

I.4

CÁLCULO DAS PROBABILIDADES

ADVERTÊNCIA PRÉVIA

Com o capítulo relativo a populações finitas, inicia-se a publicação das nossas lições de Cálculo das Probabilidades no Instituto Superior de Agronomia. Os capítulos seguintes serão publicados à medida que a experiência nos indicar qual a melhor orientação a seguir no ensino desta matéria, tendo sempre em conta a finalidade prática desse ensino e as condições especiais em que tem de ser realizado.

Na preparação das nossas lições serviram-nos de base, principalmente, as seguintes obras:

G. CASTELNUOVO - *Calcolo delle probabilita*, Zanichelli, Bologna, 1933.

D. J. FINNEY - *An introduction to statistical science in Agriculture*, John Wiley Sons, New York, 1953.

CRAMÉR - *Mathematical methods of Statistics*, Princeton University Press, Princeton, 1951.

G. UDN YULE and M. G. KENDALL - *An introduction to the Theory of Statistics*, Charles Griffin, Londres, 1937.

P. de VARNES E MENDONÇA - *Noções de Cálculo das Probabilidades*, Instituto Superior de Agronomia, Lisboa, 1950.

Convirá, no entanto, precisar, desde já, que tanto a orientação geral do curso, como muitos dos pormenores didáticos têm carácter pessoal.

Lisboa, Maio de 1955

J. Sebastião e Silva

I.4.1

INTRODUÇÃO AO CÁLCULO DAS PROBABILIDADES: POPULAÇÕES FINITAS

A – Frequências

1. Primeiros exemplos

Comecemos por um exemplo que nos é bem familiar. Suponhamos que, numa dada época de exames finais duma cadeira, se apresentaram a exame 189 alunos e os resultados foram os que constam da seguinte tabela:

TABELA N.º 1

Classificações	N.º de Alunos	Classificações	N.º de Alunos
0	0	11	45
1	0	12	28
2	0	13	17
3	0	14	9
4	2	15	12
5	1	16	7
6	5	17	5
7	11	18	2
8	7	19	1
9	0	20	0
10	37		

Diremos então que, nestas provas, a *frequência* da classificação 3 foi 0, a *frequência* da classificação 10 foi 37, etc.

Todavia, para ter uma ideia de conjunto destes resultados, nas suas relações mútuas, é preferível indicar a frequência de cada classificação em *percentagem*, tal como se observa na seguinte tabela deduzida da precedente:

TABELA N.º 2

Classificações	Percentagens	Classificações	Percentagens
0	0,0%	10	19,6%
1	0,0%	11	23,8%
2	0,0%	12	14,8%
3	0,0%	13	9,0%
4	1,1%	14	4,8%
5	0,5%	15	6,3%
6	2,6%	16	3,8%
7	5,8%	17	2,6%
8	3,7%	18	1,1%
9	0,0%	19	0,5%
		20	0,0%

No primeiro caso, diz-se que se trata de *frequências absolutas* e, no segundo, de *frequências relativas*. Assim, a frequência relativa da classificação 2 foi 0%, a frequência relativa da classificação 10 foi 19,6%, etc.

É claro que, designando por n o número total de alunos e por v o número de alunos que obtiveram uma dada classificação x (frequência absoluta de x), a frequência relativa de x , em percentagem, será o número $fr(x)$ dado pela fórmula

$$fr(x) = \frac{100v}{n} \% .$$

Mas uma percentagem não é mais do que uma maneira especial, usada na prática, de designar um número, geralmente compreendido entre 0 e 1. Em considerações puramente teóricas será preferível

designar esse número à maneira usual, e então, a fórmula que dá a frequência relativa toma o aspecto:

$$\text{fr}(x) = \frac{v}{n} .$$

Observe-se ainda que, na prática, o cálculo das percentagens conduz geralmente a dízimas, das quais não interessam todos os algarismos decimais. Escolhem-se valores aproximados desses números, por defeito ou por excesso, com o grau de aproximação que se considere suficiente. Mas, por coerência lógica, convirá escolher esses valores de modo que a sua soma seja exactamente 100. (No exemplo anterior tomaram-se valores aproximados a menos de 0,1).

Os conceitos de frequência absoluta e de frequência relativa são usados a cada passo na vida corrente. Indicaremos mais alguns:

- 1) – Numa dada cidade com 23.528 habitantes, 9.253 pertencem ao sexo masculino e os restantes 14.275 ao sexo feminino. Estes dois últimos números dão, pois, as frequências absolutas do sexo masculino e do sexo feminino entre os habitantes da referida cidade. As frequências relativas dos mesmos atributos, serão, respectivamente, de 39% e de 61%.
- 2) – Num dado país, 82% dos habitantes têm os cabelos castanhos ou pretos, 14% têm os cabelos louros e 4% têm os cabelos ruivos. São estas, pois, as frequências relativas daqueles atributos entre os habitantes do referido país.
- 3) – Numa série de 5.000 viagens efectuadas por uma dada companhia de aviação, houve desastres em 2 viagens. Nesta série, a frequência relativa dos desastres foi, portanto, de 0,04%, ou seja, de 0,0004.
- 4) – Numa série de competições, um dado clube desportivo teve 10 vitórias, 2 derrotas e 0 empates. Nesta série, as frequências relativas das vitórias, derrotas e empates foram, pois, respectivamente, iguais a 0,8, a 0,2 e a 0 (a menos de uma décima).

2. Populações. Álgebra dos atributos

Entre os exemplos anteriores, distinguem-se dois grupos principais. Num dos grupos apresentam-se nos frequências de *atributos*; no outro, aparecem nos frequências de *acontecimentos*. Em qualquer dos casos as considerações restringem-se a um determinado conjunto, que pode ser constituído por entidades das mais diversas naturezas (seres humanos, competições desportivas, viagens de aviação, etc.). Este conjunto fundamental, a que se refere um dado inquérito estatístico, é chamado, conforme os autores, *população*, *universo lógico*, *universo do discurso* ou, simplesmente, *universo*; os seus elementos são chamados os *indivíduos* (desse universo)⁽¹⁾.

Consideremos uma dada população U . Em U podem estar definidos diferentes *atributos* (em vez do termo “atributo” usam-se, ainda, como sinónimos, “propriedade”, “predicado”, “carácter”, etc.). A cada atributo α corresponde um determinado *conjunto* ou *classe* A , constituída por todos os indivíduos (elementos de U), que possuem o atributo α .

A presença simultânea de dois atributos α , β num mesmo indivíduo constitui um novo atributo, que se chama *conjunção* (ou *produto lógico*) dos primeiros e se representa por $\alpha \cap \beta$ ou, simplesmente, por $\alpha\beta$. Claro está que, se for A o conjunto dos indivíduos com o atributo α , e se for B o conjunto dos indivíduos com o atributo β , será $A \cap B$ (*intersecção* de A com B) o conjunto dos indivíduos com o atributo $\alpha\beta$ (ler “ α e β ”). Por exemplo, pode acontecer que, numa dada população de seres humanos, o atributo “masculino” e o atributo “ruivo” coexistam num mesmo indivíduo: a totalidade dos indivíduos que os possuem simultaneamente será a intersecção do conjunto dos indivíduos do sexo masculino com o conjunto dos indivíduos ruivos; mas também pode acontecer que, numa outra população, esta intersecção seja o conjunto vazio, isto é, que não haja nessa população homens ruivos.

