x^2 = \langle x^2 \rangle - \langle x \rangle^2 \quad (2.14)$$

onde

$$\langle x^2 \rangle = \sum_{i=1}^N x_i^2 P(x_i) \quad (2.15)$$

é a média dos quadrados dos valores da coleção $\{x_i\}$

6

$$V(x) = E[(x - \mu)^2] = E(x^2 - 2\mu x + \mu^2) = \underbrace{E(x^2)}_{\langle x^2 \rangle} - 2\mu \underbrace{E(x)}_{\mu} + \underbrace{\mu^2}_{\langle x \rangle^2} = \langle x^2 \rangle - \langle x \rangle^2$$

Assim, se a probabilidade associada as ocorrências de cada face de um dado é $1/6$, o valor médio ou valor esperado dos resultados é dada por

$$\mu = \langle x \rangle = \sum_{i=1}^6 i \frac{1}{6} = \underbrace{(1 + 2 + 3 + 4 + 5 + 6)}_{21} \frac{1}{6} = 3.5$$

a média dos quadrados por

$$\langle x^2 \rangle = \sum_{i=1}^6 i^2 \frac{1}{6} = \underbrace{(1 + 4 + 9 + 16 + 25 + 36)}_{91} \frac{1}{6} = \frac{91}{6}$$

e a variância, por

$$\sigma^2 = \frac{91}{6} - (3.5)^2 = 2.917 \quad \Rightarrow \quad \sigma = 1.71$$

A expressão, satisfeita pela totalidade das possibilidades de ocorrências dos dados,

$$\boxed{\sum_{i=1}^{\infty} P(x_i) = 1} \tag{2.16}$$

a denominada condição de normalização, é utilizada como uma equação de vínculo que deve ser satisfeita por função que representa uma distribuição de probabilidades associada a uma dada população de dados.

O valor médio e a variância dos dados de uma população são denominados **parâmetros** de uma distribuição de probabilidades. Por isso, a procura ou determinação do valor esperado de uma grandeza aleatória e sua incerteza denomina-se **estimação de parâmetros**.

2.4. Distribuições de probabilidades contínuas

Para o caso de grandezas cujas variações são hipoteticamente contínuas, como as coordenadas (x) de uma pedra atirada aleatoriamente sobre um canaleta, os dados observados, que são as medidas das coordenadas, podem ser agrupados em intervalos discretos $[x_k, x_{k+1})$, de amplitudes $\Delta x_k = x_{k+1} - x_k$.⁷ Desse modo, se n_k é a frequência das medidas associadas ao intervalo (x_k, x_{k+1}) , obtém-se o histograma das coordenadas mostrado na Fig. 2.2.

Assim, a probabilidade de ocorrência de uma medida da coordenada x em um dado intervalo $[x_k, x_{k+1})$ é dada pela razão entre a área $n_k \cdot \Delta x_k$ da parte do histograma correspondente ao intervalo dado e a área total $\left(\sum_{i=1}^N n_i \cdot \Delta x_i = A \right)$ do histograma, ou seja,

$$P(x_k \leq x < x_{k+1}) = \frac{n_k \cdot \Delta x_k}{\sum_{i=1}^N n_i \cdot \Delta x_i} = \frac{n_k}{A} \cdot \Delta x_k$$

o que implica

$$\sum_{k=1}^N P(x_k \leq x < x_{k+1}) = 1$$

onde N é o número total de medidas.

⁷ Em geral, mas não necessariamente, esses intervalos possuem a mesma amplitude.

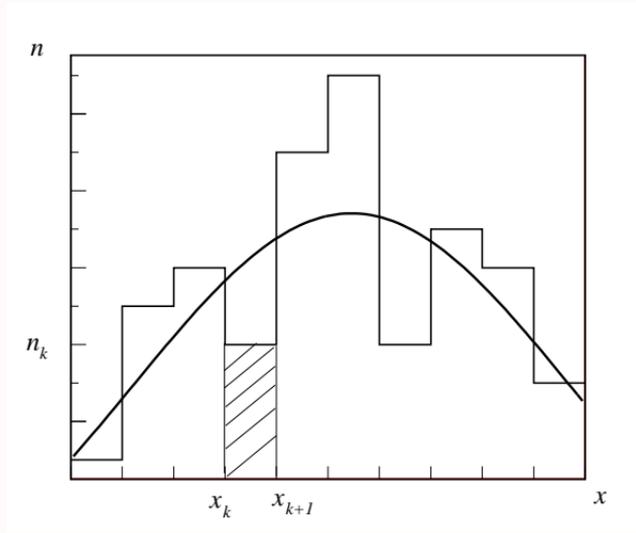


Figura 2.2: Histograma (hipotético) das coordenadas (x) de uma pedra e a distribuição de frequência limite.

A representação gráfica dos termos $\rho(x_k) = n_k/A$ associados aos intervalos $[x_k, x_{k+1})$ é denominada distribuição normalizada de frequência.

No limite de um grande número de observações ($N \gg 1$), os intervalos $[x_k, x_{k+1})$ ou classes de frequência podem ser tais que a amplitude Δx_k de cada classe seja tão pequena quanto se queira. Nessas condições, os termos limites

$$\rho(x)|_{x_k} = \lim_{\Delta x_k \rightarrow 0} \frac{P(x_k < x < x_{k+1})}{\Delta x_k} = \lim_{N \gg 1} \frac{n_k}{A} \quad k = 1, 2, 3, \dots, N$$

definem uma distribuição contínua normalizada de frequências relativas por unidade de medida da coordenada x , denominada **função densidade de probabilidade** ou **pdf**,⁸ $\rho(x)$, para os possíveis resultados contínuos de x , tal que a probabilidade associada à medida de x entre dois valores a e b é dada por

$$P(a < x < b) = \int_a^b \rho(x) \, dx \quad (2.17)$$

e a chamada condição de normalização da distribuição por

$$\int_{x_{\min}}^{x_{\max}} \rho(x) \, dx = 1 \quad (2.18)$$

onde x_{\min} e x_{\max} são, respectivamente, o mínimo e o máximo valores possíveis para a medida de x , que delimitam o domínio $D = (x_{\min}, x_{\max})$ de valores possíveis para as ocorrências de x .

⁸Acrônimo do inglês *probability density function*.

Assim, para variáveis contínuas, como as coordenadas ou momenta de uma partícula, ou os intervalos de tempo de vida de uma partícula instável, ou as massas ou alturas dos alunos de uma turma, não se pode atribuir uma medida de probabilidade como no caso de variáveis discretas, pois a soma de uma quantidade não enumerável de números positivos não pode satisfazer a condição de normalização.

Em vez de se atribuir, como no caso discreto, probabilidades aos valores de uma variável contínua, associam-se probabilidades a intervalos de valores da variável contínua, a partir de uma função densidade de probabilidade, análoga à função densidade de massa de um corpo rígido.

Nesse sentido, a densidade de probabilidade é análoga à densidade de massa de uma barra de massa unitária, para qual $\int_D \rho(x) dx = 1$, e, portanto, o valor médio $E(x) = \mu = \langle x \rangle$ de uma variável x , associada a uma densidade de probabilidade $\rho(x)$, em um domínio D , dada por

$$E(x) = \mu = \langle x \rangle = \int_D x \rho(x) dx \quad (2.19)$$

equivale ao centro de massa de uma barra de massa unitária distribuída de acordo com a função densidade $\rho(x)$.

Assim, o desvio-padrão é dado por

$$E[(x - \mu)^2] \sigma_x = \sqrt{\langle x^2 \rangle - \langle x \rangle^2} \quad (2.20)$$

Além de estar associada à distribuição de valores da variável x , a densidade de probabilidade $\rho(x)$ determina também a expectativa de ocorrência dos valores de qualquer outra função $f(x)$. Por exemplo, se $x \in D = (0, \infty)$ representa as possíveis medidas para o raio de um círculo, os possíveis valores para a área $f(x) = \pi x^2$

estarão associadas à mesma distribuição de probabilidade $\rho(x)$, de modo que o valor médio $\langle f \rangle$ será dado por

$$\langle f \rangle = \int_D f(x) \rho(x) dx \quad (2.21)$$

e a respectiva variância (σ_f^2) por

$$\sigma_f^2 = \langle f^2 \rangle - \langle f \rangle^2 \quad (2.22)$$

Para uma pdf $\rho(x)$ definida no domínio $(-\infty, \infty)$, a função

$$F(t) = P(x \leq t) = \int_{-\infty}^t \rho(x) dx \quad (2.23)$$

que indica a probabilidade associada à medida de x para valores menores que t é denominada **função de distribuição acumulada (fd)**.

Algumas das propriedades da função de distribuição são:

$$\triangleright P(x > t) = 1 - P(x \leq t) = 1 - F(t)$$

$$\begin{aligned} \triangleright F(-t) &= \int_{-\infty}^{-t} \rho(x) dx = - \int_{-t}^{-\infty} \rho(x) dx \\ &= \int_t^{\infty} \rho(x) dx = P(x > t) = 1 - F(t) \end{aligned}$$

$$\triangleright P(a \leq x \leq b) = F(b) - F(a)$$

$$\triangleright P(|x| \leq t) = P(-t \leq x \leq t) = F(t) - F(-t) = 2F(t) - 1$$

As distribuições de probabilidades utilizadas nos experimentos de Física deveriam ser determinadas *a posteriori*, uma vez que a evolução dos fenômenos e o próprio processo de medição são aleatórios. No entanto, a definição *a priori* de probabilidade é também amplamente usada na fundamentação de estudos teóricos, nos quais são utilizados conceitos probabilísticos, como no desenvolvimento da Mecânica Estatística (Boltzmann, Gibbs, Pathria). Nesses casos, distribuições de probabilidades especiais (Maxwell-Boltzmann, Planck, Fermi-Dirac e Bose-Einstein), adequadas à descrição de sistemas de muitas partículas quase-independentes, são deduzidas e mostram-se compatíveis com os comportamentos experimentais observados em diversos sistemas como os gases moleculares a baixa pressão, os elétrons em metais e semicondutores e, ainda, em diversos fenômenos, como a variação do calor específico dos sólidos a baixas temperaturas e a radiação de corpo negro.

Uma abordagem mais geral da Mecânica Estatística, iniciada pelo físico austríaco Ludwig Boltzmann (1844-1906), em 1877, e sintetizada pelo físico norte-americano Josiah Willard Gibbs (1839-1903), em 1901, ao utilizar, também, o conceito *a priori* de probabilidade, permite o estabelecimento de distribuições de probabilidades gerais (microcanônica, canônica e gran-canônica) não associadas apenas a gases, pois são expressas em termos de variáveis dinâmicas que caracterizam qualquer sistema de partículas. Nesse caso, para cada classe de problema, a partir das formas gerais dessas distribuições, é possível a obtenção de uma distribuição de probabilidades específica.

Até o primeiro quarto do século XX, a utilização na Física do conceito *a priori* de probabilidade, estava associada apenas à impossibilidade prática da caracterização simultânea do estado de todas as partículas de um sistema com um grande número de graus de liberdade. Ou seja, o problema resultava da complexidade dos sistemas observados. Com o surgimento da teoria da Mecânica Quântica, que utiliza também um conceito *a priori* de probabilidade, a aleatoriedade passa a ser uma característica intrínseca da evolução de qualquer

fenômeno ou sistema, mesmos aqueles com poucos graus de liberdade.

2.5. Os axiomas de Kolmogorov

A concepção frequentista, apesar de mais próxima a intuição física e estatística do que a concepção de Laplace, sempre foi rejeitada por muitos físicos e matemáticos.

A simulação de 1000 lançamentos de uma moeda (Fig. 2.3) mostra que o limite experimental da frequência relativa de ocorrências de caras não é uma operação matematicamente bem definida.

O conceito de probabilidade foi apropriadamente axiomatizado logo após o primeiro quarto do século XX, em 1933, pelo matemático russo Andrei Nikolaevich Kolmogorov (1903-1987), ao estabelecer que qualquer conjunto discreto de números não-negativos $\{P_i\}$, que esteja associado a uma coleção discreta de elementos $S = \{x_1, x_2, x_3, \dots\}$, como os possíveis valores para as medidas de uma grandeza, tal que

$$\boxed{\sum_i P(x_i) = 1} \quad (\text{condição de normalização})$$

e, para elementos distintos,⁹

$$\boxed{P(x_i \text{ ou } x_j) = P_i + P_j} \quad (x_i \neq x_j)$$

é denominado um conjunto de probabilidades $\{P_i\}$, associado à coleção S .

⁹ Na linguagem estatística, eventos mutuamente excludentes.

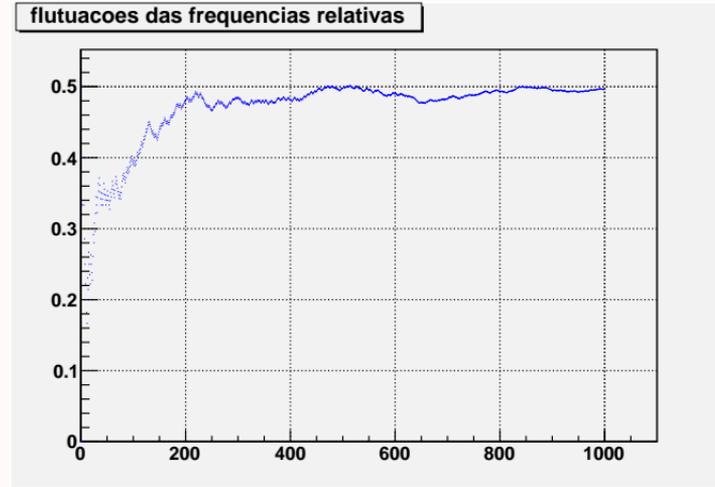


Figura 2.3: Frequência relativa do número de caras em termos do número de lançamentos.

Ou seja, do ponto de vista axiomático, não é necessário atribuir nenhum significado ao conceito de probabilidade, como a expectativa ou a crença na ocorrência de um evento; basta que se associe aos possíveis resultados independentes de um experimento, como as medidas de uma grandeza, um número positivo menor que a unidade, cuja adição sobre todos os possíveis eventos (igual à unidade), expressa a certeza em se obter um dos possíveis resultados (condição de normalização).

Nesse sentido, tanto a concepção clássica *a priori* de Laplace (expectativa de ocorrência), como a concepção frequentista *a posteriori* de von Mises satisfazem aos chamados axiomas de Kolmogorov para probabilidades discretas.

Durante a primeira metade do século XX, o ponto de vista frequentista foi adotado de forma quase unânime pelos principais estatísticos e matemáticos como Venn, von Mises, Reichenback, Gosset, Neyman, Fisher, Pearson, Cramér e Borel.

A concepção de Laplace teve dois desdobramentos:

- ▷ o defendido pela escola da lógica matemática, como Keynes, Jeffreys e Carnap, considerando a probabilidade como uma relação lógica entre duas proposições – a evidência e a hipótese – que determina o grau de implicação da hipótese pela evidência;
- ▷ o defendido por Ramsey, de Finetti e Savage – subjetivo – considerando a probabilidade como uma avaliação do grau de credibilidade que uma pessoa atribui a uma dada hipótese, na posse de uma evidência.

A Estatística Clássica, em geral, utiliza o conceito frequentista *a posteriori* de probabilidade, e suas inferências são obtidas unicamente por amostragem, a partir dos dados experimentais.

Por outro lado, a chamada Estatística Bayesiana utiliza o conceito subjetivo *a priori* de probabilidade para caracterizar certas hipóteses e, a partir da fórmula de Bayes, os dados amostrais são ponderados.

De qualquer modo, a componente amostral ou experimental é comum a ambos os métodos estatísticos: os clássicos e os bayesianos.

2.6. Distribuições limites de frequências

A tabela de frequências a seguir (Tab. 2.2), apresenta as idades dos 72 calouros que ingressaram no curso de Física da UERJ, no segundo semestre de 2001, com idades entre 17 e 37 anos distribuídas em 21 intervalos de 1 ano, representadas também no histograma da Fig. 2.4.

onde n_i é o número de alunos (frequência) que têm idades no intervalo $[x_i, x_{i+1})$, tal que $x_1 = 17$, $x_{i+1} = x_i + 1$ e $\sum_{i=1}^m n_i = N$ ($m = 21$ e $N = 72$). Assim, a média de idades é igual 21,33.

Desse modo, ao se escolher casualmente algum aluno da turma, a probabilidade (p_i) de que a idade desse aluno esteja no intervalo $[x_i, x_{i+1})$ é dada por

$$p_i = \frac{n_i}{N}$$

uma vez que dentre os N alunos apenas n_i deles têm idades entre x_i e x_{i+1} . Ou seja, a probabilidade, nesse caso, é igual à frequência relativa de idades no intervalo $[x_i, x_{i+1})$, pois, de acordo com a relação $\sum_{i=1}^m n_i = N$, esses números satisfazem à condição de normalização,

$$\sum_{i=1}^m p_i = \sum_{i=1}^m \frac{n_i}{N} = 1$$

Essa relação expressa simplesmente a certeza de que a idade de um aluno casualmente escolhido esteja no intervalo $[17, 37]$.¹⁰

¹⁰Esse intervalo, ou seja, o coleção de idades entre 17 e 37 anos é a população das idades dos alunos da turma.

idade (x_i - anos)	n_i
17 – 18	2
18 – 19	9
19 – 20	15
20 – 21	14
21 – 22	12
22 – 23	5
23 – 24	2
24 – 25	2
25 – 26	5
26 – 27	1
27 – 28	1
28 – 29	0
29 – 30	0
30 – 31	0
31 – 32	0
32 – 33	0
33 – 34	1
34 – 35	1
35 – 36	0
36 – 37	0
37 – 38	2

Tabela 2.2: Distribuição de frequências de idades de uma turma de 72 alunos.

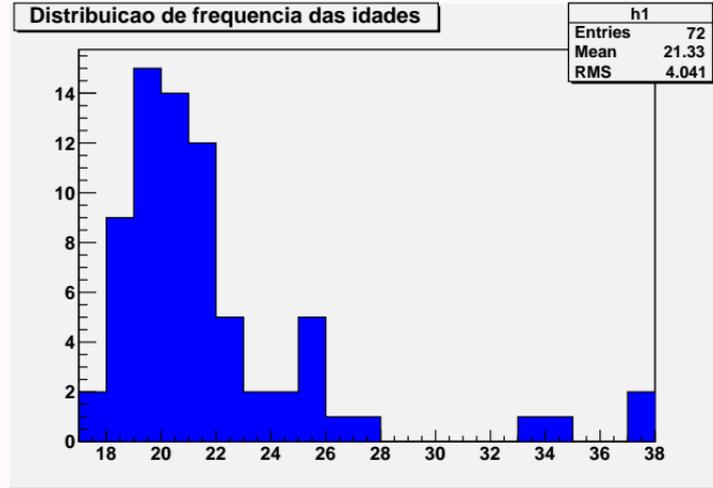


Figura 2.4: Histograma da Tab. 2.2.

De modo análogo, a probabilidade de que a idade (x) de um aluno escolhido pertença a um dos intervalos distintos $[x_i, x_{i+1})$ e $[x_j, x_{j+1})$ é dada por

$$p \{x \in [x_i, x_{i+1}) \text{ ou } x \in [x_j, x_{j+1})\} = \frac{n_i + n_j}{N} = p_i + p_j$$

uma vez que dentre as N idades possíveis apenas $(n_i + n_j)$ pertencem aos intervalos $[x_i, x_{i+1})$ e $[x_j, x_{j+1})$.

Se, após o sorteio do primeiro aluno, sorteia-se um segundo aluno, sem excluir a possibilidade de escolhê-lo novamente, a probabilidade de que a idade do primeiro sorteado (x^1) pertença ao intervalo $[x_j, x_{j+1})$, e a idade do segundo (x^2) ao intervalo $[x_k, x_{k+1})$ é dada por

$$p \{x^1 \in [x_i, x_{i+1}) \text{ e } x^2 \in [x_k, x_{k+1})\} = p(x_i^1 \text{ e } x_k^2) = p_i \cdot p_k$$

uma vez que para cada uma das n_i idades no intervalo $[x_i, x_{i+1})$ correspondem n_k idades no intervalo $[x_k, x_{k+1})$, e que existem N idades possíveis para a primeira escolha e para cada uma delas N possíveis escolhas para a segunda, ou seja,

$$\frac{n_i \cdot n_k}{N \cdot N} = p_i \cdot p_k$$

Ao se observar as idades (x) de um número cada vez maior de alunos ($N \gg 1$), como, por exemplo, as idades de todos os alunos de uma Universidade, ou de todas as Universidades de uma grande cidade ou de um país, os intervalos de cada classe de idade podem ser diminuídos até que a distribuição de frequências limite, $n(x)$, seja proporcional a uma distribuição contínua de probabilidades, a pdf $\rho(x)$, no intervalo ($a = 17, b = 37$),

$$\rho(x) = \frac{n(x)}{\int_a^b n(x) dx} \geq 0 \quad \Rightarrow \quad \int_a^b \rho(x) dx = 1$$

onde os limites de integração delimitam o domínio (D) de valores possíveis para as ocorrências de x , e a probabilidade de ocorrência de valores de x entre a e b é dada por ¹¹

$$P(a < x < b) = \int_a^b \rho(x) dx \quad (2.24)$$

Assim, qualquer distribuição de frequências devidamente normalizada representa uma distribuição de probabilidades para os possíveis valores da variável envolvida.

¹¹Nem todas as distribuições limites são contínuas. Distribuições limites contínuas derivam do fato de a variável observada estar associada a um domínio contínuo de valores. No caso do lançamento de dados ou de moedas, a distribuição limite também é discreta (Fig. 2.1).

2.7. A lei dos grandes números

Ao se escolher aleatoriamente idades de M alunos, por exemplo, uma amostra k com 16 alunos ($M = 16$), não excluindo as idades observadas previamente, a média da amostra k (\bar{x}_k) das idades é dada por

$$\bar{x}_k = \sum_{i=1}^M \frac{x_j}{M} \quad (\text{média amostral}) \quad (2.25)$$

Em geral, essa média amostral ou experimental \bar{x}_k (Fig. 2.5), será diferente da média esperada $\langle x \rangle$, ou seja,

$$\bar{x}_k = 19,88 \neq \langle x \rangle = 21,33$$

No entanto, no limite de um número muito grande de amostras ($N \rightarrow \infty$), a média das médias amostrais se aproxima da média esperada $\langle x \rangle$ das idades da turma (Fig. 2.6), isto é,

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{k=1}^N \bar{x}_k \rightarrow \langle x \rangle \quad (2.26)$$

Desse modo, a lei dos grandes números mostra que a média amostral, segundo o critério de Fisher (Cap. ??), é um estimador consistente e não tendencioso para o valor esperado de qualquer variável aleatória associada a uma distribuição de probabilidades, não necessariamente gaussiana.

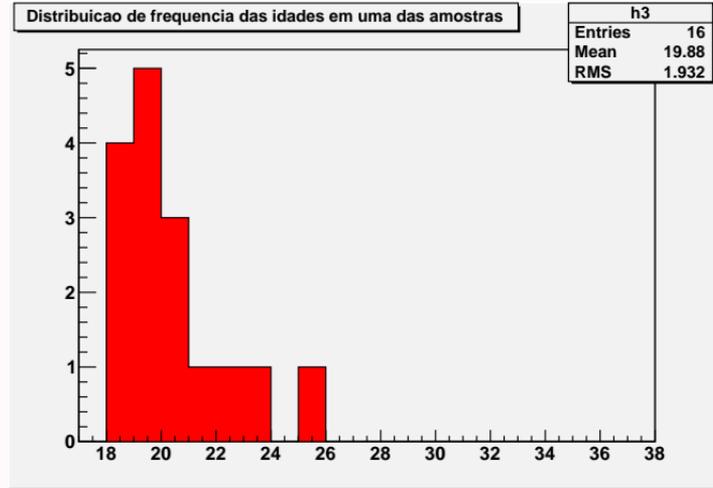


Figura 2.5: Distribuição de idades de uma amostra (simulada) de 16 idades da Tab. 2.2.

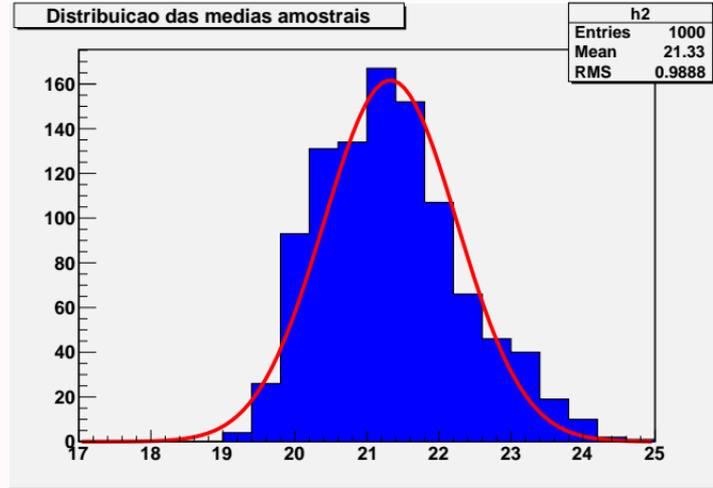


Figura 2.6: Distribuição das médias de 1000 amostras de 16 idades da Tab. 2.2.

2.8. A distribuição uniforme

Além da simulação direta de fenômenos simples, como o lançamento de dados ou de sorteio de números, os dois principais resultados da teoria de probabilidades – a lei dos grandes números e o teorema do limite central – podem ser obtidos a partir de simulações que utilizam um gerador de números aleatórios uniformemente distribuídos em um dado intervalo.¹²

A distribuição uniforme (Fig. 2.7), em um intervalo (a, b) , de uma variável aleatória contínua x no intervalo $(-\infty, \infty)$, é definida por

$$\rho(x|a, b) = \begin{cases} 0 & x \leq a \\ \frac{1}{b-a} & a < x < b \\ 0 & x \geq b \end{cases} \quad (2.27)$$

Assim, a probabilidade associada a qualquer intervalo (α, β) , de mesma amplitude $\beta - \alpha = \delta$, é constante, ou seja,

$$P(\alpha < x < \beta) = \frac{\delta}{\Delta} \quad (\Delta = b - a)$$

¹²Métodos baseados nesse procedimento são denominados métodos de simulação a Monte Carlo.

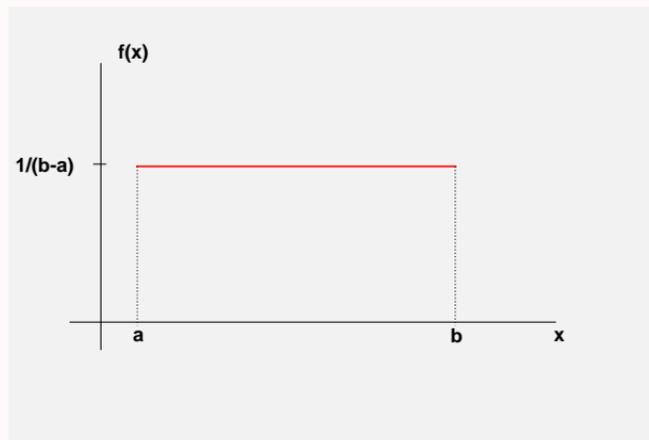


Figura 2.7: Distribuição uniforme de números aleatórios, em um intervalo (a, b) .

Os principais parâmetros de uma distribuição uniforme são:

$$\left\{ \begin{array}{l} E(x) = \langle x \rangle = \frac{1}{b-a} \int_a^b x \, dx = \frac{1}{2}(a+b) \quad (\text{valor esperado}) \\ V(x) = \sigma_x^2 = \frac{1}{b-a} \int_a^b [x - \langle x \rangle]^2 \, dx = \frac{1}{12}(b-a)^2 \quad (\text{variância}) \end{array} \right.$$

As Fig. 2.8 e 2.9 mostram a distribuição de números e a correlação entre os sucessivos números aleatórios obtidos a partir do programa gerador de números aleatórios utilizado pelo pacote de análise de dados do CERN, o ROOT (versão 5).

