Elementos de Teoria da Probabilidade

Oscar J. P. Éboli e J. Fernando Perez

Instituto de Física
Universidade de São Paulo

Fevereiro de 2004

APÊNDICE A

Breve Introdução à Teoria da Probabilidade

A Mecânica Quântica traz em sua estrutura objetos que, *independente de qualquer postulado interpretativo*, são descritos em Matemática no contexto da Teoria das Probabilidades. Os postulados interpretativos tornam-se naturais do ponto de vista conceitual ao constatarmos esse fato. Este apêndice tem por objetivo sumarizar e familiarizar o leitor com esse instrumental matemático.

A.1 Conceitos básicos

A teoria da probabilidade é o capítulo da Matemática que fornece as ferramentas conceituais para a análise de fenômenos que em Ciências Naturais são classificados como *aleatórios*. A palavra *aleatório* é usada para contrastar com a palavra *determinístico* no seguinte sentido. Consideremos um conjunto $\mathcal{C}$ de condições sob as quais um determinado experimento é realizado. Um determinado evento A é denominado *certo*, *seguro* ou *necessário* se cada vez que o experimento for realizado observando-se exatamente as mesmas condições $\mathcal{C}$, verificar-se a ocorrência do evento A. Se ao contrário sob as mesmas condições $\mathcal{C}$ o evento A nunca ocorre, ele é declarado *impossível*. Um sistema físico é dito determinístico quando cada evento pode ser apenas necessário ou impossível.

Exemplo: Se um objeto apenas sob a ação da força da gravidade, num local onde a aceleração da gravidade tem o valor g, for abandonado com velocidade inicial zero e a uma altura h do solo (conjunto $\mathcal{C}$ de condições), então necessariamente o objeto atingirá o solo com velocidade $v = \sqrt{2gh}$ (evento A). De maneira geral, todos os fenômenos descritos no âmbito da Mecânica Clássica, quando o conjunto $\mathcal{C}$ de condições fornece com precisão absoluta as posições e velocidades iniciais de todas as partículas envolvidas bem como as forças atuantes, são determinísticos, isto é, todo evento é necessário ou impossível.

Há porém fenômenos para os quais a especificação de um conjunto $\mathcal{C}$ de condições pode ser insuficiente para que um determinado evento A seja **apenas** necessário ou impossível, *i.e.* o evento A as vezes ocorre e as vezes não. Tais fenômenos são alcunhados de *aleatórios*.

Exemplo: Os exemplos mais típicos são os lançamentos de um mesmo dado ou de uma mesma moeda (conjunto $\mathcal{C}$ de condições) e os eventos referem-se aos resultados, por exemplo, $A =$ dar cara no lançamento da moeda ou dar o número 4 no lançamento do dado. Note que se o conjunto $\mathcal{C}$ de condições fornecer exatamente a posição inicial e velocidade inicial de todos os pontos dos sólidos envolvidos então esses dois fenômenos são de natureza determinística: é só a nossa ignorância sobre as condições iniciais bem como a dificuldade de precisá-las de forma absoluta que determinam o caráter aleatório desses fenômenos. Uma descoberta relativamente recente mostra porém que os chamados sistemas caóticos, embora descritos no âmbito da Mecânica Clássica e portanto determinísticos, por apresentarem extrema sensibilidade em relação às condições iniciais, devem ser tratados como aleatórios já que mínimas incertezas com relação às condições iniciais implicam num grau imprevisibilidade quase que absoluto.

A.1.1 Espaço amostral

Dado um fenômeno aleatório, tal como o resultado de uma roleta ou de um dado, definimos o **ESPAÇO AMOSTRAL** S como sendo o conjunto de todos os resultados possíveis.

Exemplos:

- Para descrever o resultado do lançamento de uma moeda temos um espaço amostral com 2 elementos, a saber, $S_0 = \{-1,\ +1\}$ onde o ponto -1 representa $C \equiv$ cara e $+1$ representa $K \equiv$ coroa. É claro que o conjunto S_0 pode ser substituído por qualquer conjunto com exatamente dois elementos.

- No caso do lançamento de 1 dado o espaço amostral pode ser representado pelo conjunto $S_1 = \{1, 2, 3, 4, 5, 6\}$.

- Para o lançamento de 2 dados o espaço amostral é dado pelo produto cartesiano $S_2 = S_1 \times S_1 = \{(1,1), (1,2), (1,3), \ldots, (6,5), (6,6)\} = \{x = (x_1, x_2) \mid x_i \in S_1,\ i = 1, 2\}$ com 36 elementos. Analogamente o lançamento de N dados pode ser descrito por $S = (S_1)^N = \{x = (x_1, x_2, \ldots, x_N) \mid x_i \in S_1,\ i = 1, 2, ..., N\}$ ou por qualquer conjunto com 6^N elementos.

- No lançamento de 3 moedas podemos utilizar

$$\begin{aligned} S_3 &= \{(CCC),\ (CCK),\ (CKC),\ (KCC), \\ &\quad\ (CKK),\ (KCK),\ (KKC),\ (KKK)\} \\ &= \{C, K\} \times \{C, K\} \times \{C, K\}\ , \end{aligned}$$

  com 2^3 elementos, ou simplesmente $S_3 = (S_0)^3 = \{x = (x_1, x_2, x_3) \mid x_i \in S_0,\ i = 1, 2, 3\}$. De novo, para o lançamento de N moedas podemos tomar como espaço amostral o produto cartesiano $S = (S_0)^N = \{x = (x_1, x_2, \ldots, x_N) \mid x_i \in S_0,\ i = 1, 2, \ldots, N\}$.

- De maneira geral se S é o espaço amostral que descreve um determinado fenômeno aleatório, o espaço adequado para descrever N repetições *dependentes ou independentes* do experimento é o produto cartesiano $(S)^N = \{x = (x_1, x_2, \ldots, x_N) \mid x_i \in S,\ i = 1, 2, \ldots, N\}$.

- O espaço amostral que representa a posição num dado instante de uma partícula executando movimento Browniano pode ser descrito por $S = \mathbb{R}^3$.

- A vida útil de lâmpadas em horas está associada ao espaço amostral $S = \{x \in \mathbb{R} \mid 0 < x \leq 5000 \text{ horas}\}$.

A.1.2 Eventos

Definimos um **EVENTO** (E) como sendo qualquer subconjunto do espaço amostral. Nas aplicações práticas um evento refere-se a algum aspecto observável num determinado experimento com espaço amostral S dado.

Exemplos:

- Dado S_1 acima, podemos considerar os seguintes eventos:

  $E_1 = \{1, 3, 5\}$, isto é, o resultado é ímpar.

  $E_2 = \{2, 4, 6\}$, isto é, o resultado é par.

  $E_3 = \{1, 2, 3, 5\}$, isto é, o resultado é primo.

- Para o espaço S_2 acima, podemos ter $E_1 = \{(2, 2),\ (1, 3),\ (3, 1)\}$ $= \{(x_1, x_2) \in S_2 \mid x_1 + x_2 = 4\}$, isto é, soma 4 em 2 lançamentos.

- No lançamento de três moedas S_3: $F_1 = \{2 \text{ ou mais caras}\} = \{(CCC), (CCK), (CKC), (KCC)\}$ correspondendo à propriedade de haver pelo menos 2 "caras " em 3 lançamentos da moeda. Alternativamente o conjunto F_1 pode ser descrito através de $F_1 = \{x = (x_1, x_2, x_3) \in S_3 | x_1 + x_2 + x_3 \leq -1\}$.