O facto de um indivíduo possuir *um, pelo menos*, dos atributos α , β (podendo ter ambos ao mesmo tempo) exprime-se por um novo

(1) – Note-se que as designações *universo lógico* e *universo do discurso* já tinham sido adoptadas em lógica matemática com um sentido equivalente. De resto, as noções de lógica matemática introduzidas em Matemáticas Gerais vão ser também aqui utilizadas.

atributo, que se chama *disjunção* (ou *soma lógica*) dos primeiros e se representa por $\alpha \cup \beta$ ou por $\alpha + \beta$ (ler “ α ou β ”). Se forem A e B os conjuntos correspondentes a α e β , será $A \cup B$ (reunião de A com B) o conjunto correspondente a $\alpha \cup \beta$. Exemplo: entre os seres humanos, o atributo “casado ou viúvo” é a disjunção dos atributos “casado” e “viúvo”.

De modo análogo, se define a conjunção e a disjunção de *vários* atributos $\alpha, \beta, \gamma, \dots$. Por exemplo, o símbolo $\alpha\beta\gamma$ significa a presença simultânea dos atributos α, β, γ num indivíduo: designa, pois, a *conjunção* desses atributos.

Note-se que as operações da conjunção e disjunção são ambas associativas e comutativas (mas não reversíveis) e qualquer delas é distributiva em relação à outra, isto é, tem-se

$$\alpha(\beta + \gamma) = \alpha\beta + \alpha\gamma, \quad \alpha + \beta\gamma = (\alpha + \beta)(\alpha + \gamma).$$

A ausência dum atributo α num dado indivíduo constitui um novo atributo, que se chama *contrário* (ou *negação*) de α e se representa por $\sim\alpha$ ou por $\tilde{\alpha}$ (ler “não α ”). É claro que o conjunto dos indivíduos com o atributo $\tilde{\alpha}$ é o *complementar* do conjunto dos indivíduos com o atributo α . Por exemplo, nos resultados dum exame, o atributo “aprovado” é contrário do atributo “reprovado”.

Tem-se, ainda, como é fácil ver:

$$\sim(\sim\alpha) = \alpha \quad (\text{Lei da dupla negação})$$

$$\sim(\alpha\beta) = (\sim\alpha) + (\sim\beta), \quad \sim(\alpha + \beta) = (\sim\alpha)(\sim\beta) \quad (1^{\text{a}} \text{ lei de Morgan}).$$

Um atributo diz-se *universal*, quando se verifica em todos os indivíduos; o conjunto correspondente é pois o universo considerado, U . Um atributo diz-se *impossível*, quando não se verifica em nenhum indivíduo (elemento de U); o conjunto correspondente é o conjunto *vazio*.

Dois atributos α e β dizem-se *incompatíveis* quando a sua conjunção é um atributo impossível, isto é, quando os conjuntos correspondentes são *disjuntos*. Assim, por exemplo, no universo dos seres humanos, o atributo “solteiro” é incompatível com o atributo “casado”.

Dois atributos α e β contrários são sempre incompatíveis. Mas dois atributos podem ser incompatíveis sem serem contrários; exemplos: os atributos “solteiro” e “casado” entre seres humanos, os atributos “mediocre” e “bom” nos resultados dum exame, etc. Para dois atributos serem contrários, é necessário e suficiente que a sua conjunção seja um atributo impossível e a sua disjunção um atributo universal.

Dados dois atributos α , β , diz-se que α *implica* β e escreve-se $\alpha \subset \beta$, quando todos os indivíduos que possuem α também possuem β . Exemplos: “mediocre” implica “reprovado”, “casado” implica “não solteiro”. Os atributos α e β dizem-se *equivalentes* quando $\alpha \subset \beta$ e $\beta \subset \alpha$; exemplo: “mediocre ou mau” equivale a “reprovado”.

De tudo o que precede ressalta que a álgebra dos atributos é perfeitamente análoga à álgebra das funções proposicionais (numa variável) e ambas se traduzem directamente na álgebra dos conjuntos. É que, a bem dizer, entre atributos e funções proposicionais existe apenas uma diferença de *forma*. Seja, por exemplo, a função proposicional

“ x é agrónomo ou silvicultor”

definida entre indivíduos portugueses; trata-se, como se vê, duma fórmula que se converte em proposição verdadeira para certas determinações de x e em proposição falsa para outras determinações de x . Exprime, pois, um *atributo* que se concretiza na reunião de duas classes: a dos engenheiros agrónomos e a dos engenheiros silvicultores (classes não necessariamente disjuntas).

Uma classe pode ser determinada de dois modos diversos: – ou dando um sistema de atributos ou regras que a caracterizem completamente (*ponto de vista da compreensão*); – ou indicando um por um os indivíduos que a compõem (*ponto de vista de extensão*). Conforme o ponto de vista adoptado, assim o conceito de classe se aproxima do conceito de atributo ou do conceito de conjunto. A distinção entre tais conceitos tem sempre um carácter mais ou menos subjectivo.

Note-se que, na linguagem comum, os *adjectivos* exprimem normalmente atributos, enquanto os *substantivos comuns* se referem a *classes* e os *substantivos próprios* a indivíduos.

Convém, ainda, lembrar que os atributos (ou as classes) que a Natureza nos apresenta nunca são nitidamente definidos, como os conceitos abstractos da Matemática. Entre um atributo e os atributos vizinhos há geralmente zonas fronteiriças, em que a distinção acaba sempre por ser feita de modo mais ou menos artificial ou arbitrário. Por exemplo, ao classificar os habitantes duma cidade segundo a cor dos cabelos, muitas vezes se hesita entre os atributos “castanho” e “loiro” ou “ruivo”, para um dado indivíduo.

3. Álgebra dos acontecimentos

Recordemos que, entre os exemplos do n.º 1, alguns se referem, não propriamente a atributos, mas sim a *acontecimentos*. Não pretendemos aqui definir “acontecimento”, assim como não tentámos definir “atributo”: tratam-se de conceitos psicologicamente primitivos. O que será possível e conveniente é esclarecer tais conceitos e precisar, tanto quanto possível, a terminologia que lhes diz respeito.

Comecemos por notar que, em vez de “acontecimento”, se usam com significado semelhante os termos “facto”, “fenómeno”, “eventualidade”, etc.

Imaginemos uma prova ou experiência, que se possa repetir várias vezes em condições idênticas, conduzindo, de cada vez, a um ou mais resultados, entre vários que são possíveis. É a cada um desses resultados que, em cálculo das probabilidades e em estatística, se costuma dar o nome *acontecimento*. Assim, antes de efectuar a prova, vários acontecimentos são de esperar: cada um deles é, então, apenas uma *eventualidade* ou *hipótese* (*acontecimento em potência*); efectuada a prova, apenas alguma ou algumas dessas hipóteses se realizam (*acontecimento em acto*).