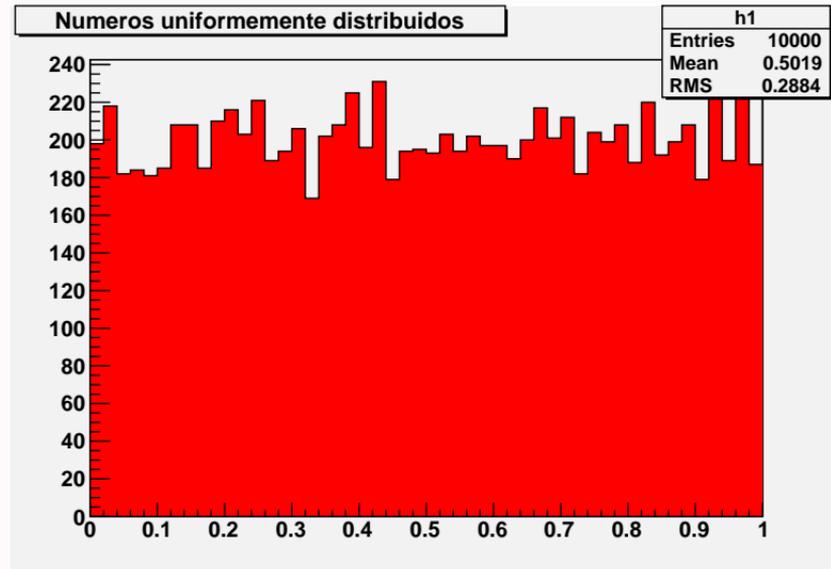


Figura 2.8: Distribuição uniforme de 10000 números aleatórios no intervalo (0, 1), com o gerador do ROOT.

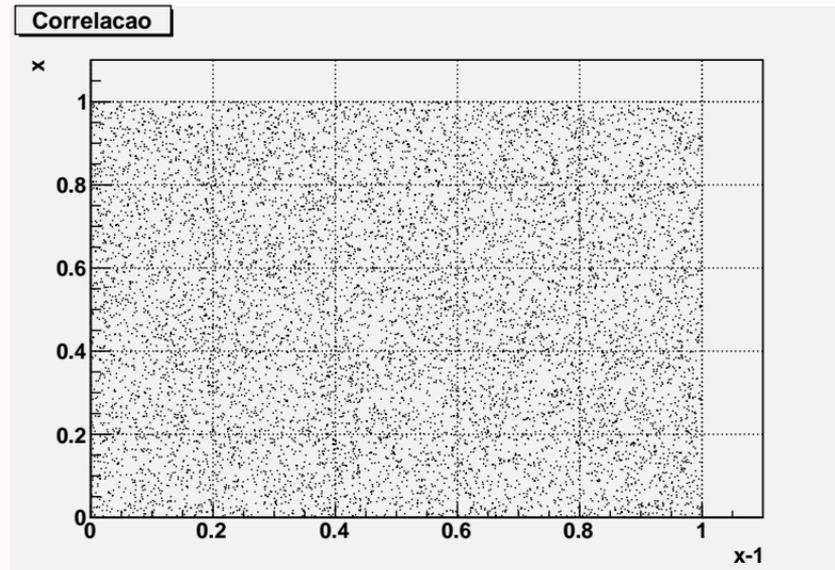


Figura 2.9: Correlação dos 10000 números aleatórios gerados com o ROOT, no intervalo (0, 1).

2.9. A distribuição binomial

A mais utilizada distribuição limite de probabilidade é a chamada distribuição normal, ou distribuição de Gauss (de Moivre - 1733 e Gauss - 1809), por representar a distribuição de frequência limite das medidas diretas de uma grandeza física, quando o processo de medição está sujeito a várias pequenas parcelas de incertezas.

Outras distribuições associadas a fenômenos aleatórios de baixa expectativa de ocorrência, como a distribuição de Poisson (1837), ou a situações dicotomizadas, como a distribuição binomial (Bernoulli - 1713), podem ser estabelecidas, e possibilitam também a dedução de vários resultados estatísticos.

Essas distribuições estão relacionadas entre si, e podem ser obtidas como casos limites da distribuição binomial, a qual pode ser estabelecida e utilizada sempre que o resultado de um experimento está associado à repetição de eventos que podem ser expressos por uma dicotomia, ou seja, no estudo de processos, denominados **provas de Bernoulli**, nos quais pode-se dizer que há apenas duas possibilidades excludentes de ocorrência ou realização de uma determinada condição.

Por exemplo, o problema do cálculo das probabilidades de ocorrência de cada uma das seis faces de um dado após N lançamentos pode ser dicotomizado observando-se que a probabilidade (p) de ocorrência de uma determinada face é $1/6$, e das outras restantes $1 - p = 5/6$. Nesse caso, a probabilidade de m ocorrências de uma dada face é dada por

$$P_m = \frac{N!}{m!(N - m)!} p^m (1 - p)^{N - m}$$

O protótipo de um experimento dicotomizado é o lançamento de uma moeda, no qual apenas dois resultados são possíveis (cara ou coroa), e a probabilidade de sucesso de qualquer um dos resultados é $1/2$.

Seja A um acontecimento esperado, \bar{A} o seu complemento, $P(A) = a$ a probabilidade de ocorrência de A em cada tentativa e $P(\bar{A}) = b$ a probabilidade complementar de ocorrência de \bar{A} em cada tentativa, tal que:

$$P(A) + P(\bar{A}) = a + b = 1$$

A probabilidade $P_m^n(A)$ de m ocorrências de A em n tentativas, pode ser expressa por:

$$P_m^n(a) = C_m^n a^m (1 - a)^{n-m}$$

onde $a^m (1 - a)^{n-m}$ é a probabilidade de m ocorrências de A e $(n - m)$ ocorrências de \bar{A} em n tentativas se houvesse apenas uma configuração possível e C_m^n é a multiplicidade de cada configuração.

Utilizando-se do exemplo da moeda, onde A é a ocorrência de cara e $B = \bar{A}$ a ocorrência de coroa, pode-se notar, pela tabela 2.3, que a multiplicidade de cada configuração é dada pela fórmula de combinação:

$$C_m^n = \frac{n!}{m! (n - m)!}$$

Nesse caso, P_m^n representa a probabilidade de ocorrência de m caras em n lançamentos de uma moeda.

A fórmula do binômio de Newton,

$$(a + b)^n = \sum_{m=0}^n C_m^n a^m b^{n-m}$$

e a relação $a + b = 1$ permitem a verificação da condição de normalização,

$$\sum_{m=0}^n P_m^n(a) = 1$$

ocorrências (m)	configurações	multiplicidade (C_m^n)
4	$A A A A$	1
3	$A A A B$ $A A B A$ $A B A A$ $B A A A$	4
2	$A A B B$ $A B A B$ $B A A B$ $A B B A$ $B A B A$ $B B A A$	6
1	$A B B B$ $B A B B$ $B B A B$ $B B B A$	4
0	$B B B B$	1

Tabela 2.3: Quadro de ocorrências (m) de caras (A) em $n = 4$ lançamentos de uma moeda.

Os parâmetros característicos (média e variância) associados à distribuição binomial podem ser deduzidos a partir da definição da chamada **função geratriz**,

$$\phi(t) = (at + b)^n \rightarrow \phi(1) = 1$$

Assim,

$$\phi(t) = \sum_{m=0}^n \underbrace{C_m^n a^m b^{n-m}}_{P_m^n} t^m$$

e,

$$\frac{d\phi}{dt} = \sum_{m=1}^n m P_m^n t^{m-1} = na(at + b)^{n-1}$$

$t = 1$, implica

$$\boxed{\sum_{m=0}^n m P_m^n = \langle m \rangle = \mu = na} \quad (\text{média})$$

derivando mais uma vez,

$$\begin{aligned} \frac{d^2\phi}{dt^2} &= \sum_{m=2}^n m(m-1) P_m^n t^{m-2} \\ &= \sum_{m=2}^n (m^2 P_m^n - m P_m^n) t^{m-2} \\ &= n(n-1)a^2(at + b)^{n-2} \end{aligned}$$

$t = 1$, implica

$$\underbrace{\sum_{m=0}^n m^2 P_m^n}_{\langle m^2 \rangle} - \underbrace{\sum_{m=0}^n m P_m^n}_{\langle m \rangle = na} = n(n-1)a^2$$

$$\langle m^2 \rangle = n(n-1)a^2 + na$$

logo, a variância será dada por:

$$\sigma_m^2 = \langle m^2 \rangle - \langle m \rangle^2 = na \underbrace{(1-a)}_b$$

ou seja, $\sigma_m^2 = na(1-a)$ (variância)

Observando-se algumas representações gráficas para a distribuição binomial (figuras 2.10, 2.11, 2.12), pode-se notar que se o número de tentativas e o número de sucessos forem relativamente grandes, em quaisquer dos casos a distribuição resultante apresenta um máximo acentuado em torno de seu valor médio (limite de Gauss – figuras 2.10b e 2.11b). Entretanto, se o número de tentativas for grande mas o número de sucessos muito pequeno, a distribuição resultante é praticamente exponencial (limite de Poisson – figura 2.12).

Genericamente, qualquer processo que possui uma variável que obedece a uma distribuição binomial, ou seja, que resulte da repetição de Provas de Bernoulli, é denominado um Processo de Bernoulli e é representado por

$$B(m, n|p)$$

onde n é o número total de Provas de Bernoulli, m o número de sucessos de um evento e p a probabilidade desse evento, também denominada de parâmetro da distribuição binomial.

Um problema inverso, de inferência estatística, que surge associado à distribuição binomial é quando apesar de se saber que um processo é constituído por Provas de Bernoulli, não se sabe *a priori* a probabilidade p de

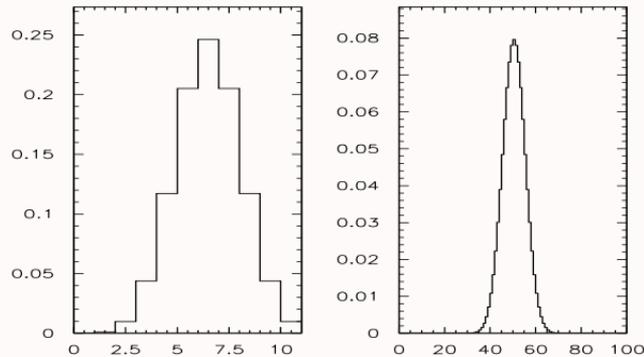


Figura 2.10: Distribuições de probabilidades de ocorrências de caras para, (a) $n = 10$ e (b) $n = 100$, lançamentos de uma moeda ($a = 0.5$).

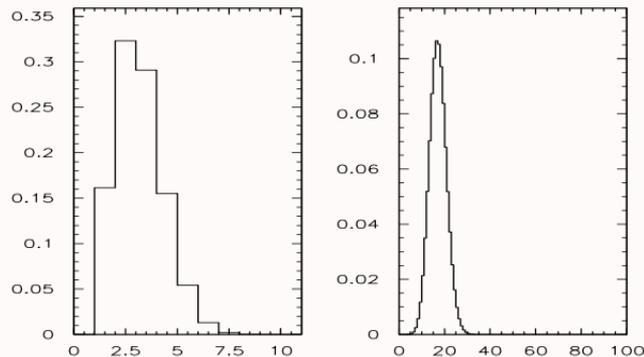


Figura 2.11: Distribuições de probabilidades de ocorrências do número 1, para, (a) $n = 10$ e (b) $n = 100$, lançamentos de um dado ($a = 1/6$).

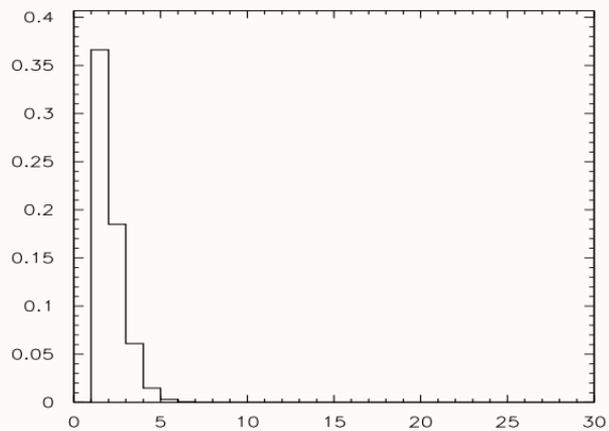


Figura 2.12: Distribuição binomial para $a = 0.01$ e $n = 100$.

cada Prova.

Por exemplo, ao se lançar cinco CD's (compact discs)¹³ ao alto 100 vezes e observar o número de vezes que a estampa do disco voltou-se para cima, obteve-se a seguinte tabela de frequências,

m	0	1	2	3	4	5
f_m	2	14	20	34	22	8

onde m é o número de possíveis êxitos (estampa para cima) em cada lançamento e f_m a frequência associada.

Nesse caso, a probabilidade p de sucesso, apesar de próxima de 0.5, não é conhecida. Entretanto, desde que $n = 5$, a média teórica seria igual a $\mu = 5p$.

Uma vez que a média observada foi de

$$\bar{m} = \sum_{m=0}^5 m f_m = \frac{284}{100} = 2.84$$

Igualando-se ambas as quantidades ($5p = \bar{m}$), a probabilidade de sucesso em cada lançamento de um CD é estimada em $p = 0.568$.

A determinação do parâmetro p de uma distribuição binomial que os dados, por hipótese, devem obedecer, é equivalente ao ajuste da função de distribuição $B(m, n|p)$ à distribuição de dados.

A figura 2.13 mostra a comparação da distribuição dos dados com a distribuição binomial, de parâmetro $p = 0.568$, ajustada.

¹³Lançar uma vez cinco CD's ou lançar um CD cinco vezes, assim com lançar uma moeda n vezes ou n moedas uma vez, são chamados eventos equivalentes, ou seja, estão associados a mesma probabilidade de ocorrência.

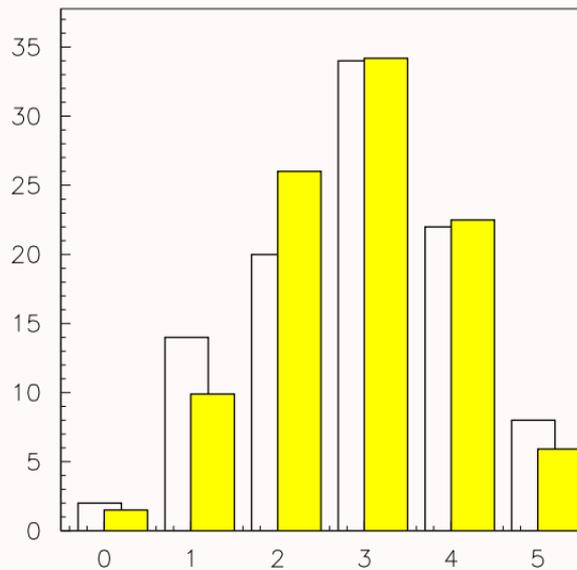


Figura 2.13: Comparação de uma distribuição de dados cuja média é 2.84 com uma distribuição binomial (pontilhada) de mesma média e $p = 0.568$.

De imediato, pode-se perguntar se apenas uma comparação visual é suficiente para garantir que houve um bom acordo entre os dados e a distribuição binomial, ou seja, a hipótese de que os dados são distribuídos binomialmente é razoável?

Um dos objetivos desse ensaio é o de proporcionar ao leitor os instrumentos necessários de avaliação de uma hipótese, como a do ajuste de uma distribuição de frequências ou probabilidades à uma coleção de dados, ou seja, estabelecer critérios que proporcionem a definição de um nível de significância ¹⁴ para esse tipo de ajuste.

O bom acordo entre essas distribuições é chamado na Estatística de aderência. Assim, o processo de verificação de um determinado nível de significância previamente escolhido, para julgar a aderência entre duas distribuições, é denominado teste de aderência.

2.10. A distribuição de Poisson

A distribuição de Poisson pode ser deduzida como um extremo ou aproximação de uma distribuição binomial $B(m, n|a)$ no caso em que o número de eventos esperados é muito menor que o número total de eventos possíveis. Ou seja, se $a \ll 1$, $m \ll n$ e $n \gg 1$, a multiplicidade pode ser aproximada por

$$\frac{\overbrace{n(n-1)\dots(n-m+1)}^{\approx n}}{m!} \approx \frac{n^m}{m!}$$

¹⁴Um nível de significância é a probabilidade de ocorrência de eventos que contrariam uma dada hipótese.

e, o fator de probabilidade por

$$a^m b^{\overbrace{n-m}^{\approx n}} \approx a^m (1-a)^n \approx a^m \underbrace{[(1-a)^{1/a}]^\mu}_{1/e} \approx a^m e^{-\mu}$$

Assim, após a combinação dos fatores a distribuição de Poisson será dada por

$$P_m(\mu) = \frac{\mu^m}{m!} e^{-\mu}$$

Uma característica fundamental da distribuição de Poisson é que ela só depende de um parâmetro, a média (μ), pois a variância pode ser expressa também, desde que $a \ll 1 \Rightarrow b \approx 1$, por

$$\sigma_m^2 = na = \mu$$

Uma das razões de seu amplo uso é que existem várias situações nas quais os valores de n e a não são bem definidos mas a média pode ser facilmente determinada.

De um ponto de vista mais amplo, qualquer fenômeno que envolve um grande número de processos idênticos com probabilidades pequenas, segue uma distribuição de Poisson. Por esse motivo, a distribuição de Poisson é chamada fórmula de probabilidades de eventos raros. Assim, em populações constituídas por um número muito grande de sistemas idênticos, se esta é a única informação, a distribuição de Poisson pode ser utilizada na extração de estimativas. Por exemplo, o decaimento de núcleos e partículas, ou a contagem de insetos, bactérias e vírus, ou a de partículas geradas em uma colisão, em intervalos de energia, momentum ou tempo, são caracterizados como processos de Poisson e representados, genericamente, por $P(m|\mu)$

A figura 2.14 mostra várias representações gráficas de distribuições de Poisson, onde pode-se observar que para $\mu > 5$ as distribuições apresentam um máximo em torno de seus valores médios (limite de Gauss).

2.11. A distribuição de Gauss

Foi estabelecida, inicialmente, por de Moivre como um limite da distribuição binomial, e por Gauss (1809) e Laplace na análise de erros de medidas.¹⁵

Como limite de uma distribuição binomial $B(m, n|a)$ ela pode ser obtida quando o número médio de ocorrências de um processo for muito grande, independentemente de sua probabilidade de ocorrência, ou seja, $n, m \rightarrow \infty$.

Considerando-se $m = x$ como uma variável contínua, a probabilidade $P_m^n(a)$ é substituída por uma distribuição de probabilidades, $\rho(x)$, que pode ser expressa por,

$$\rho(x) = \frac{n!}{x!(n-x)!} a^x b^{n-x}$$

onde $b = 1 - a$.

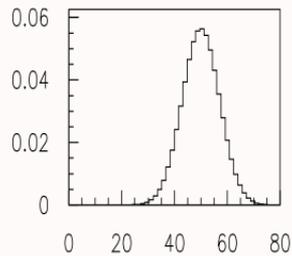
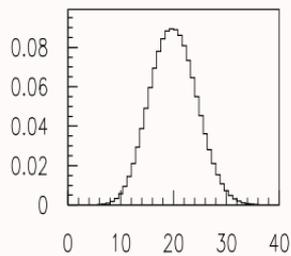
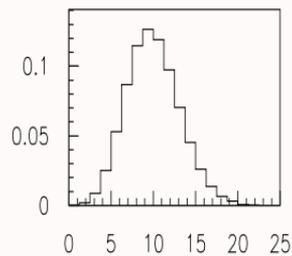
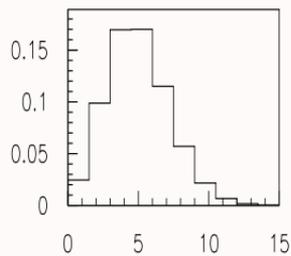
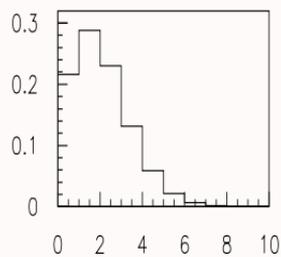
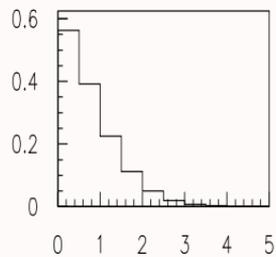
Tomando-se o logaritmo de $\rho(x)$, resulta

$$\ln n! - \ln x! - \ln(n-x)! + x \ln a + (n-x) \ln b$$

Utilizando-se a aproximação de Stirling ($x \rightarrow \infty$),

$$\ln x! \approx x \ln x \quad \text{e} \quad \frac{d \ln x!}{dx} \approx \ln x$$

¹⁵É conhecida na Estatística como distribuição Normal.



a derivada de $\ln \rho(x)$,

$$\frac{d \ln \rho(x)}{dx} = -\ln x + \ln(n-x) + \ln a - \ln b$$

mostra que, $= \ln \left(\frac{n-x}{x} \frac{a}{b} \right)$

$$\left. \frac{d \ln \rho(x)}{dx} \right|_{\hat{x}} = 0 \Rightarrow \left. \frac{a(n-x)}{bx} \right|_{\hat{x}} = 1$$

$$\hat{x} \underbrace{(a+b)}_1 = an = \mu$$

a média μ determina um valor máximo para a distribuição de probabilidade $\rho(x)$.

Expandindo-se o logaritmo de $\rho(x)$ em torno da média,

$$\ln \rho(x) = \ln \rho(\mu) + \frac{1}{2} \left. \frac{d^2 \ln \rho}{dx^2} \right|_{\mu} (x - \mu)^2$$

e, desde que,

$$\frac{d^2 \ln \rho}{dx^2} = -\frac{1}{x} - \frac{1}{n-x} = -\frac{n}{x(n-x)}$$

e,

$$\left. \frac{d^2 \ln \rho}{dx^2} \right|_{\mu} = -\frac{n}{na \underbrace{(n-na)}_{n(1-a)}} = -\frac{1}{nab} = -\frac{1}{\sigma_x^2}$$

a distribuição de probabilidade de ocorrência de um valor x pode ser expressa por

$$\rho(x|\mu, \sigma_x^2) = \rho(\mu) e^{-\frac{(x - \mu)^2}{2\sigma_x^2}}$$

O valor máximo $\rho(\mu)$ pode ser determinado pela condição de normalização ¹⁶, com os limites estendidos de $-\infty$ a ∞ ,

$$\int_{-\infty}^{\infty} \rho(x) dx = 1 = \rho(\mu) \sigma_x \sqrt{2} \underbrace{\int_{-\infty}^{\infty} e^{-u^2} du}_{\sqrt{\pi}}$$

¹⁶Integrais do tipo $I = \int_{-\infty}^{\infty} e^{-\alpha x^2} dx$

podem ser calculadas como:

$$\begin{aligned} I^2 &= \left(\int_{-\infty}^{\infty} e^{-\alpha x^2} dx \right) \left(\int_{-\infty}^{\infty} e^{-\alpha y^2} dy \right) \\ &= \int_{x=-\infty}^{\infty} \int_{y=-\infty}^{\infty} e^{-\alpha(x^2 + y^2)} dx dy \end{aligned}$$

Definido-se as coordenadas polares

$$\begin{cases} x = r \cos \theta \\ y = r \sin \theta \end{cases}$$

a integral $\int \int_R f(x, y) dx dy$

onde,

Assim,

$$u = \frac{x - \mu}{\sigma_x \sqrt{2}} \Rightarrow du = \frac{dx}{\sigma_x \sqrt{2}}$$
$$\rho(\mu) = \frac{1}{\sigma_x \sqrt{2\pi}}$$

de modo que a distribuição normalizada de Gauss é dada por

pode ser transformada em

$$\int \int_R f[x(r, \theta), y(r, \theta)] \underbrace{\left| \frac{\partial(x, y)}{\partial(r, \theta)} \right|}_r dr d\theta$$

Assim,

$$I^2 = \int_0^{2\pi} \int_0^\infty e^{-\alpha r^2} dr d\theta$$
$$= 2\pi \int_0^\infty e^{-\alpha r^2} dr \quad (u = r^2)$$
$$= \pi \int_0^\infty e^{-\alpha u} du = \frac{\pi}{\alpha}$$

e,

$$I = \sqrt{\frac{\pi}{\alpha}}$$

$$\rho_G(x|\mu, \sigma_x^2) = \frac{1}{\sigma_x \sqrt{2\pi}} e^{-\frac{(x - \mu)^2}{2\sigma_x^2}}$$

A distribuição de Gauss associada a uma variável aleatória x depende de 2 parâmetros (média e variância) e, usualmente, é representada também por

$$\mathcal{N}_x(\mu, \sigma_x^2)$$

Em relação à variável reduzida adimensional

$z = \frac{x - \mu}{\sigma_x}$, obtém-se a chamada **distribuição normal padrão**,

$$\mathcal{N}_z(0, 1)$$

cuja variável aleatória associada z tem valor médio nulo e desvio padrão unitário.

A distribuição de Gauss é uma densidade de probabilidade (fdp) e, as probabilidades de ocorrências dos

valores da variável reduzida z , no intervalo $(-\infty, t)$, ou seja, os valores de sua fd $\Phi(t)$ ¹⁷, dadas por

$$P_G(z \leq t) = \Phi(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^t e^{-\frac{z^2}{2}} dz$$

são encontradas em diversas tabelas, uma vez que essa integral não é imediata.

Utilizando-se as aproximações, para grandes números, para as distribuições binomial,

$$P_m^n(a) = \frac{1}{\sqrt{2\pi na(1-a)}} e^{-\frac{1}{2} \frac{(m-na)^2}{na(1-a)}}$$

¹⁷Por convenção a função de distribuição acumulada (fd) de uma distribuição normal padrão é representada por $\Phi(t)$.

Algumas de suas propriedades são:

- $P_G(x \leq t) = F(t) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^t e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$
 $= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{(t-\mu)/\sigma} e^{-\frac{z^2}{2}} dz = \Phi\left(\frac{t-\mu}{\sigma}\right)$
- $\Phi(-t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{-t} e^{-\frac{z^2}{2}} dz = 1 - \Phi(t)$

e de Poisson,

$$P_m(\mu) = \frac{1}{\sqrt{2\pi\mu}} e^{-\frac{1}{2} \frac{(m - \mu)^2}{\mu}}$$

pode-se compará-las com as gaussianas de mesma média e desvio padrão (figura 2.15).

Apesar de deduzida a partir de um limite da distribuição binomial, essa é apenas uma das inúmeras situações em que a distribuição gaussiana pode ocorrer. Gauss e Laplace foram levados à mesma por uma linha de raciocínio diferente, na investigação da lei de distribuição que os erros ou desvios em uma medição devem obedecer de modo que a média aritmética das medidas seja o valor mais provável para o valor de referência de uma grandeza medida.

A tendência das distribuições amostrais de alguns parâmetros à distribuição gaussiana, quando são consideradas grandes amostras,¹⁸ ou seja, que a média e alguns parâmetros tendem, usualmente, a serem distribuídos gaussianamente, quer os dados originais sejam ou não, fez com que a distribuição gaussiana, no início das aplicações estatísticas, fosse amplamente utilizada nos mais diversos campos de estudos e situações.