- Duas pessoas combinam de se encontrar em um determinado local no intervalo de tempo que vai das 2 até as 3 horas da tarde, com a condição de que uma só espera pela outra por no máximo 10 minutos. Supondo que as pessoas chegarão aleatoriamente ao local dentro do intervalo de tempo combinado, o espaço amostral adequado pode ser descrito como $S_4 = [0,60]^2 = \{x = (x_1, x_2) \mid x_i \in [0,60],\ i = 1,2\}$. O evento E correspondente à ocorrência do encontro é dado por

$$E = \{x \in S_4 \mid |x_1 - x_2| < 10\} .$$

É conveniente introduzir dois eventos triviais:

$E = S \equiv$ evento certo pois sempre ocorre;

$E = \emptyset$ (vazio) $\equiv$ evento impossível visto que nunca ocorre.

Mais ainda, dados dois ou mais eventos E e F , podemos gerar novos eventos através das operações usuais de conjuntos: $E \cup F$, $E \cap F$, $E \backslash F$. Assim, dado um evento E, definimos o seu evento complementar $E^c = S \backslash E$.

Definição: dizemos que dois eventos E e F são **eventos mutuamente exclusivos** se $E \cap F = \emptyset$.

Exemplos:

- Em S_1 os eventos E_1 e E_2 acima, são mutuamente exclusivos dado que $E_1 \cap E_2 = \emptyset$.
- Em S_3 definindo $F_2 = \{KKK\}$, segue que F_1 e F_2 são mutuamente exclusivos já que $F_1 \cap F_2 = \emptyset$.

A.2 Probabilidades

Existem fenômenos aleatórios que apresentam uma característica notável: dado o conjunto $\mathcal{C}$ de condições, embora elas sejam insuficientes para que um determinado evento A seja certo ou impossível, repetindo-se o experimento N vezes observa-se que a fração

$$\frac{n_A}{N} ,$$

onde n_A é o número de vezes em que o evento A ocorreu, tende a estabilizar-se quando N cresce, convergindo para um número que denotaremos por

$$p_A = \lim_{N\to\infty} \frac{n_A}{N} \ .$$

Quando um fenômeno aleatório apresenta esta notável propriedade para todo evento A, dizemos que esse sistema é *probabilístico ou estocástico* e que o número p_A é a *probabilidade* do evento A. Para um tal sistema o número p_A representa a freqüência de ocorrência do evento A .

Essas considerações juntamente com as propriedades de freqüências **motivam** a introdução de definição abaixo.

Definição: A cada evento E de um espaço amostral S associamos um número real $\mathbf{P}(E)$, chamado probabilidade, o qual satisfaz as seguintes propriedades:

1. $0 \leq \mathbf{P}(E) \leq 1$,

2. $\mathbf{P}(S) = 1$,

3. Se os eventos E_i, $i = 1, 2, ...$ são mutuamente exclusivos então $\mathbf{P}(\cup_i E_i) = \sum_i \mathbf{P}(E_i)$.

Exemplos:

- No lançamento de um dado honesto, a probabilidade de uma dada face é 1/6, *i.e.* para o espaço amostral S_1 se $E = \{x\}$ é um conjunto com um único elemento $x \in S_1$, então $\mathbf{P}(E) = \frac{1}{6}$. A probabilidade de um evento qualquer é obtida a partir destas utilizando a propriedade (3) acima.[1]

- Uma classe muito grande de exemplos pode ser descrita através de espaços amostrais finitos, *i.e.*, S é uma coleção finita

$$S = \{x_1, ..., x_N\}$$

[1] **Exercício:** Mostre este fato, e compute a probabilidade do evento associado ao resultado ser primo.

de elementos. Nesse caso a probabilidade $\mathbf{P}(E)$ de um evento qualquer $E \subset S$ pode, como no exemplo acima, ser completamente definida a partir de uma coleção de N números $\{p_1, ..., p_N\}$ satisfazendo

$$p_i \geq 0 \ , \ \sum_{i=1}^{N} p_i = 1 \ ,$$

onde

$$\mathbf{P}(x_i) = p_i$$

e

$$\mathbf{P}(E) = \sum_{x_i \in E} p_i \ .$$

Por exemplo, se todos os resultados possíveis x_i tem a mesma freqüência de ocorrência temos

$$p_i = \frac{1}{N} \ ,$$

em cujo caso a probabilidade de um evento E é dada por

$$\mathbf{P}(E) = \frac{n_E}{N} \ ,$$

onde n_E é o número de elementos no conjunto E.

- No exemplo das 3 moedas lançadas independentemente, *i.e.* para o espaço amostral S_3 a probabilidade de uma seqüência $x \in S_3$ é $p_x = 1/8$.

- Consideremos o lançamento de dois dados que é descrito pelo espaço amostral S_2. A probabilidade de no segundo lançamento sair 2 ou 3, *i.e.* a probabilidade do evento $E = \{x \mid x_2 \in \{2, 3\}\}$ é dada por $\mathbf{P}(E) = \frac{1}{3}$. Justifique esse resultado intuitivamente óbvio a partir da definição de $\mathbf{P}(E)$.

- Considere o espaço amostral $S_4 = [0, 60]^2$ introduzido acima para descrever o problema do encontro de suas pessoas. A probabilidade de um evento qualquer $F \subset S_4$ é dada pela razão

$$\mathbf{P}(F) = \frac{\text{área de } F}{\text{área de } S_4} \ .$$

Exercício: Calcule a probabilidade $\mathbf{P}(E)$ de ocorrer o encontro.

- De maneira geral a escolha aleatória de um número real no intervalo $S_5 = [a, b]$, onde $a < b$, tem a probabilidade do número estar em $E \subset S_5$ dada por

$$\mathbf{P}(E) = \frac{|E|}{b-a} \ ,$$

onde $|E|$ é o comprimento do conjunto E.

A.2.1 Propriedades da probabilidade

Pode-se demonstrar[2] que a definição dada para a probabilidade $\mathbf{P}$ conduz às seguintes propriedades:

1. $\mathbf{P}(\phi) = 0$ (Dica: use $S = S \cup \phi$ e o axioma (3))
2. $\mathbf{P}(E^c) = 1 - \mathbf{P}(E)$ (Use que $S = E + E^c \,, \ E \cap E^c = \phi$)
3. $\mathbf{P}(E \cup F) = \mathbf{P}(E) + \mathbf{P}(F) - \mathbf{P}(E \cap F)$
4. $\mathbf{P}(E \cup F) = \mathbf{P}(E) + \mathbf{P}(E^c \cap F)$

A.2.2 Distribuições de probabilidade

Para espaços amostrais contínuos, como nos dois exemplos anteriores, a situação mais típica é quando $S \subset \mathbb{R}^n$ e a probabilidade de eventos é definida através de uma função $\rho : S \to \mathbb{R}$ com as seguintes propriedades:

a) $\rho(x) \geq 0$, para todo $x = (x_1, ..., x_n) \in S$ e
b) $\int_S \rho(x)\, d^n x = 1$.

A probabilidade de evento $E \subset S$ é então expressa através de

$$\mathbf{P}(E) = \int_E \rho(x)\, d^n x \ .$$

[2] Demonstre estas propriedades!

Por essa razão a função ρ é denominada de *densidade de probabilidade*.

Exercício: Determine as densidades de probabilidade nos dois exemplos anteriores.

Exemplos:

- Na reta real, $S = \mathbb{R}$ algumas densidades de probabilidade freqüentes são:

  1. Distribuição gaussiana, definida através de 2 parâmetros $\sigma > 0$ e $x_0 \in \mathbb{R}$
  $$\rho(x) = \frac{1}{\sqrt{2\pi}\ \sigma} \exp\left[-\frac{(x - x_0)^2}{2\sigma^2}\right] \ ;$$
  2. Distribuição de Cauchy com parâmetro $\gamma > 0$
  $$\rho(x) = \frac{\gamma}{\pi} \frac{1}{x^2 + \gamma^2} \ ;$$
  3. Distribuição exponencial com parâmetro $\gamma > 0$
  $$\rho(x) = \begin{cases} \gamma e^{-\gamma x} \text{ se } x > 0 \\ 0 \text{ se } x \leq 0 \end{cases} \quad .$$

Exercício: Verifique que as 3 funções ρ acima definem densidades de probabilidade.