São inúmeros os exemplos que, neste sentido, se podem apresentar. Por enquanto, recordaremos apenas os que foram citados no n.º 1: desastre ou ausência de desastre numa viagem aérea, resultados duma prova desportiva, etc.

Mas convém desde já observar que, muitas vezes, os acontecimentos se exprimem por meio de atributos (e vice-versa). Tal é, por exemplo, o caso dos resultados dum exame (ou o caso análogo duma competição desportiva); assim, ao atributo “reprovado” corresponde o acontecimento “ficar reprovado”, ao atributo “vitorioso” corresponde

o acontecimento “vencer” (na linguagem comum, os acontecimentos são geralmente expressos por *verbos*, enquanto os atributos são expressos por *adjectivos*).

Ainda aqui, portanto, há, de certo modo, uma questão de ponto de vista psicológico. Podemos, portanto, desde já, prever que para os acontecimentos, exista uma álgebra perfeitamente análoga à dos atributos. Com efeito, sejam α e β dois acontecimentos a esperar numa dada prova \mathcal{P} :

- 1) - Chama-se *conjunção* (ou *produto lógico*) de α e β , e representa-se por $\alpha\beta$, o acontecimento que consiste na realização simultânea de α e de β .
- 2) - Chama-se *disjunção* (ou *soma lógica*) de α e β , e representa-se por $\alpha + \beta$, o acontecimento que consiste em se realizar *um, pelo menos*, dos acontecimentos α e β .
- 3) - Chama-se *contrário* (ou *negação*) de α , e representa-se $\sim \alpha$ ou por $\tilde{\alpha}$, o acontecimento que consiste em não se realizar α .
- 4) - Diz-se que α *implica* β , e escreve-se $\alpha \subset \beta$, quando β se realiza todas as vezes que α se realiza.
- 5) - Dois acontecimentos α , β dizem-se *equivalentes*, e escreve-se $\alpha \equiv \beta$, quando $\alpha \subset \beta$ e $\beta \subset \alpha$.
- 6) - Um acontecimento diz-se *certo* quando já se sabe *a priori*, com certeza absoluta, que se realizará na prova \mathcal{P} , todas as vezes que esta for efectuada. Um acontecimento diz-se *impossível*, quando é certo o seu contrário. Por exemplo, no exame final dum aluno é certo o acontecimento “aprovação ou reprovação” (excluem-se os casos de falta ou desistência) e é impossível o acontecimento “obter 21 valores” (em escolas portuguesas).
- 7) - Dois acontecimentos dizem-se *incompatíveis*, quando a sua conjunção é um acontecimento impossível.

As operações de conjunção, disjunção e negação entre acontecimentos gozam das mesmas propriedades que entre atributos, tornando-se, pois, supérfluo mencioná-las aqui.

4. Acontecimentos expressos em forma proposicional

A maior parte dos acontecimentos estudados em cálculo das probabilidades apresentam-se, naturalmente, sob a forma de funções proposicionais. Vejamos dois exemplos:

a) Imaginemos uma urna que contenha bolas brancas e bolas pretas. Designando por U o universo das bolas contidas na urna, por B o conjunto das bolas brancas e por P o conjunto das bolas pretas, teremos duas funções proposicionais

$$x \in B, \quad x \in P$$

definidas em U . A variável x toma, pois, cada um dos seus valores em U . Ora, o valor de x pode ser determinado, extraindo *ao acaso* uma bola de U . Nesta prova (extracção da bola) realiza-se, então, um dos seguintes *acontecimentos contrários*: $x \in B$, (a bola x que sai é branca), $x \in P$ (a bola x que sai é preta). Vê-se, pois, como as funções proposicionais consideradas passam a exprimir acontecimentos.

b) Seja x a classificação dum aluno numa prova \mathcal{P} a realizar. Os resultados “mau”, “medíocre”, “suficiente”, “bom”, “distinto” e “muito bom”, segundo convenções frequentemente adoptadas, são expressos, respectivamente, pelas funções proposicionais seguintes:

$$x < 6, \quad 6 \leq x < 10, \quad 10 \leq x < 14, \quad 14 \leq x < 16, \\ 16 \leq x < 18, \quad 18 \leq x.$$

O resultado “aprovado”, soma lógica de “suficiente”, “bom”, “distinto” e “muito bom”, é expresso pela função proposicional $x \geq 10$.

Assim, nestes casos, os acontecimentos a que pode dar lugar a prova \mathcal{P} considerada aparecem directamente sob a forma de funções proposicionais $\alpha(x)$ definidas num dado universo U , sendo o valor da variável x (elemento de U) determinado em cada realização da prova \mathcal{P} . A variável x recebe, nos casos como o considerado em a), o nome de *variável casual* ou *variável aleatória*, atendendo a que na sua determinação, intervém, mais ou menos, o *acaso*⁽¹⁾. É claro

(1) – É impossível definir logicamente o termo “acaso”. Numa primeira aproximação, poderíamos dizer que este termo designa ausência de *causa*, ou, pelo menos, de *causa conhecida*. Para certos autores, o acaso consistiria na acumulação de um grande número de pequenas causas desconhecidas que actuam em diversos sentidos, tornando praticamente impossível a previsão do efeito global. Sobre este ponto, aconselhamos a leitura do prefácio do livro de CASTELNUOVO, citado na Bibliografia.

que a conjunção de dois acontecimentos α e β será expressa pela conjunção das funções proposicionais que exprimem α e β e analogamente para a disjunção, para a negação, etc.

Deste modo, a cada acontecimento $\alpha(x)$ ficará a corresponder um determinado subconjunto A de U : o conjunto de indivíduos que verificam $\alpha(x)$; reciprocamente, a cada conjunto A de U corresponde o acontecimento expresso pela função proposicional $x \in A$. (A dois acontecimentos corresponde o mesmo conjunto, quando, e só quando, os acontecimentos são equivalentes). *A álgebra dos acontecimentos poderá, pois, traduzir-se na álgebra dos conjuntos.*

Mesmo no caso geral, a redução dos acontecimentos à forma de função proposicional é sempre possível, pelo menos de modo indirecto.

5. Frequência dum atributo numa população

Chama-se *frequência absoluta* dum atributo α (num dado universo U) ao número de indivíduos que possuem esse atributo, isto é, ao número de elementos do conjunto A correspondente ao atributo α . Designaremos esse número por (α) .

Chama-se *frequência relativa* do atributo α ao quociente de (α) pelo número total de indivíduos (elementos de U). Designaremos por N esse número e por $\text{fr}(\alpha)$ a frequência relativa de α . Ter-se-á, pois, por definição:

$$\text{fr}(\alpha) = \frac{(\alpha)}{N} .$$

A frequência relativa costuma também ser expressa em *tantos por cento* e, algumas vezes, em *tantos por mil*, devendo usar-se, então, respectivamente, as fórmulas:

$$\text{fr}(\alpha) = \frac{100 \cdot (\alpha)}{N} \% , \quad \text{fr}(\alpha) = \frac{1.000 \cdot (\alpha)}{N} \text{‰} .$$

Já no n.º 1 demos vários exemplos de frequências absolutas e de frequências relativas.