Assim adquiriu-se um sentimento de que ela seria o ideal para o qual todas as demais distribuições deveriam aproximar-se. Daí a origem do cognome *normal*. Entretanto, ao final do século XIX e no início do século XX, principalmente, com os trabalhos de Pearson e Fisher, foi se constatando que a ocorrência de uma curva normal em muitos casos, longe de ser esperada, é que seria uma situação abnormal.

¹⁸Esse fato é justificado pelo chamado Teorema do Limite Central, que estabelece que se os desvios de uma variável aleatória, como os erros na medição de uma grandeza, decorrem de uma soma de um grande número de desvios elementares independentes, eles são distribuídos normalmente (apêndice 5.6). Por exemplo, uma quantidade que decorre de efeitos acumulativos de variáveis aleatórias independentes, ou seja, que é distribuída de forma aproximadamente normal é a altura das pessoas de um grupo. Entretanto, o peso das pessoas de um grupo não obedece à uma distribuição normal.

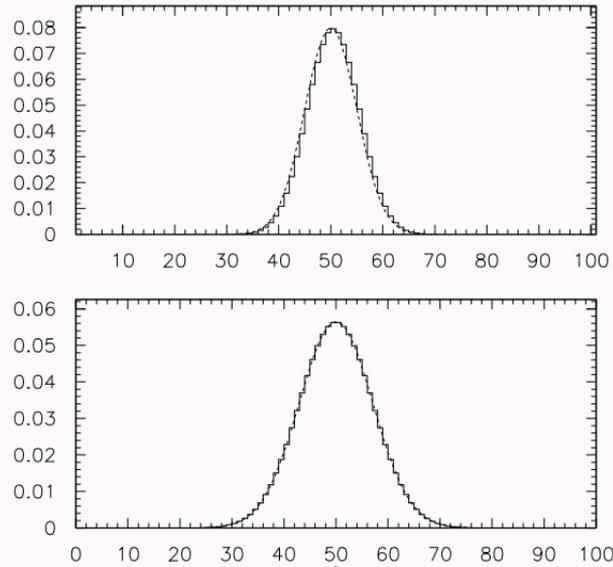


Figura 2.15: Comparação das distribuições binomial e de Poisson com gaussianas de mesma média e desvio padrão.

Mesmo na teoria dos erros a crença na validade da distribuição gaussiana, segundo Poincaré em seu *Calcul des Probabilités* é que: “*Todos acreditam na lei dos erros, o experimentador porque pensa que é uma lei matemática e, o matemático porque pensa que é um fato experimental.*”

É interessante notar que, segundo F. N. David ¹⁹, o registro mais antigo sobre um estudo de uma distribuição amostral deve-se a Galileu (1623).

O caso abordado por ele foi o da distribuição da média em um grande número N de lançamentos simultâneos de 3 dados. Ampliando-se o estudo para lançamentos de 1, 2, 3 e 6 dados pode-se observar (figura 2.16) a evolução do comportamento da média, ou seja, a tendência de distribuir-se normalmente a medida que o número de dados aumenta, o que, em essência, constitui o Teorema do Limite Central.

- **Tamanho de amostras**

O processo de amostragem, isto é, de seleção de amostras com distribuições semelhantes à distribuição parental pode ser praticamente impossível, se a distribuição parental for totalmente desconhecida. No caso de estimativas de valores esperados de uma grandeza, a partir da média de uma coleção de medidas, uma vez que a tanto a distribuição amostral quanto a parental são gaussianas, o problema da amostragem é simplificado pela determinação apenas do tamanho da amostra necessária para a estimativa.

Assim, se uma distribuição parental gaussiana, associada a uma grandeza x , tem valor esperado μ e desvio padrão σ_x , a probabilidade de que uma estimativa (média) \bar{x}_n , calculada a partir de uma amostra de tamanho n , difira de μ , no mínimo de um valor δ , seja menor ou igual a um valor previamente escolhido a é expressa

¹⁹Em *Dicing and Gaming (on the History of Probability)*, Biometrika, V42 (1955) [?]

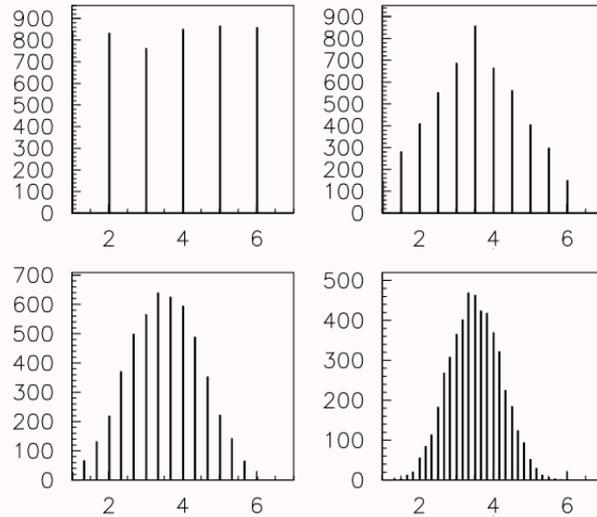


Figura 2.16: Distribuições da média no lançamento simultâneo de (a) 1 dado, (b) 2 dados, (c) 3 dados e (d) 6 dados

por

$$P(|\bar{x}_n - \mu| \geq \delta) \leq a$$

E, a partir da condição

$$\begin{aligned} a &= P(\bar{x}_n - \mu \geq \delta) + P(-\bar{x}_n + \mu \geq \delta) \\ &= P(\bar{x}_n \geq \mu + \delta) + P(\bar{x}_n \leq \mu - \delta) \\ &= 1 - P(\bar{x}_n \leq \mu + \delta) + P(\bar{x}_n \leq \mu - \delta) \\ &= 1 - \Phi(\delta/\sigma_x\sqrt{n}) + \underbrace{\Phi(-\delta/\sigma_x\sqrt{n})}_{1-\Phi(\delta/\sigma_x\sqrt{n})} \\ &= 2 \left[1 - \Phi\left(\frac{\delta\sqrt{n}}{\sigma_x}\right) \right] \end{aligned}$$

ou seja,

$$\Phi\left(\frac{\delta\sqrt{n}}{\sigma_x}\right) = 1 - \frac{a}{2}$$

determina-se um valor z_a tal que

$$\Phi(z_a) = 1 - \frac{a}{2}$$

e, o tamanho n mínimo da amostra será dado por

$$n = \left(\frac{\sigma_x z_a}{\delta} \right)^2$$

Por exemplo, segundo um fabricante, o fator de amplificação de um tipo de foto-multiplicadoras distribui-se com desvio padrão igual a 10% de seu valor nominal A ($\sigma = 0.1A$). Para estimar-se o número de tubos necessários à realização de um teste de qualidade no qual a probabilidade de que o fator de amplificação médio difira de seu valor nominal, por mais de 6% ($\delta = 0.06A$), seja menor que 0.05 (α), desde que z_a é dado por

$$\Phi(z_a) = 1 - \frac{0.05}{2} \implies z_a = 1.96$$

o número n de tubos necessários é dado por

$$n = \frac{0.1 \times 1.96}{0.06} = 15.3 \implies n = 16 \text{ tubos}$$

2.12. Exercícios

- 2.12.1) Em uma Universidade, 60% dos estudantes matriculados em seus cursos são homens. Dentre eles, 45% são de cursos da área tecnológica, enquanto somente 15% das alunas são dessa área.

Determine a probabilidade de que uma aluna seja sorteada em um lote de fichas de estudantes dos cursos da área tecnológica.

- 2.12.2) Em uma certa noite, um táxi atropelou uma pessoa e fugiu. Apenas dois tipos de táxis operam na cidade, 15% são azuis e o restante laranja.

Uma testemunha identificou o táxi envolvido no acidente como sendo azul, e a polícia constatou que, nas mesmas circunstâncias da noite do acidente, a fração de acertos da testemunha na identificação de cores foi de 80%.

Determine a probabilidade de ter sido realmente azul o táxi envolvido no acidente.

- 2.12.3) Enquanto 7% das mamografias acusam o câncer quando ele não existe (taxa de falsos-positivos), 10% não acusam a doença quando ela existe (taxa de falsos-negativos). Sabendo de que a incidência sobre a população é cerca de 0.8%, determine a probabilidade de que uma mulher esteja com câncer ao receber um resultado de teste positivo.

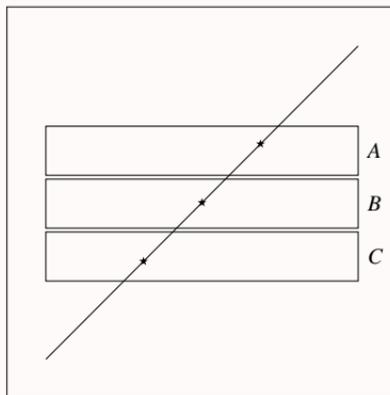
- 2.12.4) Três urnas têm a seguinte composição: a primeira contém 5 bolas brancas e 6 pretas, a segunda contém 4 brancas e 5 pretas, e a terceira 4 brancas e 4 pretas.

Após escolher por acaso uma urna e se retirar uma bola preta, determine a probabilidade de que a sorteada tenha sido extraída da terceira urna.

2.12.5) Seja x uma variável contínua, como as possíveis posições de uma partícula confinada em uma região de dimensão a , cuja densidade de probabilidade $\rho(x)$ é proporcional a função $\sin \frac{\pi}{a}x$.

Determine o valor esperado médio e variância.

2.12.6) A figura abaixo representa um sistema de detecção de múons incidentes, constituído por três tubos (PDT)²⁰ A , B e C .



Se a localização do ponto de ionização em cada tubo é determinada com 60% de eficiência, ou seja, com

²⁰PDT (Proportional Drift Tubes) são câmaras de fios, onde um gás é ionizado pela passagem de uma partícula carregada como o múon e, através da medida do tempo de deriva dos íons até o fio, no interior do tubo, localiza-se o ponto onde a partícula incidente ionizou o gás.

probabilidade igual a 0.6, e a reconstrução da trajetória de um múon requer a determinação de pelo menos três pontos em câmaras distintas, a eficiência do sistema é dada por

$$B(3, 3|0.6) \implies 21.6\%$$

Determine as eficiências para sistemas compostos por quatro e cinco câmaras.

- 2.12.7) Qual a probabilidade de que dentre 720 pessoas, exatamente duas aniversariem num mesmo dia? (compare binomial e Poisson)
- 2.12.8) Se a probabilidade de se fazer 13 pontos numa extração da loteria esportiva é $(1/3)^{13}$. Qual a expectativa de que de um total de 6 milhões de apostadores ao menos um acerte os 13 pontos, numa extração?
- 2.12.9) Cada uma das 15 questões de um teste tem 4 alternativas e apenas uma delas é correta. Desse modo, a probabilidade (p), *a priori*, de acerto ao acaso de uma questão é $1/4$.
- Determine a distribuição de probabilidades de acertos ao acaso de m das 15 questões.

$$P_0 = P_6 = P_{11} =$$

$$P_1 = P_7 = P_{12} =$$

$$P_2 = P_8 = P_{13} =$$

$$P_3 = P_9 = P_{14} =$$

$$P_4 = P_{10} = P_{15} =$$

$$P_5 =$$

- Represente em um histograma
- Se 1000 alunos fizerem o teste respondendo as questões ao acaso, quantos, em média, acertarão pelo menos 3 questões?

2.12.10) Uma máquina produz peças das quais 1% são defeituosas. Em um lote de 400 peças, qual a expectativa de que menos de duas sejam defeituosas?

2.12.11) Uma caixa de banco atende cerca de 150 clientes por hora. Qual a probabilidade de que ela atenda:

a) nenhum cliente em 4 minutos? (e^{-10})

b) no máximo 2 clientes em 2 minutos? (0.125)

2.12.12) Um problema clássico envolvendo a distribuição de Poisson é o experimento de Rutherford-Geiger, da contagem, em intervalos de 7.5s, num total de 2608 intervalos, do número de partículas α emitidas por uma amostra de polônio.

A tabela abaixo, mostra as frequências f_m correspondentes ao número m de contagens em cada intervalo.

m	f_m
0	57
1	203
2	383
3	525
4	532
5	408
6	273
7	139
8	45
9	27
10	10
11	4
12	2
13	0
14	0
$\sum_{m=0}^{14} f_m = 2608$	

- Determine o número médio de contagens em cada intervalo de 7.5s.
- Supondo que a distribuição de frequências obedece a um processo de Poisson, compare a distribuição dos dados com a distribuição de Poisson de parâmetro $\mu = 3.87$.
- Há uma boa aderência entre as distribuições?

2.12.13) Em um grupo de pessoas a altura média é de 170cm com desvio padrão de 5cm. Calcule a altura acima

da qual estão os 10% mais altos.

- 2.12.14) A média dos diâmetros dos rolamentos de esfera produzidos por uma determinada máquina é de 0.482cm com desvio padrão de 0.004cm. Uma peça é considerada defeituosa se tiver mais que 0.491cm ou menos que 0.473cm. Qual a porcentagem de peças defeituosas produzidas?

Métodos de simulação a Monte Carlo

Todo experimento deve começar com um modelo que estime os resultados esperados.

R. Fisher

Experimentos em Física de Altas Energias, como aqueles realizados em grandes aceleradores de partículas, como o Tevatron do Fermilab,²¹ nos EUA, ou o LHC do CERN,²² na Suíça, constituem ambientes nos quais os métodos estatísticos são utilizados em toda sua plenitude, desde a simulação de eventos e de detectores, até a extração e a comparação de resultados experimentais e teóricos.

Métodos probabilísticos de simulação de eventos a Monte Carlo ou, simplesmente, métodos Monte Carlo, são técnicas numéricas construídas a partir de distribuições uniformes de números aleatórios, que permitem a geração de amostras de eventos aleatórios com distribuições de probabilidades conhecidas *a priori*, ou a determinação de quantidades (como integrais de funções), associadas ou não à processos aleatórios.

A designação alude às roletas do famoso cassino Monte Carlo, em Mônaco, como geradora de números aleatórios distribuídos uniformemente em um dado intervalo.

²¹ Anel de colisão (Tevatron) de feixes de prótons (p) e antiprótons (\bar{p}), onde a estrutura e a criação de partículas são estudadas através da detecção de outras partículas resultantes de colisões inelásticas entre os feixes a energias da ordem de 1.0 TeV, do Fermi National Accelerator Laboratory (Fermilab).

²² Anel de colisão de feixes de prótons contra prótons (Large Hadron Collider - LHC) a energia de 7 TeV, do European Laboratory for Particle Physics (CERN).

3.1. O experimento de Buffon

A origem dos métodos a Monte Carlo remonta a 1777, quando o francês Georges Buffon determinou o valor de π a partir de lançamentos de uma agulha de comprimento ℓ sobre uma folha de papel, onde foram traçadas linhas paralelas separadas por uma distância $d \geq \ell$ (Fig. 3.1).

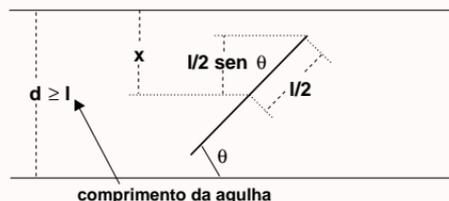


Figura 3.1: O experimento da agulha de Buffon.

De acordo com o esquema da Fig. 3.1, o ângulo θ e a distância x , do ponto médio da agulha à linha mais próxima, são as coordenadas que caracterizam a configuração espacial da agulha.

Desse modo, pode-se analisar o experimento no chamado espaço de configurações do sistema (Fig. 3.2), onde o lugar geométrico para os possíveis pares de coordenadas (θ, x) é um retângulo de lados π e $d/2$, de área $A = \pi d/2$.

Nesse contexto, as interseções da agulha em uma linha qualquer só ocorrerão se a condição $x \leq (l/2) \sin \theta$ for satisfeita, ou seja, se θ e x forem as coordenadas de um ponto sob a curva definida pela função

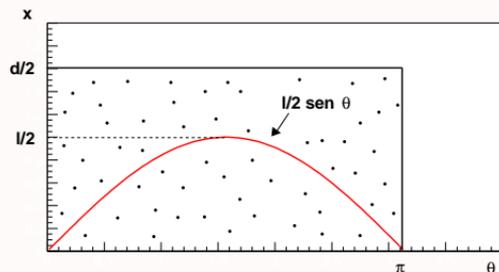


Figura 3.2: Espaço de configurações do experimento de Buffon.

$f(x) = (\ell/2) \text{sen}\theta$, no espaço de configurações.

Considerando que os lançamentos da agulha são uniformemente aleatórios, esses lançamentos correspondem a pontos que se distribuem uniformemente na região (retangular) acessível do espaço de configurações.

O experimento de Buffon é análogo também às tentativas de acertar uma região alvo, a partir de disparos aleatórios de um rifle, distribuídos uniformemente, sobre uma região retangular ao redor do alvo. Assim, de maneira alternativa, cada interceptação pode ser considerada também como um disparo de sucesso (evento) e, a probabilidade p (*a posteriori*) de interceptação da agulha, após N lançamentos (tentativas), é dada por

$$p = m/N$$

onde m é o número de sucessos.

Do ponto de vista geométrico, a probabilidade p (*a priori*) pode ser expressa pela razão entre a área I

sob a curva senoidal e a área A do retângulo de lados π e $d/2$, de modo que

$$p = \frac{I}{A} = \frac{2\ell}{\pi d} = \frac{m}{N}$$

uma vez que a área (I) sob a curva senoidal é dada pela integral $\int_0^\pi \frac{\ell}{2} \text{sen}\theta \, d\theta = \ell$.

Desse modo, a partir de N lançamentos de uma agulha, pode-se estimar experimentalmente, o valor de π por

$$\pi = \left(\frac{2N}{m}\right) \left(\frac{\ell}{d}\right)$$

Em vez do valor de π , pode-se estimar o valor da integral I por

$$I = A \frac{m}{N} = \frac{\pi d}{2} \left(\frac{m}{N}\right)$$

Para N_{exp} repetições de um experimento de Buffon com N lançamentos (tentativas), as frequências (f_m^{esp}) esperadas para os números (m) de interseções obedecerão à seguinte distribuição binomial (Fig. 3.3)

$$f_m^{\text{esp}} = N_{\text{exp}} \times \frac{N!}{(N-m)!m!} p^m (1-p)^{(N-m)}$$

com média $\mu = \langle m \rangle = Np = N2/\pi = 6.3662$ e desvio padrão $\sigma_m = \sqrt{Np(1-p)}$.

Assim, a incerteza do estimador do valor de π , para um grande número ($N \gg 1$) de lançamentos (ou tentativas), é dada por

$$\sigma_\pi = \pi \frac{\sigma_m}{m} = \frac{\pi}{Np} \sqrt{Np(1-p)}$$

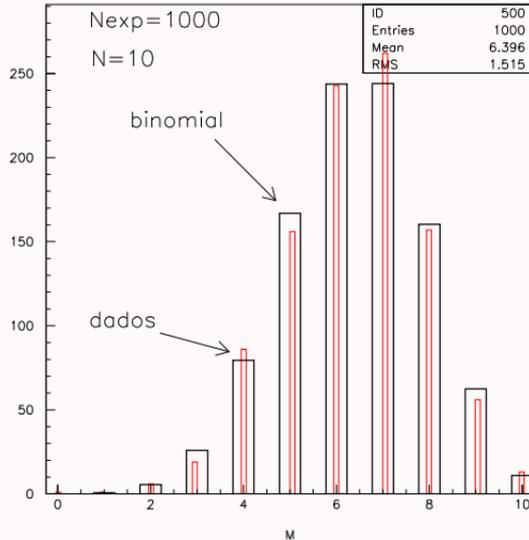


Figura 3.3: Histograma dos resultados (dados) da simulação de 1000 (N_{exp}) experimentos de Buffon, cada qual com 10 (N) lançamentos, e da distribuição binomial esperada para os números (m) de intercepções, para $d = \ell = 1$.

ou seja,

$$\sigma_{\pi} = \frac{\pi}{\sqrt{N}} \sqrt{\frac{1}{p} - 1} = \pi \sqrt{\frac{1}{m} - \frac{1}{N}}$$

A Fig. (3.4) mostra a concordância entre a estimativa da incerteza teórica, dada por $\sigma_{\pi} = 2.3735/\sqrt{N}$, e a estimativa a partir da simulação de um experimento de Buffon, no qual $d = \ell = 1$, e $p = 2/\pi = 0.63662$, para o cálculo do valor de π .

Assim, a incerteza no valor da integral é dada por

$$\sigma_I = \frac{A}{N} \sigma_m = \frac{A}{N} \sqrt{Np(1-p)}$$

ou seja,

$$\sigma_I = \frac{A}{\sqrt{N}} \sqrt{\epsilon(1-\epsilon)} = \frac{A}{\sqrt{N}} \sqrt{\frac{m}{N} \left(1 - \frac{m}{N}\right)}$$

onde a probabilidade de sucesso $p = m/N = \epsilon$ é chamada também de eficiência do experimento.

Um outro resultado, derivado do experimento de Buffon, é que a sequência de números $\{\theta_i\}$, correspondentes aos pares (θ, x) que satisfazem a condição de interceptação, estarão distribuídos segundo a função $f(\theta) = \sin \theta$ (Fig. 3.5).

Essa é a característica do experimento na qual se baseia o chamado método da rejeição para a geração de eventos distribuídos segundo uma dada pdf.

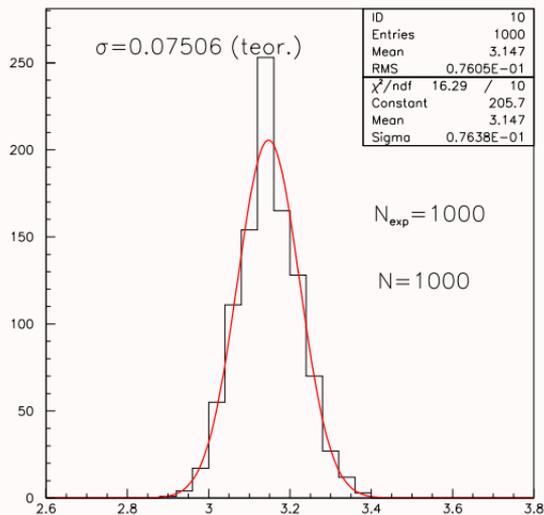


Figura 3.4: Distribuição de valores de π , a partir da simulação de 1000 (N_{exp}) experimentos de Buffon, cada qual com 1000 (N) lançamentos, cuja incerteza esperada é dada por $2.3735/\sqrt{1000} = 0.07506$, para $d = \ell = 1$, e $p = 2/\pi = 0.63662$.

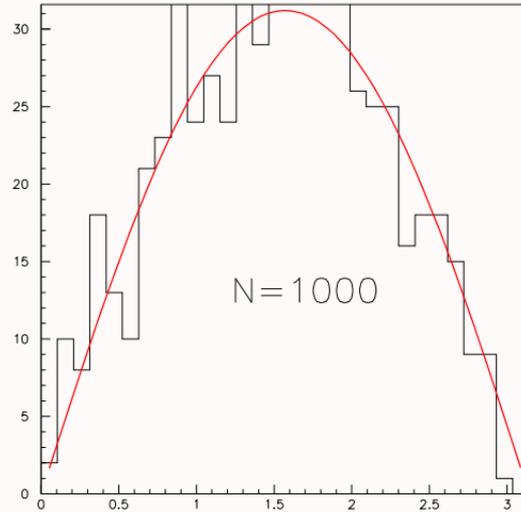


Figura 3.5: Histogramas das coordenadas dos eventos de sucesso gerados a partir da simulação de um experimento de Buffon, com 1000 (N) tentativas, para o qual $d = \ell = 1$.

3.2. Geração de eventos a Monte Carlo

Em Física de Altas Energias, a geração de eventos a partir de simulações diretas é muito restritiva, tendo em vista a enorme multiplicidade de partículas produzidas em colisões em altas energias, os processos (criação, aniquilação, bremsstrahlung) envolvidos são tão complexos que a simulação de eventos, em geral, não pode ser realizada diretamente. Daí a importância e a necessidade dos chamados métodos a Monte Carlo.

Apesar de ter sido utilizado por Fermi, em 1934, no estudo da difusão de nêutrons em materiais físséis, os primeiros algoritmos de Monte Carlo foram desenvolvidos por von Neuman, Ulam e Metropolis²³, durante a construção da bomba atômica, aproveitando-se da capacidade dos computadores em realizar grandes quantidades de operações, para a simulação tanto de problemas probabilísticos como determinísticos.

Na prática, a construção de um algoritmo simples para um método a Monte Carlo, é iniciado a partir da geração de uma sequência $\{r_1, r_2, \dots\}$ de números aleatórios distribuídos uniformemente no intervalo unitário $(0, 1)$, segundo $u(r)$ (Fig. 3.6).

Em seguida, são adotados alguns procedimentos (integração, inversão, rejeição, transformação, termalização de Metropolis) para determinar e gerar uma outra sequência $\{x_1, x_2, \dots\}$ de números aleatórios distribuídos de acordo com uma dada distribuição $f(x)$, em um intervalo (a, b) (Fig. 3.7).

3.2.1. Método de inversão

Uma vez que as sequências $\{r_i\}$ e $\{x_i\}$ são conjuntos de eventos equivalentes, a probabilidade de ocorrência de um valor r_i em um intervalo $(r, r + dr)$ é igual à probabilidade de ocorrência de um valor x_i no intervalo

²³N. Metropolis & S. Ulam, *The Monte Carlo Method*, Journal of the American Statistical Association, 44 (247), 335-341 (1949).

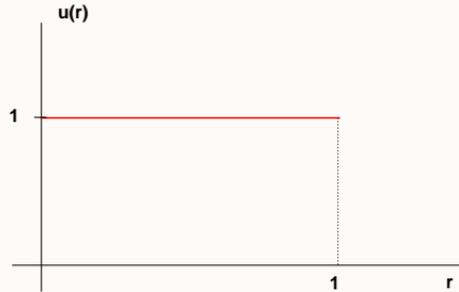


Figura 3.6: Distribuição de probabilidades uniforme no intervalo $(0, 1)$.

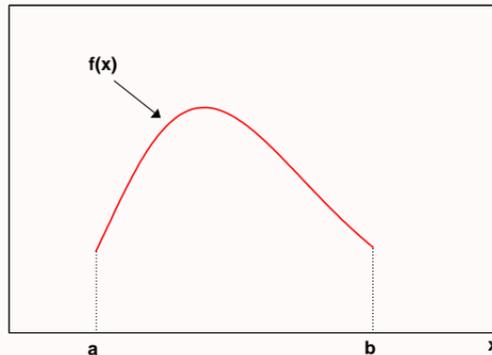


Figura 3.7: Distribuição genérica $f(x)$ de probabilidades em um intervalo (a, b) .

correspondente $[x(r), x(r) + dx]$, ou seja,

$$u(r) dr = f(x) dx$$

Assim, a probabilidade de que r_i seja menor que um valor genérico r é igual à probabilidade de que x_i seja menor que um correspondente $x(r)$,

$$r = \int_a^{x(r)} f(x') dx' \in (0, 1)$$

Desse modo, os valores de $x(r)$ gerados no intervalo (a, b) , a partir da inversão da integral de $f(x)$, estarão distribuídos segundo $f(x)$.²⁴

Por exemplo, uma sequência $\{x_i\}$ de números aleatórios distribuídos de acordo com uma pdf exponencial,

$$f(x) = \frac{1}{\lambda} e^{-x/\lambda} \quad (0, \infty)$$

pode ser gerada (Fig. 3.8) a partir da relação

$$x(r) = -\lambda \log(1 - r)$$

²⁴Se $f(x)$ não for uma pdf, ou uma função normalizada, deve-se normalizá-la e, gerar os eventos a partir de

$$r = \frac{1}{Z} \int_a^{x(r)} f(x') dx' \in (0, 1)$$

onde $Z = \int_a^b f(x) dx$

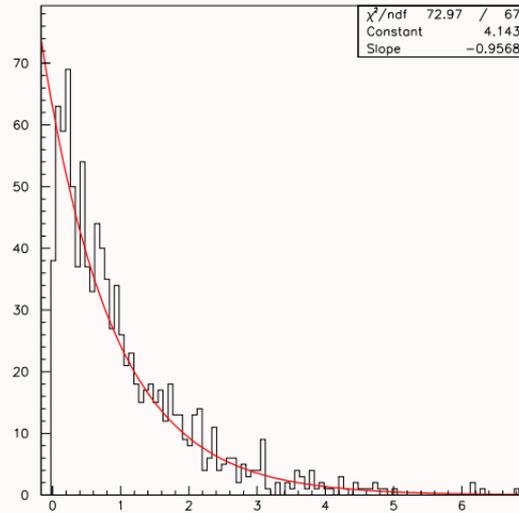


Figura 3.8: Histograma da geração de 1000 eventos aleatórios distribuídos exponencialmente, segundo $f(x) = e^{-x}$.

onde r são números distribuídos uniformemente entre 0 e 1.