- Se uma partícula em $\mathbb{R}^3$ parte da origem no instante $t = 0$ e executa um movimento Browniano num meio com constante de difusão D, a probabilidade da partícula no instante $t \geq 0$ encontrar-se no conjunto $E \subset \mathbb{R}^3$ é dada através da densidade de probabilidade
$$\rho_t(\vec{x}) = \frac{1}{(4\pi t D)^{\frac{3}{2}}} \exp\left[-\frac{\vec{x}^2}{4tD}\right] \ .$$

Exercício: Verifique que a função ρ_t tem as propriedades características de uma densidade de probabilidade.

A.2.3 Distribuições mistas

Espaços amostrais contínuos podem também acomodar densidades de probabilidade concentradas em pontos, com isso sendo possível tratar espaços amostrais discretos através de densidades de probabilidade. Por exemplo, consideremos o espaço discreto $S_d = \{a_1, a_2, ..., a_n \mid a_i \in \mathbb{R}\}$, com probabilidades $\mathbf{P}(a_i) = p_i \geq 0$, que satisfazem

$$\sum_{i=1}^{n} p_i = 1 \ .$$

Então, a função generalizada definida no espaço amostral $S = \mathbb{R}$

$$\rho\left(x\right) = \sum_{i=1}^{n} p_i \delta\left(x - a_i\right)$$

pode ser usada para definir probabilidades de eventos $E \subset \mathbb{R}$, através de

$$\begin{aligned} \mathbf{P}(E) &= \int_E dx \ \rho(x) \\ &= \sum_{i \in E} p_i \ , \end{aligned}$$

o que está de acordo com o que vimos anteriormente. Neste caso dizemos que a probabilidade está concentrada nos pontos $a_i, \ i = 1, \ldots, n$, pois $\mathbf{P}(E) = 0$ se o conjunto E não contiver nenhum dos pontos $a_1, \ldots, a_n$.

Exercício: Verifique que a função generalizada ρ define uma densidade de probabilidade e que $\mathbf{P}(E)$ tem as propriedades de probabilidade.

Em espaços amostrais contínuos podemos acomodar superposições de densidades contínuas com densidades concentradas em um conjunto contável de pontos. Vejamos isto através de alguns exemplos.

Exemplos:

- Considere $S = \mathbb{R}$ e

$$\rho\left(x\right) = \alpha\rho_1\left(x\right) + \beta\rho_2\left(x\right)$$

onde $\alpha \geq 0, \beta \geq 0$, $\alpha + \beta = 1$ com ρ_1 e ρ_2 sendo densidades de probabilidade. Então, ρ também é uma densidade de probabilidade; verifique! Se ρ_1 for por exemplo a densidade

$$\rho_1 = \sum_{i=1}^{n} p_i \delta \left(x - a_i\right) \ ,$$

como acima e ρ_2 for uma distribuição contínua, como por exemplo a gaussiana, de Cauchy ou qualquer outra dada por uma função positiva integrável, tem-se então uma densidade mista

$$\mathbf{P}(E) = \alpha \sum_{i \in E} p_i + \beta \int_E \rho_2 \left(x\right) dx \ .$$

- No exemplo do movimento Browniano a densidade ρ_t que é contínua para $t > 0$, entretanto, no limite $t \to 0$, produz a densidade discreta
$$\rho_0 \left(\vec{x}\right) = \lim_{t \to 0} \rho_t = \delta \left(\vec{x}\right) \ ,$$
consistente com a condição de que a partícula estava na origem em $t = 0$; justifique esta afirmativa! Sugestão: verifique antes que o mesmo efeito ocorre para as distribuições gaussiana e de Cauchy nos limites $\sigma \to 0$ e $\gamma \to 0$, respectivamente.

A.2.4 Eventos Independentes

Em um fenômeno aleatório dois eventos são ditos independentes se ocorrência ou não ocorrência de um deles não afetar a ocorrência do outro. Tipicamente tem-se em mente exemplos tais como no lançamento de 2 dados, onde um dos eventos refere-se somente ao resultado do primeiro dado e outro somente ao resultado do segundo dado. Do ponto de vista formal temos

Definição: Em um espaço de probabilidade com espaço amostral S e com probabilidade $\mathbf{P}$, dois eventos E e F são **independentes** se $\mathbf{P}(E \cap F) = \mathbf{P}(E) \times \mathbf{P}(F)$.

Exemplos:

- No lançamento de 3 moedas os eventos E = cara na primeira $= \{x \in S_3 \ |x_1 = -1\}$ e F = coroa na segunda $= \{x \in S_3 \ |qx_2 = +1\}$ = são independentes. Por outro lado, os eventos $G = \{x \in S_3 \mid x_1 + x_2 = 0\}$ e $H = \{x \in S_3 \mid x_1 - x_2 = 0\}$ não independentes. Verifique estas afirmativas.

- Exemplos típicos de eventos independentes são produzidos da seguinte forma:

  1. Considere dois espaços amostrais finitos: $S = \{x_1, ..., x_n\}$ e $R = \{y_1, ..., y_m\}$ com probabilidades definidas a partir dos números $p_i = \mathbf{P}\{x_i\}$, $i = 1, ..., n$ e $q_j = \mathbf{P}\{y_j\}$, $j = 1, ..., m$ e formemos o espaço amostral produto $T = S \times R = \{(x_i, y_j), i = 1, ..., n;\ j = 1, ..., m\}$ com probabilidade definida através de

  $$p_{ij} \ = \ \mathbf{P}\{(x_i, y_j)\} \ = \ p_i q_j \ .$$

  Nessas condições os eventos $E = A \times R$ e $F = S \times B$ são independentes qualquer que sejam $A \subset S$ e $B \subset R$.

  2. Considere dois espaços amostrais S_1 e S_2 contínuos, por exemplo $S_1 = S_2 = \mathbb{R}$ e duas densidades ρ_1 e ρ_2 em $\mathbb{R}$. A função

  $$\rho\,(x_1, x_2) = \rho_1\,(x_1)\,\rho_1\,(x_2)$$

  define então uma densidade de probabilidade em $S_1 \times S_2 = \mathbb{R}^2$. Nessas condições os eventos $E = A \times S_2$ e $F = S_1 \times B$ são independentes.

Exercício. Verifique a independência dos eventos E e F nos dois exemplos acima.

- Menos trivial do que os casos acima descritos é o seguinte exemplo. Considere em $S = \mathbb{R}^2$ a densidade gaussiana

$$\rho\,(x_1, x_2) = \frac{1}{2\pi} \exp - \left[\frac{x_1^2 + x_2^2}{2}\right]$$

e mostre que os eventos $E = \{x = (x_1, x_2) \mid (x_1 + x_2) \in A\}$ e $E = \{x = (x_1, x_2) \mid (x_1 - x_2) \in B\}$, onde $A, B \subset \mathbb{R}$ são conjuntos arbitrários, são independentes.

A.3 Variáveis Aleatórias

Em fenômenos aleatórios é natural associar a cada resultado possível de um experimento um número real. Isto é feito definindo-se uma variável aleatória f, a qual é uma função no espaço amostral S.

$$
\begin{gathered}
f : S \to \mathbb{R} \, , \\
x \in S \to f(x) \in \mathbb{R} \, .
\end{gathered}
$$

Exemplos:

- No caso do lançamento de um dado, *i.e.* no espaço amostral $S_1 = \{1, ..., 6\}$ as funções

$$
\begin{cases}
f_1(x) = x \, , \\
f_2(x) = x^2 \, ,
\end{cases}
$$

  que descrevem respectivamente o resultado de um experimento e o quadrado do resultado são variáveis aleatórias.