Resulta logo da definição que se tem $0 \leq \text{fr}(\alpha) \leq 1$, qualquer que seja o atributo α .

Em particular, será $\text{fr}(\alpha) = 1$ [isto é, $(\alpha) = N$], quando é atributo universal, e $\text{fr}(\alpha) = 0$ [ou seja $(\alpha) = 0$], quando é atributo impossível. Além disso, tem-se o

TEOREMA DA SOMA. *Dados p atributos $\alpha_1, \alpha_2, \dots, \alpha_p$, incompatíveis dois a dois, a frequência absoluta [resp. relativa] da soma lógica desses atributos é sempre igual à soma das frequências absolutas [resp. relativas] dos mesmos atributos; isto é, em fórmulas:*

$$(\alpha_1 + \alpha_2 + \dots + \alpha_p) = (\alpha_1) + (\alpha_2) + \dots + (\alpha_p),$$

$$(5.1) \quad \text{fr}(\alpha_1 + \alpha_2 + \dots + \alpha_p) = \text{fr}(\alpha_1) + \text{fr}(\alpha_2) + \dots + \text{fr}(\alpha_p).$$

Com efeito, sendo cada um dos atributos α_i incompatível com qualquer dos outros, os conjuntos correspondentes são disjuntos dois a dois e, portanto, a sua reunião, correspondente ao atributo $\alpha_1 + \alpha_2 + \dots + \alpha_p$, tem um número de elementos igual à soma dos elementos dos primeiros, isto é,

$$(\alpha_1 + \alpha_2 + \dots + \alpha_p) = (\alpha_1) + (\alpha_2) + \dots + (\alpha_p)$$

e, portanto,

$$\frac{(\alpha_1 + \alpha_2 + \dots + \alpha_p)}{N} = \frac{(\alpha_1)}{N} + \frac{(\alpha_2)}{N} + \dots + \frac{(\alpha_p)}{N},$$

que é, precisamente, o que exprime a fórmula (5.1).

Note-se que, na aplicação deste teorema, *é essencial que esses atributos considerados sejam incompatíveis, dois a dois*. Por exemplo, suponhamos que, numa certa cidade, 8% dos habitantes são empregados do Estado e 34% são empregados de empresas particulares; não se pode daí concluir, sem mais, que 42% dos habitantes são empregados (do Estado ou de empresas particulares), pois pode haver habitantes que sejam ao mesmo tempo empregados do Estado e de empresas particulares. Vemos mais adiante como proceder, de modo sistemático, em casos tais.

Entretanto, observemos que do teorema anterior (ou mesmo da definição) se deduz logo o

COROLÁRIO. *Se for fr a frequência relativa dum dado atributo, será $1 - fr$ a frequência relativa do atributo contrário; isto é, em fórmula*

$$fr(\tilde{\alpha}) = 1 - fr(\alpha).$$

Com efeito, os atributos α , $\tilde{\alpha}$ são incompatíveis e, como a sua soma lógica é atributo universal, tem-se

$$fr(\alpha) + fr(\tilde{\alpha}) = fr(\alpha + \tilde{\alpha}) = 1, \text{ donde, } fr(\tilde{\alpha}) = 1 - fr(\alpha).$$

6. Frequência dum acontecimento numa série de provas

Consideremos n realizações, $\mathcal{P}_1, \mathcal{P}_2, \dots, \mathcal{P}_n$, duma prova genérica \mathcal{P} e seja α um acontecimento a provar em cada realização de \mathcal{P} . Suponhamos que o acontecimento α se realiza em v destas provas e que, portanto, o seu contrário se realiza nas $n - v$ provas restantes. Diremos, então, que v é a *frequência absoluta* do acontecimento α na série de provas consideradas. Por outro lado, chamaremos *frequência relativa* de α na referida série ao cociente de v por n , isto é, ao número f dado pela fórmula

$$f = \frac{v}{n},$$

número este que depende não só do acontecimento, como, ainda, do número n das provas. É claro que se terá sempre

$$0 \leq f \leq 1.$$

Mas o caso $f = 1$ não habilita em absoluto a concluir que o acontecimento α é *certo*, assim como o caso $f = 0$ não habilita em absoluto concluir que o acontecimento é *impossível*.

Entretanto, importa observar que o *teorema da soma*, bem como o *respectivo corolário*, *subsistem para os acontecimentos, bastando substituir nos enunciados a palavra “atributo” pela palavra “acontecimento”*. As demonstrações são perfeitamente análogas às anteriores.

NOTA. Se tomarmos para universo lógico o conjunto $\{\mathcal{P}_1, \mathcal{P}_2, \dots, \mathcal{P}_n\}$, a função proposicional de

“ \mathcal{P} é uma prova em que se realizou α ”

exprime visivelmente um atributo, cuja frequência (absoluta ou relativa) é por definição a frequência (absoluta ou relativa) do acontecimento α na série considerada.

7. Partições

Chama-se *partição* (ou *classificação*) dum universo U a toda a decomposição de U num conjunto $\{C_1, C_2, \dots, C_m\}$ de conjuntos disjuntos dois a dois, cuja reunião seja U , isto é, de conjuntos C_i tais que

$$C_i C_k = \emptyset \text{ para } i \neq k \text{ e } \sum_{i=1}^m C_i = U.$$

Os conjuntos C_i dir-se-ão *elementos* ou *células* da partição⁽¹⁾.

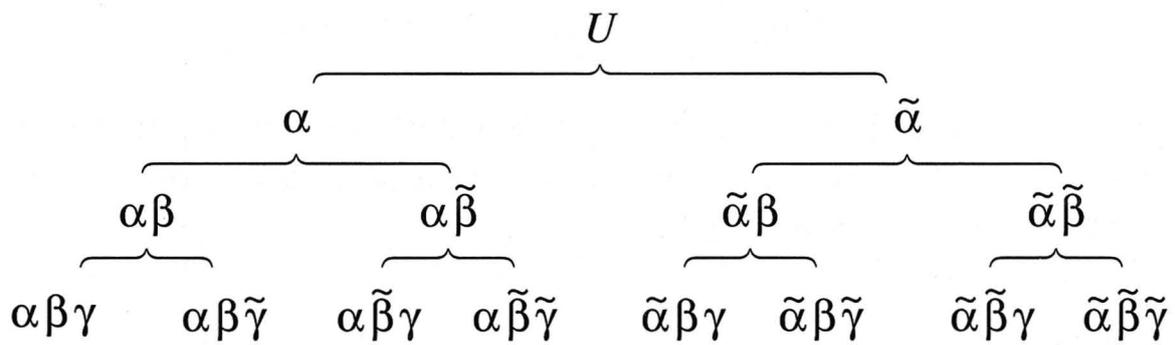
Paralelamente, chamaremos *partição em atributos* a todo o conjunto $\{\alpha_1, \alpha_2, \dots, \alpha_n\}$ de atributos incompatíveis, dois a dois, cuja soma lógica seja o atributo universal.