Como $(1 - r)$ também são números distribuídos uniformemente entre 0 e 1, a sequência $\{x_i\}$ pode ser calculada também por

$$x(r) = -\lambda \log r$$

3.2.2. Método de rejeição simples

Um outro procedimento geral de Monte Carlo para a simulação e geração de eventos, a partir de uma sequência $\{r_1, r_2, \dots\}$ de números aleatórios distribuídos uniformemente em um intervalo $(0, 1)$, desenvolvido por Von Neumann, apoia-se no experimento de Buffon.

De acordo com o experimento, tanto o problema da geração de eventos, segundo uma pdf $f(x)$, como o cálculo da integral de uma função genérica positiva $f(x)$, em um intervalo (a, b) , podem ser encarados como uma sequência de tentativas de acertar um alvo (uma certa região do espaço), a partir de disparos aleatórios distribuídos uniformemente em uma região retangular de área $A = f_{\max}(b - a)$ que engloba o alvo (Fig. 3.9).

Assim, baseando-se no experimento de Buffon, a partir de duas sequências $\{r_1^i\}$ e $\{r_2^j\}$ de números aleatórios distribuídos uniformemente no intervalo $(0, 1)$, geram-se outras duas sequências $\{x_1, x_2, \dots, x_N\}$ e $\{y_1, y_2, \dots, y_N\}$, uniformes nos intervalos (a, b) e $(0, f_{\max})$, respectivamente, a partir de $x = r_1(b - a) + a$ e $y = f_{\max}r_2$.

Nesse caso, a condição para que um disparo acerte o alvo, ou seja, que um ponto genérico (x, y) esteja na região limitada pela curva $f(x)$, pelas retas $x = a$ e $x = b$ e pelo eixo das abscissas, é dada por

$$\boxed{y \leq f(x)}$$

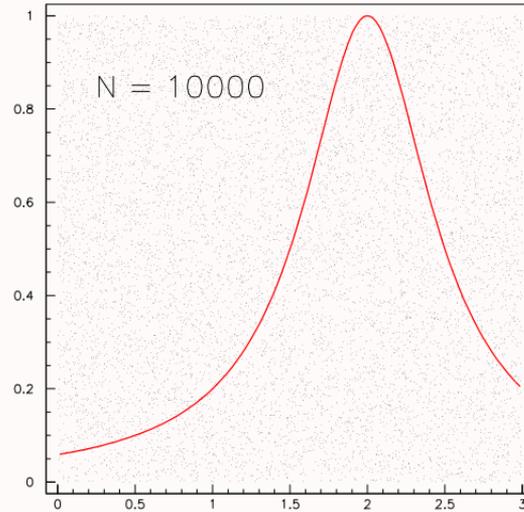


Figura 3.9: Espaço de configuração para a geração de eventos distribuídos segundo uma pdf do tipo Breit-Wigner $\frac{\Gamma^2}{(x - \mu)^2 + \Gamma^2}$, de parâmetros $\mu = 2$ e $\Gamma = 1/2$ (FWHM), via método da rejeição, a partir de duas sequências aleatórias de 10000 (N) valores distribuídos de maneira uniforme, respectivamente, entre os intervalos $(0, f_{\max} = 1)$ e $(a = 0, b = 3)$.

Desse modo, as coordenadas x dos eventos de sucesso (acerto ao alvo) que satisfazem a condição $y \leq f(x)$, constituem uma sequência $\{x_i\}$ de valores distribuídos segundo $f(x)$ (Fig. 3.10).

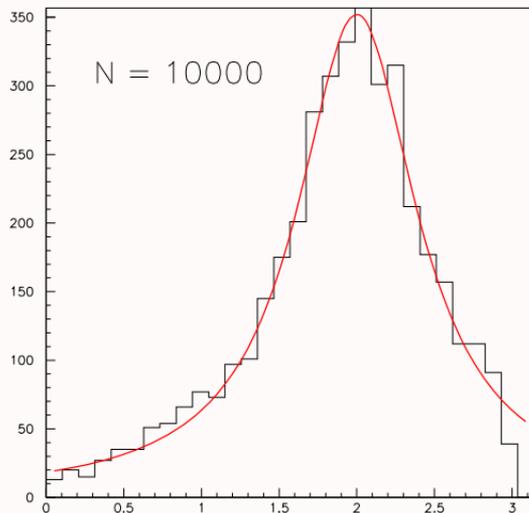


Figura 3.10: Histograma de eventos aleatórios, gerados via método de rejeição, distribuídos segundo a pdf da Fig. 3.9.

3.3. Cálculo de integrais e médias

Associados aos problemas de geração de eventos, estão também os cálculos de integrais ou médias.

3.3.1. Método de rejeição simples

Se, na simulação por um método de rejeição simples, m é o número de eventos de sucesso, para um grande número (N) de tentativas, uma estimativa para a integral de $f(x)$ no intervalo (a, b) é dada por

$$I \pm \sigma_I = A \frac{m}{N} \pm \frac{A}{\sqrt{N}} \sqrt{\frac{m}{N} \left(1 - \frac{m}{N}\right)}$$

Para a pdf da Fig. (3.9), uma estimativa para a integral da Breit-Wigner,

$$I = \int_0^3 \frac{\Gamma^2}{(x - \mu)^2 + \Gamma^2} dx$$

é dada por $I = 1.2042 \pm 0.0147$.

Esse valor, para $N = 10000$, foi calculado a partir do fragmento de algoritmo reproduzido a seguir.

```
double fun(double x)
    {return (.5^2)/((x-2)^2 + .5^2);}

main()
{
    fmax=1.;
    m=0;
```

```

A=3.;
for (int i=1; i<=N; i++)
{
    // rejeicao s/peso
    y = fmax*random(1);
    x = A*random(1);

    if (y < fun(x)) m++;
}

efic=float(m)/N;
I=A*efic;
sig=(A/sqrt(N))*sqrt(efic*(1-efic));
}

```

Uma vez que essa integral é conhecida também a priori ($I = 1.21255$), a eficiência teórica ϵ , nesse caso, é igual a $\epsilon = I/A = 0.4041833$, o que implica uma incerteza teórica da ordem de

$$\sigma_I = \frac{A}{\sqrt{N}} \sqrt{\epsilon(1-\epsilon)} = 0.014722$$

para $A = 3$ e $N = 10000$.

3.3.2. Método direto

Integrais definidas, em um intervalo (a, b) , como

$$I = \int_a^b f(x) dx$$

também podem ser estimadas diretamente, a partir de distribuições uniformes de números aleatórios em um intervalo $(0, 1)$.

Realizando-se as transformações $x' = (x - a)/(b - a)$, a integral pode ser escrita como uma integral no intervalo unitário $(0, 1)$.

$$I = (b - a) \underbrace{\int_0^1 f[x'(b - a) + a] dx'}_{I'}$$

Assim, se $\{x_1, x_2, \dots\}$ representa uma sequência de números aleatórios que se distribuem uniformemente no intervalo $(0, 1)$, segundo $u(x)$, a integral I' pode ser escrita como o valor médio de uma função $h(x) = f[x(b - a) + a]$,

$$I' = \int_0^1 u(x) h(x) dx = \langle h \rangle$$

tal que seu valor seja dado pela média aritmética da sequência $\{h(x_1), h(x_2), \dots\}$ gerada a partir de uma grande amostra de $\{x_i\}$,

$$\langle h \rangle = \frac{1}{N} \sum_i^N h(x_i)$$

com variância

$$\sigma_h^2 = \sqrt{\langle h^2 \rangle - \langle h \rangle^2}$$

Desse modo, a integral I de $f(x)$ pode ser estimada por

$$I = \frac{(b-a)}{N} \sum_{i=1}^N f[x_i(b-a) + a]$$

com incerteza

$$\sigma_I = (b-a)\sigma_h/\sqrt{N}$$

onde $\{x_i\}$ é uma sequência de números aleatórios distribuídos uniformemente no intervalo (0, 1).

O fragmento de algoritmo, a seguir, com $N = 3000$, foi utilizado para estimar o valor da integral,

$$\int_0^1 \frac{1}{1+x^2} dx = 0.785398$$

```
double fun(double x) {return 1/(1+x^2);}
```

```
main()
{
  a=0.;
  b=1.

  for (int i=1; i<=N; i++)
  {
    // direto s/peso
    x = random(1);
    hx=fun(x*(b-a)+a);
    sum = sum + hx;
    sumq= sumq + hx*hx;
  }
  h = sum/N;
  hq = sumq/N;
```

```
sigh = sqrt(hq - h*h);  
  
I = (b-a)*h;  
sigI = (b-a)*sigh/sqrt(N);  
}
```

que resultou em $I = 0.785319 \pm 0.000293$.

3.4. Redução de incertezas

Tanto no método direto, quanto no método de rejeição simples, a incerteza na estimativa de uma integral depende do tamanho (N) da amostra da sequência de números gerados aleatoriamente, segundo

$$\frac{1}{\sqrt{N}}$$

Devido a esse tipo de dependência, ao aumentarmos o tamanho da amostra, a incerteza é gradualmente reduzida.

Para problemas unidimensionais, essa dependência implica que as incertezas estimadas a partir dos métodos de Monte Carlo são maiores do que as estimadas por quaisquer métodos numéricos não-probabilísticos. Entretanto, quando são simuladas distribuições, ou calculadas integrais multidimensionais, as incertezas dos métodos tradicionais tornam-se maiores, enquanto a dependência da incerteza nos métodos a Monte Carlo permanece a mesma.

Por outro lado, para problemas que envolvem um número de variáveis da ordem de 6 ou mais, os métodos não-probabilísticos tornam-se impraticáveis, ou nem mesmo existem.

3.4.1. Transformação e inversão

Uma forma alternativa, e mais eficaz, para a redução da incerteza na estimativa de uma integral $I = \int_0^1 f(x) dx$, consiste em multiplicar e dividir o integrando por um fator de peso $w(x)$,

$$I = \int_0^1 w(x) \frac{f(x)}{w(x)} dx = Z \int_0^1 \frac{w(x)}{Z} h(x) dx$$

onde $Z = \int_0^1 w(x) dx$, de modo que o novo integrando $h(x) = f(x)/w(x)$ seja praticamente constante no intervalo de integração, o que pode ser conseguido se $w(x)$ segue as variações de $f(x)$.

Em seguida, são geradas amostras de valores de x distribuídos segundo o fator de peso, a partir da inversão de

$$r = \frac{1}{Z} \int_0^{x(r)} w(x') dx' \in (0, 1)$$

de modo que a integral possa ser estimada por

$$I = Z \times \langle h \rangle = \frac{Z}{N} \sum_{i=1}^N h(x_i)$$

com incerteza

$$\sigma_I = Z\sigma_h/\sqrt{N}$$

menor do que a da correspondente estimativa direta sem o fator de peso.

Por exemplo, a integral anterior, $I = 0.785398$, com $N = 3000$, calculada pelo algoritmo de transformação e inversão a seguir,

```
double fun(double x)
    {return 1/(1 + x^2);}
double w(double x)
    {return (4-2*x)/3.;}
double x(double y)
    {return 2.-sqrt(4-3*y);}

main()
{
    Z=1.;
    for (int i=1; i<=N; i++)
        {
            // direto c/peso
            r = random(1);
            y = x(r);
            hx = fun(y)/w(y);
            sum = sum + hx;
            sumq= sumq + hx*hx;
        }
    h = sum/N;
    hq = sumq/N;
    sigh = sqrt(hq - h*h);

    I = Z*h;
    sigI = Z*sigh/sqrt(N);
}
```

resultou em $I = 0.785369 \pm 0.000037$.

3.4.2. Métodos de rejeição

Para os métodos de rejeição, a incerteza na estimativa de uma integral $I = \int_0^1 f(x) dx$, pode ser obtida aumentando-se a eficiência do método, ou seja, a fração de eventos de sucesso, englobando a pdf original $f(x)$ por uma outra curva $w(x)$, tal que $w(x) \geq f(x)$, para a qual pode-se gerar uma sequência $\{x_i\}$ de N números aleatórios distribuídos segundo a mesma, a partir de uma sequência uniforme $\{r_1^i\}$,

$$r_1 = \frac{1}{Z} \int_0^{x(r_1)} w(x') dx' \in (0, 1)$$

onde $Z = \int_0^1 w(x) dx$.

Em seguida, a partir de uma outra sequência uniforme $\{r_2^i\}$, é gerada uma correspondente sequência $\{y_i\}$ de N números aleatórios distribuídos uniformemente entre 0 e $w(x_1)$,

$$y_i = w(x_i) \times r_2^i$$

A partir, então, da contagem dos pares (x_i, y_i) de valores que satisfazem a condição $y \leq f(x)$, a integral pode ser estimada por

$$I = A \frac{m}{N}$$

onde m é o número total de pares (x_i, y_i) não rejeitados pela condição $y \leq f(x)$, N é o número total de pares gerados e $A = Z$.

A mesma integral anterior, $I = 0.785398$, com $N = 3000$, calculada pelo algoritmo de rejeição com peso a seguir,

```

double fun(double x)
    {return 1/(1 + x^2);}
double w(double x)
    {return (4-2*x)/3.;}
double x(double y)
    {return 2.-sqrt(4-3*y);}

main()
{
    m=0;
    Z=1.;
    A=Z;
    for (int i=1; i<=N; i++)
        {
            // rejeicao c/peso
            r1 = random(1);
            x1 = x(r1);

            r2 = random(1);
            y = w(x1)*r2;

            if (y < fun(x)) m++;
        }

    efic=float(m)/N;
    I=A*efic;
    sig=(A/sqrt(N))*sqrt(efic*(1-efic));
}

```

resultou em $I = 0.786323 \pm 0.0007482$.

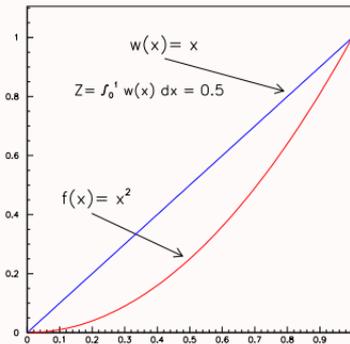
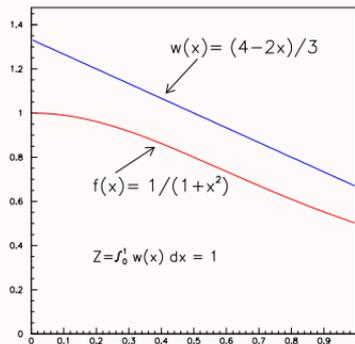
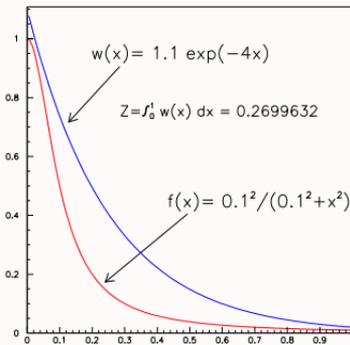
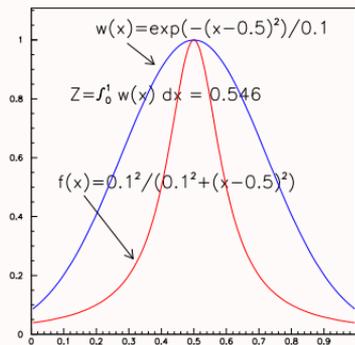


Figura 3.11: Comportamentos típicos de algumas distribuições $f(x)$ e possíveis fatores de peso $w(x)$.

A Fig. 3.11 mostra algumas distribuições $f(x)$ e, correspondentes, possíveis fatores de peso $w(x)$ ou envoltórias.

A geração de eventos, a partir do método de rejeição simples, e a determinação de integrais, a partir dos métodos direto e de rejeição simples, não dependem de integrações de funções adicionais. Por outro lado, a utilização de métodos para a redução de incertezas, quando se fazem transformações do integrando, ou a eficiência nos métodos de rejeição é aumentada, pressupõem a determinação e a inversão de integrais de funções adicionais.

Assim, a escolha de um método para simulação, além da redução de incertezas, depende do compromisso entre o tempo de processamento, a eficiência e a simplicidade do método. Em geral, em um mesmo problema são utilizados vários métodos.

3.5. O método de Metropolis

Um outro procedimento de simulação, que tanto pode ser utilizado na geração de eventos como na integração de funções, é o chamado **Método de Metropolis**.

O método apoia-se na comparação da transformação que pode ser realizada em uma integral para a redução da incerteza,

$$I = \int f(x) dx = \int \frac{w(x)}{Z} \underbrace{\frac{f(x)}{w(x)/Z}}_{h(x)} dx$$

onde $Z = \int w(x) dx$, com o cálculo da média de uma grandeza (h), associada a um sistema em equilíbrio

térmico, à temperatura T .

$$\langle h \rangle = \int \frac{\overbrace{e^{-\varepsilon(x)/kT}}^{w(x)}}{Z} h(x) dx$$

onde x representa um conjunto $\{x_i\}$ de variáveis que caracterizam o sistema, $\varepsilon(x)$ é a energia de cada configuração e $Z = \int \underbrace{e^{-\varepsilon(x)/kT}}_{w(x)} dx$ é a função de partição.

Baseando-se nessa analogia, Metropolis introduz um algoritmo, a partir do qual a integral pode ser calculada por um método probabilístico.

Em equilíbrio térmico, as transições (Q) do sistema entre suas várias configurações podem ser relacionadas por ²⁵

$$w(x) Q(x \rightarrow x') = w(x') Q(x' \rightarrow x)$$

onde $Q(x \rightarrow x')$ é a taxa de transição de um estado de configuração x para um estado x' .

Do ponto de vista probabilístico, diz-se que o processo (estocástico) de transição entre os estados constitui uma cadeia de Markov.

Denotando-se $\beta = 1/kT \implies w(x) = e^{-\beta\varepsilon(x)}$, a condição de equilíbrio entre as transições do sistema pode ser expressa por

$$\frac{T(x \rightarrow x')}{T(x' \rightarrow x)} = e^{-\beta(\overbrace{\varepsilon - \varepsilon'}^{\Delta\varepsilon})}$$

²⁵Denominado princípio do balanceamento detalhado.

Se x e x' são configurações do sistema tais que

$$\begin{cases} \varepsilon > \varepsilon' & (\Delta\varepsilon < 0) \\ \varepsilon > \varepsilon' & (\Delta\varepsilon > 0) \end{cases} \quad \text{e} \quad e^{-\beta\Delta\varepsilon} > r \in (0, 1)$$

a transição é aceita.

Como primeiro passo para a construção de um algoritmo para implementar o cálculo de uma integral segundo o método de Metropolis, seleciona-se uma configuração inicial x_o e calcula-se $w(x_o)$.

A seguir, por meio de uma variação aleatória (Δx) entre $-\Delta$ e Δ ,

$$\Delta x_i = \Delta(2r_i - 1)$$

determina-se uma nova possível configuração,

$$x_i = x_o + \Delta x_i$$

O tamanho da variação é determinado pela taxa de aceitação desejada. Um valor grande para Δ resulta pequena taxa de aceitação.

A nova configuração é aceita com probabilidade

$$p = \frac{w(x_i)}{w(x_o)}$$

A nova configuração é aceita se $p \geq r_i \in (0, 1)$, caso contrário, a configuração anterior passa a ser considerada como nova.

Em geral, a geração de novas possíveis configurações é realizada após N_{term} vezes (termalização) para evitar a influência da configuração inicial.

Além disso, o número (N_{siz}) de configurações possíveis, $\{x_k\}$, é escolhido após a realização de um certo número (N_{step}) passos para evitar correlações entre configurações sucessivas.

Assim, o resultado da integração de uma função $h(x)$ é dada por

$$I = \frac{Z}{N_{\text{siz}}} \sum_{k=1}^{N_{\text{siz}}} h(x_k)$$

com incerteza

$$\sigma_I = Z\sigma_h/\sqrt{N}$$

Para minimizar possíveis correlações entre configurações sucessivas e determinar o número de passos adequados, deve-se realizar um estudo sobre a variação da função de autocorrelação, expressado por

$$C_l = \frac{\langle h_{n+l} h_n \rangle - \langle h_n \rangle^2}{\langle h_n^2 \rangle - \langle h_n \rangle^2}$$

Um algoritmo para a simulação do método de Metropolis, para a integral anterior, $I = 0.785398$, com a possibilidade de se determinar a função de autocorrelação, pode ser implementado a partir do fragmento:

```
double xm, rw, wm, delta=.4, Z=.7469;
int n_acept;
```

```
double fun(double x) {return 1/(1 + x^2);}
double wx(double x) {return (exp(-x*x));}
```

```

void metropolis(double &xm, double &wm, int &n_acept)
{
    double xsav, wtry, rxm;
    xsav = xm;
    rxm = random(1);

    xm=xm+2*delta*(rxm-.5);
    if ((xm < 0.)||(xm >1.)) xm=xsav;

    wtry = wx(xm);

    rw = random(1);

    if (wtry > (wm*rw))
    {
        wm = wtry;
        n_acept++;
    }
    else xm = xsav;
}

main()
{
    int N_siz=3000, N_amos=100, N_step=15, lj, N_term=1000;

    for (int j=1; j<=N_amos; i++)
    {
        xm = random(1);
        wm = wx(xm);
        n_acept=0;
    }
}

```

```

//termalizacao
for (int l=1; l<=N_term; l++) metropolis(xm,wm,n_acept);

sum = 0.;
sumq = 0.;
n_acept=0;

fsb=0.;
fs1=0.;
fs2=0.;
lj=N_step+1;

for (int i=1; i<=N_siz*N_step; i++)
{
  if (i==1) fxb = fun(xm)/wx(xm);
  if (i%lj == 0)
    fxb = fun(xm)/wx(xm);

  metropolis(xm,wm,n_acept);

  if (j%N_step == 0)
  {
    sum = sum + fun(xm)/wx(xm);
    sumq = sumq + (fun(xm)/wx(xm))
              * (fun(xm)/wx(xm));
    fxa=fun(xm)/wx(xm);
    fsb=fsb+fxb;
    fs1=fs1+fxb*fxb;
    fs2=fs2+fxa*fxb;
  }
}

```

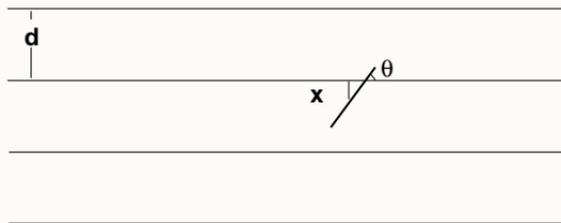
```
corr = (fs2-fsb*fsb/N_siz)/(fs1-fsb*fsb/N_siz);
I     = Z*sum/N_siz;
sigsq = sqrt(sumq - sum^2);
sig    = Z*sigsq/sqrt(N_siz);
accept = float(n_acept)/(N_siz*N_step);
}
```

o resultado, para $N_{\text{siz}} = 3000$ e $N_{\text{step}} = 15$, foi de $I = 0.785321 \pm 0.000749$.

3.6. Exercícios

1) O problema de Buffon

Uma agulha de comprimento ℓ é lançada sobre uma folha de papel onde foram traçadas linha paralelas separadas por uma distância $d > \ell$.



a) Faça um programa (em C++) que realize 5000 (N_{exp}) simulações diretas do problema de Buffon, com a agulha sendo lançada 10000 (N) vezes em cada simulação.

b) Determine o valor de π , em cada uma das 5000 simulações, armazenando as diversas estimativas de π em um histograma (utilizando o ROOT).

c) Observe o efeito nas estimativas ao se variar os números de simulações (N_{exp}) e de lançamentos (N), comparando os desvios padrões das diversas estimativas.

d) Para $N_{\text{exp}} = 1$:

- armazene todos os pares de valores (θ, x) gerados,
- faça o diagrama de dispersão $x \times \theta$,
- histografe os valores de θ que passam pelo teste $x \leq \ell/2 \text{ sen}\theta$.

2) Decaimento radioativo

Estudos sobre as transformações espontâneas (radioatividade), que alteram a composição ou a energia de alguns materiais, conduziram Rutherford (1902) à descoberta do radônio (Rn), à famosa lei do decaimento radioativo e, posteriormente (Rutherford e Geiger), ao estabelecimento do fenômeno como um processo de Poisson.

Rutherford verificou que após um período de $T_{1/2} = 3.8$ dias, a atividade (número de decaimentos ou partículas alfas detectadas em um pequeno intervalo de tempo) inicial A_0 do radônio era, praticamente, reduzida à metade, $A(T_{1/2}) = A_0/2$. Após outros 3.8 dias o fenômeno se repetia, com a atividade novamente reduzida quase à metade, $A(2T_{1/2}) = A_0/2^2$, e assim, sucessivamente, até que, após um intervalo de tempo t , a atividade seria dada aproximadamente por

$$A(t) = \frac{A_0}{2^{t/T_{1/2}}} = \frac{A_0}{e^{\log 2^{t/T_{1/2}}}} = \frac{A_0}{e^{t \left(\frac{\log 2}{T_{1/2}} \right)}} = A_0 e^{-\lambda t}$$

onde $\lambda = (\log 2)/T_{1/2}$ é a denominada constante de decaimento, e $T_{1/2}$ é a meia-vida²⁶ do isótopo.

Tal fenômeno foi interpretado como devido a um decréscimo do número inicial N_0 de núcleos radioativos, de acordo com uma lei similar,

$$N(t) = N_0 e^{-\lambda t}$$

Entretanto, a variação fracional do número de decaimentos não é a mesma para um mesmo intervalo de tempo, ou seja, o fenômeno do decaimento é um processo aleatório, tal que a lei do decaimento é uma lei probabilística que descreve as frequências (ou contagens) esperadas de decaimentos.

Encarando o processo como aleatório, a dependência exponencial da lei decorre da hipótese que a probabilidade $q(t_1, t_2)$, de que um núcleo que não tenha decaído até t_1 não decaia até t_2 , não depende dos instantes anteriores à t_1 , ou seja,

$$q(t_1, t_2) = q(t_1 - t_2)$$

Desse modo, a probabilidade de um único núcleo não haver decaído até t_2 deve obedecer à condição

$$q(t_2) = q(t_1 - t_2) q(t_1)$$

a qual é satisfeita se a dependência temporal for do tipo exponencial ($e^{-\lambda t}$).

Portanto, a probabilidade de ocorrência de um decaimento ($p = 1 - q$) em um pequeno intervalo de tempo dt é dada por

$$p = \lambda dt$$

Essa aproximação é válida também para intervalos finitos, se $\lambda \ll 1$, ou seja, se o número médio de decaimentos em um intervalo T for muito menor que o número de núcleos que decaem.