- No lançamento de 3 moedas, *i.e.* para o espaço amostral $S = (S_0)^3$ as funções

$$
\begin{cases}
f_1(x) = x_1 \, , \quad f_2(x) = x_2 \, , \quad f_3(x) = x_3 \, , \\
g(x) = \frac{1}{3}(x_1 + x_2 + x_3) \, , \\
h(x) = \frac{1}{3}(x_1^2 + x_2^2 + x_3^2) \, ,
\end{cases}
$$

  com $x = (x_1, x_2, x_3) \in S$, definem variáveis aleatórias, onde f_i corresponde ao resultado do i-ésimo lançamento, g à média dos resultados dos 3 lançamentos e h à média dos quadrados dos resultados dos 3 lançamentos.

- No lançamento de N moedas, $S = (S_0)^N$ podemos definir de forma análoga as variáveis aleatórias

$$\begin{cases} f_n(x) = x_n, n = 1, ..., N \\ g(x) = \frac{1}{N}(x_1 + ... + x_N) = \frac{1}{N}\sum_{n=1}^{N} x_n \\ h(x) = \frac{1}{N}(x_1^2 + ... + x_N^2) = \frac{1}{N}\sum_{n=1}^{N} x_n^2 \end{cases}$$

com as mesmas interpretações do caso $N = 3$.

- Um exemplo de variável aleatória em um espaço de probabilidade é a função indicatriz χ_E de um evento E, definida por

$$\chi_E(x) = \begin{cases} 1 \text{ se } x \in E \\ 0 \text{ se } x \notin E \end{cases} .$$

- No caso do movimento Browniano em $\mathbb{R}^3$ podemos definir por exemplo as seguintes variáveis aleatórias:

$$\begin{array}{l} f_i(\vec{x}) = x_i \text{ para } i = 1, 2, 3 \ , \\ r(\vec{x}) = \sqrt{x_1^2 + x_2^2 + x_3^2} \ , \\ g(\vec{x}) = r(\vec{x})^2 = x_1^2 + x_2^2 + x_3^2 \ , \end{array}$$

que representam respectivamente as coordenadas da partícula, a distância e o quadrado da distância da partícula à origem.

A.3.1 Distribuição de probabilidades

Consideremos um espaço amostral finito, *i.e.* $S = \{x_i, \ i = 1, 2, \ldots, N\}$ com probabilidades definidas a partir dos números $p_i = \mathbf{P}\{x_i\}, \ i = 1, \ldots, N$ e uma variável aleatória f com valores no conjunto imagem $f(S) \equiv \{a_1, a_2, ..., a_K\}$, onde $K \leq N$ pois podemos ter $f(x_i) = f(x_j)$ embora $x_i \neq x_j$. Podemos então definir probabilidades no conjunto imagem $\{a_1, a_2, ..., a_K\}$ através de

$$q_j = \mathbf{P}\{f^{-1}(a_j)\} \ , \ j = 1, ..., K$$

i.e. os números q_j representam as probabilidades dos eventos $E_j = \{x \in S \mid f(x) = a_j\}$, ou seja, a probabilidade da função f assumir o valor a_j. Por isso escrevemos também

$$q_j = \mathbf{P}\{f = a_j\} \ .$$

A coleção $\{q_1, \ldots, q_K\}$ é a *distribuição de probabilidades* da variável aleatória f. Note que

$$\begin{array}{l} q_j \geq 0 \text{ e que} \\ \sum_{j=1} K q_j = 1 \ , \end{array}$$

logo, definindo probabilidades no conjunto imagem $f(S) \equiv \{a_1, \ldots, a_K\}$.

Exemplos:

- No espaço amostral S_1, correspondente a um dado, a variável aleatória f_2 definida acima tem um conjunto imagem

$$f_2(S_1) = \left\{a_1 = \frac{1}{4}, a_2 = \frac{9}{4}, a_3 = \frac{25}{4}\right\}$$

possuindo a seguinte distribuição de probabilidade

$$q_1 = q_2 = q_3 = \frac{1}{3} \ .$$

- Para o lançamento de três moedas $S = (S_0)^3$, e as variáveis aleatórias f_1 e g definidas acima temos $f_1(S) = \{a_1 = -1, a_2 = +1\}$, e as probabilidades

$$q_1 = \mathbf{P}\{f = a_1\} = q_2 = \mathbf{P}\{f = a_2\} = \frac{1}{2},$$

e $g(S) = \{b_1 = -1, b_2 = -\frac{1}{3}, b_3 = +\frac{1}{3}, b_4 = +1\}$, com

$$\begin{array}{l} q_1 = \mathbf{P}\{f = a_1\} = \frac{1}{8} \ , \\ q_2 = \mathbf{P}\{f = b_2\} = \frac{3}{8} \ , \\ q_3 = \mathbf{P}\{f = b_3\} = \frac{3}{8} \ , \\ q_4 = \mathbf{P}\{f = b_4\} = \frac{1}{8} \ . \end{array}$$

Caso contínuo

Consideremos agora um espaço amostral S contínuo no qual temos uma probabilidade $\mathbf{P}$ definida. Dada uma variável aleatória f em S, *i.e.* uma função com imagem A contida em $\mathbb{R}$, podemos de maneira análoga ao que foi feito no caso de espaços amostrais finitos definir uma distribuição de probabilidade $\mathbf{P}_f$ em $\mathbb{R}$ através da fórmula

$$\mathbf{P}_f\,(A) = \mathbf{P}(f^{-1}\,(A)) = \mathbf{P}\,(\{x \in S \mid f\,(x) \in A\})\ ,$$

i.e. $\mathbf{P}_f\,(A)$ é a probabilidade da função f tomar valores em A, e por isso escrevemos também

$$\mathbf{P}_f\,(A) = \mathbf{P}(f \in A)\ .$$

Exemplo: Consideremos em R as variáveis aleatórias $f_1\,(x) = x$ e $f_2\,(x) = x^2$ e consideremos em $\mathbb{R}$ a probabilidade $\mathbf{P}$. Podemos então calcular

$$\mathbf{P}_{f_1}\,(A) = \mathbf{P}(A)\ ,$$

e

$$\mathbf{P}_{f_2}\,(A) = \mathbf{P}(x^2 \in A)\ .$$

Se por exemplo a probabilidade $\mathbf{P}$ for definida por uma densidade ρ e $A = [0, a]$ onde $a > 0$ então

$$\mathbf{P}_{f_1}\,([0, a]) = \int_0^a \rho\,(x)\,dx\ ,$$

e

$$\mathbf{P}_{f_2}\,([0, a]) = \int_0^{\sqrt{a}} \rho\,(x)\,dx\ .$$

De maneira geral a distribuição de probabilidade de uma variável aleatória f pode ser dada por uma densidade ρ_f definida por

$$\mathbf{P}_f\,(A) = \mathbf{P}(f \in A) = \int_A \rho_f\,(x)\,dx\ .$$

A função ρ_f é a densidade de probabilidade da variável aleatória f, sendo $\rho_f\,(x)\,dx$ a probabilidade de a variável f tomar valores entre x e $x + dx$. A densidade ρ_f tem as propriedades:

$$\begin{aligned} \rho_f(x) &\geq 0 , \\ \text{e} \qquad \int \rho_f(x)\, dx &= 1 . \end{aligned}$$

Exemplo: Nos dois exemplos acima podemos calcular ρ_{f_1} e ρ_{f_2}, resultando que

$$\rho_{f_1}(x) = \rho(x) ,$$

e

$$\rho_{f_2}(x) = \begin{cases} \frac{1}{2\sqrt{x}}\rho(x) \text{ se } x \geq 0 \\ 0 \text{ se } x < 0 \end{cases} .$$

Definição: Dizemos que uma variável aleatória f é *discreta* se o conjunto $f(S)$ de valores que ela pode tomar for um *conjunto enumerável.* Quando os valores assumidos por uma variável aleatória f formam um conjunto *não enumerável* dizemos que f é uma *variável aleatória contínua.*