Uma partição diz-se *dupla*, *tripla*, etc., conforme o número m dos seus elementos for 2, 3, Sempre que esse número é superior a dois, a partição diz-se *múltipla* (excluimos o caso sem interesse $m=1$).

O caso mais simples será o das classificações duplas, também chamadas *classificações dicotómicas* ou *dicotomias*. Estas podem sempre ser determinadas em U , por um atributo α e pelo seu contrário $\tilde{\alpha}$ (é claro que não interessa o caso em que α é impossível ou universal).

Dados vários atributos $\alpha, \beta, \gamma, \dots$, sobre U , as operações de negação e de conjunção efectuadas sobre esses atributos e sobre os resultados obtidos conduzem necessariamente a uma partição em atributos. Ora, essa partição pode sempre ser atingida por sucessivas dicotomias, como vamos ver. Para fixar ideias, limitemo-nos ao caso de 3 atributos α, β, γ . As sucessivas dicotomias podem ser esquematizadas do seguinte modo:

(1) – Supõe-se nesta definição que o universo é finito, mas é claro que a definição se pode estender imediatamente ao caso dos universos infinitos e das partições com uma infinidade de células.

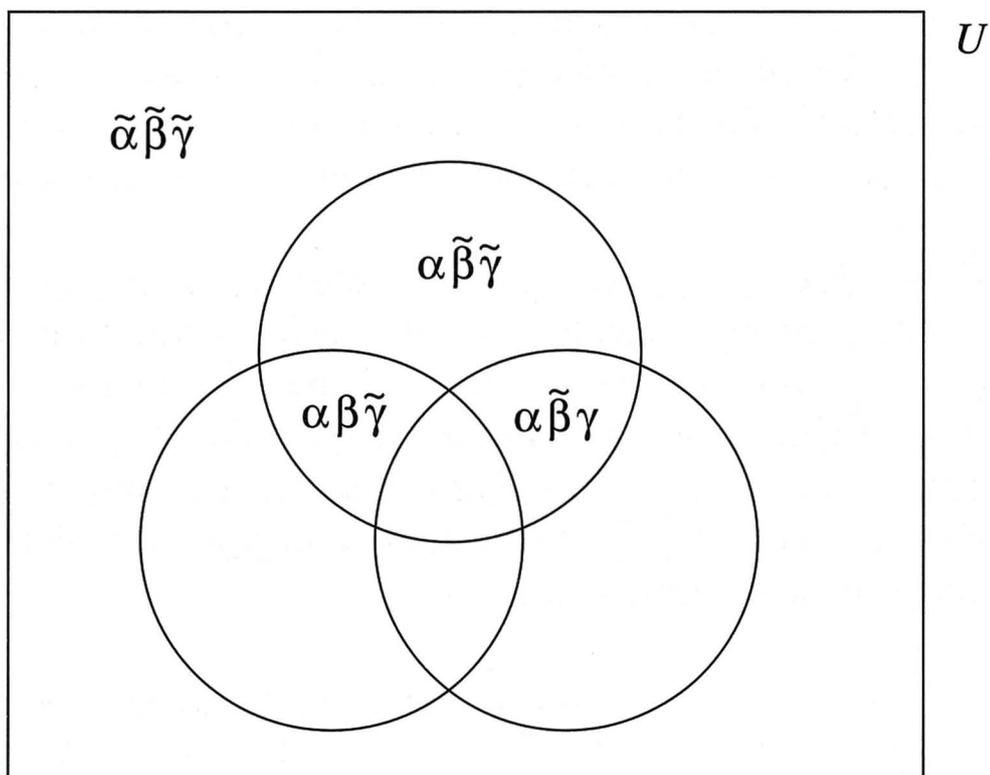


Obtivemos assim, como é fácil verificar, uma partição constituída pelos 2^3 atributos

$$\alpha\beta\gamma, \alpha\tilde{\beta}\gamma, \alpha\tilde{\beta}\tilde{\gamma}, \tilde{\alpha}\beta\gamma, \tilde{\alpha}\beta\tilde{\gamma}, \tilde{\alpha}\tilde{\beta}\gamma, \tilde{\alpha}\tilde{\beta}\tilde{\gamma}, \alpha\beta\tilde{\gamma},$$

entre os quais, porém, pode haver algum repetido ou impossível, o que reduz o seu efectivo.

Estes resultados estão sugestivamente figurados no diagrama junto, em que as classes correspondentes aos atributos α , β , γ são representadas por círculos e o universo por um rectângulo.



Suponhamos, por exemplo, que o universo U é constituído pelos portugueses, sendo α o atributo “casado”, β o atributo “empregado” e γ o atributo “com menos de 35 anos”. Então, $\alpha\beta\gamma$ será o atributo “casado, empregado, com menos de 35 anos”, $\tilde{\alpha}\beta\tilde{\gamma}$ o atributo “não casado, empregado, com idade não inferior a 35 anos”, etc.

Toda a partição U^* dum universo U (em conjuntos ou em atributos) pode ser tomada para novo universo, dando-se, por assim dizer, uma *condensação* do universo primitivo. Os novos indivíduos serão, é claro, as células da partição U^* . Este processo de condensação é muitas vezes usado em Estatística para resumir dados. Por exemplo, na Tabela n.º 1 fez-se uma partição dos alunos em classes, conforme as notas obtidas no exame, substituindo-se o universo dos alunos pelo universo das notas; este, por sua vez, pode ser condensado no universo constituído pelos atributos “mau”, “medíocre”, “suficiente”, “bom”, “distinto” e “muito bom”, o qual, por sua vez, pode ser substituído pelo universo {“aprovado”, “reprovado”}.

Estas considerações aplicam-se *mutatis mutandis* ao caso dos acontecimentos. Chamaremos *partição dum acontecimento certo* a todo o sistema de acontecimentos incompatíveis, dois a dois, cuja soma lógica seja um acontecimento certo. (Esta definição, torna-se mais sugestiva, se em vez do termo “acontecimento” usarmos o termo “eventualidade”).

8. Corpos de conjuntos, corpos de atributos, corpos de acontecimentos

Consideremos um universo U (finito). Diz-se que uma família $\{C_1, C_2, \dots, C_r\}$ de conjuntos de elementos de U é um *corpo*, quando se verificam as duas seguintes condições: 1) – a soma ou o produto lógico de dois (ou mais) conjuntos da família ainda é um conjunto da família; 2) – o complemento de cada conjunto da família ainda pertence à mesma.

Analogamente, se define “corpo de atributos” e “corpo de acontecimentos”. Neste último caso, devemos supor que se trata dum sistema *finito* de acontecimentos a prever numa dada prova \mathcal{P} .