²⁶ $T_{1/2}$ varia de 3×10^{-7} s ($^{212}_{84}\text{Po}$) até 5×10^{15} anos ($^{144}_{60}\text{Nd}$).

Assim, o número (m) esperado de partículas α que resultam do decaimento de N_0 núcleos iniciais, durante um intervalo de tempo T , obedece à distribuição binomial

$$B(m|N_0, p) = P_m(N_0, p) = \frac{N_0!}{m! (N_0 - m)!} p^m (1 - p)^{N_0 - m}$$

onde $p = \lambda T$.

Devido à condição $\lambda \ll 1$, essa distribuição tende à distribuição de Poisson

$$P_m(\mu) = \frac{\mu^m}{m!} e^{-\mu}$$

onde $\mu = N_0 \lambda T$.

A partir da simulação direta de 1000 (N_{exp}) decaimentos, como um processo de Bernoulli, obtenha as distribuições do número (m) de partículas α que resultam do decaimento de 500 (N_0) núcleos iniciais, durante um intervalo de tempo (T) de 100s, para constantes (λ) de decaimentos iguais a:

- a) $\lambda = 0.01\text{s}^{-1}$
- b) $\lambda = 2 \times 10^{-4}\text{s}^{-1}$
- c) $\lambda = 4 \times 10^{-5}\text{s}^{-1}$

3) Experimento de Rutherford-Geiger

Utilizando-se dos dados de Rutherford-Geiger, das frequências das contagens do número (m) de partículas α emitidas por uma amostra de polônio, em 2608 intervalos de $T = 7.5\text{s}$ de duração,

m	f_m
0	57
1	203
2	383
3	525
4	532
5	408
6	273
7	139
8	45
9	27
10	10
11	4
12	2
13	0
14	0
$\sum_{m=0}^{14} f_m = 2608$	

escolha uma constante de decaimento (λ) e um número (N_0) inicial de núcleos, compatíveis com os dados, que permitam a simulação direta do experimento de Rutherford-Geiger como um processo de Poisson cuja

probabilidade (p) de cada decaimento é dada por $p = \lambda T$.

4) Fragmentação de quarks pesados

Inputs: Nevmax (numero maximo de eventos).

ptmin (momentum transverso minimo do quark b - 5GeV).

ptmax (momentum transverso maximo do quark b - 50GeV)

Outputs: histogramas dos momenta transversos, longitudinais,
energia e rapidity dos quarks b, do mesons B e dos muons.

Faça um programa (em C++) que simule a geração de múons provenientes da fragmentação de quarks b , no Tévatron do Fermilab, conforme os seguintes passos:

a) O primeiro passo consiste na geração de quarks b ($m_b = 4.9$ GeV), com momenta transversos (p_t) à direção do feixe de prótons,²⁷ entre 5 e 50 GeV, segundo

$$\sigma = \frac{\sigma_0}{(p_t + p_0)^\alpha} \text{ nb}$$

onde $\sigma_0 = 2.76 \times 10^{12}$, $p_0 = 14.0$ GeV e $\alpha = 7.25$, uniformemente distribuídos segundo o ângulo azimutal e normalmente no intervalo de rapidity (y) (-1.2, 1.2).

A partir desses dados as componentes dos *quadri-momenta* podem ser calculadas.

b) Uma vez gerados os quarks b , a fragmentação em mésons B pode ser simulada a partir da função de probabilidade de Peterson,²⁸

$$D(z) \sim \frac{1}{z[1 - 1/z - \epsilon/(1 - z)]}$$

²⁷Field & Feynman, Phys. Rev. D, 15, 2590 (1977).

²⁸Peterson et. al., Phys. Rev. D, 27, 105 (1983).

onde $\epsilon = 0.006 \pm 0.003$ e $z = \frac{E^B + p_{//}^B}{E^b + p^b}$ é a variável (fração da energia-momentum carregada pelo hádron gerado) segundo a qual a fragmentação é parametricamente expressa.

Assim, após determinar os limites de z (z_{\min} e z_{\max}), e considerar que os mésons B são gerados na mesma direção dos correspondentes quarks b , pode-se determinar as componentes dos quadri-momenta dos mésons B .

c) O último passo consiste na simulação do decaimento de 3 corpos, $B \rightarrow D\mu\nu$. Consulte o livro de Barger & Phillips (Collider Physics, Addison-Wesley Pub. Company, 1987)

Estimadores de máxima verossimilhança

O objetivo de um teste estatístico é alcançado pela construção de uma hipotética população infinita, da qual os dados constituem uma amostra aleatória.

R. Fisher

4.1. As origens dos métodos estatísticos

As origens dos métodos e testes estatísticos modernos remontam aos trabalhos dos ingleses Francis Galton (1822-1911), Karl Pearson (1857-1936), William Gosset (1876-1937), Ronald Fisher (1890-1962) e do polonês Jerzy Neyman (1894-1981).

Em sua obra clássica, *The grammar of science* (1892), Pearson estabelece os modelos estatísticos como alternativa à visão determinística do século XIX, com base nos seguintes princípios:

- ▷ todo experimento está sujeito a efeitos imprevistos e não observáveis;
- ▷ os resultados de um experimento obedecem a certas distribuições estatísticas que são caracterizadas por alguns parâmetros – valor esperado, variância, assimetria e curtose.

Para Pearson, as distribuições limites de probabilidades descreviam verdadeiramente a coleção de dados (medidas) resultante de um experimento e, partir de um grande número de medições, poderiam ser determinados os parâmetros da distribuição real das medidas ou dos dados dos experimentos.

Por outro lado, Fisher, em seus trabalhos, sintetizados nos textos *Statistical methods for research workers* (1925) e *The design of experiments* (1935), considera que os dados constituem uma amostra aleatória de uma população hipotética e, a partir de um experimento, obtêm-se apenas os estimadores dos parâmetros da distribuição hipotética dos dados. Ao contrário dos parâmetros (hipotéticos), os estimadores são aleatórios e devem ser avaliados, tanto segundo as distribuições limites de probabilidades, quanto aos seguintes critérios:

- ▷ **consistência** – quanto maior o número (N) de dados em uma amostra, mais próximo um estimador \hat{a} deve estar do valor (a) do parâmetro

$$\lim_{N \rightarrow \infty} \hat{a} = a$$

- ▷ **eficiência** – quanto menor a variância associada, mais eficiente é o estimador

$$V(\hat{a}_1) < V(\hat{a}_2) \implies \hat{a}_1 \text{ mais eficiente}$$

- ▷ **não-tendenciosidade** – o valor esperado de um estimador $E(\hat{a})$ deve ser igual ao valor (a) do parâmetro

$$E(\hat{a}) = a$$

Para Pearson e para Fisher os parâmetros (mesmo que sejam hipotéticos) são fixos e os estimadores, aleatórios; para a escola bayesiana, no entanto, tanto os parâmetros como os estimadores são aleatórios.

4.2. O método da máxima verossimilhança de Fisher

Seja $p(x|\theta)$ a distribuição de probabilidades para as medidas de uma grandeza x , onde θ é o parâmetro da distribuição. Para uma amostra (x_1, x_2, \dots, x_N) de N medidas independentes de x , a probabilidade associada a essa sequência particular de medidas é dada por

$$\prod_{i=1}^N p(x_i|\theta)$$

Se apenas a forma funcional da distribuição de probabilidades for conhecida *a priori*, isto é, o parâmetro for desconhecido, a função definida por

$$\mathcal{L}(x_1, x_2, \dots, x_N; \theta) = K \prod_{i=1}^N p(x_i|\theta) \quad (4.28)$$

onde K é uma constante arbitrária, denominada **função de verossimilhança**, quantifica o quão verossímil é qualquer hipótese relativa ao valor do parâmetro.

Nesse sentido, para uma dada amostra de dados, se θ_A e θ_B representam dois possíveis valores o parâmetro, e

$$\mathcal{L}(\theta_A) > \mathcal{L}(\theta_B)$$

diz-se que θ_A é um estimador mais verossímil para o parâmetro do que θ_B .

Como a determinação dos parâmetros de uma distribuição determina a dependência explícita funcional da distribuição, o procedimento é usualmente chamado também de **ajuste de funções**.

Tome-se como exemplo, uma caixa contendo 10 bolas, entre vermelhas e azuis, da qual é extraída uma amostra com reposição de 3 bolas vermelhas e uma azul. Deseja-se estimar quantas bolas azuis há na caixa.

Para se estimar a quantidade x de bolas azuis contidas na caixa, a função de verossimilhança $\mathcal{L}(x)$ correspondente à amostra obtida será dada pela probabilidade binomial de que em 4 tentativas ($N = 4$) de extração de bolas azuis haja apenas um sucesso ($m = 1$).

$$\mathcal{L}(x) \propto B(m = 1, N = 4, p = x/10)$$

ou seja,

$$\mathcal{L}(x) = \frac{x}{10} \left(1 - \frac{x}{10}\right)^3 = \frac{x(10 - x)^3}{10000}$$

De acordo com a Tab. 4.1, representada na Fig. 4.1, que mostra os valores de $\mathcal{L}(x)$ para todos os possíveis valores de x , e segundo, ainda, o princípio de Fisher, o valor 3 é a melhor estimativa para a quantidade de bolas azuis.

Aumentando-se o número de amostras de quatro tentativas de extração, por exemplo, para duas amostras de um sucesso ($m = 1$) e uma de dois sucessos ($m = 2$), a função de verossimilhança é dada por (Fig. 4.2)

$$\mathcal{L}^3(x) = B(1, 4, x/10)^2 \times B(2, 4, x/10)$$

o que implica, igualmente, o valor 3 como a melhor estimativa para a quantidade de bolas azuis.

Tabela 4.1: Função de verossimilhança para uma amostra de um sucesso ($m = 1$) em 4 tentativas ($N = 4$) de extração de bolas azuis de uma caixa contendo 10 bolas.

x	$\mathcal{L}(x) \times 10000$
0	0
1	729
2	1024
3	1029
4	864
5	625
6	384
7	189
8	64
9	9
10	0

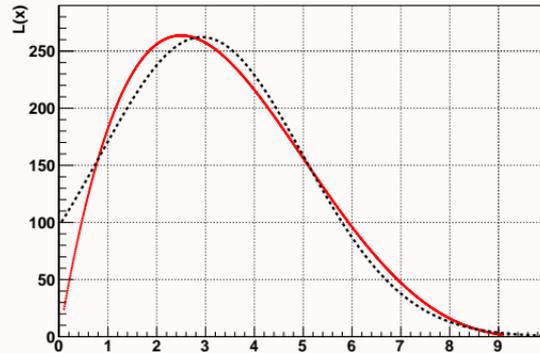


Figura 4.1: Função de verossimilhança (linha contínua) para uma amostra de x um sucesso em 4 tentativas de extração de bolas azuis de uma caixa contendo 10 bolas. A linha pontilhada é uma distribuição gaussiana ajustada à função de verossimilhança.

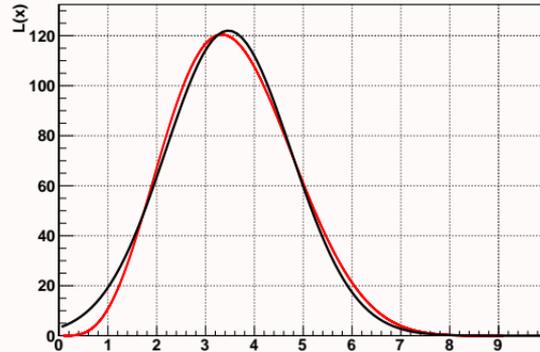


Figura 4.2: Função de verossimilhança (linha contínua) para 2 amostras de um sucesso, e uma de dois sucessos em 4 tentativas de extração de bolas azuis de uma caixa contendo 10 bolas. A linha pontilhada é uma distribuição gaussiana ajustada à função de verossimilhança.

Desse modo, um estimador de máxima verossimilhança²⁹ ($\hat{\theta}$) para um parâmetro é aquele que maximiza a função de verossimilhança, ou seja,

$$\left. \frac{\partial \mathcal{L}}{\partial \theta} \right|_{\hat{\theta}} = 0 \quad (4.29)$$

Uma vez que o logaritmo de \mathcal{L} atinge seu máximo para o mesmo valor de θ e \mathcal{L} , em geral, a condição de máxima verossimilhança é expressa como

$$\left. \frac{\partial \ln \mathcal{L}}{\partial \theta} \right|_{\hat{\theta}} = 0 \quad (4.30)$$

ou seja, o estimador de máxima verossimilhança é raiz da eq. 4.30.

Se o domínio da variável aleatória x for contínuo, como usualmente as medidas de uma grandeza, as amostras constituem sequências discretas de valores organizados em M classes de intervalos Δ_i , como

$$\left[(x_1, x_1 + \Delta_1); (x_2, x_2 + \Delta_2); \dots; (x_M, x_M + \Delta_M) \right]$$

Nesse caso, a probabilidade associada a essa sequência particular é dada por

$$\int_{x_1}^{x_1 + \Delta_1} \int_{x_2}^{x_2 + \Delta_2} \dots \int_{x_M}^{x_M + \Delta_M} \rho(x_1 | \theta) \rho(x_2 | \theta) \dots \rho(x_M | \theta) \, dx_1 \, dx_2 \dots \, dx_M$$

e define-se a função de verossimilhança por

$$\mathcal{L}(x_1, x_2, x_3 \dots x_M; \theta) = K \prod_{i=1}^M \rho(x_i | \theta) \quad (4.31)$$

²⁹Maximum likelihood estimator.

4.3. Limite da função de verossimilhança

A Fig. 4.3 mostra que, à medida que cresce o número de amostras, a função de verossimilhança se aproxima de uma gaussiana³⁰ cujo valor esperado é igual ao estimador de máxima verossimilhança.

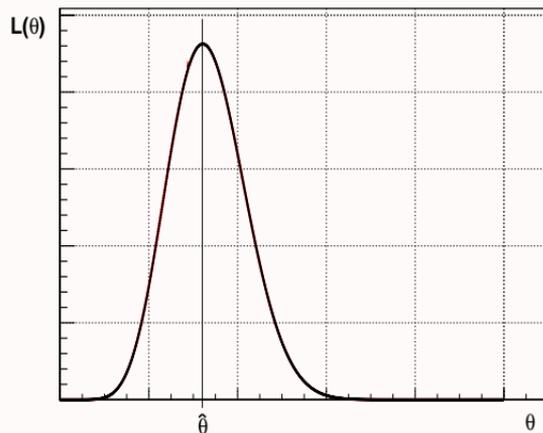


Figura 4.3: Limite da função de verossimilhança do exemplo de extração de bolas (Sec. 4.2), para 30 amostras de um sucesso e 1 amostra de dois sucessos de extração de bolas azuis.

³⁰Essa é uma consequência do teorema do limite central.

Com efeito, expandindo-se o logaritmo da função de verossimilhança em torno de seu máximo,

$$\ln \mathcal{L}(\theta) = \ln \mathcal{L}(\hat{\theta}) - \frac{1}{2} \left| \frac{\partial^2 \ln \mathcal{L}}{\partial \theta^2} \right|_{\hat{\theta}} (\theta - \hat{\theta})^2 + \dots$$

para um número “suficientemente grande” de amostras, pode-se considerar a chamada aproximação parabólica, de modo que a função de verossimilhança pode ser aproximada por uma expressão do tipo gaussiana (eq. 4.32).

$$\mathcal{L}(\theta) \propto e^{-\frac{1}{2} \left| \frac{\partial^2 \ln \mathcal{L}}{\partial \theta^2} \right|_{\hat{\theta}} (\theta - \hat{\theta})^2} \quad (4.32)$$

Para a escola clássica, os parâmetros não são quantidades aleatórias: portanto, a função de verossimilhança não representa a probabilidade de ocorrências dos parâmetros. No entanto, como o estimador é aleatório pode-se considerar a eq. 4.32 como proporcional à densidade de probabilidade para o estimador,

$$\rho(\hat{\theta}|\theta) \propto e^{-\frac{1}{2} \frac{(\hat{\theta} - \theta)^2}{2\sigma_{\hat{\theta}}^2}} \quad (4.33)$$

com valor esperado $E(\hat{\theta}) = \theta$, variância dada por

$$V(\hat{\theta}) = \sigma_{\hat{\theta}}^2 = \left| \frac{\partial^2 \ln \mathcal{L}}{\partial \theta^2} \right|_{\hat{\theta}}^{-1} = \langle (\hat{\theta} - \theta)^2 \rangle \quad (4.34)$$

e incerteza $\sigma_{\hat{\theta}}$ por

$$\sigma_{\hat{\theta}} = \left| \frac{\partial^2 \ln \mathcal{L}}{\partial \theta^2} \right|_{\hat{\theta}}^{-\frac{1}{2}} \quad (4.35)$$

Para a escola bayesiana, no entanto, os parâmetros também são aleatórios e estão associados a distribuições de probabilidades *a priori*. Desse modo, a escola bayesiana interpreta a eq. 4.32 como a probabilidade condicional de ocorrência da sequência aleatória (x_1, x_2, \dots, x_N) para um dado valor do parâmetro de θ ,

$$\rho(x_1, x_2, x_3 \dots x_N | \theta) \propto \mathcal{L}(x_1, x_2, x_3 \dots x_N; \theta)$$

Assim, a probabilidade condicional *a posteriori* $\rho(\theta | x_1, x_2, x_3 \dots x_N)$ de ocorrência de um valor θ para uma dada sequência (x_1, x_2, \dots, x_N) é determinada por

$$\rho(\theta | x_1, x_2, x_3 \dots x_N) \propto \mathcal{L}(x_1, x_2, x_3 \dots x_N; \theta) \rho(\theta)$$

onde $\rho(\theta)$ é a distribuição de probabilidades *a priori* para o parâmetro θ .

Nesse sentido, se o domínio da distribuição de probabilidades *a priori* para o parâmetro está restrito à região para a qual a função de verossimilhança apresenta uma variação muito grande, em torno de seu valor máximo, a função de verossimilhança pode ser interpretada também como uma pdf para o parâmetro.

4.4. Estimadores para a distribuição de Poisson

Se x é uma variável aleatória associada a um processo de Poisson, como o tempo de decaimento de uma partícula, cujo valor esperado é μ , ou seja,

$$P(x|\mu) = \frac{\mu^x}{x!} e^{-\mu}$$

a função de verossimilhança para uma sequência (x_1, x_2, \dots, x_N) de N valores independentes de x será dada por

$$\mathcal{L}(\mu) = \prod_{i=1}^N \frac{\mu^{x_i}}{x_i!} e^{-\mu} = e^{-N\mu} \prod_{i=1}^N \frac{\mu^{x_i}}{x_i!}$$

e seu logaritmo por

$$\ln \mathcal{L} = -N\mu + \ln \mu \sum_{i=1}^N x_i + \dots$$

Desse modo, se o valor do parâmetro μ for desconhecido, a condição de máximo,

$$\left. \frac{\partial \ln \mathcal{L}}{\partial \mu} \right|_{\hat{\mu}} = -N + \frac{1}{\hat{\mu}} \sum_{i=1}^N x_i = 0$$

implica que os estimadores para o valor esperado e sua incerteza, para um processo de Poisson, a partir de

uma amostra de N dados, são³¹

$$\hat{\mu} = \frac{1}{N} \sum_{i=1}^N x_i = \bar{x}$$

$$\sigma_{\hat{\mu}} = \sqrt{\frac{\bar{x}}{N}} = \sigma_{\bar{x}}$$

4.5. Estimadores para a distribuição de Gauss

Seja x_i uma variável aleatória que obedece a uma distribuição gaussiana, $\mathcal{N}_{x_i}(\mu, \sigma_i^2)$, com valor esperado μ e variância σ_i^2 , ou seja,

$$\rho_G(x_i|\mu, \sigma_i^2) = \frac{1}{\sigma_i^2 \sqrt{2\pi}} e^{-\frac{(x_i - \mu)^2}{2\sigma_i^2}}$$

³¹Para a distribuição de Poisson,

$$V(\hat{\mu}) = V\left(\frac{1}{N} \sum_{i=1}^N x_i\right) = \frac{1}{N^2} \sum_{i=1}^N \underbrace{V(x_i)}_{\mu} = \frac{\bar{x}}{N}$$

A função de verossimilhança para uma sequência de N dessas variáveis é dada por

$$\mathcal{L}(\mu) = \prod_{i=1}^N \frac{1}{\sigma_i \sqrt{2\pi}} e^{-\frac{(x_i - \mu)^2}{2\sigma_i^2}}$$

e seu logaritmo por

$$\ln \mathcal{L} = -\frac{1}{2} \sum_{i=1}^N \frac{(x_i - \mu)^2}{\sigma_i^2} - \sum_{i=1}^N \ln \sigma_i + \dots$$

Assim, supondo que σ_i^2 seja conhecida e μ o parâmetro a ser estimado, a condição de máximo,³²

$$\left. \frac{\partial \ln \mathcal{L}}{\partial \mu} \right|_{\hat{\mu}} = - \sum_{i=1}^N \frac{(x_i - \hat{\mu})}{\sigma_i^2} = 0$$

³²A condição de máximo para $\ln \mathcal{L}$ equivale a condição de mínimo para a quantidade

$$\chi^2 = \frac{1}{2} \sum_{i=1}^N \frac{(x_i - \mu)^2}{\sigma_i^2}$$

Esse procedimento, utilizado pelo matemático alemão Carl F. Gauss (1777-1855), desde 1795, foi sistematizado pelo francês Adrien M. Legendre, em 1805, e tornou-se conhecido como o **método dos mínimos quadrados**. Para o método dos mínimos quadrados, a minimização de χ^2 é considerada como a hipótese fundamental do processo de ajuste de funções que dependem de alguns parâmetros a uma amostra (x_1, x_2, \dots, x_N) de N medidas de uma grandeza x , mesmo que as medidas não se distribuam gaussianamente. Por exemplo, se $f(x|\theta_1, \theta_2, \dots, \theta_p)$ é uma função que depende de p parâmetros, o método consiste em minimizar a quantidade

$$\chi^2 = \frac{1}{2} \sum_{i=1}^N \left[\frac{x_i - f(x_i|\theta_1, \theta_2, \dots, \theta_p)}{\sigma_i} \right]^2$$

onde σ_i é a incerteza associada cada uma das medidas x_i , com relação a cada um dos p parâmetros.

implica que os estimadores para o valor esperado e sua incerteza, a partir de uma amostra de N dados normalmente distribuídos, são³³

$$\hat{\mu} = \frac{\sum_{i=1}^N \frac{x_i}{\sigma_i^2}}{\sum_{i=1}^N \frac{1}{\sigma_i^2}} \quad (\text{média})$$
$$\sigma_{\hat{\mu}} = \frac{1}{\sqrt{\sum_{i=1}^N \frac{1}{\sigma_i^2}}} \quad (\text{erro da média})$$

³³Para distribuições gaussianas,

$$V(\hat{\mu}) = \sigma_{\hat{\mu}}^2 = \left(\sum_{i=1}^N \frac{1}{\sigma_i^2} \right)^{-2} \sum_{i=1}^N V\left(\frac{x_i}{\sigma_i^2} \right) = \left(\sum_{i=1}^N \frac{1}{\sigma_i^2} \right)^{-2} \underbrace{\sum_{i=1}^N \frac{1}{\sigma_i^4} V(x_i)}_{\sum_{i=1}^N \frac{1}{\sigma_i^2}} = \left(\sum_{i=1}^N \frac{1}{\sigma_i^2} \right)^{-1}$$

Se (x_1, x_2, \dots, x_N) são as medidas de uma grandeza física x , nas mesmas condições, ou seja, tal que $\sigma_i = \sigma$, obtém-se

$$\hat{\mu} = \sum_{i=1}^N \frac{x_i}{N} = \bar{x}$$

$$\sigma_{\hat{\mu}} = \frac{\sigma}{\sqrt{N}} = \sigma_{\bar{x}}$$

4.6. Estimadores para parâmetros de outras distribuições

A maximização da função de verossimilhança para outras distribuições, ou seja, a determinação das raízes da eq. 4.2, pode não ser analiticamente tão simples e direta como para as distribuições de Poisson e de Gauss. No entanto, sempre se pode utilizar um procedimento numérico exaustivo para a determinação de um estimador para o parâmetro de uma dada distribuição.

Considerando que o estimador $\hat{\theta}$ do parâmetro θ de uma pdf qualquer tenha sido numericamente determinado, de acordo com a aproximação parabólica,

$$\ln \mathcal{L}(\theta) = \ln \mathcal{L}_{\max} - \frac{(\theta - \hat{\theta})^2}{2\sigma_{\hat{\theta}}^2}$$

a incerteza $\sigma_{\hat{\theta}}$ associada ao estimador pode ser avaliada pelos limites (θ_1 e θ_2) para os quais a função de verossimilhança satisfaz a eq. 4.36.

A eq. 4.36 pode ser utilizada para a determinação da incerteza do estimador, mesmo que o $\ln \mathcal{L}$ não seja exatamente parabólica. Em geral, os intervalos em torno do estimador não são simétricos, de tal modo que o estimador e suas incertezas são representados como

$$\ln \mathcal{L}(\theta \pm \sigma_{\hat{\theta}}) \Big|_{\theta_1, \theta_2} = \ln \mathcal{L}_{\max} - \frac{1}{2} \quad (4.36)$$

$$\hat{\theta} \begin{matrix} +\sigma_2 \\ -\sigma_1 \end{matrix}$$

Essa é a origem das estimativas de incertezas assimétricas.³⁴

³⁴No limite de grandes amostras, quando a pdf for gaussiana, as incertezas são simétricas.

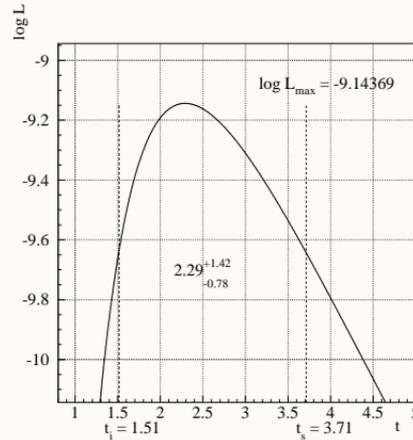


Figura 4.4: Determinação dos limites de um estimador a partir da aproximação parabólica da função de verossimilhança.

4.7. Estimadores para distribuições multiparamétricas

Se a função de verossimilhança depende de M parâmetros,

$$\mathcal{L}(\theta) \quad \text{onde } \theta = (\theta_1, \theta_2, \dots, \theta_M)$$

as estimativas desses parâmetros ainda são dadas pelas condições de máximo,

$$\left. \frac{\partial \ln \mathcal{L}}{\partial \theta_i} \right|_{\hat{\theta}_i} = 0 \quad (i = 1, 2, \dots, M)$$

Entretanto, a avaliação das variâncias associadas a cada parâmetro só será análoga ao caso uniparamétrico, se os estimadores forem independentes, ou seja,

$$E\left[(\hat{\theta}_i - \theta_i)(\hat{\theta}_j - \theta_j)\right] = \langle (\hat{\theta}_i - \theta_i)(\hat{\theta}_j - \theta_j) \rangle = 0$$

se as covariâncias forem nulas.