No caso de uma variável aleatória f discreta temos a opção de considerar o conjunto imagem $f(S) = \{a_i,\ i = 1, 2, \ldots\}$ como feito no caso de um espaço finito e definir uma distribuição de probabilidade $q_i = \mathbf{P}(f = a_i)$, ou então considerar o conjunto imagem $f(S)$ como um subconjunto de $\mathbb{R}$ e definir a sua densidade de probabilidade utilizando

$$\rho_f(x) = \sum_i q_i \delta(x - a_i) .$$

Exemplo: Considere a variável aleatória

$$f : S \to \mathbb{R} ,$$

que assume somente os valores inteiros não negativos $k = 0, 1, \ldots$. f descreve, por exemplo, o número de vezes que ocorre determinado fato. Caso façamos a escolha

$$q_k = \mathbf{P}(f = k) = \frac{e^{-\lambda}\lambda^k}{k!}$$

onde $\lambda > 0$, dizemos que a variável aleatória f tem uma distribuição de Poisson com densidade λ.[3] Esta variável aleatória também pode ser caracterizada por

$$\rho_f(x) = \sum_{k=1}^{\infty} \frac{e^{-\lambda}\lambda^k}{k!}\, \delta(x-k)\ .$$

Consideremos simultaneamente duas variáveis aleatórias f_1 e f_2 num espaço amostral, ou seja, consideremos a probabilidade de *simultaneamente* f_1 e f_2 assumirem valores em conjuntos A_1e A_2 respectivamente. Essas probabilidades determinam probabilidades $\mathbf{P}_{f_1,f_2}$ em $\mathbb{R}^2$ através da seguinte fórmula.

$$\mathbf{P}_{f_1,f_2}(A_1 \times A_2) = \mathbf{P}\left(f_1^{-1}(A_1) \cap f_2^{-1}(A_2)\right) = \mathbf{P}\left(f_1 \in (A_1)\,, f_2 \in (A_2)\right)$$

Analogamente obtemos a distribuição $(\rho_{f_1,f_2}(x_1,x_2))$ de probabilidade conjunta definida pela relação

$$\mathbf{P}_{f_1,f_2}(B) = \int\int_B \rho_{f_1,f_2}(x_1,x_2)\, dx_1 dx_2\ .$$

A função ρ_{f_1,f_2} tem as propriedades de uma densidade de probabilidade em $\mathbb{R}^2$, e $\rho_{f_1,f_2}(x_1,x_2)\, dx_1 dx_2$ é a probabilidade de f_1 estar entre x_1 e $x_1 + dx_1$ **e** de f_2 estar entre x_2 e $x_2 + dx_2$.

Definição: Em um espaço de probabilidade com espaço amostral S e probabilidade $\mathbf{P}$, duas variáveis aleatórias f_1 e f_2 são ditas **independentes** se os eventos $E_1 = f_1^{-1}(A_1)$ e $E_2 = f_2^{-1}(A_2)$ são independentes, *i.e.* $\mathbf{P}(E_1 \cap E_2) = \mathbf{P}(E_1) \cdot \mathbf{P}(E_2)$, qualquer que seja a escolha dos conjuntos A_1e $A_2 \subset \mathbb{R}$. Em outras palavras se

$$\mathbf{P}(f_1 \in (A_1)\,, f_2 \in (A_2)) = \mathbf{P}(f_1 \in A_1) \cdot \mathbf{P}(f_2 \in (A_2))\ ,$$

ou ainda, todos os eventos que se referem somente à variável aleatória f_1 são independentes de todos os eventos que referem somente à variável

[3]Verifique que a coleção de números q_k realmente definem uma distribuição de probabilidades.

aleatória f_2. Se f_1 e f_2 são independentes então a distribuição conjunta ρ_{f_1,f_2} satisfaz a propriedade de fatorização

$$\rho_{f_1,f_2}(x_1,x_2) = \rho_{f_1}(x_1)\,\rho_{f_2}(x_2)\ .$$

Exemplo: Consideremos o espaço amostral $S = \mathbb{R}^2$ com a distribuição gaussiana de probabilidade

$$\rho(x_1,x_2) = \frac{1}{2\pi}\exp\left(-\frac{x_1^2+x_2^2}{2}\right)\ .$$

As variáveis aleatórias $f_1 = x_1$ e $f_2 = x_2$ são independentes. Mais ainda, as variáveis aleatórias $f_3 = x_1 + x_2$ e $f_4 = x_1 - x_2$ também o são. Verifique estas afirmativas!

A.4 Momentos

Definiremos alguns parâmetros, chamados momentos, que servem para caracterizar uma distribuição de probabilidade discreta ou contínua de uma variável aleatória.

A.4.1 Média

A média, ou esperança matemática, denotada por $\langle f \rangle$, da variável aleatória f, definida em um espaço amostral S com probabilidade $\mathbf{P}$, é definida da seguinte maneira:

a) Se S for um conjunto enumerável, finito ou infinito, $S = \{x_1, x_2, \ldots\}$ com $\mathbf{P}(x_i) = p_i$, então

$$\langle f \rangle = \sum_{i\geq 1} f(x_i)p_i\ .$$

De um ponto de vista heurístico, é razoável esperar que este parâmetro estime a média dos valores obtidos para a variável aleatória em uma seqüência grande de “sorteios” repetidos. Esse tipo de afirmativa será mais esclarecida quando discutirmos a “lei dos grandes números ”.

Uma maneira alternativa para calcular é somando-se sobre os possíveis valores f e não sobre os pontos no espaço amostral. Assim suponha que variável aleatória f possa assumir os valores no conjunto (finito ou infinito) $\{y_k,\ k = 1, 2, \dots\}$ com probabilidades $q_k = \mathbf{P}(f = y_k)$. A expressão acima pode então ser escrita da seguinte forma:

$$\langle f \rangle = \sum_{i \geq 1} f(x_i) p_i = \sum_{k \geq 1} \left[y_k \sum_{i | f(x_i) = y_k} p_i \right] .$$

Ora

$$\sum_{i | f(x_i) = y_k} p_i = q_k \ ,$$

e portanto

$$\langle f \rangle = \sum_{i \geq 1} f(x_i) p_i = \sum_{k \geq 1} y_k q_k \ .$$

Exemplos:

- Consideremos o lançamento de um dado, *i.e.* $S_1 = \{1, ..., 6\}$, $p_i = \frac{1}{6}$ e as variáveis aleatórias $f_1 = i$ e $f_2 = (i - \langle f_1 \rangle)^2$. Podemos então calcular

  $$\langle f_1 \rangle = \sum_{i=1}^{6} i \frac{1}{6} = \frac{1}{6} (1 + 2 + 3 + 4 + 5 + 6) = \frac{7}{2} \ ,$$

  e

  $$\langle f_2 \rangle = \sum_{i=1}^{6} \frac{1}{6} \left(i - \frac{7}{2} \right)^2 = \frac{1}{3} \left(\frac{1}{2} \right)^2 + \frac{1}{3} \left(\frac{3}{2} \right)^2 + \frac{1}{3} \left(\frac{5}{2} \right)^2 \ .$$

  No cálculo de $\langle f_2 \rangle$ as duas fórmulas correspondem às duas maneiras possíveis de calcular o valor médio, como descrito acima.

