Experimento 1

Eventos Aleatórios

Introdução

Medidas físicas não são exatas. Devido às limitações dos aparelhos de medição ou dos processos adotados, o resultado de qualquer medida é diferente do "valor verdadeiro" da grandeza. Parte dos desvios que se observam resulta de fenômenos que intervêm no processo de medição e são incontroláveis, gerando interferências *aleatórias* (= ao acaso). A magnitude das flutuações de origem aleatória pode ser estimada, por exemplo, com a repetição da medição diversas vezes e a consequente obtenção de resultados diferentes. Normalmente, quando se realiza um experimento em condições controladas, é possível encontrar regras para a *probabilidade* de se obter um resultado dentro de certo intervalo de valores. O objetivo deste experimento é mostrar a existência de flutuações aleatórias em resultados de medições de uma mesma grandeza, flutuações essas que têm origem em condições *incontroláveis* por quem realiza a experiência e podem, por isso, ser interpretados dentro do quadro da teoria da Estatística, cujos conceitos básicos desenvolveremos ao longo do semestre, de uma maneira muito ligada ao trabalho experimental em um laboratório.

A distribuição binomial

A função de probabilidade binomial $P_{n,p}(i)$ aplica-se às situações em que a *variável aleatória* i é o número inteiro de *sucessos* em n tentativas independentes quando a probabilidade de sucesso p é constante e igual em cada tentativa.

Os conceitos tornam-se mais claros através do exemplo específico desta experiência: lançaremos repetidamente oito dados cúbicos, cada um com duas faces marcadas com um ponto e quatro faces sem nenhuma marca. Quando lançarmos os dados sem arrumá-los no copinho nem esparramá-los cuidadosamente – ou seja, se jogarmos honestamente – não se poderão controlar as faces que os dados exibirão para cima, o que fará com que a quantidade de dados que caem com a face marcada para cima seja um *número aleatório* entre 0 e 8. Ao definirmos *sucesso* como o evento: *o dado cai com a face marcada para cima*, então a probabilidade[1] de sucesso é $p = 2/6 = 1/3$. Os 8 dados lançados são interpretados como um número de tentativas $n = 8$, que são independentes, porque a face que um dado exibe para cima não influi na que outro dado mostra. Considerando os dados iguais, não importa, para efeitos de estatística, se os 8 dados são lançados simultaneamente ou em sequência, ou se um único dado é lançado 8 vezes[2]. Nesse caso, o número de sucessos i é uma variável aleatória, que pode assumir qualquer

---

[1]Probabilidade é um conceito muito abrangente e pode ter significados diferentes conforme a aplicação. Duas propriedades, porém, são essenciais para que uma grandeza possa ser chamada de probabilidade: ser definida não-negativa e a soma das probabilidades de ocorrência de todos os resultados possíveis igualar 1.

[2]Esta ideia é muito importante em física. Frequentemente, n sistemas físicos idênticos e independentes são descritos como n repetições de um mesmo processo. Por exemplo, tanto faz medirmos as velocidades de n átomos de um gás quanto medirmos a velocidade de um átomo n vezes, em instantes suficientemente separados no tempo para que tenha ocorrido ao menos uma colisão no intervalo.

[3]Uma maneira de entender este resultado é fazer uma tabela com todos os eventos possíveis com dois dados e perceber que apenas 1/9 deles corresponde ao evento "os dois com faces marcadas para cima".

valor inteiro entre 0 e o número de tentativas n = 8, com probabilidade $P_n(i)$ dada pela fórmula

$$P_n(i) = \binom{n}{i} p^i (1-p)^{(n-i)} \tag{1.1}$$

Vamos deduzir esta fórmula, que serve para qualquer n, nesse caso específico em que n = 8. Começaremos, então, com a análise de um evento muito particular: em certo lançamento dos 8 dados *em sequência*, isto é, dado a dado e não todos eles no copinho, ocorrem inicialmente 2 sucessos e depois 6 insucessos (isto é, seis faces vazias). A probabilidade de ocorrência deste evento composto pode ser calculada usando que a probabilidade de dois eventos independentes ocorrerem é o produto das probabilidades de cada evento isolado. A probabilidade do primeiro sucesso é 1/3, bem como a do segundo, portanto a sequência de dois sucessos seguidos tem probabilidade[3]:

$$1/3 \times 1/3 = (1/3)^2 = 1/9 = p^2$$

O terceiro e todos os demais dados saíram com as faces vazias. A probabilidade de ocorrência q de um insucesso é o complemento para 1 da probabilidade do sucesso, uma vez que esses dois eventos são os únicos possíveis e mutuamente exclusivos:

$$q = (1 - p) = (1 - 1/3) = 2/3$$

A probabilidade de ocorrência de 6 fracassos seguidos seria, portanto,

$$(1 - p)^6 = (2/3)^6$$

A probabilidade de ocorrência de 2 sucessos seguida de 6 fracassos é, então:

$$P' = p^2(1 - p)^6$$

Esta não é, contudo, a probabilidade de se obter i = 2, uma vez que existem outras sequências possíveis de eventos que têm no total 2 faces para cima. O problema de contar o número de sequências diferentes possíveis de 2 sucessos e 6 fracassos é resolvido pela análise combinatória e é igual a $\frac{8!}{[2!\cdot(8-2)!]}$, que é o número de permutações de 8 dados, corrigido pelo fato que permutações dos dados com faces marcadas para cima ou dos outros dados não devem ser contadas múltiplas vezes, ou seja, o número de combinações é o binomial de 8 sobre 2: $\binom{8}{2}$. Assim, como existem $\binom{8}{2}$ maneiras **distintas** de ocorrer o resultado i = 2, cada uma delas com probabilidade igual a $P' = p^2(1-p)^6 = (1/3)^2(2/3)^6$, a probabilidade total de i = 2 é:

$$P_8(2) = \binom{8}{2}\left(\frac{1}{3}\right)^2\left(\frac{2}{3}\right)^6 = 0{,}2731...$$

Esta fórmula corresponde ao caso particular da distribuição binomial para n = 8, p = 1/3 e i = 2, cuja generalização (fórmula 1) pode ser inferida substituindo os valores dessas grandezas nesse exemplo particular pelos símbolos que as representam.

A média e o desvio-padrão da distribuição binomial

A média de uma função de probabilidade pode ser calculada diretamente pela definição. Assim, a média da variável aleatória *i* é simplesmente a média ponderada pela probabilidade $P(i)$, que neste caso se escreve [3]:

$$\langle i \rangle = \sum_{i=0}^{n} i\, P_n(i) = \sum_{i=0}^{n} i \frac{n!}{i!(n-i)!} p^i (1-p)^{n-i} \qquad (1.2)$$

Como no cálculo de qualquer média ponderada, é necessário dividir pela soma dos pesos, mas neste caso ela *sempre* vale 1:

$$\sum_{i=0}^{n} P_n(i) = \sum_{i=0}^{n} \frac{n!}{i!(n-i)!} p^i (1-p)^{n-i} = (p+1-p)^n = 1^n = 1$$

Nessa dedução, usamos a fórmula do binômio de Newton. Toda função de probabilidade é normalizada para que sua soma seja 1, o que facilita sua interpretação: algo acontece sempre, de modo que a soma de todas as probabilidades tem que dar 100%, ou seja 1; quando dizemos que a probabilidade de um evento é 1, é porque ele vai acontecer, com toda certeza, e não há alternativa possível. A somatória em (2) pode ser efetuada e dá o resultado analítico simples

$$<i> = np \qquad (1.3)$$

O desvio-padrão, σ, também pode ser obtido pela raiz quadrada da média ponderada dos quadrados dos desvios:

$$\sigma = +\sqrt{\sigma^2}$$

$$\sigma^2 = \sum_{i=0}^{n} (i - <i>)^2 P_n(i) = np(1-p) \qquad (1.4)$$

As demonstrações dos resultados (1.3) e (1.4) são deixadas como exercícios.

Histogramas e a determinação experimental dos parâmetros estatísticos

A distribuição de probabilidade pode ser obtida experimentalmente realizando-se um número muito grande *N* de jogadas de *n* dados e verificando em cada jogada quantas (*i*) faces marcadas estão voltadas para cima. Ao longo da atividade, ficará claro o que significa *N* muito grande, na prática.

A aproximação experimental da probabilidade para a ocorrência de *i* sucessos com *n* dados é a frequência relativa, que é dada por:

$$F_n(i) = \frac{N_i}{N} \qquad (1.5)$$

[3] J.H. Vuolo, Fundamentos da Teoria de Erros, 2ª edição, Editora Edgard Blücher, São Paulo (1996).

onde N_i é o número de jogadas nas quais ocorreram i sucessos, isto é, foram observadas i faces marcadas voltadas para cima.