Levando-se em conta as covariâncias, no limite de grandes amostras, a função de verossimilhança pode ser aproximada por

$$\ln \mathcal{L}(\theta) = \ln \mathcal{L}(\hat{\theta}) - \frac{1}{2} \sum_{i,j=1}^M H_{ij} \delta_i \delta_j$$

$$\text{onde } \delta_i = \hat{\theta}_i - \theta \quad \text{e} \quad H_{ij} = - \left. \frac{\partial^2 \ln \mathcal{L}}{\partial \theta_i \partial \theta_j} \right|_{\hat{\theta}}$$

Desse modo, a inversa da matriz hessiana,

$$H = \begin{pmatrix} H_{11} & H_{12} & \dots & H_{1M} \\ H_{21} & H_{22} & \dots & H_{2M} \\ \vdots & \vdots & \ddots & \vdots \\ H_{M1} & H_{M2} & \dots & H_{MM} \end{pmatrix}$$

determina a matriz de covariância V ,

$$V = H^{-1} \quad \Leftrightarrow \quad (V)_{ij} = (H^{-1})_{ij}$$

- **Estimadores para medidas diretas de uma grandeza**

Em geral, tanto o valor esperado ($\mu = \theta_1$) como a variância ($\sigma = \theta_2$) são parâmetros desconhecidos na medição de uma grandeza x . Assim, a função de verossimilhança para uma amostra de N medidas de x depende de dois parâmetros,

$$\mathcal{L}(\theta_1, \theta_2) = \prod_{i=1}^N \frac{1}{\theta_2} e^{-\frac{(x_i - \theta_1)^2}{2\theta_2^2}}$$

e seu logaritmo é dado por

$$\ln \mathcal{L} = -\frac{1}{2} \sum_{i=1}^N \frac{(x_i - \theta_1)^2}{\theta_2^2} - N \ln \theta_2 + \dots$$

Desse modo, de acordo com as condições de máximo,

$$\left\{ \begin{array}{l} \frac{\partial \ln \mathcal{L}}{\partial \theta_1} \Big|_{\hat{\theta}} = \frac{1}{\hat{\theta}_2^2} \sum_{i=1}^N (x_i - \hat{\theta}_1) = 0 \quad \Rightarrow \quad \boxed{\hat{\theta}_1 = \frac{1}{N} \sum_{i=1}^N x_i = \bar{x}} \\ \frac{\partial \ln \mathcal{L}}{\partial \theta_2} \Big|_{\hat{\theta}} = \frac{1}{\hat{\theta}_2^2} \sum_{i=1}^N (x_i - \hat{\theta}_1)^2 - \frac{N}{\hat{\theta}_2} = 0 \quad \Rightarrow \quad \boxed{\hat{\theta}_2 = \sqrt{\sum_{i=1}^N \frac{(x_i - \hat{\theta}_1)^2}{N}} = \sigma_x} \end{array} \right.$$

Calculando-se os elementos da matriz hessiana

$$\left\{ \begin{array}{l} H_{11} = - \left. \frac{\partial^2 \ln \mathcal{L}}{\partial \theta_1^2} \right|_{\hat{\theta}} = \frac{N}{\hat{\theta}_2^2} \\ H_{12} = H_{21} = - \left. \frac{\partial^2 \ln \mathcal{L}}{\partial \theta_1 \theta_2} \right|_{\hat{\theta}} = 0 \\ H_{22} = - \left. \frac{\partial^2 \ln \mathcal{L}}{\partial \theta_2^2} \right|_{\hat{\theta}} = \frac{3}{\hat{\theta}_2^4} \underbrace{\sum_{i=1}^N (x_i - \hat{\theta}_1)^2}_{N\hat{\theta}_2^2} - \frac{N}{\hat{\theta}_2^2} = \frac{2N}{\hat{\theta}_2^2} \end{array} \right.$$

pode-se escrever

$$H = \begin{pmatrix} N/\hat{\theta}_2^2 & 0 \\ 0 & 2N/\hat{\theta}_2^2 \end{pmatrix} \Rightarrow H^{-1} = V = \begin{pmatrix} \sigma_x^2/N & 0 \\ 0 & \sigma_x^2/2N \end{pmatrix}$$

Assim, as incertezas associadas aos estimadores, \bar{x} e s_x , do valor esperado μ e do desvio-padrão σ são dadas por³⁵

$$\sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{N}} \quad (\text{erro da média})$$

$$\sigma_{\sigma_x} = \frac{\sigma_x}{\sqrt{2N}} \quad (\text{erro do erro})$$

³⁵O chamado erro do erro da média pode ser calculado por

$$\sigma_{\sigma_{\bar{x}}} = \frac{\sigma_{\sigma_x}}{\sqrt{N}} = \frac{1}{\sqrt{N}} \frac{\sigma_x}{\sqrt{2N}} = \frac{\sigma_{\bar{x}}}{\sqrt{2N}}$$

4.8. Propriedades dos estimadores de máxima verossimilhança

Uma vez que os estimadores de máxima verossimilhança para o valor esperado de variáveis (x) que obedecem às distribuições de Poisson ou de Gauss são dados pela média, segundo a lei dos grandes números (Sec. 2.7), esses estimadores são consistentes e não-tendenciosos (visto que $E(\bar{x}) = E(x) = \mu$) e eficientes (uma vez que para grandes amostras a variância desses estimadores é praticamente nula).

No entanto, para variáveis que obedecem à distribuição de Gauss, como uma grande amostra de medidas diretas de uma grandeza, os estimadores de máxima verossimilhança das variâncias, apesar de serem consistentes e eficientes, são ligeiramente tendenciosos.

Expressando-se o estimador de máxima verossimilhança da variância de uma distribuição gaussiana como³⁶

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

³⁶Para grandes amostras,

$$\left\{ \begin{aligned} \sigma_x^2 &= \frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^2 E[(x - \bar{x})^2] = E\left\{[(\bar{x} - \mu)^2 - (\bar{x} - \mu)]^2\right\} \\ &= E[(x - \mu)^2] + E[(\bar{x} - \mu)^2] + \underbrace{E(x\bar{x} - x\mu - \mu\bar{x} + \mu^2)}_0 \end{aligned} \right.$$

seu valor esperado pode ser calculado por

$$E(\sigma_x^2) = \frac{1}{N} \sum_{i=1}^N \overbrace{E[(x_i - \mu)^2]}^{N\sigma^2} - \underbrace{E[(\bar{x} - \mu)^2]}_{\sigma^2/N} = \frac{(N-1)}{N} \sigma^2$$

ou seja, seu valor esperado é diferente do valor esperado para a variância. Assim, σ_x é um estimador tendencioso para o desvio padrão σ .

Para grandes amostras, mesmo que os parâmetros não estejam associados a distribuições gaussianas, diz-se que a incerteza associada a um estimador de máxima verossimilhança define um intervalo de confiança de 68.3%. Para a escola clássica, o nível de confiança 0.683 representa a fração de vezes que o intervalo (aleatório) de confiança $(\bar{x} - \sigma_{\bar{x}}, \bar{x} + \sigma_{\bar{x}})$ conterá o valor (fixo) do parâmetro.

Para pequenas amostras, o estimador não-tendencioso usualmente utilizado para o desvio padrão σ é dado por

$$s_x = \sqrt{\frac{N}{N-1}} \sigma_x = \sqrt{\sum_{i=1}^N \frac{(x_i - \bar{x})^2}{N-1}} \Rightarrow s_{\bar{x}} = \frac{s_x}{\sqrt{N}}$$

e o intervalo $(\bar{x} - s_{\bar{x}}, \bar{x} + s_{\bar{x}})$ não está mais associado a um nível de confiança de 0.683, uma vez que a pdf associada ao estimador \bar{x} é a distribuição de Student com $(N-1)$ graus de liberdade (Ap. A.10).

Nesse caso, intervalos de confiança de 68.3% podem ser determinados integrando-se a distribuição de Student entre dois limites tais que

$$P[(\mu - t_\nu s_{\bar{x}}) < \bar{x} < (\mu + t_\nu s_{\bar{x}})] = 0.683$$

onde o fator t_ν depende do número de graus de liberdade ($\nu = N - 1$) e, portanto do número de medidas da amostra (Tab. 4.2).

Como as relações $[(\mu - t_\nu s_{\bar{x}}) < \bar{x} < (\mu + t_\nu s_{\bar{x}})]$ e $[(\bar{x} - t_\nu s_{\bar{x}}) < \mu < (\bar{x} + t_\nu s_{\bar{x}})]$ são equivalentes, a probabilidade associada ao intervalo $(\bar{x} - t_\nu s_{\bar{x}}, \bar{x} + t_\nu s_{\bar{x}})$ também é igual a 0.683.

Tabela 4.2: Relação entre números de medidas, graus de liberdade e níveis de confiança, associados à distribuição de Student.

Número (N) de medidas	Graus de liberdade (ν)	Níveis de confiança (CL)		
		68,3%	95%	99%
2	1	1,84	12,71	63,66
3	2	1,32	4,30	9,92
4	3	1,20	3,18	5,84
5	4	1,14	2,78	4,60
6	5	1,11	2,57	4,03
7	6	1,09	2,45	3,71
8	7	1,08	2,36	3,50
9	8	1,07	2,31	3,36
10	9	1,06	2,26	3,25
15	14	1,04	2,14	2,98
20	19	1,03	2,09	2,86
∞	∞	1,00	1,96	2,58

4.9. Ajuste de funções a histogramas

A determinação da função de verossimilhança torna-se impraticável para uma grande quantidade de dados. Em tais casos, é conveniente a organização dos dados em classes de intervalos, representando-os em um histograma (Fig. 4.5).

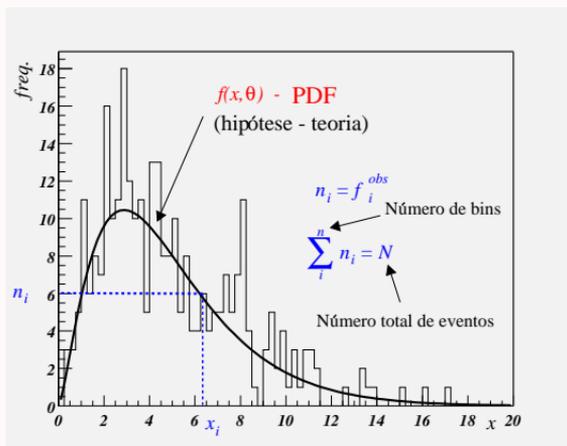


Figura 4.5: Ajuste de funções a histogramas.

Nesses casos, as frequências esperadas (f_i) das N medidas de uma grandeza x associada a uma pdf $\rho(x|\theta)$,

em cada uma das M classes de intervalos $[x_i, x_i + \Delta_i)$ são dependentes do parâmetro θ , e dadas por

$$f_i(\theta) = N \int_{x_i}^{x_i + \Delta_i} \rho(x|\theta) dx$$

Como as frequências (n_i) de ocupação das classes de intervalos em um histograma, com probabilidade de ocupação de cada classe dada por $p_i = f_i/N$, obedece a uma distribuição multinomial, a função de verossimilhança para uma determinada sequência de frequências (f_1, f_2, \dots, f_N) pode ser expressa como

$$\mathcal{L}(\theta) = \frac{N!}{n_1! n_2! \dots n_N!} \left(\frac{f_1}{N}\right)^{n_1} \left(\frac{f_2}{N}\right)^{n_2} \dots \left(\frac{f_N}{N}\right)^{n_N}$$

e seu logaritmo por

$$\ln \mathcal{L}(\theta) = \sum_{i=1}^N n_i \ln f_i(\theta)$$

uma vez que tanto as frequências esperadas (f_i) como as observadas (n_i) obedecem a relação de vínculo

$$\sum_{i=1}^M f_i = \sum_{i=1}^M n_i = N$$

Esse é o chamado método de máxima verossimilhança para ajuste de funções a distribuições de frequências, ou histogramas.

As incertezas associadas aos parâmetros da função ajustada podem ser determinadas pela eq. 4.36 (Sec. 4.6).

Para histogramas nos quais $n_i \gg 1$, a distribuição de frequências em cada classe pode ser aproximada por uma gaussiana de variância igual $f_i(\theta)$, e a função de verossimilhança pode ser expressa como

$$\mathcal{L}(\theta) = \sum_{i=1}^M \frac{1}{\sqrt{f_i}} e^{-\frac{[n_i - f_i(\theta)]^2}{2f_i}}$$

Nesse caso, a maximização de $\ln \mathcal{L}$ equivale à minimização da quantidade

$$\chi^2 = \sum_{i=1}^M \frac{[n_i - f_i(\theta)]^2}{f_i}$$

Nessas circunstâncias, o procedimento é chamado de **método dos mínimos quadrados modificado**, ou **método de χ^2 modificado**.

Em geral, do ponto de vista computacional, o procedimento torna-se mais simples se a quantidade χ^2 é aproximada por

$$\chi^2 = \sum_{i=1}^M \frac{[n_i - f_i(\theta)]^2}{n_i} = \left[\sum_{i=1}^M \frac{f_i^2(\theta)}{n_i} \right] - N$$

Testes estatísticos paramétricos

Mas até aqui não pude de descobrir a causa dessas propriedades da gravidade a partir dos fenômenos, e não faço nenhuma hipóteses (hypotheses non fingo).

I. Newton

Uma vez determinado um valor μ_{obs} para o estimador de um parâmetro μ , a partir de uma amostra (x_1, x_2, \dots, x_N) de N medidas de uma grandeza x , caso exista algum valor de referência (μ_o),³⁷ os testes estatísticos estabelecem critérios que permitem avaliar se esses valores são compatíveis.

Em geral, esses valores apresentam uma discrepância tal que $|\mu_{\text{obs}} - \mu_o| \neq 0$. Desse modo, um teste estatístico deve auxiliar a decisão sobre a casualidade da discrepância, ou seja, a escolha sobre a discrepância ser apenas fortuita, ou dever-se a algum efeito que decorra de alguma causa desconhecida, que não tenha sido previamente identificada?

³⁷ Em geral uma estimativa teórica, ou o resultado de um experimento mais confiável.

O procedimento inicial para um teste paramétrico consiste na escolha de uma pdf $\rho(\mu|\mu_0)$ que associa possíveis valores para o estimador do parâmetro μ , e para a qual o valor de referência é um parâmetro, como as distribuições de Gauss ou de Student para a média das medidas diretas de uma grandeza.

Por outro lado, considerando o problema como um ajuste de funções, pode-se perguntar: Quão bem a pdf $\rho(x|\mu_0)$ se ajusta aos dados?

5.1. O teste de significância de Fisher

De acordo com Fisher, o objetivo de qualquer experimento é testar evidências contra a aceitação da compatibilidade entre um valor observado e um valor de referência, e uma medida para essa evidência é dada pelo chamado nível descritivo (*p-value*), o qual expressa a probabilidade (p) associada a um resultado cujo valor seja maior que o observado (Fig. 5.1),

$$p = P(\mu \geq \mu_{\text{obs}}) = \int_{\mu_{\text{obs}}}^{\infty} \rho(\mu|\mu_o) d\mu$$

Quanto menor o nível descritivo, menor a compatibilidade entre os valores observado e de referência. Ou seja,

$$p \rightarrow 0 \quad \implies \quad \text{valor observado não compatível com valor de referência}$$

Por exemplo, se $p < 0.05$, diz-se que a evidência contra a compatibilidade não é significativa ao nível de 5%.

O fato de se encontrar um valor muito baixo para p significa apenas uma evidência estatística contra a hipótese de compatibilidade entre o valor observado e o valor de referência. Segundo Fisher, o mais importante em um experimento não é o teste estatístico, mas a existência de um argumento lógico sobre a correção da distribuição de probabilidades admitida para os estimadores.

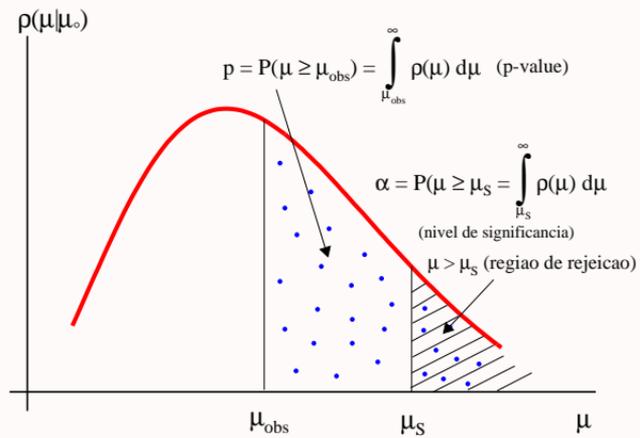


Figura 5.1: p-value

5.2. Os testes de hipóteses de Neyman-Pearson

De acordo com Neyman e Pearson, o problema deve ser colocado em termos de um teste para o qual a hipótese (H_0), de que o valor do estimador seja igual a um valor de referência (μ_0),

$$H_0 : \mu = \mu_0$$

deva ser rejeitada ou não. Essa hipótese é denominada **hipótese de nulidade**.

Nesse sentido, um teste de significância a partir de um nível descritivo expressa a evidência do resultado de um experimento contra a hipótese de nulidade.

Segundo Neyman e Pearson, deve-se escolher *a priori* um conjunto de valores possíveis para o estimador que definam uma região crítica de rejeição ($\mu \leq \mu_1$ e $\mu \geq \mu_5$) associada a uma probabilidade bem pequena, de modo que a hipótese de nulidade deva ser rejeitada se o valor observado do estimador esteja contido nessa região crítica (Fig. 5.2). A região complementar à região de rejeição é denominada região de aceitação ($\mu_1 < \mu < \mu_5$).

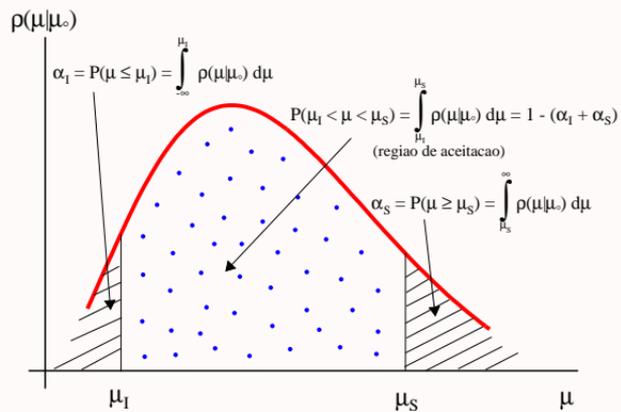


Figura 5.2: Regiões de rejeição e de aceitação da hipótese de nulidade ao nível de confiança (CL) de $[1 - (\alpha_1 + \alpha_S)] \times 100 \%$

Em geral, os limites dessas regiões são determinados pelos chamados níveis de significâncias (α_1 e α_s), arbitrariamente escolhidos,

$$\left\{ \begin{array}{l} \alpha_s = P(\mu \geq \mu_s) = \int_{\mu_s}^{\infty} \rho(\mu|\mu_o) d\mu \\ \alpha_1 = P(\mu \leq \mu_1) = \int_{-\infty}^{\mu_1} \rho(\mu|\mu_o) d\mu \end{array} \right.$$

Se o valor observado está contido na região de aceitação, diz-se que o resultado é compatível com a hipótese de nulidade ao nível de significância $(\alpha_1 + \alpha_s) \times 100\%$, ou com nível de confiança de $[1 - (\alpha_1 + \alpha_s)] \times 100\%$.

A aceitação de uma hipótese de nulidade não significa ser essa hipótese verdadeira, significa apenas ser, do ponto de vista estatístico, consistente com os dados de uma dada amostra. Outras hipóteses alternativas podem mostrar-se igualmente consistentes. Todavia, a rejeição de uma hipótese de nulidade não significa que ela seja falsa.

Como a escolha dos níveis de significância é arbitrária, deve-se ter em mente que, para pequenos níveis de significâncias, ou seja, para uma região de aceitação grande, menores são as chances de rejeição de uma hipótese de nulidade verdadeira, ou de se cometer um erro do tipo I. Esse é o tipo de erro cometido ao se anunciar uma falsa descoberta, ou seja, ao se tentar estabelecer um efeito que em realidade não existe. No entanto, se uma hipótese de nulidade falsa é aceita, cometendo-se um erro tipo II, pode-se perder a chance de se estabelecer um novo efeito. Grandes descobertas podem ocorrer quando uma hipótese de nulidade há longo tempo estabelecida é rejeitada em um experimento.

5.3. Intervalos de confiança

Um intervalo de confiança para o valor de referência (μ_o) de um parâmetro só será igual a região de aceitação (μ_I, μ_S) para as distribuições de Gauss ou de Student.

Como (μ_I e μ_S) dependem do parâmetro da pdf $\rho(\mu|\mu_o)$ (Fig. 5.3),

$$P(\mu_I < \mu < \mu_S) = \int_{\mu_I}^{\mu_S} \rho(\mu|\mu_o) d\mu = 1 - (\alpha_I + \alpha_S)$$

para um dado μ_o^* , o valor observado μ_{obs} só será compatível com a condição

$$\mu_I(\mu_o^*) < \mu_{obs} < \mu_S(\mu_o^*)$$

de que o valor observado esteja na região de aceitação, se o intervalo $[a(\mu_{obs}), b(\mu_{obs})]$ contiver o valor μ_o^* , ou seja,

$$a(\mu_{obs}) < \mu_o^* < b(\mu_{obs})$$

Desse modo, os intervalos (μ_I, μ_S) e (a, b) são probabilisticamente equivalentes e, portanto, associados ao mesmo nível de confiança.³⁸

Segundo Neyman, $[1 - (\alpha_I + \alpha_S)]$ é uma estimativa da fração de vezes que o intervalo (a, b) conterá um dado valor de referência μ_o .

³⁸Confidence level (CL).

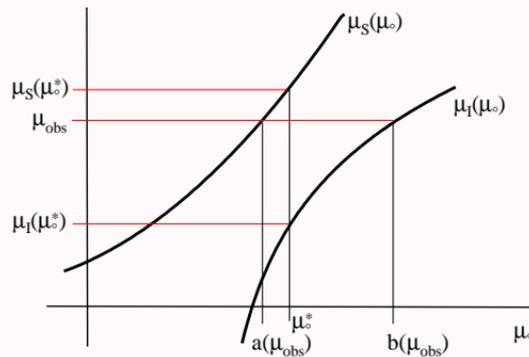


Figura 5.3: Intervalos de confiança correspondentes ao nível de confiança (CL) $[1 - (\alpha_I + \alpha_S)] \times 100\%$

• Intervalos de confiança para as distribuições de Gauss e de Student

Para uma distribuição gaussiana, associada às medidas de uma grandeza x ($\mu_{\text{obs}} = \bar{x}$) cujo valor de referência é μ_o , e um nível de probabilidade igual a 0.683 associado ao intervalo $(\mu_o - \sigma, \mu_o + \sigma)$, ou seja,

$$P[(\mu_o - \sigma) < \mu < (\mu_o + \sigma)] = 0.683$$

Como os valores de $\mu_1 = \mu_o - \sigma$ e $\mu_s = \mu_o + \sigma$, em função de μ_o , são dados por duas retas paralelas com inclinação de $\pi/4$ rad, os intervalos (μ_1, μ_s) e (a, b) possuem a mesma amplitude (Fig. 5.4),

$$b - a = \mu_s - \mu_1 \quad \Leftrightarrow \quad a = \bar{x} - \sigma \quad \text{e} \quad b = \bar{x} + \sigma$$

Assim, os intervalos $(\mu_o - \sigma, \mu_o + \sigma)$ e $(\bar{x} - \sigma, \bar{x} + \sigma)$ são intervalos de confiança de 0.683%.

O mesmo ocorrerá para a distribuição de Student, uma vez que os valores dos limites da região de aceitação, em função de μ_o , para qualquer nível de confiança são definidos também por retas paralelas de inclinação igual a $\pi/4$ rad,

$$\begin{cases} \mu_1 = \mu_o - t_\nu s_{\bar{x}} \\ \mu_s = \mu_o + t_\nu s_{\bar{x}} \end{cases} \quad \Rightarrow \quad \begin{cases} a = \bar{x} - t_\nu s_{\bar{x}} \\ b = \bar{x} + t_\nu s_{\bar{x}} \end{cases}$$

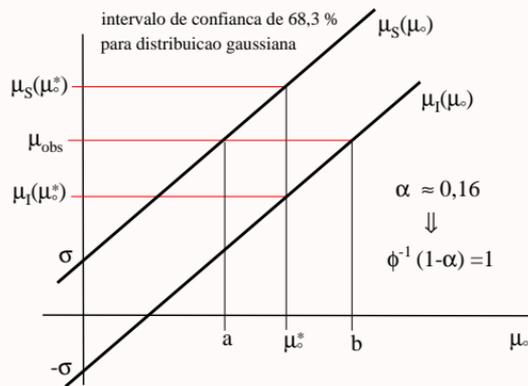


Figura 5.4: Intervalos correspondentes ao nível de confiança de 68.3% .

5.4. O teste de χ^2 de Pearson

O teste mais utilizado para se determinar a significância da estimativa do valor esperado de uma grandeza cujas medidas obedecem à distribuição de Gauss foi estabelecido por Pearson, em 1900.

Uma vez que a probabilidade associada às N medidas independentes de uma grandeza x , entre os valores (x_1, x_2, \dots, x_N) e $(x_1 + dx_1, x_2 + dx_2, \dots, x_N + dx_N)$, é dada por

$$dp \propto e^{-\frac{1}{2} \sum_{i=1}^N z_i^2} dx_1 dx_2 \dots dx_M$$

onde $z_i = \frac{x_i - \mu}{\sigma_i}$, μ é o valor esperado e σ_i é o desvio-padrão da distribuição gaussiana, considerando z_i como

coordenada em um espaço N -dimensional, a variável $\chi^2 = \sum_{i=1}^N z_i^2$ está associada ao raio (r) de uma esfera nesse espaço,

$$\chi^2 = \sum_{i=1}^N z_i^2 = r^2 \quad \Rightarrow \quad dr = (\chi^2)^{-1/2} d\chi^2$$

cujo elemento de volume ($dV = dx_1 dx_2 \dots dx_M \propto r^{N-1} dr$) é proporcional a

$$(\chi^2)^{(N/2-1)} d\chi^2$$

Logo, a probabilidade associada a uma dada sequência de medidas,

$$dp \propto (\chi^2)^{N/2-1} e^{-\frac{1}{2}\chi^2} d\chi^2$$

é determinada por uma pdf $\rho(\chi^2|\nu)$ para a quantidade χ^2 , denominada **distribuição de χ^2** , dada por (Ap. A.9)

$$\rho(\chi^2|\nu) \propto (\chi^2)^{\nu/2-1} e^{-\chi^2/2}$$

Nesse contexto, o número de medidas N representa o parâmetro ν da distribuição de χ^2 , denominado **número de graus de liberdade (dof)**.

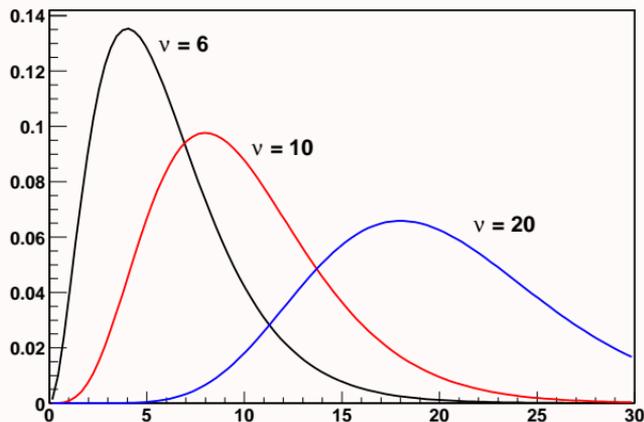


Figura 5.5: Distribuições de χ^2 para $\nu = 2, 6, 10, 20$ e 30 graus de liberdade.