- No lançamento de 3 moedas, *i.e.* $S = (S_0)^3$ com $p_x = \left(\frac{1}{2}\right)^3$ para todo $x = (x_1, x_2, x_3) \in S$, podemos calcular o valor médio das funções f_i, g e h definidas anteriormente:

$$\langle f_i \rangle = \sum_{x \in S} \left(\frac{1}{2}\right)^3 x_i = \left(\frac{1}{2}\right)(+1) + \left(\frac{1}{2}\right)(-1) = 0 \ ,$$

$$\langle g \rangle = 0 \ ,$$

$$\langle h \rangle = 1 \ .$$

b) Se a variável aleatória estiver definida num espaço amostral S contínuo, como por exemplo $S = \mathbb{R}^n$ ($x = (x_1, \ldots, x_n)$), e se a probabilidade for definida através de uma densidade de probabilidade $\rho(x)$, então o valor médio da variável aleatória $f : S \to \mathbb{R}$ é dado por

$$\langle f \rangle = \int_{S=\mathbb{R}^n} f(x) \, \rho(x) \, d^n x \ .$$

A exemplo do que foi feito no caso de um espaço discreto, a média da variável aleatória f pode também ser calculada através da distribuição de probabilidades dos valores da função f, integrando-se sobre o conjunto (H) dos valores possíveis de f ponderado adequadamente pela densidade de probabilidade $\rho_f(y)$:

$$\langle f \rangle = \int_{S=R^n} f(x) \, \rho(x) \ d^n x = \int_H y \rho_f(y) \ dy \ .$$

Essa fórmula pode ser trivialmente verificada para funções do tipo

$$f = \sum_{i=1}^{N} c_i \chi_{E_i} \ ,$$

i.e., combinações lineares de funções indicatrizes de eventos E_i, $i = 1, \ldots, N$. O caso geral é obtido passando-se ao limite $N \to \infty$.

Exemplos:

- Se uma variável aleatória f tem uma distribuição uniforme concentrada no intervalo $[0, 1]$, *i.e.*

$$\rho_f(x) = \begin{cases} 1 \text{ se } x \in [0,1] \\ 0 \text{ se } x \notin [0,1] \end{cases}$$

então
$$\langle f\rangle = \int_0^1 dx\, x\, \rho_f(x) = \int_0^1 dx\, x = \frac{1}{2} \; .$$

- Se uma variável aleatória tem uma distribuição gaussiana, $\rho_f(x) = \frac{1}{\sqrt{2\pi}\,\sigma}\, e^{-\frac{(x-x_0)^2}{2\sigma^2}}$, é fácil ver que
$$\langle f\rangle = \int_{-\infty}^{+\infty} dx\, x\, \rho_f(x) = x_0 \; .$$
De maneira geral, se $\rho_f(x)$ for uma função simétrica por reflexão em torno do ponto x_0 , ou seja,
$$\rho_f(x) = \rho_f(-x+2x_0)$$
então $\langle f\rangle = x_0$. Justifique!

- Se f exibir uma distribuição exponencial
$$\rho_f(x) = \begin{cases} \lambda\, e^{-\lambda x} \text{ , se } x \geq 0 \\ 0 \text{ , se } x < 0 \end{cases}$$
onde $\lambda > 0$, então
$$\langle f\rangle = \frac{1}{\lambda} \; .$$

Propriedades da média[4]

Se k é uma constante e f e g são variáveis aleatórias então temos que

1. $\langle kf\rangle = k\langle f\rangle$;
2. $\langle f+g\rangle = \langle f\rangle + \langle g\rangle$;
3. Se f e g são variáveis aleatórias *independentes* então
$$\langle fg\rangle = \langle f\rangle\langle g\rangle \; ,$$

[4] **Exercício:** Demonstre estas propriedades.

ou seja, a média do produto de variáveis aleatórias independentes é igual ao produto das médias. Verifiquemos essa propriedade no caso em que f e g são variáveis discretas, onde f admite valores $\{x_i,\ i=1,...\}$ com probabilidades $\{p_i,\ i=1,...\}$ e g admite valores $\{y_i,\ i=1,...\}$ com probabilidades $\{q_j,\ j=1,...\}$, respectivamente. Assim

$$\langle fg\rangle = \sum_{i\geq 1}\sum_{j\geq 1}(x_i y_j)\,\mathbf{P}(f=x_i, g=y_j)\ ,$$

como f e g são variáveis aleatórias *independentes* temos que

$$\mathbf{P}(f=x_i, g=y_j) = \mathbf{P}(f=x_i)\mathbf{P}(g=y_j) = q_i p_j\ .$$

Portanto,

$$\langle fg\rangle = \sum_{i\geq 1}\sum_{j\geq 1}(x_i q_i)\,(y_j p_j) = \left(\sum_{i\geq 1} x_i q_i\right)\left(\sum_{j\geq 1} y_j p_j\right) = \langle f\rangle\langle g\rangle\ .$$

Para o caso de variáveis com distribuições contínuas use o fato de que se f e g são variáveis aleatórias *independentes*, então a função de distribuição conjunta $\rho_{f,g}(x,y)$ é dada através de $\rho_{f,g}(x,y) = \rho_f(x)\ \rho_g(y)$.

A.4.2 Moda

A *moda* é definida como sendo o ponto de maior probabilidade, no caso discreto, ou de maior densidade de probabilidade, no caso contínuo.

Exemplos:

- Para a distribuição Gaussiana $\rho_f(x) = \frac{1}{\sqrt{2\pi}\sigma}\ e^{-\frac{(x-x_0)^2}{2\sigma^2}}$ a moda assume o valor x_0.

- Para a distribuição exponencial $f(x) = \lambda\ e^{-\lambda x}$ a moda vale zero.

Observação: Note a partir do segundo exemplo acima que em geral a moda não é igual à média.

A.4.3 Variância

Um bom indicador de como uma variável aleatória f "flutua" em torno de sua média $\langle f\rangle$ é o seu *desvio padrão* σ_f, cujo quadrado σ_f^2 é chamado de *variância* de f, a qual é definida através de

$$\sigma_f^2 = \left\langle (f - \langle f\rangle)^2 \right\rangle \ .$$

Mais explicitamente, para variáveis contínuas sua expressão é dada por

$$\sigma_f^2 = \int dx\, \rho_f(x)\, (x - \langle f\rangle)^2 \ ,$$

enquanto que para variáveis discretas ela toma a forma

$$\sigma_f^2 = \sum_{i\geq 1} (x_i - \langle f\rangle)^2\, p_i \ .$$

Podemos ainda utilizar as propriedades da média ($\langle f\rangle$) para reescrever σ_f^2, obtendo que[5]

$$\sigma_f^2 = \left\langle f^2 - 2f\langle f\rangle + \langle f\rangle^2 \right\rangle = \left\langle f^2\right\rangle - \langle f\rangle^2 \ .$$

Exemplos:

- No caso do lançamento de um dado e da variável aleatória $f = i$ temos que

$$\begin{aligned}
\sigma_f^2 &= \sum_{i=1}^{6} \frac{1}{6}\,(i - \langle x\rangle)^2 \\
&= \sum_{i=1}^{6} \frac{1}{6} i^2 - \left(\frac{1}{6}\sum_{i=1}^{6} i\right)^2 \\
&= \frac{91}{6} - \left(\frac{21}{6}\right)^2 = \frac{35}{12} \ .
\end{aligned}$$

[5] **Exercício:** Mostre esta relação.