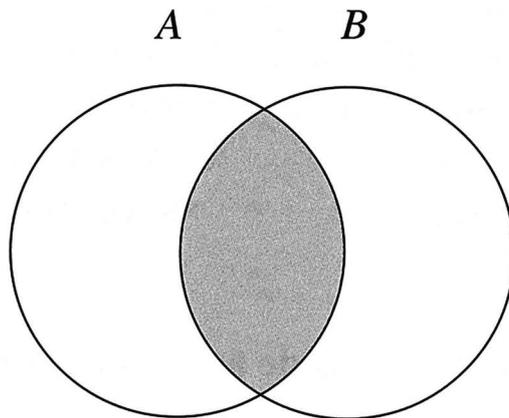
A família de todos os possíveis conjuntos de elementos de U é, manifestamente, um corpo de conjuntos. Mas há outros exemplos de corpos.

Consideremos um corpo \mathcal{R} de conjuntos e seja C um desses conjuntos. Então, \tilde{C} também pertence a \mathcal{R} . Logo, também pertencem a \mathcal{R} o conjunto $C + \tilde{C} = U$ e o conjunto $C\tilde{C} = \emptyset$; isto é, *entre os conjuntos dum corpo figuram sempre o universo e o conjunto vazio*.

Chamam-se *células* do corpo \mathcal{R} os conjuntos não vazios de \mathcal{R} que não contêm nenhum outro conjunto de \mathcal{R} , a não ser o conjunto vazio. Dado um indivíduo a , a intersecção de todos os conjuntos de \mathcal{R} que contêm a é manifestamente uma célula de \mathcal{R} ; o menor conjunto de \mathcal{R} que contém a . Tem-se, além disso, o seguinte

TEOREMA. *Sejam A e B duas células de \mathcal{R} . Então, de duas uma: ou os conjuntos A e B coincidem ou são disjuntos.*

Com efeito, suponhamos que os conjuntos A , B não coincidem, isto é, que $A \neq B$, mas que a intersecção AB não é vazia.



Então, visto ser $A \neq B$, podemos garantir que um, pelo menos, destes conjuntos não estará contido no outro: seja, por exemplo, A esse conjunto. Nesta hipótese, tanto AB como $A\tilde{B}$ são conjuntos de \mathcal{R} diferentes de A e contidos em A e, como $A = AB + A\tilde{B}$, nenhum deles será vazio. Mas isto é impossível, porque então não seria A uma célula. Logo, ou A e B são disjuntos (isto é, $AB = \emptyset$) ou coincidem (isto é, $A = B$).

Daqui se deduzem logo os seguintes corolários:

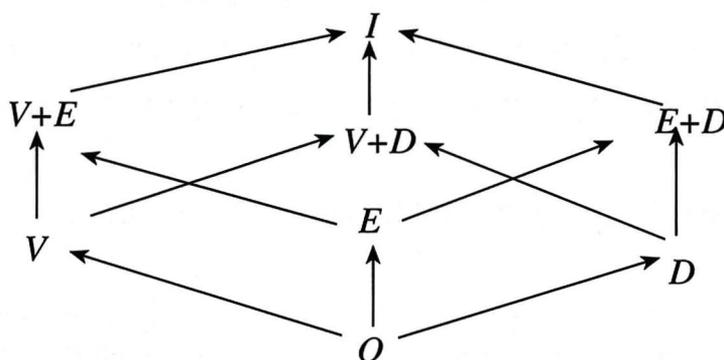
COROLÁRIO I. *As células dum corpo de conjuntos formam uma partição do universo.*

COROLÁRIO II. *Para todo o conjunto A pertencente a um dado corpo \mathfrak{R} de conjuntos, existe um e só um conjunto de células de \mathfrak{R} , de que A é soma lógica.*

Posto isto, consideremos uma classe \mathfrak{F} qualquer de subconjuntos de P . Efectuando operações de intersecção, reunião e passagem ao complementar, de todos os modos possíveis, a partir dos conjuntos de \mathfrak{F} , os resultados obtidos constituem um corpo de conjuntos; ao menor corpo que contem \mathfrak{F} , dá-se-lhe o nome de *corpo de conjuntos gerado por \mathfrak{F}* . É claro que as células deste corpo são as células da partição determinada por \mathfrak{F} , a qual se pode obter por sucessivas dicotomias como foi indicado no n.º anterior. Os restantes conjuntos do corpo (além do conjunto vazio) são reuniões de células, como indica o Corolário II.

Todas estas considerações se aplicam, *mutatis mutandis*, aos corpos de atributos e de acontecimentos.

Consideremos, por exemplo, entre os resultados normais dum desafio de futebol entre dois grupos A e B , os dois seguintes: *vitória de A* (que designaremos por V) e *empate* (que designaremos por E). O corpo de acontecimentos gerado por V e E é formado pelos seguintes elementos: $V + E$ (vitória de A ou empate), $\overline{V + E}$ (derrota de A – acontecimento que designaremos por D), $V + D$ (vitória ou derrota de A), $E + D$ (empate ou derrota de A), $V + D + E$ (vitória ou derrota de A , ou empate – acontecimento certo, que designaremos por I), VD (vitória e derrota de A – acontecimento impossível, que designaremos por O). Este corpo de eventualidades pode ser representado pelo seguinte esquema, em que usamos a seta como símbolo de implicação:



É claro que V, E, D (células do corpo considerado) ainda podem exprimir-se como somas de acontecimentos elementares (indecomponíveis), *não pertencentes a este corpo*. Com efeito, todo o resultado dum jogo se pode exprimir por um par ordenado (m, n) de números inteiros não negativos, sendo m o número de golos marcados por A e n o número de golos marcados por B . Então, por exemplo, o acontecimento V será a soma dos acontecimentos representados pelos pares (m, n) com $m > n$. Por sua vez, os acontecimentos E e D são expressos, respectivamente, pelas funções proposicionais $m = n$ e $m < n$.

Dum modo geral, a totalidade dos acontecimentos a considerar numa dada prova ou experiência \mathcal{P} constitui um corpo, desde que se convencie (como é uso) incluir nessa totalidade o acontecimento impossível⁽¹⁾.

Esta convenção é adoptada para comodidade de linguagem. Na mesma ordem de ideias se convencia que *o acontecimento impossível implica qualquer outro acontecimento* – convenção, de resto, natural, pois que, só admitindo-a, será inteiramente válida a seguinte propriedade intuitiva:

$$\text{Se } \alpha \subset \beta, \text{ então } \tilde{\beta} \subset \tilde{\alpha}$$

Com efeito, designando por I o acontecimento certo, tem-se $\alpha \subset I$, qualquer que seja α , e portanto, $\tilde{I} \subset \tilde{\alpha}$, qualquer que seja $\tilde{\alpha}$.

Se isto não fosse verdade, a propriedade anterior não seria válida.

Consideremos agora um corpo de atributos e sejam $\alpha_1, \alpha_2, \dots, \alpha_m$ as suas células. Qualquer outro atributo α do corpo se exprime como soma de células. Ora, como estas são incompatíveis duas a duas, deduz-se do teorema da soma, este outro

TEOREMA. *A frequência (absoluta ou relativa) de qualquer atributo α dum corpo \mathcal{R} de atributos é igual à soma das frequências (absolutas ou relativas) das células de \mathcal{R} cuja soma é α .*

Conclusão análoga para os corpos de acontecimentos.