A figura 5.5 mostra diversas distribuições de probabilidades de χ^2 para vários graus de liberdade ν , e evidencia que as distribuições apresentam máximos para os valores de χ^2 em torno de ν , o qual também é o valor esperado de χ^2 ,³⁹ ou seja,

$$\langle \chi^2 \rangle = \nu$$

Se os parâmetros μ e σ^2 da distribuição gaussiana são desconhecidos, e os estimadores tiverem sido determinados pela maximização de $\ln \mathcal{L} = -\frac{1}{2}\chi^2$, ou pela minimização de χ^2 , a partir das condições

$$\begin{cases} \frac{\partial \chi^2}{\partial \mu} = 0 \\ \frac{\partial \chi^2}{\partial \sigma^2} = 0 \end{cases}$$

essas duas equações constituem relações de vínculo entre as medidas que reduzem o número de termos independentes na expressão de χ^2 para $\nu = N - 2$. Nesse caso, a quantidade χ^2 não mais obedece à distribuição de χ^2 , uma vez que as medidas agora obedecem à distribuição de Student (Ap. A.10). Como para um número grande de graus de liberdade a distribuição de Student tende à distribuição de Gauss, pode-se considerar que

³⁹Como as medidas x_i e, portanto, $z_i = (x_i - \mu)/\sigma_i$ se distribuem gaussianamente, o valor esperado de $\chi^2 = \sum_{i=1}^{\nu} z_i^2$ pode ser calculado como

$$\langle \chi^2 \rangle = E(\chi^2) = \sum_{i=1}^{\nu} E(z_i^2) = \sum_{i=1}^{\nu} E \left[\left(\frac{x_i - \mu}{\sigma_i} \right)^2 \right] = \sum_{i=1}^{\nu} \frac{1}{\sigma_i^2} \underbrace{E[(x_i - \mu)^2]}_{\sigma_i^2} = \nu$$

a probabilidade associada a uma amostra de N medidas de uma grandeza é dada por uma distribuição de χ^2 com $\nu = N - 2$ graus de liberdade.⁴⁰

Se as N medidas são distribuídas em M classes cujas frequências (n_i) obedecem a uma distribuição multinomial, para um grande número de dados, tal que $p_i = f_i(\theta)/N \ll 1$ e $n_i \gg 1$, a distribuição em cada classe pode ser aproximada por uma gaussiana de variância igual a f_i . Nesse caso, a quantidade χ^2 dada por

$$\chi^2 = \sum_{i=1}^M \left[\frac{n_i - f_i(\theta)}{f_i} \right]^2$$

obedece a uma distribuição de χ^2 com $\nu = M - p - 1$ graus de liberdade. Ou seja, o número de graus de liberdade é reduzido pelo número de parâmetros p da pdf ajustada, e pela relação de vínculo $\sum_{i=1}^M f_i = \sum_{i=1}^M n_i = N$.

Testes de χ^2 associados a histogramas são chamados também de testes de aderência entre distribuições de frequências.

Como o valor máximo da distribuição de χ^2 , para $\nu > 2$, é igual ao número de graus de liberdade, a idéia do teste de χ^2 é comparar o valor calculado χ_{obs}^2 , a partir de uma amostra de dados, com o parâmetro (ν) da distribuição. Assim, se a razão χ^2/ν , denominada χ^2 reduzido, for próxima à unidade,

$$\frac{\chi^2}{\nu} \simeq 1$$

⁴⁰ Analogamente, no caso do ajuste de uma função que depende de p parâmetros, pode-se considerar que, para uma grande amostra de dados, a quantidade χ^2 obedece a uma distribuição de χ^2 com $\nu = N - p$ graus de liberdade.

diz-se que os estimadores dos parâmetros são consistentes com os dados, ou que o ajuste de função é satisfatório.

A quantificação apropriada do teste de χ^2 , para a avaliação da evidência dos resultados da amostra contra a hipótese de nulidade, de que os dados são distribuídos gaussianamente, pode ser realizada por um teste de significância, calculando-se o nível descritivo (p) da probabilidade de que o valor de χ^2 seja maior que o valor observado,

$$p(\chi^2 \geq \chi_{\text{obs}}^2) = \int_{\chi_{\text{obs}}^2}^{\infty} \rho(\chi^2|\nu) d\chi^2$$

ou por um teste de hipóteses, determinando-se uma região de aceitação para o valor observado, a partir da escolha arbitrária dos níveis de significância α_1 e α_s .

A Tab. 5.1 apresenta valores de ν , χ_1^2/ν e χ_s^2/ν , para os quais os níveis de significância $\alpha_1 = p(\chi^2 \leq \chi_1^2)$ e $\alpha_s = p(\chi^2 \geq \chi_s^2)$ são, respectivamente, 0.01 e 0.99.

Representando-se graficamente os dados da Tab. 5.1 (Fig. 5.6), pode-se estabelecer que para $\nu = 10$ a probabilidade de que χ^2/ν ocorra entre 0.26 e 2.32 é igual a 98%, ou seja, o nível de confiança do intervalo $0.26 < \chi^2/\nu < 2.32$ é de 98%.

Assim, para $\nu > 2$, valores de χ^2/ν em torno de 1 certamente estarão dentro de um intervalo de confiança de 98%. Enquanto valores muito pequenos indicam que as incertezas foram subestimadas, ou seja, indicam um acordo bom demais entre a função de ajuste e os dados, valores grandes, maiores que 3, indicam que o ajuste é muito ruim e podem ser resultantes de estimativas incorretas das incertezas dos dados.

Cabe reforçar a observação de que o teste de χ^2 , como qualquer teste estatístico, não valida nenhuma hipótese física com relação à função de ajuste, apenas estima a sua significância estatística. Ou seja, nada indica acerca da melhor relação funcional $y = f(x)$ entre duas variáveis x e y , apenas sobre as melhores estimativas de parâmetros da função a ser ajustada aos dados.

Tabela 5.1: Valores de ν , χ_1^2/ν e χ_5^2/ν , tais que a integral da pdf de χ^2 seja igual a 0.01 e 0.99.

ν	χ_1^2/ν	χ_5^2/ν
1	6.630	0.0002
2	4.601	0.010
3	3.780	0.060
4	3.318	0.073
5	3.016	0.100
6	2.802	0.142
7	2.639	0.179
8	2.510	0.208
9	2.407	0.232
10	2.320	0.255
11	2.247	0.276
12	2.184	0.297
13	2.129	0.315
14	2.082	0.332
15	2.038	0.348
16	2.000	0.363
17	1.965	0.376
18	1.933	0.389
19	1.904	0.402
20	1.878	0.413
21	1.853	0.423
22	1.831	0.433
23	1.810	0.443
24	1.790	0.452
25	1.771	0.460
26	1.753	0.468

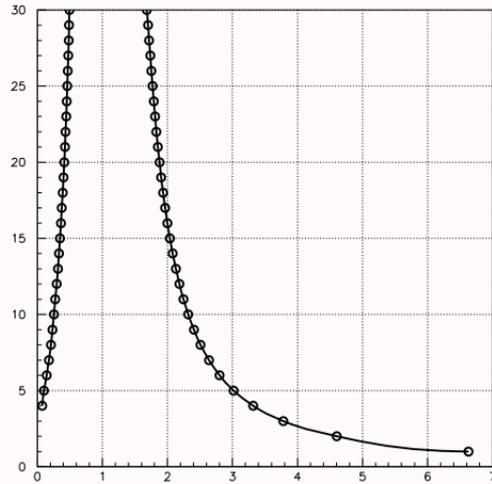


Figura 5.6: Probabilidades relativas à Tab. 5.1.

Assim, a Física não se baseia em testes estatísticos para o descobrimento e estabelecimento de suas leis. Qualquer nova lei, para ser validada, deve ser compatibilizada com outras leis prévias, teorias e princípios já bem estabelecidos. Só assim são extrapoladas de seu domínio inicial e, então, testadas em outras situações. Esse processo, consolidado por Galileu no séc. XVI, permitiu à Física estabelecer um corpo de conhecimentos sobre a natureza não compartilhado por nenhuma outra área da Ciência.

5.5. Lançamento de dados

Exemplo Lançamento de dados

ocorrência de uma face, \iff processo aleatório
valor ou evento (i) (probabilidade - p_i)

$$\triangleright \text{a posteriori} \rightarrow p_i = \lim_{N \rightarrow \infty} \left(\frac{n_i}{N} \right)$$

$$\triangleright \text{a priori} \rightarrow p_i = \frac{1}{6}$$

$$\triangleright \sum_{i=1}^{n=6} p_i = 1 \text{ (normalização)}$$

$$\triangleright \langle i \rangle = \sum_{i=1}^{n=6} i \times p_i = \frac{1}{6} \underbrace{(1 + 2 + 3 + 4 + 5 + 6)}_{21} = 3.5 \text{ (média)}$$

Amostra de N lançamentos

$$N = 120$$

i	1	2	3	4	5	6
n_i	16	19	27	17	23	18
ϵ_i	20	20	20	20	20	20

freq. observadas ($f_i^{\text{obs}} = n_i$) $\rightarrow \{n_1, \dots, n_6\}$

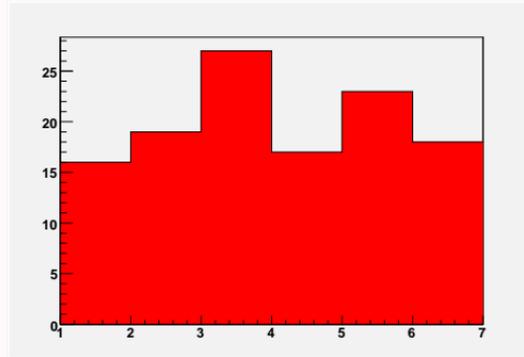
$$\sum_{i=1}^{n=6} n_i = N$$

freq. esperadas ($f_i^{\text{esp}} = N p_i = \epsilon_i$) $\rightarrow \{\epsilon_1, \dots, \epsilon_6\}$

$$\sum_{i=1}^{n=6} \epsilon_i = N \underbrace{\sum_{i=1}^{n=6} p_i}_1 = N p_i = \frac{1}{6}$$

Amostra de N lançamentos

i	n_i	ϵ_i	n_i^2
1	16	20	254
2	19	20	361
3	27	20	729
4	17	20	289
5	23	20	589
6	18	20	324
			2488



- ▷ As diferenças são significativas?
- ▷ Como caracterizar a discrepância?

Medidas de discrepância

- ▷ $(n_i - \epsilon_i) \neq 0$ (desvio)

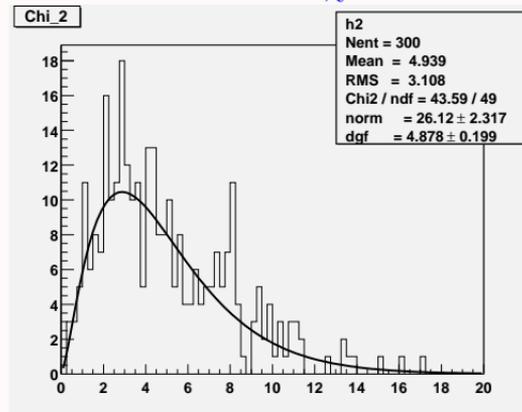
$$\triangleright \sum_{i=1}^n (n_i - \epsilon_i) = \underbrace{\sum_{i=1}^n n_i}_N - \underbrace{\sum_{i=1}^n \epsilon_i}_N = 0$$

$$\triangleright \sum_{i=1}^n (n_i - \epsilon_i)^2 \text{ (não suficiente)}$$

$$\triangleright \chi^2 = \sum_{i=1}^n \frac{(n_i - \epsilon_i)^2}{\epsilon_i} = \sum_{i=1}^n \frac{n_i^2}{\epsilon_i} - N = 4.4$$

χ^2 depende de $(n - 1)$ termos independentes
 $\nu = n - 1$ (número de graus de liberdade)

Teste de χ^2



$$\begin{cases} \nu = 5 \\ \epsilon_i = N/6 \end{cases} \Rightarrow \chi^2 = \frac{1}{N/6} \left(\sum_{i=1}^n n_i^2 \right) - N = 4.4$$

No caso de testes de hipóteses, envolvendo medições de grandezas físicas, como na estimativa de um parâmetro, as hipóteses alternativas estão associadas com distribuições de probabilidades do mesmo tipo que as associadas à hipótese de nulidade, ou seja, distribuições gaussianas.

Nesses casos, supondo que uma dada hipótese alternativa H_a seja verdadeira. Por exemplo, a hipótese de que o valor esperado seja igual a um outro valor μ_a ,

$$H_a : E(x) = \mu_a$$

a expectativa de não-rejeição de uma hipótese de nulidade falsa é dada por

$$P_a(|t| \leq t_c) = \beta$$

é denominada probabilidade de ocorrência (β) de erro do tipo II.

Representando-se as fdp's associadas às hipóteses de nulidade (ρ_0) e alternativa (ρ_a) em um mesmo gráfico (figura 5.7), nota-se que, para um dado valor crítico \bar{x}_c ⁴¹, existe um compromisso entre a escolha de α e β , devido a justaposição das duas distribuições, que pode ser expresso por uma relação de complementariedade entre α e β . Portanto, a maneira de reduzir as chances de ocorrências de ambos os tipos de erros é diminuir a zona de justaposição das duas distribuições, reduzindo a incerteza associada à média da amostra, ou seja, aumentando-se o tamanho da amostra.

⁴¹ $\bar{x}_c = \mu + t\Delta\bar{x}$

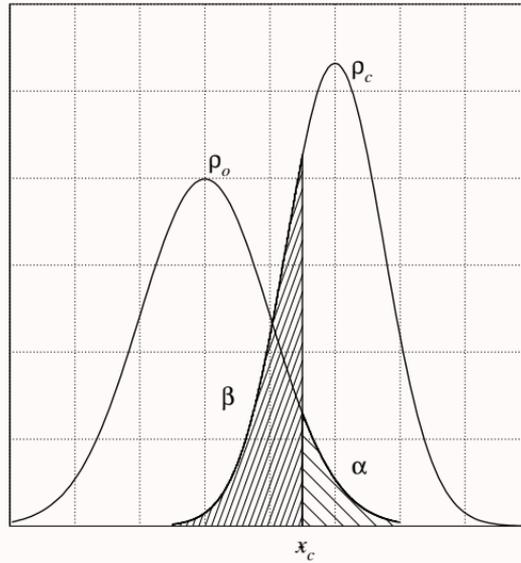


Figura 5.7: Distribuições de probabilidades associadas a uma hipótese de nulidade H_0 e a uma hipótese alternativa H_a .

Testes de hipóteses envolvem a comparação entre duas distribuições e fornecem suporte para avaliar a compatibilidade dessas distribuições. O critério para essa avaliação é a escolha arbitrária de um nível de significância e, para limitar o risco de erros do tipo I, em geral, esse nível é fixado em 0.05 e, para decisões mais conservativas em 0.01.

Testes de significância estatística nos quais as hipóteses testadas estão associadas aos parâmetros de uma distribuição de probabilidades são também denominados testes **paramétricos**, em contraste com os testes que são baseados em hipóteses sobre a forma de uma distribuição, os quais são chamados testes **não-paramétricos**.

Nesse sentido, os ajustes de funções são testes compostos nos quais fazem-se hipóteses sobre a forma de distribuições e envolvem também a determinação de um ou mais parâmetros.

5.6. Exercícios

- 5.6.1) O desvio padrão de uma população é igual a 22. Se uma amostra de 100 elementos dessa população fornece a média $\bar{x} = 115,8$, pode-se afirmar que a média da população é inferior a 120, ao nível de 5% de significância?
- 5.6.2) As notas $\{x_i\}$ de mecânica dos alunos de uma faculdade, constitui uma população distribuída normalmente com média 3.0 e desvio padrão 2.0.
A média de uma nova turma de 100 alunos foi de 3.4. Essa nova turma é superior às outras ao nível de 5% ($\alpha = 0.05$) de significância?
- 5.6.3) Se a frequência de ocorrências dos números 5 e 6 em $n = 500000$ lançamentos de um dado foi de $f_{5 \text{ ou } 6}^{\text{obs}} = 168350$, esse dado pode ser considerado “honesto”?

Apêndices

Apêndice A

As loterias são formas de taxaço (aceitas livremente) das camadas menos privilegiadas da sociedade.

D. Ruelle

A.1. Eventos equivalentes

Se duas variáveis aleatórias x e y estão relacionadas por $y = f(x)$ e Y é um conjunto de condições (eventos) sobre a variável y ⁴² e $X = \{x \mid f(x) \in Y\}$, diz-se que os conjuntos X e Y são probabilisticamente independentes, ou que são eventos equivalentes, no sentido de que suas probabilidades de ocorrência são idênticas,

$$P(Y) = P(X)$$

⁴²Por exemplo, uma relação do tipo $y < a$, onde a é uma constante.

A.1.1. Distribuições de probabilidades de eventos equivalentes

Se a distribuição de probabilidades (**pdf**) associada à variável x é $f(x)$ e $y = h(x)$ é uma outra variável aleatória, por exemplo, $y = 1 + x = h(x)$ e $f(x) = x/2$ para $0 < x < 2$, a função de distribuição acumulada (**fd**) $G(t)$ associada à y será dada por

$$\begin{aligned}G(t) &= P(y \leq t) = P(1 + x \leq t) \\&= P(x \leq t - 1) = \int_0^{t-1} f(x) dx \\&= \int_0^{t-1} \frac{x}{2} dx = \frac{(t-1)^2}{4}\end{aligned}$$

de modo que,

$$g(y) = \frac{dG}{dy} = \frac{y-1}{2}$$

é a pdf de y .

Uma vez que

$$G(t) = P(x \leq \underbrace{t-1}_{h^{-1}(t)}) = F[\underbrace{h^{-1}(t)}_{x(t)}]$$

implica

$$g(y) = \underbrace{\frac{dF}{dx}}_{f[x(y)]} \frac{dx(y)}{dx} = \frac{y-1}{2}$$

Esse resultado pode ser sistematizado como,

$$\boxed{g(y) = f[x(y)] \quad \left| \frac{dx(y)}{dy} \right|}$$

se $y = h(x)$ for uma função monótona. Pois,

a) f crescente

$$\begin{aligned} G(t) &= P(y \leq t) = P[f(x) \leq t] \\ &= P[(x \leq h^{-1}(t))] = F[h^{-1}(t)] \end{aligned}$$

$$\begin{aligned} &\Downarrow \\ \underbrace{\frac{dG}{dy}}_{g(y)} &= \underbrace{\frac{dF}{dx}}_{f(x)} \frac{dx}{dy} \quad \left(\frac{dx}{dy} > 0 \right) \end{aligned}$$

b) f decrescente

$$\begin{aligned} G(t) &= P(y \leq t) = P[(x > h^{-1}(t))] \\ &= 1 - P[(x \leq h^{-1}(t))] \end{aligned}$$

\Downarrow

$$g(y) = -f(x) \frac{dx}{dy} \quad \left(\frac{dx}{dy} < 0 \right)$$

$$= f(x) \left| \frac{dx}{dy} \right|$$

A.1.2. Valor médio de função de uma variável aleatória

A princípio, para a determinação do valor esperado de uma função $y = h(x)$ de uma variável aleatória x de pdf $f(x)$, seria necessário primeiro, a determinação da pdf $g(y)$ de y e, então calculá-lo por

$$\langle y \rangle = \int_{-\infty}^{\infty} y g(y) dy$$

Entretanto, desde que

$$g(y) = f[x(y)] \frac{dx}{dy}$$

$$\Downarrow$$

$$\langle y \rangle = \int_{-\infty}^{\infty} y f[x(y)] \left| \frac{dx}{dy} \right| dy$$

$$= \int_{-\infty}^{\infty} y f[x(y)] \frac{dx}{dy} dy$$

$$\langle h(x) \rangle = \int_{-\infty}^{\infty} h(x) f(x) dx$$

A.2. Variáveis aleatórias multidimensionais

Se a um determinado conjunto (sistema físico) estão associados alguns atributos (grandezas) $\{x_1, x_2 \dots x_n\}$, e as ocorrências de seus valores (medidas) são aleatórias, a n -úpla $X = (x_1, x_2 \dots x_n)$ é denominada **variável aleatória n-dimensional** ou **vetor aleatório n-dimensional**.

A pdf das ocorrências conjuntas de cada combinação de valores, associada a uma variável aleatória n -dimensional, é uma função de n variáveis $f(x_1, x_2 \dots x_n)$ tal que

$$\underbrace{\int \int \dots \int}_R f(x_1, x_2 \dots x_n) dx_1 dx_2 \dots dx_n = 1$$

onde R é o domínio de definição das variáveis.

A partir da pdf conjunta $f(x_1, x_2 \dots x_n)$, a pdf de qualquer uma das variáveis x_k , denominada **pdf marginal** de x_k , é definida por

$$\underbrace{\int \dots \int f(x_1 \dots x_n) dx_1 \dots dx_{k-1} dx_{k+1} \dots dx_n}_{f(x_k)} = 1$$

uma vez que $\int f(x_k) dx_k = 1$

A.2.1. Variáveis independentes

Duas variáveis aleatórias x e y , associadas às pdf $\rho(x)$ e $\lambda(y)$ são ditas independentes se a pdf conjunta $f(x, y)$ for dada por

$$f(x, y) = \rho(x)\lambda(y)$$

ou seja, se a pdf conjunta for fatorável.

Se duas variáveis aleatórias x e y são independentes,

$$E(xy) = E(x) E(y)$$

pois,

$$\begin{aligned} E(xy) &= \int \int xy f(x, y) dx dy \\ &= \underbrace{\int x \rho(x) dx}_{E(x)} \underbrace{\int y \lambda(y) dy}_{E(y)} \end{aligned}$$

A.2.2. Funções de várias variáveis aleatórias

Se x e y são variáveis aleatórias com pdf conjunta $f(x, y)$ e $u = U(x, y)$, a pdf $\rho(u)$ pode ser determinada como uma pdf marginal, a partir de uma pdf conjunta $g(u, w)$ de u com outra função auxiliar arbitrária ⁴³

⁴³Em geral, a mais simples possível.

$w = W(x, y)$, ou seja,

$$\rho(u) = \int_{-\infty}^{\infty} g(u, w) dw$$

Assim, se as funções U e W admitem inversas,

$$\int \int f(x, y) dx dy = \int \int f[x(u, w), y(u, w)] \left| \frac{\partial(x, y)}{\partial(u, w)} \right| du dw$$

onde $\frac{\partial(x, y)}{\partial(u, w)} = \det \begin{pmatrix} \partial x / \partial u & \partial x / \partial w \\ \partial y / \partial u & \partial y / \partial w \end{pmatrix} = J(u, w)$

é o chamado **jacobiano** da transformação

$$(x, y) \rightarrow (u, w)$$

Desse modo,

$$\boxed{g(u, w) = f[x(u, w), y(u, w)] |J(u, w)|}$$

e,

$$\boxed{\rho(u) = \int_{-\infty}^{\infty} f[x(u, w), y(u, w)] |J(u, w)| dw}$$

Exemplos típicos da determinação de pdf de funções de duas variáveis x e y com pdf $\rho(x)$ e $\lambda(y)$.

$$\triangleright \quad \begin{cases} u = xy \\ w = x \end{cases} \implies \begin{cases} x = w \\ y = u/w \end{cases}$$

Assim,

$$J = \begin{vmatrix} 0 & 1 \\ 1/w & -u/w^2 \end{vmatrix} = -\frac{1}{w}$$

e,

$$\begin{aligned} g(u, w) &= \rho(w) \lambda(u/w) |1/w| \\ g(u) &= \int_{-\infty}^{\infty} \rho(w) \lambda(u/w) |1/w| dw \\ &= \int_0^{\infty} \rho(w) \lambda(u/w) \frac{1}{w} dw + \\ &\quad - \int_{-\infty}^0 \rho(w) \lambda(u/w) \frac{1}{w} dw \end{aligned}$$

$$\triangleright \quad \begin{cases} u = y/x \\ v = x \end{cases} \implies \begin{cases} x = v \\ y = uv \end{cases}$$

Assim,

$$J = \begin{vmatrix} 0 & 1 \\ v & u \end{vmatrix} = -v$$

e,

$$g(u, v) = \rho(v) \lambda(uv) |v|$$

$$g(u) = \int_0^{\infty} \rho(v) \lambda(uv) v \, dv +$$

$$- \int_{-\infty}^0 \rho(v) \lambda(uv) v \, dv$$

$$\triangleright \quad \begin{cases} u = x + y \\ v = x \end{cases} \implies \begin{cases} x = v \\ y = u - v \end{cases}$$

Assim,

$$J = \begin{vmatrix} 0 & 1 \\ 1 & -1 \end{vmatrix} = -1$$

e,

$$g(u, v) = \rho(v) \lambda(u - v)$$

$$g(u) = \int_{-\infty}^{\infty} \rho(v) \lambda(u - v) \, dv$$

A integral acima é denominada integral de convolução.

Por outro método, se $u = x + y$

$$G(t) = P(u \leq t) = P(x + y \leq t)$$

$$= \underbrace{\int \int}_R \rho(x) \lambda(y) \, dx dy$$

onde $R = \{(x, y) \mid x + y \leq t\}$

Assim,

$$G(t) = \int_{-\infty}^{\infty} dx \rho(x) \int_{-\infty}^{t-x} dy \lambda(y)$$

e,

$$G'(u) = g(u) = \int_{-\infty}^{\infty} \rho(x) \lambda(u-x) dx$$

A.3. A desigualdade de Markov

Se x é uma variável aleatória não-negativa, com valor esperado $\mu = E(x)$ e ε é qualquer real positivo,

$$\boxed{P(x \geq \varepsilon) \leq \mu/\varepsilon = E(x)/\varepsilon}$$

Uma vez que

$$E(x) = \int_0^{\infty} x \rho(x) dx = \int_0^{\varepsilon} x \rho(x) dx + \int_{\varepsilon}^{\infty} x \rho(x) dx$$
$$\mu \geq \int_{\varepsilon}^{\infty} x \rho(x) dx \geq \varepsilon \int_{\varepsilon}^{\infty} \rho(x) dx = \varepsilon P(x \geq \varepsilon)$$

A.4. A desigualdade de Chebyshev

Para qualquer variável aleatória x , com média μ e variância σ^2 , e qualquer real ε ,

$$P(|x - \mu| \geq \varepsilon) \leq \sigma^2/\varepsilon^2$$

ou,

$$P(|x - \mu| < \varepsilon) \leq 1 - \sigma^2/\varepsilon^2$$

pois, substituindo-se na desigualdade de Markov x por $(x - \mu)^2$,

$$P[(x - \mu)^2 \geq \varepsilon^2] \leq \frac{E[(x - \mu)^2]}{\varepsilon^2} = \sigma^2/\varepsilon^2$$

segue-se a desigualdade, desde que $(x - \mu)^2 \geq \varepsilon^2$ é equivalente ⁴⁴ a $|x - \mu| \geq \varepsilon$.

Essas desigualdades, obtidas a partir de pouquíssimas hipóteses sobre o comportamento de uma variável aleatória ⁴⁵, permitem estabelecer limites para probabilidades associadas a determinados intervalos de definição da variável.

A.5. A lei dos grandes números

Uma aplicação imediata da desigualdade de Chebyshev é na prova da chamada lei dos grandes números ou do limite da média das médias, que estabelece:

A média \bar{x}_N de N variáveis aleatórias independentes x_1, x_2, \dots, x_N com mesma média e variância tende ao valor esperado μ de cada uma das variáveis.

⁴⁴Definem eventos equivalentes.

⁴⁵Apenas a existência da média e da variância.

ou seja,

$$\lim_{N \rightarrow \infty} P[|\bar{x}_N - \mu| < \varepsilon] = 0$$

onde $\bar{x}_N = \frac{1}{N} \sum_{i=1}^N x_i$ e $\varepsilon > 0$

Uma que $V(x) = \sigma^2 \implies V(\bar{x}_N) = \sigma^2/N$, segue-se, da desigualdade de Chebyshev,

$$P[|\bar{x}_N - \mu| < \varepsilon] \leq 1 - \frac{\sigma^2}{N\varepsilon^2}$$

ou seja,

$$\lim_{N \rightarrow \infty} \bar{x}_N = \mu$$

Assim, a lei dos grandes números pode ser expressa como

$$\bar{x} = \lim_{N \rightarrow \infty} \sum_{i=1}^N \frac{x_i}{N} = \int_{-\infty}^{\infty} x \rho(x) dx = \langle x \rangle = \mu$$

↓

$$\bar{f}(x) = \lim_{N \rightarrow \infty} \sum_{i=1}^N \frac{f(x_i)}{N} = \int_{-\infty}^{\infty} f(x) \rho(x) dx = \langle f(x) \rangle$$

onde $\rho(x)$ é a pdf comum às variáveis aleatórias x e qualquer função de x , $f(x)$.