- Para uma variável contínua com distribuição uniforme entre $(0, 1)$, sua variância é dada por

$$\sigma_f^2 = \int_0^1 dx \left(x - \frac{1}{2}\right)^2 = \frac{1}{3} - \frac{1}{4} = \frac{1}{12}\ .$$

- Para uma variável com distribuição exponencial

$$\sigma_f^2 = \int_0^\infty dx\ \left(x - \frac{1}{\lambda}\right)^2 \lambda\ e^{-\lambda x} = \frac{1}{\lambda^2}\ .$$

- No caso de uma variável contínua com distribuição Gaussiana $\rho_f(x) = \frac{1}{\sqrt{2\pi}\ \sigma}\ e^{-\frac{(x-x_0)^2}{2\sigma^2}}$, a variância é dada por

$$\sigma_f^2 = \int_{-\infty}^{+\infty} dx\ \frac{(x-x_0)^2}{\sqrt{2\pi}\ \sigma}\ e^{-\frac{(x-x_0)^2}{2\sigma^2}} = \sigma\ .$$

  É interessante notar que

$$\mathbf{P}\left(-\sigma \le x \le \sigma\right) = \int_{-\sigma}^{\sigma} dx\ \frac{1}{\sqrt{2\pi}\sigma}\ e^{-\frac{x^2}{2\sigma^2}} = 0.6826\ .$$

  De maneira mais geral, podemos estimar para $a > 0$ a probabilidade $\mathbf{P}(x > a)$ através de

$$\begin{aligned}\mathbf{P}(x > a) &= \frac{1}{\sqrt{2\pi}\ \sigma}\int_{x>a} dx\ e^{-\frac{x^2}{2\sigma^2}} \le \frac{1}{\sqrt{2\pi}\ \sigma}\int_{x>a} dx\ \frac{x}{a} e^{-\frac{x^2}{2\sigma^2}} \\ &= \frac{\sigma}{\sqrt{2\pi}\ a} e^{-\frac{a^2}{2\sigma^2}}\ .\end{aligned}$$

Propriedades da variância

Partindo da definição da variância podemos demonstrar as seguintes propriedades. Se k é uma constante e f e g são variáveis aleatórias *independentes*, então

1. $\sigma_k^2 = 0$;

2. $\sigma^2_{f\pm g} = \sigma^2_f + \sigma^2_g$;

3. $\sigma^2_{kf} = k^2\sigma^2_f$;

4. $\sigma^2_{(f\pm k)} = \sigma^2_f$.

Uma conseqüência notável das propriedades da variância é o fato de que se considerarmos uma coleção $\{f_{1,}, f_2, \dots\}$ variáveis aleatórias independentes e igualmente distribuídas, em particular com médias iguais a $\langle f_i \rangle = m$ e variâncias iguais a $\sigma^2_{f_i} = \sigma^2$, então a variância da soma

$$S_N = \sum_{i=1}^{N} f_i \ ,$$

é dada por

$$\sigma^2_{S_N} = \sum_{i=1}^{N} \sigma^2_{f_i} = N\sigma^2 \ ,$$

ou seja, o desvio padrão $\sigma_{S_N} = \sqrt{N}\sigma$. Isto mostra que embora a média

$$\langle S_N \rangle = \sum_{i=1}^{N} \langle f_i \rangle = Nm \ ,$$

cresça proporcionalmente a N, as flutuações como medidas pelo desvio padrão σ_{S_N} crescem apenas proporcionalmente a $\sqrt{N}$!

Momentos de uma distribuição

A informação sobre a média e a variância de uma variável aleatória é, em geral, de pouca valia como ingrediente para caracterizar completamente a sua distribuição de probabilidade, pois há infinidades de distribuições de probabilidade com mesmas médias e variâncias.

A média $\langle f \rangle$ e a média do quadrado $\langle f^2 \rangle$ de uma variável aleatória f são apenas casos particulares dos chamados momentos de f: para $n = 1, 2, \dots$, o n-ésimo momento de f é o valor médio de f^n, *i.e.* $\langle f^n \rangle$. O conhecimento de todos os momentos $\langle f^n \rangle = a_n$, $n = 1, 2, \dots$

permite calcular o valor médio de qualquer função $F(f)$ que admita uma expansão convergente série de potências. De fato, se

$$F(f) = \sum \alpha_n f^n ,$$

então

$$\langle F(f) \rangle = \sum \alpha_n \langle f^n \rangle = \sum \alpha_n a_n .$$

Exemplo: Considere uma variável aleatória f com uma distribuição uniforme no intervalo $[0, 1]$. Neste caso

$$\langle f^n \rangle = \int_0^1 x^n dx = \frac{1}{n+1} .$$

Mais ainda, se por exemplo $F(f) = e^f$ então

$$\langle F(f) \rangle = \sum_{n=0}^{\infty} \frac{1}{n!} \frac{1}{n+1} = \int_0^1 e^x dx = e - 1 .$$

O conhecimento de *todos* os momentos de uma variável aleatória possibilita uma completa reconstrução desta, *i.e.* podemos não só obter as probabilidades, mas também os valores que a variável aleatória assume! Para entendermos isso, consideremos o caso de uma variável aleatória que assume um conjunto finito de valores distintos $\{x_k,\ k = 1, \ldots, N\}$ com probabilidades p_k. Podemos determinar as respectivas probabilidades $p_k = \mathbf{P}(f = x_k)$ e os valores x_k a partir do conhecimento dos momentos $\langle f^n \rangle = a_n$ uma vez que as igualdades

$$a_n = \sum_{k=1}^{N} x_k^n p_k$$

podem ser consideradas como um sistema de equações para determinar os valores de p_k e x_k, para $k = 1, \ldots, N$.

A.5 Função Característica de uma Variável Aleatória

Podemos determinar todos os momentos $\langle f^n \rangle$ e a própria distribuição de probabilidade ρ_f de uma variável aleatória f a partir da função característica da variável aleatória f definida por

$$C_f(t) \equiv \left\langle e^{itf} \right\rangle \ ,$$

onde t é uma variável real. Caso f seja uma variável aleatória contínua temos que

$$C_f(t) = \int e^{itx} \rho_f(x)\, dx \ ,$$

enquanto que para variáveis discretas

$$C_f(t) = \sum_k p_k \, e^{itx_k} \ .$$

A função C_f também é chamada de função geratriz dos momentos da variável aleatória já que podemos calcular o n-ésimo momento de f através da derivada n-ésima de $C_f(t)$ ponto $t = 0$

$$\langle f^n \rangle = \frac{1}{i^n} \left. \frac{\partial^n C_f}{\partial t^n} \right|_{t=0} \ .$$

Exemplo: Considere a distribuição gaussiana

$$\rho(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-x_0)^2}{2\sigma^2}} \ .$$

Sua função característica $C_f(t)$ é dada por

$$C_f(t) = e^{itx_0} e^{-\frac{1}{2}\sigma^2 t^2} \ .$$

Teorema Importante: A função característica $C_f(t)$ determina completamente a distribuição de probabilidade ρ_f da variável aleatória através da expressão

$$\rho_f(x) = \int \frac{e^{-itx}}{2\pi} \ C_f(t)\, dt \ ,$$

a qual é a transformada de Fourier de $C_f(t)$.

A.5.1 Propriedades da função característica

A função característica $C_f(t)$ possui as seguintes propriedades:

1. Se k é uma constante e f é uma variável aleatória, então a função característica da variável aleatória kf é dada por

$$C_{kf}(t) = C_f(kt) \ .$$

2. Se f e g, são variáveis aleatórias *independentes* então

$$C_{f+g}(t) = C_f(t)\, C_g(t) \ .$$

Em particular, se $\{f_1, \ldots, f_N\}$ são variáveis aleatórias independentes e igualmente distribuídas, *i.e.* $\rho_{f_1} = \cdots = \rho_{f_N}$, então

$$C_{f_1+\ldots+f_N}(t) = \left[C_{f_1}(t)\right]^N \ .$$

A.6 A Lei dos Grandes Números

Um fenômeno aleatório foi definido como sendo de natureza probabilística se ao serem efetuadas N ensaios com o mesmo conjunto $\mathcal{C}$ de condições, a freqüência relativa de ocorrência de um determinado evento E, n_E/N tende a estabilizar-se para grandes valores de N em um valor $\mathbf{P}(E)$, ou seja,

$$\frac{n_E}{N} \to \mathbf{P}(E) \ ,$$

e esse valor é identificado com a probabilidade do evento.