O valor experimental para a probabilidade de sucesso p é calculado a partir do valor médio dos valores observados para i,

$$< i >_{\exp} = \sum_{i=0}^{n} iF_n(i) \qquad (1.6)$$

usando a expressão (1.6), de modo que

$$p_{\exp} = \frac{< i >_{\exp}}{n} = \frac{1}{n} \sum_{i=0}^{n} iF_n(i) \qquad (1.7)$$

O valor médio do número de sucessos da relação (1.3) também poderia ser calculado diretamente a partir dos números de sucessos nas N jogadas. Assim, vamos numerar cada lance dos N realizados, de modo que i_k seja o número de sucessos na k-ésima jogada,

$$\langle i \rangle = \frac{1}{N} \sum_{k=1}^{N} i_k \quad .$$

A incerteza desse valor médio é medida pelo *desvio padrão da média*, que é:

$$\sigma_m = \frac{\sigma}{\sqrt{N}} \qquad (1.8)$$

Procedimento Experimental

A experiência consiste em realizar N = 10, 50 e 200 jogadas com 8 dados, anotando o número de sucessos em cada jogada, e comparar os resultados obtidos com as expectativas calculadas.

Será utilizado um conjunto de dados de faces opostas e outro conjunto e dados de faces consecutivas. Cada aluno da dupla trabalhará com apenas um conjunto de dados.

Deverá ser construído um histograma – recomenda-se fortemente a construção do histograma DURANTE a aula. Um histograma é um resumo gráfico de um conjunto de dados que nos permite ver certos comportamentos que são difíceis de observar em uma simples tabela numérica.

Sabemos que os valores variam em um conjunto de dados experimentais. O propósito da análise de um histograma é, por um lado, identificar e classificar o comportamento da grandeza e sua variação, e, por outro lado, desenvolver uma explicação razoável e relevante desse comportamento. A explicação deve ser baseada em conhecimentos gerais e na observação da situação específica. Naturalmente, apenas a análise de um histograma não é suficiente para confirmar ou refutar uma teoria.

Aula 1:

Deverão ser entregues os dados brutos e a tabela-resumo que apresente i, N_i, $F_n(i)$ para os dois conjuntos: N = 10, N = 50. Apresente os resultados numéricos da probabilidade teórica com um dígito significativo a mais que o do valor experimental correspondente.

Aula 2:

Deverão ser entregues os dados brutos e a tabela-resumo que apresente i, N_i, $F_n(i)$ para os dois conjuntos: N = 200. Apresente os resultados numéricos da probabilidade teórica com um dígito significativo a mais que o do valor experimental correspondente. Apresente os histogramas de frequência relativa com respectivos valores de média e desvio padrão.

Síntese:

A síntese a ser entregue deverá conter os seguintes itens:

- Descrição experimental
- Dados experimentais: valores experimentais para a probabilidade p obtidas a partir dos três conjuntos com N = 10, N = 50 e N = 200 de cada conjunto de dados, e N=400. Não se esqueça de indicar os valores obtidos com suas respectivas incertezas.
- Resultados e gráficos
  - Cálculo das probabilidades esperadas $P_n(i)$ de i sucessos ($0 \leq i \leq 8$).
  - Tabela-resumo que apresente i, N_i, $F_n(i)$ para os três conjuntos: N = 10, N = 50 e N = 200, em comparação com $P_n(i)$. Apresente os resultados numéricos da probabilidade teórica com um dígito significativo a mais que o do valor experimental correspondente.
  - Médias dos números de sucessos, $\langle i \rangle$, observados em cada um dos conjuntos com N = 10, 50 e 200 lançamentos.
  - Desvios-padrão experimentais (σ) para os três conjuntos com N = 10, N = 50 e N= 200, calculados usando a expressão abaixo e comparados ao teórico[4].

$$\sigma^2 = \frac{1}{N-1}\sum_{i=0}^{n}(i - \langle i \rangle)^2 N_i$$

  - Desvios-padrão das médias (σ_m) dos números de sucessos nos três conjuntos com N = 10, N = 50 e N = 200, e comparação com os desvios padrão da média esperados (eq. (1.8)).
  - Gráficos, FEITOS A MÃO, da probabilidade teórica (binomial) e da frequência relativa (estimativa experimental da probabilidade) em função da variável i para a amostra com N = 200.
- Discussão: discussão das diferenças observadas entre suas estimativas de p do item anterior com o valor esperado 1/3; comente sua dependência com o número

[4]Nas estimativas experimentais do desvio padrão, entra (N-1) no denominador e não N.

N de lançamentos do conjunto de dados. Breve discussão e comparação dos histogramas de N=200 de cada conjunto de dados (faces opostas e consecutivas) e de N=400 (referente a junção dos dados de N=200 de cada conjunto de dados).