A.6. Funções características

A transformada de Fourier (T.F.) de uma pdf $\rho(x)$,

$$\text{T.F.}[\rho(x)] = \rho(t) = \int_{-\infty}^{\infty} e^{ixt} \rho(x) dx$$

onde t é um número real, é chamada **função característica**, ou seja,

$$\boxed{\rho(t) = E(e^{ixt})}$$

Expandindo-se a exponencial e^{ixt} em série de Taylor,

$$e^{ixt} = 1 + ixt + \frac{(ixt)^2}{2} + \frac{(ixt)^3}{3!} + \frac{(ixt)^4}{4!} + \dots$$

implica

$$\begin{aligned} E(e^{ixt}) &= \underbrace{E(1)}_1 + it \underbrace{E(x)}_{\langle x \rangle = \mu} + \\ &\quad - \frac{t^2}{2} \underbrace{E(x^2)}_{\langle x^2 \rangle = \sigma_x^2 + \mu^2} - \frac{t^3}{3!} \underbrace{E(x^3)}_{\langle x^3 \rangle} + \dots \end{aligned}$$

ou seja,

$$\rho(t) = 1 + it\mu - \frac{t^2}{2}(\sigma_x^2 + \mu^2) + \dots$$

- Propriedades da função característica

Se x e y são variáveis independentes, $\rho_x(t)$ e $\rho_y(t)$ suas respectivas funções características e α e β são números reais, a função característica de $z = f(x, y)$ será dada por

$$\triangleright \boxed{\rho_z(t) = \rho_x(t)\rho_y(t)} \quad \text{se } z = x + y$$

uma vez que

$$\rho_z(t) = E(e^{izt}) = E(e^{ixt}e^{iyt}) = \underbrace{E(e^{ixt})}_{\rho_x(t)} \underbrace{E(e^{iyt})}_{\rho_y(t)}$$

$$\triangleright \boxed{\rho_y(t) = e^{i\beta t}\rho_x(\alpha t)} \quad \text{se } y = \alpha x + \beta$$

pois,

$$\rho_y(t) = E(e^{iyt}) = E[e^{i(\alpha x + \beta)t}] = e^{i\beta t} \underbrace{E(e^{i\alpha x t})}_{\rho_x(\alpha t)}$$

\triangleright se a pdf de x é gaussiana,

$$\rho_G(x) = \frac{1}{\sigma_x \sqrt{2\pi}} e^{-\frac{1}{2} \left(\frac{x-\mu}{\sigma_x} \right)^2}$$

e,

$$\rho_G(t) = \int_{-\infty}^{\infty} e^{ixt} \rho_G(x) dx$$

Fazendo-se $z = \frac{x - \mu}{\sigma_x} \implies dz = \frac{dx}{\sigma_x}$

Assim,

$$\begin{aligned}\rho_G(t) &= \frac{e^{i\mu t}}{2\pi} \int_{-\infty}^{\infty} e^{i\sigma_x z t - z^2/2} dz \\ &= \frac{e^{i\mu t - \sigma_x^2 t^2/2}}{2\pi} \underbrace{\int_{-\infty}^{\infty} e^{-\frac{(z - i\sigma_x t)^2}{2}} dz}_{\int_{-\infty}^{\infty} e^{-\frac{w^2}{2}} dw = \sqrt{2\pi}}\end{aligned}$$

ou seja,

$$\boxed{\rho_G(t) = e^{i\mu t - \sigma_x^2 t^2/2}}$$

\Updownarrow

$$\log \rho_G(t) = i\mu t - \sigma_x^2 t^2/2$$

Para uma distribuição normal padrão, $\mathcal{N}_t(0, 1)$,

$$\boxed{\phi(t) = e^{-t^2/2}}$$

A.7. O teorema do limite central

Se $\{x_i\}$ é um conjunto de N variáveis aleatórias independentes associadas à uma mesma pdf $\rho(x)$, não necessariamente normal, com média e variância dados por

$$\begin{cases} E(x_i) = \mu \\ V(x_i) = \sigma^2 \end{cases}$$

A média $\bar{x} = \sum_{i=1}^N x_i/N$ é também uma variável aleatória de mesmo valor esperado $E(\bar{x}) = \mu$ e variância $V(\bar{x}) = \sigma^2/N$.

Nessas circunstâncias, a variável reduzida $\bar{z}_N = \frac{\bar{x} - \mu}{\sigma/\sqrt{N}}$ tende à distribuição normal padrão $\mathcal{N}_{\bar{z}_N}(0, 1)$, no sentido de que

$$\lim_{N \rightarrow \infty} P(\bar{z}_N \leq t) = \phi(t)$$

Uma vez que $\rho_{\bar{x}}(t) = [\rho(t/N)]^N$,

$$\begin{aligned}\rho_{\bar{z}_N}(t) &= e^{-\frac{i\mu t}{\sigma\sqrt{N}}} \rho_{\bar{x}}(\sqrt{N}t/\sigma) \\ &= e^{-\frac{i\mu t}{\sigma\sqrt{N}}} \left[\rho\left(t/\sqrt{N}\sigma\right) \right]^N\end{aligned}$$

Tomando-se o logaritmo,

$$\log \rho_{\bar{z}_N}(t) = -\frac{i\mu t}{\sigma\sqrt{N}} + N \log \rho\left(t/\sqrt{N}\sigma\right)$$

e, levando-se em conta que,

$$\begin{aligned}\rho\left(t/\sqrt{N}\sigma\right) &= 1 + \frac{i\mu t}{\sqrt{N}\sigma} - \frac{(\sigma^2 + \mu^2)}{2N\sigma^2}t^2 + \dots \\ \log \rho_{\bar{z}_N}(t) &= -\frac{i\mu t}{\sigma\sqrt{N}} + N \log \left[1 + \frac{i\mu t}{\sqrt{N}\sigma} - \frac{(\sigma^2 + \mu^2)}{2N\sigma^2}t^2 + \dots \right] \\ &= -\frac{i\mu t}{\sigma\sqrt{N}} + N \left[\frac{i\mu t}{\sqrt{N}\sigma} - \frac{(\sigma^2 + \mu^2)}{2N\sigma^2}t^2 + \dots \right] + \\ &\quad - \frac{1}{2} \left(\frac{-\mu^2}{N\sigma^2}t^2 - \frac{i\mu t^3}{\sqrt{N}} \frac{(\sigma^2 + \mu^2)}{N} + \frac{(\sigma^2 + \mu^2)}{4N^2\sigma^4}t^4 + \dots \right)\end{aligned}$$

Assim,

$$\lim_{N \rightarrow \infty} \log \rho_{\bar{z}_N}(t) = -\frac{t^2}{2}$$

↓

$$\mathcal{N}_{\bar{z}_N}(0, 1)$$

A.8. A distribuição gama

Se z é uma variável aleatória distribuída normalmente, segundo $\mathcal{N}(0, 1)$ e,

$$u(z) = \frac{1}{2} z^2$$

a pdf associada a u é dada por

$$g(u) = \frac{1}{\sqrt{\pi}} u^{-\frac{1}{2}} e^{-u}$$

Uma vez que a função $u(z)$ não é monótona, sua pdf deve ser determinada como

$$\begin{aligned}G(t) &= P(u \leq t) = P(1/2z^2 \leq t) \\&= P(z^2 \leq 2t) = P(-\sqrt{2t} \leq z \leq \sqrt{2t}) \\&= F(\sqrt{2t}) - F(-\sqrt{2t}) = \int_{-\sqrt{2t}}^{\sqrt{2t}} \rho_G(z) dz \\&\Downarrow \\ \frac{dG}{du} &= g(u) = \frac{1}{\sqrt{2\pi}} 2(2u)^{-1/2} e^{-u}\end{aligned}$$

Sabendo-se que $\Gamma(1/2) = \sqrt{\pi}$,⁴⁶ $g(u)$ pode ser escrita como

$$g(u) = \frac{u^{-1/2}}{\Gamma(1/2)} e^{-u}$$

⁴⁶A função gama é definida por

$$\Gamma(x) = \int_0^{\infty} t^{x-1} e^{-t} dt$$

e algumas de suas propriedades são:

$$\triangleright \Gamma(x+1) = \int_0^{\infty} t^x e^{-t} dt = \left| -t^x e^{-t} \right|_0^{\infty} + x \underbrace{\int_0^{\infty} t^{x-1} e^{-t} dt}_{\Gamma(x)} = x \Gamma(x)$$

Essa distribuição é um caso particular da chamada distribuição gama,

$$\gamma(u|\alpha) = \frac{u^{\alpha-1}}{\Gamma(\alpha)} e^{-u}$$

onde α é o parâmetro da distribuição denominado número de graus de liberdade.

$$\triangleright \Gamma(1) = \int_0^{\infty} e^{-t} dt = 1$$

$$\triangleright \Gamma(n+1) = n!$$

$$\begin{aligned} \triangleright \Gamma(1/2) &= \int_0^{\infty} t^{-1/2} e^{-t} dt \quad (t = u^2) \\ &= 2 \int_0^{\infty} e^{-u^2} du = \int_{-\infty}^{\infty} e^{-u^2} du = \sqrt{\pi} \end{aligned}$$

$$\triangleright \Gamma(3/2) = \frac{1}{2}\Gamma(1/2)$$

A função característica de uma distribuição gama é dada por

$$\begin{aligned}\gamma(t) &= \frac{1}{\Gamma(\alpha)} \int_{-\infty}^{\infty} e^{-\overbrace{u(1-it)}^v} u^{\alpha-1} du \\ &= \frac{(1-it)^{-\alpha}}{\Gamma(\alpha)} \underbrace{\int_{-\infty}^{\infty} e^{-v} v^{\alpha-1} dv}_{\Gamma(\alpha)}\end{aligned}$$

ou seja,

$$\boxed{\gamma(t) = (1-it)^{-\alpha}}$$

A.9. A distribuição de χ^2

Se $\{x_i\}$ é um conjunto de N variáveis independentes distribuídas normalmente com média μ_i e desvio padrão σ_i , a variável

$$z_i = \frac{x_i - \mu_i}{\sigma_i}$$

distribuir-se-á normalmente segundo $\mathcal{N}(0, 1)$ e,

$$\xi = \sum_{i=1}^N \frac{z_i^2}{2} = \sum_{i=1}^N u_i$$

estará associada à função característica

$$\begin{aligned}g_{\xi}(t) &= \prod_{i=1}^N g_{u_i}(t) = g_{u_i}^N(t) = (1 - it)^{-N\alpha} \\ &= (1 - it)^{-N/2} \quad (\alpha = 1/2)\end{aligned}$$

cuja pdf é dada por

$$g(\xi|N) = \frac{\xi^{N/2-1} e^{-\xi}}{\Gamma(N/2)} = \gamma(\xi|n/2)$$

Assim, a variável $\chi^2 = 2\xi$ ($\xi \geq 0$) distribuir-se-á segundo

$$\rho(\chi^2) = g[\xi(\chi^2)] \frac{d\xi}{d\chi^2}$$

ou seja,

$$\rho(\chi^2|N) = \frac{(\chi^2/2)^{N/2-1} e^{-\chi^2/2}}{2\Gamma(N/2)}$$

Essa pdf é denominada **distribuição de χ^2** com N graus de liberdade.

O valor esperado de χ^2 pode ser calculado como

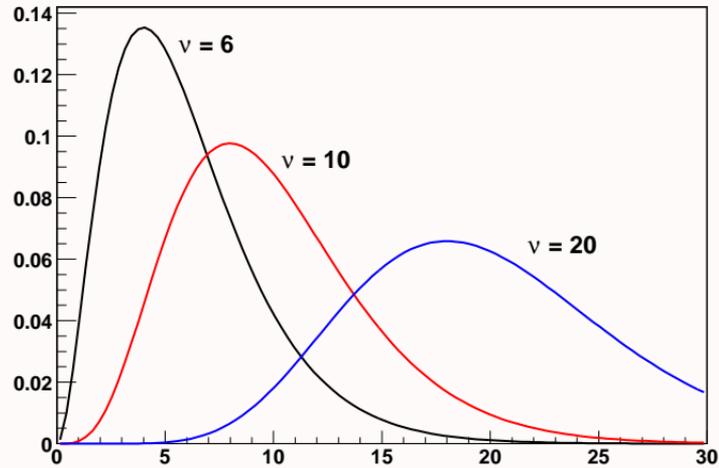


Figura A.8: Distribuições de χ^2 , para diversos graus de liberdade.

$$\begin{aligned}
\langle \chi^2 \rangle &= \int_0^\infty \chi^2 \rho(\chi^2) d\chi^2 \\
&= \frac{(1/2)^{\nu/2}}{\Gamma(\nu/2)} \int_0^\infty \underbrace{\chi^2 (\chi^2)^{\nu/2-1}}_{(x^2)^{\nu/2}} e^{-\chi^2/2} d\chi^2 \\
&\qquad\qquad\qquad \Gamma(\nu/2+1)/(1/2)^{\nu/2+1}
\end{aligned}$$

Uma vez que $\Gamma(x+1) = x\Gamma(x)$, obtém-se

$$\langle \chi^2 \rangle = \nu$$

A.10. A distribuição de Student

Se x, x_1, x_2, \dots, x_n são $(n+1)$ variáveis independentes e distribuídas normalmente segundo $\mathcal{N}(0, \sigma)$, a variável

$$t = \frac{x}{\sqrt{\frac{\sum_{i=1}^n x_i^2}{n}}}$$

obedecerá a chamada **distribuição de Student**.

Pois, uma vez que a variável

$$\lambda = \sum_{i=1}^n \left(\frac{z_i^2}{2} \right)$$

onde $z_i = x_i/\sigma$, possui pdf igual a

$$\rho(\lambda) = \gamma(\lambda|n/2) = \frac{\lambda^{n/2} e^{-\lambda}}{\Gamma(n/2)}$$

a variável

$$u = \sum_{i=1}^n x_i^2 = 2\sigma^2 \sum_{i=1}^n \frac{1}{2} \left(\frac{x_i^2}{\sigma} \right)^2 = 2\sigma^2 \sum_{i=1}^n \frac{z_i^2}{2} = 2\sigma^2 \lambda$$

terá pdf

$$\begin{aligned} \rho(u) &= \frac{d\lambda(u)}{du} \gamma[\lambda(u)|n/2] \\ &= \frac{1}{2\sigma} \gamma\left(\frac{u}{2\sigma^2} \middle| \frac{n}{2}\right) = \frac{1}{2\sigma} \left(\frac{u}{2\sigma^2}\right)^{n/2-1} \frac{e^{-u/(2\sigma^2)}}{\Gamma(n/2)} \end{aligned}$$

Além disso, a variável

$$\eta = \left(\frac{\sum_{i=1}^n x_i^2}{n} \right)^{1/2} = \frac{1}{n^{1/2}} (2\sigma^2)^{1/2} \left[\sum_{i=1}^n \frac{1}{2} \left(\frac{x_i}{\sigma} \right)^2 \right]^{1/2}$$

ou seja,

$$\eta = \left(\frac{2\sigma^2}{n} \right)^{1/2} \lambda^{1/2} \quad (\eta > 0)$$

terá pdf dada por

$$\rho(\eta) = \rho[\lambda(\eta)] \frac{d\lambda}{d\eta}$$

ou seja,

$$\begin{aligned}\rho(\eta) &= \frac{n\eta}{\sigma^2} \gamma\left(\frac{n}{2} \frac{\eta^2}{\sigma^2} \middle| \frac{n}{2}\right) \\ &= \frac{n\eta}{\sigma^2} \frac{\left(\frac{n}{2} \frac{\eta^2}{\sigma^2}\right)^{n/2-1}}{\Gamma(n/2)} e^{-\frac{n}{2} \frac{\eta^2}{\sigma^2}} \\ &= \frac{2(n/2)^{n/2} \eta^{n-1} e^{-\frac{n}{2} \frac{\eta^2}{\sigma^2}}}{\sigma^n \Gamma(n/2)}\end{aligned}$$

Assim, a pdf conjunta de $t = x/\eta$ e η será dada por

$$\rho(x/\eta, \eta) = \rho[x(\eta)] \rho(\eta) |\eta|$$

ou seja,

$$\rho(x/\eta, \eta) = \frac{\sqrt{\frac{2}{\pi}} \left(\frac{n}{2}\right)^{n/2}}{\sigma^{n+1} \Gamma(n/2)} \eta^{n-1} e^{-\frac{(x^2 + n\eta^2)}{2\sigma^2}} |\eta|$$

Fazendo-se

$$\frac{\sqrt{\frac{2}{\pi}} \left(\frac{n}{2}\right)^{n/2}}{\sigma^{n+1} \Gamma(n/2)} = c_n$$

e

$$\frac{(x^2 + n\eta^2)}{2\sigma^2} = \frac{\eta^2}{2\sigma^2} [(x/\eta)^2 + n] = \eta^2 \frac{(t^2 + n)}{2\sigma^2} = a_n$$

A pdf de t será dada por

$$\rho(t) = c_n \int_0^\infty \eta^n e^{-a_n \eta^2} d\eta$$

Fazendo-se

$$\begin{cases} a_n \eta^2 = v \\ 2a_n \eta d\eta = dv \end{cases} \implies \begin{cases} \eta^n = (v/a_n)^{n/2} \\ d\eta = \frac{dv}{2a_n^{1/2} v^{1/2}} \end{cases}$$

$$\rho(t) = \frac{c_n}{2} a_n^{\left(\frac{n+1}{2}\right)} \underbrace{\int_0^\infty v^{\frac{n-1}{2}} e^{-v} dv}_{\Gamma\left(\frac{n+1}{2}\right)}$$

Desde que

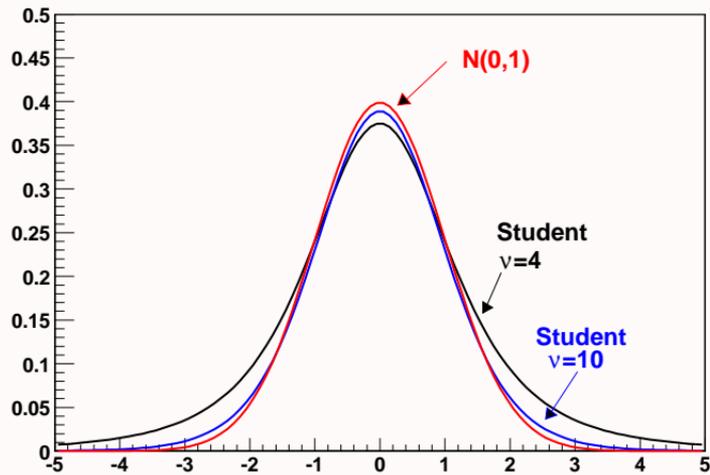
$$\begin{aligned} a_n \left(\frac{n+1}{2} \right) &= \left(\frac{t^2+n}{2\sigma^2} \right)^{-\frac{n+1}{2}} \\ &= 2 \left(\frac{n+1}{2} \right) \sigma^{n+1} n^{-\left(\frac{n+1}{2}\right)} \left(1 + \frac{t^2}{n} \right)^{-\left(\frac{n+1}{2}\right)} \end{aligned}$$

e

$$\frac{c_n}{2} = \frac{2^{-1/2} n^{n/2} 2^{-n/2}}{\sqrt{\pi} \sigma^{n+1} \Gamma(n/2)} = \frac{2^{-\left(\frac{n+1}{2}\right)} n^{n/2}}{\sqrt{\pi} \sigma^{n+1} \Gamma(n/2)}$$

a chamada distribuição de Student com n graus de liberdade é dada por

$$\rho(t|n) = \frac{1}{\sqrt{n\pi}} \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)} \left(1 + t^2/n \right)^{-\left(\frac{n+1}{2}\right)}$$



$$\rho(t|\nu) = A (1 + t^2/\nu)^{-(\nu + 1)/2}$$

$$\sigma^2 = \frac{\nu}{\nu - 2} \quad (\nu - \text{número de graus de liberdade})$$

Apêndice B

O fenômeno da radioatividade é o exemplo mais simples de um processo físico aleatório que pode ser utilizado para ilustração de um método de simulação direta.

Se a probabilidade (p) de que um núcleo de um isótopo radioativo decaia em outro núcleo, emitindo uma partícula α , em um pequeno intervalo de tempo dt é dada por

$$p = \lambda dt$$

e, inicialmente, existem N_0 núcleos, o número médio de núcleos restantes $\langle N \rangle_t$, após um intervalo de tempo t , é dado pela lei de decaimento exponencial,

$$\langle N \rangle_t = N_0 e^{-\lambda t}$$

onde λ , a chamada constante de decaimento está relacionada com a meia-vida ⁴⁷ ($T_{1/2}$) do isótopo por $\lambda = (\log 2)/T_{1/2}$.

O fragmento de algoritmo abaixo pode ser utilizado para simular esse comportamento.

```
N = NO (numero inicial de nucleos)
LOOP de t=0 a T, step dt
  Nr = N
  LOOP sobre nucleos restantes (Nr)
```

⁴⁷Intervalo de tempo em que o número médio de núcleos que decaem é reduzido à metade. $T_{1/2}$ varia de 3×10^{-7} s (${}_{84}\text{Po}^{212}$) até 5×10^{15} anos (${}_{60}\text{Nd}^{144}$).

```
IF [random(1) < p] N = N - 1
END LOOP sobre nucleos
WRITE t,N
END LOOP temporal
```

Esse algoritmo pressupõe o conhecimento da quantidade inicial de núcleos, informação que não é experimentalmente factível.

O Experimento de Rutherford-Geiger

Experimentalmente, para isótopos com meia-vidas da ordem de horas, ou seja, constantes de decaimentos da ordem de 10^{-4}s^{-1} , as partículas α emitidas durante certos intervalos de tempo ($T \approx 10\text{s}$) podem ser detectadas e contadas. Esse foi o procedimento realizado por E. Rutherford e Geiger, que observaram o decaimento de uma amostra de polônio (Po), em um certo número (2608) de intervalos de tempo (7.5s) pré-determinados.

Uma vez que a quantidade de núcleos é da ordem de 10^{23} , pode-se considerar que, em cada intervalo de tempo pré-determinado (T), muito menor que a meia-vida do isótopo, a probabilidade (p) de detecção de uma partícula α , além de ser extremamente pequena, é constante e igual a $p = \lambda T$, pois o número de núcleos praticamente não se altera.

Desse modo, o número (m) de partículas α detectadas obedece a uma distribuição binomial que tende a uma distribuição de Poisson.

Referências Bibliográficas

- [1] D. J. Bennett. *Aleatoriedade*, Martins Fontes, 2003.
- [2] P. R. Bevington, D. K. Robinson. *Data Reduction and Error Analysis for the Physical Sciences*, 2nd. edition, McGraw-Hill, 1992.
- [3] E. Borel. *Probabilidade e Certeza*, Difusão Européia do Livro, 1966.
- [4] C. B. Boyer. *História da Matemática*, Edgard Blücher., 1974.
- [5] H. Cramér. *Elementos da Teoria da Probabilidade (e algumas de suas aplicações -1949)*, Mestre Jou, 1973.
- [6] C. A. B. Dantas. *Probabilidade: um curso introdutório*, Gradiva, 2006.
- [7] F. N. David. *Games, Gods and Gambling: A History of probability and statistical Ideas (1962)*, Dover, 1998.
- [8] P. A. M. Dirac. *The Principles of Quantum Mechanics*, Oxford Univ. Press, 1958.
- [9] R. P. Feynman. *Statistical Mechanics*, W.A. Benjamin, 1972.
- [10] R. A. Fisher. *Statistical Methods for Research Workers (1925)*, 13th. edition, Oliver & Boyd, 1958.
- [11] R. A. Fisher. *Statistical Methods and Scientific Inference (1925)*, 2nd. edition, Oliver & Boyd, 1959.
- [12] B. V. Gnedenko. *The Theory of Probability*, Mir Publishers, 1978.
- [13] J. W. Gibbs. *Elementary Principles in Statistical Mechanics (1904)*, Ox Bow Press, 1981.

- [14] D. L. Goodstein. *State of Matter*, Dover, 1975.
- [15] V. R. Vanin & P. Gouffon. *Tópicos Avançados em Tratamento Estatístico de Dados em Física Experimental*, LAL - IFUSP, 1996.
- [16] J. M. de Sá. *O acaso: a vida do jogo e o jogo da vida*, Gradiva, 2006.
- [17] M. G. Kendall & A. Stuart. *The Advanced Theory of Statistics*, Vol. 1, Distribution Theory, Griffin & Co. Ltd., 1958.
- [18] M. G. Kendall & A. Stuart. *The Advanced Theory of Statistics*, Vol. 2, Inference and Relationship, Griffin & Co. Ltd., 1961.
- [19] A. N. Kolmogorov. *Foundations of the Theory of Probability (1933)*, Chelsea Pub., 1956.
- [20] L. D. Landau, E. M. Lifshitz. *Statistical Physics*, Pergamon Press, 1968.
- [21] P. S. Laplace. *Ensaio Filosófico sobre as probabilidades*, Contraponto, 2010.
- [22] R. B. Lindsay & H. Margenau. *Foundations of Physics*, Ox Bow Press, 1981.
- [23] L. Lyons. *Statistical for Nuclear and Particle Physics*, 3rd. Edition, Cambridge Univ. Press, 1986.
- [24] M. N. Nascimento & A. C. P. de Lima *Noções de Probabilidade e Estatística*, 6ª ed. Edusp, 2007.
- [25] F. Mandl. *Statistical Physics*, John Wiley, 1971.
- [26] J. Mandel. *The Statistical Analysis of Experimental Data*, Dover, 1964.
- [27] P. L. Meyer. *Probabilidade: aplicações à Estatística*, 2ª ed., LTC, 1983
- [28] R. von Mises. *Probability, Statistics and Truth*, 2nd. ed., Dover, 1981.
- [29] J. Orear. *Notes on Statistics for Physicists*, Revised. Laboratory for Nuclear Studies, Cornell University, 1982.
- [30] A. Papoulis. *The Meaning of Probability*, IEEE Transactions on Education, June- September, 19xx.
- [31] M. Planck. *The Theory of Heat Radiation (1913)*, Dover, 1991.
- [32] J. V. Plato. *Creating Modern Probability*, Cambridge Univ. Press, 1994.
- [33] S. Ross. *Probabilidade: um curso moderno com aplicações*, Gradiva, 2006.
- [34] D. Ruelle. *Chance and Chaos*, Princeton Univ. Press, 1991.
- [35] D. E. Smith. *An Source Book in Mathematics*, Dover, 1959.

- [36] G. L. Squires. *Practical Physics*, 3rd. Edition, Cambridge Univ. Press, 1985.
- [37] S. M. Stigler. *The History of Statistics*, Harvard Univ. Press, 1986.
- [38] H. Tijms. *Understanding Probability*, Cambridge Univ. Press, 2004.
- [39] J. Varela. *O século dos quanta*, Gradiva, 1996.
- [40] J. H. Vuolo. *Fundamentos da Teoria de Erros*, Egdard Blücher, 1992.
- [41] G. U. Yale & M. G. Kendall. *An Introduction to the Theory of Statistics*, Griffin & Co. Ltd., 1968.