A teoria da probabilidade é capaz de adequadamente expressar essa propriedade. De fato, consideremos o espaço amostral S_0 que descreve um determinado fenômeno. Para sermos mais concretos podemos considerar, por exemplo, o lançamento de um dado $S_0 = \{-1, +1\}$. Se considerarmos N repetições independentes do lançamento, o espaço amostral relevante é dado por $S = (S_0)^N = \{x = (x_1, \ldots, x_N),\ x_i \in S_0\}$ com probabilidade $p_x = \frac{1}{2^N}$. Intuitivamente esperamos que a freqüência relativa de ocorrência de "caras" *i.e.* $x_i = -1$ estabilize-se em $\frac{1}{2}$, pois esse foi o valor adotado para definir $p_x = \frac{1}{2^N}$! Para investigar esse fato, consideremos a variável aleatória

$$f_N(x) = \frac{1}{N}\sum_{i=1}^{N}\chi_i(x) \ ,$$

onde

$$\chi_i(x) = \begin{cases} 1 \text{ se } x_i = -1 \\ 0 \text{ se } x_i = +1 \end{cases} .$$

A variável aleatória representa exatamente a freqüência relativa de ocorrência de "caras" em N tentativas. O que esperamos, de acordo com o que foi dito acima, é que $f_N(x) \to \frac{1}{2}$, em palavras, a variável aleatória deve convergir para uma constante! Note inicialmente que a média de f_N é dada por

$$\langle f_N \rangle = \frac{1}{2} \ ,$$

e que a sua variância é dada por[6]

$$\sigma^2_{f_N} = \frac{N}{N^2}\sigma^2_{\chi_i} = \frac{1}{2N} \ .$$

O fato relevante é que a variância $\sigma^2_{f_N} \to 0$ quando $N \to \infty$, o que significa que a variável f_N está convergindo para uma constante! Para fazer uma afirmativa ainda mais precisa vamos inicialmente deduzir uma importante, pelo número de aplicações práticas, desigualdade devida a Chebyschev. Se f é uma variável aleatória então, se $a > 0$:

$$\begin{aligned} \langle f^2 \rangle &= \int x^2 \rho_f(x)\, dx \geq \int_{x^2 > a^2} x^2 \rho_f(x)\, dx \\ &\geq a^2 \int_{x^2 > a^2} \rho_f(x)\, dx = a^2 \mathbf{P}(|f| \geq a) \ , \end{aligned}$$

ou seja,

$$\mathbf{P}(|f| \geq a) \leq \frac{\langle f^2 \rangle}{a^2} \ .$$

[6]Verifique essas afirmativas, notando que as variáveis χ_i, são mutuamente independentes.

Em particular para uma variável f de média zero,

$$\mathbf{P}(|f| \geq a) \leq \frac{\sigma_f^2}{a^2} \ .$$

Retornando então à nossa variável f_N , podemos aplicar a última desigualdade à variável $g_N \equiv f_N - \langle f_N \rangle$, que tem média zero e com variância $\sigma_{g_N}^2 = \sigma_{f_N}^2$, para obter

$$\mathbf{P}(|f_N - \langle f_N \rangle| \geq a) \leq \frac{\sigma_{f_N}^2}{a^2} = \frac{1}{2a^2 N} \to 0 \text{ quando } N \to \infty \ .$$

Esse limite mostra exatamente a consistência interna da teoria da probabilidade nos seus aspectos interpretativos: é a lei dos grandes números em ação.

Repetindo todos os passos da discussão acima podemos apresenta versão muito mais geral do limite acima obtido.

Teorema: Sejam χ_i $(i \geq 1)$ variáveis aleatórias independentes em um espaço amostral S com médias e variâncias iguais $\langle \chi_i \rangle = m$ e $\sigma_{\chi_i}^2 = \sigma^2$ e seja também $f_N = \frac{1}{N} \sum_{i=1}^{N} \chi_i$. Então,

$$\mathbf{P}(|f_N - m| \geq a) \leq \frac{\sigma_{f_N}^2}{a^2} = \frac{\sigma^2}{a^2 N} \to 0 \text{ quando } N \to \infty \ ,$$

o que significa que para grandes valores de N os desvios da média tornam-se cada vez mais improváveis.

A.7 Teorema Central do Limite ou Importância da Gaussiana

Da seção anterior concluímos que a variável aleatória

$$f_N = \frac{1}{N} \sum_{i=1}^{N} \chi_i \ ,$$

onde χ_i $(i \geq 1)$ são variáveis aleatórias independentes e igualmente distribuídas, em particular com médias e variâncias iguais $\langle \chi_i \rangle = m$ e

$\sigma^2_{\chi_i} = \sigma^2$, converge " em probabilidade" para o valor médio comum m, isto é

$$\mathbf{P}(|f_N - m| \geq a) \to 0 \text{ quando } N \to \infty \ .$$

Queremos agora obter uma estimativa mais precisa de como a variável f_N flutua em torno do seu valor médio, para N grande. Essa informação nos é fornecida pelo chamado Teorema Central do Limite, que diz que para N grande essas flutuações tem uma distribuição gaussiana com variância σ^2.

$$f_N - m = \frac{1}{N}\left(\sum_{i=1}^{N} \chi_i - Nm\right) \simeq \frac{1}{\sqrt{N}} G \ ,$$

ou equivalentemente,

$$g_N \equiv \sqrt{N}\,(f_N - m) = \frac{1}{\sqrt{N}}\left(\sum_{i=1}^{N} \chi_i - Nm\right) \to G \text{ quando } N \to \infty \ ,$$

onde G é uma variável aleatória com uma distribuição Gaussiana com média zero e variância igual à variância comum $\sigma^2 = \sigma^2_{\chi_i}$.

Uma forma matematicamente precisa de se enunciar e demonstrar esse resultado 'gravissimum" da teoria da probabilidade é através do cálculo da função característica $C_{g_N}(t)$ da variável g_N. Por simplicidade apenas, suponhamos que a média comum m seja nula, $m = 0$. Usando as propriedades das funções características podemos calcular:

$$C_{g_N}(t) = C_{\sqrt{N} f_N}(t) = \left[C_{\frac{\chi_1}{\sqrt{N}}}(t)\right]^N = \left[C(\frac{t}{\sqrt{N}})\right]^N \ ,$$

onde $C(t) \equiv C_{\chi_1}(t) = \cdots = C_{\chi_N}(t)$ é a função característica comum às variáveis aleatórias $\chi_i, i = 1, \ldots, N$. O lado direito pode ser então calculado.

$$C_{g_N}(t) = \left[C(\frac{t}{\sqrt{N}})\right]^N = \left\langle e^{i\frac{t}{\sqrt{N}}\chi_1}\right\rangle^N \ .$$

Para N grande, podemos expandir a exponencial

$$e^{i\frac{t}{\sqrt{N}}\chi_1} = 1 + i\frac{t}{\sqrt{N}}\chi_1 - \frac{t^2}{2N}\chi_1^2 + O\left(\frac{1}{N^2}\right) \ ,$$

e portanto

$$\left\langle e^{i\frac{t}{\sqrt{N}}\chi_1}\right\rangle^N = \left[1 - \frac{t^2}{2N}\sigma^2 + O\left(\frac{1}{N^2}\right)\right]^N \ ,$$

onde usamos o fato de que $\langle\chi_i\rangle = 0$ e $\sigma_{\chi_i}^2 = \sigma^2$. Tomando agora o limite $N \to \infty$, obtemos

$$\lim_{N\to\infty} C_{g_N}(t) = \lim_{N\to\infty}\left[1 - \frac{t^2\sigma^2}{2N} + O\left(\frac{1}{N^2}\right)\right]^N = e^{-\frac{t^2\sigma^2}{2}}; \ ,$$

em palavras: a função característica $C_{g_N}(t)$ converge para a função característica de uma gaussiana de média zero e variância igual à variância comum σ^2. Isto implica para a distribuição de probabilidade ρ_{g_N} a propriedade

$$\lim_{N\to\infty}\int_E \rho_{g_N}(x)\,dx = \int_E \frac{e^{-\frac{x^2}{2\sigma^2}}}{\sqrt{2\pi}\sigma}dx \ .$$