(1) – Subentende-se que a totalidade é finita. Note-se, porém, que o conceito de corpo se pode generalizar ao caso dos conjuntos infinitos (de conjuntos, de atributos ou de acontecimentos).

NOTA. É preciso não confundir um *conjunto de conjuntos* com a *reunião* desses conjuntos. Por exemplo, o conjunto V dos vertebrados está classificado em vários conjuntos parciais: mamíferos, aves, etc.; porém, V não é o *conjunto* desses conjuntos, mas sim a sua *reunião* ou *soma*.

Ora, vimos atrás que, para todo o conjunto C pertencente a um corpo \mathcal{R} de conjuntos, existe um, e um só, conjunto C^* de células de \mathcal{R} cuja soma é C ; reciprocamente, todo o conjunto C^* de células de \mathcal{R} determina um conjunto $C \in \mathcal{R}$. Portanto, se designarmos por U^* a partição de U constituída pelas células de \mathcal{R} , fica, assim, estabelecida uma correspondência biúnivoca $C \leftrightarrow C^*$ entre os conjuntos C de \mathcal{R} e os subconjuntos de U^* . Além disso, é fácil ver que à soma $C_1 + C_2$ de dois conjuntos C_1, C_2 de \mathcal{R} corresponde, deste modo, a soma $C_1^* + C_2^*$ dos subconjuntos correspondentes de U^* , ao produto $C_1 C_2$ corresponde o produto $C_1^* C_2^*$, e ao complementar de C corresponde o complementar de C^* . Exprime-se este facto dizendo que a referida correspondência é um *isomorfismo* entre o corpo \mathcal{R} e o corpo dos subconjuntos de U^* ; diz-se, também, que esses dois corpos são *isomorfos*.

Analogamente, se conclui o seguinte facto:

Se \mathcal{S} é um corpo (finito) de acontecimentos e \mathcal{E} o conjunto das células de \mathcal{S} , o corpo \mathcal{S} é isomorfo a um corpo de conjuntos, que é precisamente o corpo dos subconjuntos de \mathcal{E} .

Por exemplo, o corpo gerado pelos acontecimentos V, E atrás considerados é isomorfo ao corpo dos subconjuntos de $\{V, E, D\}$, que são: este próprio conjunto, o conjunto vazio e, ainda, os conjuntos $\{V\}, \{E\}, \{D\}, \{V, E\}, \{V, D\}, \{E, D\}$.

Ficam assim completadas as considerações do n.º 3: o isomorfismo em questão traduz exactamente a álgebra dos acontecimentos na álgebra dos conjuntos.

9. Distribuição em universos finitos

Consideremos um corpo \mathcal{R} de conjuntos (num universo U finito). Suponhamos que, a cada conjunto $C \subset \mathcal{R}$, se faz corresponder um determinado número real e designemos esse número por $\Phi(C)$: fica assim definida em \mathcal{R} a *função* $\Phi(C)$, do conjunto variável C .

Pois bem, chamaremos *distribuição em \mathcal{R}* toda a função de conjunto, $\Phi(C)$, definida em \mathcal{R} , que verifique as duas seguintes condições:

- 1) – $\Phi(C) \geq 0$, para todo o $C \in \mathcal{R}$.
- 2) – Se C_1, C_2 são conjuntos *disjuntos* de \mathcal{R} , então

$$\Phi(C_1 + C_2) = \Phi(C_1) + \Phi(C_2).$$

Em particular, se \mathcal{R} é formado por todos os subconjuntos de U , diremos, então, que $\Phi(C)$ é uma *distribuição sobre U* .

É fácil reconhecer a veracidade da seguinte proposição:

TEOREMA. *Sejam C_1, C_2, \dots, C_m as células do corpo \mathcal{R} . Se associarmos a cada conjunto C_i , arbitrariamente, um número real $\gamma_i \geq 0$, existe sempre uma, e uma só, distribuição $\Phi(A)$ em \mathcal{R} que, para cada conjunto C_i , toma o valor γ_i ($i = 1, 2, \dots, m$). O valor de $\Phi(A)$ para cada $A \subset \mathcal{R}$ é, então, a soma dos valores que a função toma nas células cuja soma é A .*

Em particular, se \mathcal{R} é formado por todos os subconjuntos de U , as suas células reduzem-se aos indivíduos (elementos de U) e tem-se:

COROLÁRIO. *Toda a função não negativa definida em U é prolongável numa (e numa só) distribuição sobre U .*

Daqui o chamar-se também, algumas vezes, *distribuição em U* a qualquer função real não negativa definida em U .

É claro que o conceito de “distribuição num corpo de conjuntos” se estende imediatamente ao caso dos corpos de atributos e dos corpos de acontecimentos.

Exemplos de distribuições são-nos oferecidos pelas frequências, tais como atrás foram definidas, tendo em vista o teorema da soma. Mas, além das distribuições de frequência, apresentam-se ainda na prática muitos outros exemplos de distribuições: distribuições de probabilidade (que estudaremos mais adiante), distribuições de massa, distribuições de carga eléctrica, etc., etc.

Chamaremos *distribuição relativa* (num corpo \mathcal{R}) a toda a distribuição $\Phi(A)$ em \mathcal{R} que verifique a condição suplementar seguinte:

3) – $\Phi(U) = 1$ (sendo U o universo, como já foi dito).

É fácil ver que, de cada distribuição $\Phi(A)$ em \mathcal{R} , se deduz uma distribuição relativa, $\phi(A)$, em \mathcal{R} , pondo

$$\phi(A) = \frac{\Phi(A)}{\Phi(U)}, \text{ sendo } U \text{ o universo.}$$

Do teorema anterior deduz-se o seguinte

COROLÁRIO. *Se $\Phi(A)$ é uma distribuição relativa, tem-se sempre*

$$\Phi(\tilde{A}) = 1 - \Phi(A).$$

A dedução faz-se exactamente como no caso particular das frequências relativas (n.º 5).

10. Soma de conjuntos não disjuntos (atributos ou acontecimentos compatíveis)

Seja $\mu(A)$ uma distribuição definida num corpo \mathcal{R} de conjuntos. Sendo A, B dois conjuntos de \mathcal{R} , o uso da fórmula

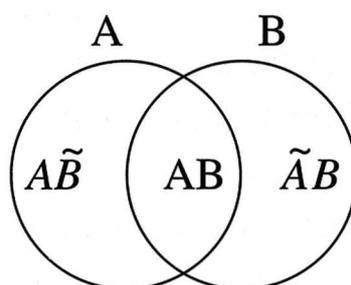
$$\mu(A + B) = \mu(A) + \mu(B)$$

exige a restrição de que A e B sejam disjuntos. Como proceder, porém, no caso geral? Já atrás (n.º 5) se nos pôs esta questão a propósito dum exemplo comezinho.

Basta notar que se tem, necessariamente,

$$A = A\tilde{B} + AB, \quad B = AB + \tilde{A}B, \quad A + B = A\tilde{B} + AB + \tilde{A}B.$$

Como os conjuntos $A\tilde{B}, AB, \tilde{A}B$ são disjuntos dois a dois, ter-se-á:



$$\begin{aligned}\mu(A) &= \mu(A\tilde{B}) + \mu(AB), \quad \mu(B) = \mu(AB) + \mu(\tilde{A}B), \\ \mu(A + B) &= \mu(A\tilde{B}) + \mu(AB) + \mu(\tilde{A}B)\end{aligned}$$

donde $\mu(A\tilde{B}) = \mu(A) - \mu(AB)$, $\mu(\tilde{A}B) = \mu(B) - \mu(AB)$ e, portanto,

$$\mu(A + B) = \mu(A) + \mu(B) - \mu(AB).$$

Para o caso de 3 conjuntos A, B, C , bastará aplicar esta fórmula duas vezes:

$$\begin{aligned}\mu(A + B + C) &= \mu[A + (B + C)] = \mu(A) + \mu(B + C) - \mu[A(B + C)] \\ &= \mu(A) + \mu(B) + \mu(C) - \mu(BC) - \mu(AB + AC).\end{aligned}$$

Como

$$\mu(AB + AC) = \mu(AB) + \mu(AC) - \mu(ABC),$$

virá, por último

$$\begin{aligned}\mu(A + B + C) &= \mu(A) + \mu(B) + \mu(C) - \mu(AB) - \\ &\quad - \mu(AC) - \mu(BC) + \mu(ABC).\end{aligned}$$

Consideremos agora, em geral, p conjuntos quaisquer, A_1, A_2, \dots, A_p , de \mathcal{R} . Por indução, chega-se à fórmula seguinte:

$$\begin{aligned}\mu\left(\sum_i^p A_i\right) &= \sum_i \mu(A_i) - \sum_{i < j} \mu(A_i A_j) + \sum_{i < j < k} \mu(A_i A_j A_k) - \dots \\ &\quad \dots + (-1)^{p-1} \mu(A_1 A_2 \dots A_p)\end{aligned}$$

que, no seu domínio inicial foi descoberta pelo matemático português DANIEL DA SILVA, que a indica no seu trabalho "*Propriedades gerais e resolução directa das congruências binómias: Introdução ao estudo da teoria dos números*".

É claro que estes resultados se aplicam, *mutatis mutandis*, ao caso dos atributos ou acontecimentos compatíveis.

11. Atributos quantitativos

Entre os atributos considerados em inquéritos estatísticos, há que distinguir duas espécies principais – os que são grandezas e os que o não são. Os primeiros são traduzíveis em números, resultados de medições, enquanto para os segundos não faz sentido falar de medida. Estes são chamados *atributos qualitativos ou qualidades*, e aqueles, *atributos quantitativos ou quantidades* (ou, ainda, grandezas).

Como sempre sucede, quando se trata de entes do mundo empírico, a fronteira entre as duas espécies é imprecisa. Há casos intermédios, dúbios, em que se fala de *avaliação* ou *correção* dum atributo (em vez de medição): tal é o caso da escala de Mohs relativa à dureza dos minerais, tal é, ainda, o caso das cotações que se atribuem aos alunos nos exames etc.

De resto, dada uma partição em atributos (finita), qualquer que ela seja, é sempre possível, com critério mais ou menos convencional, distinguir esses atributos por meio de números. Em particular, no caso da partição dicotómica $\{\alpha, \tilde{\alpha}\}$, formada por um atributo e pelo seu contrário, é natural designar $\tilde{\alpha}$ por 0 (ausência) e α por 1 (presença) ou por -1 e $+1$, etc.

Em correspondência às duas espécies de atributos, apresentam-se, naturalmente, duas modalidades de estatística: a *qualitativa* e a *quantitativa*. A primeira é, também, chamada *estatística de atributos* e a segunda, *estatística de variáveis*; mas estas designações não condizem com a nossa terminologia precedente.

Entre os atributos quantitativos (ou, melhor, entre as partições em tais atributos) há, ainda, que distinguir duas categorias: a dos que formam escalas contínuas (*variáveis contínuas*) e a dos que se dispõem conforme a sucessão de números naturais (*variáveis discretas*). Pertencem à primeira categoria, por exemplo, os comprimentos, as massas, as densidades, etc. Pertencem à segunda categoria por exemplo, os números de pétalas de uma flor, de grãos de uma espiga, etc., etc.

O caso das variáveis contínuas só poderá ser estudado devidamente nos capítulos seguintes, em que trataremos de universos infinitos. Todavia, numa primeira fase, que é a da recolha e classificação dos dados, trabalha-se ainda com universos finitos, tal como vamos ver.

Consideremos uma variável contínua x (variável real), atributo quantitativo definido numa dada população U finita: por exemplo, a *altura* dos indivíduos. Em vez de considerar a frequência de cada um dos valores de x em U (valores estes que são em número infinito), o que há a fazer, na prática, é agrupar esses valores em classes e considerar a frequência de cada uma dessas classes. Por outras palavras, o que há a fazer é uma *classificação* dos valores possíveis de x , ou seja, uma *partição* do conjunto dos números reais. De resto, sendo a população U finita, nem é necessário considerar toda a recta real, visto que, para além de certos limites, a frequência se torna nula. Por exemplo, sendo x a variável “altura” numa população de seres humanos e tomando para unidade o metro, não será necessário, geralmente, sair do intervalo $[0, 2]$, a não ser em casos excepcionais de gigantismo.

Quanto às células da partição, convém que sejam, evidentemente, intervalos, em número finito. Tudo se resume, portanto, na partição dum intervalo limitado $[a, b]$ em intervalos, por meio dum número finito de pontos, $a < x_1 < x_2 < \dots < x_{n-1} < b$. Todavia, para que se trate efectivamente duma partição, não podem os intervalos ser todos abertos ou todos fechados. Adoptaremos, em regra, intervalos fechados à esquerda e abertos à direita (excepto o último que poderá ser fechado). Assim, para o intervalo $[a, b]$, dividido pelos pontos x_1, x_2, \dots, x_n , teremos a partição

$$[a, x_1[, [x_1, x_2[, \dots, [x_{n-2}, x_{n-1}[, [x_{n-1}, b].$$

A tais intervalos dá-se, em Estatística, o nome de *intervalos de classe* ou simplesmente, *classes*. Haverá vantagem, é claro, em que todos esses intervalos tenham o mesmo comprimento e que os pontos de divisão correspondam a números inteiros da unidade escolhida ou dum submúltiplo decimal da mesma.

Posto isto, consideremos, de novo, o atributo quantitativo x relativo à população U . Supondo que se efectuou a medição dessa grandeza em cada um dos indivíduos, os resultados serão recolhidos numa primeira tabela, em que, à frente do número de ordem de cada indivíduo, se indica o valor correspondente de x , por exemplo, a altura desse indivíduo com o grau de aproximação adoptado.

Em seguida, feita a escolha dos intervalos de classe, as observações podem ser classificadas contando, para cada intervalo, o número total de indivíduos a que correspondem valores de x situados nesse intervalo (*frequência absoluta*). Organiza-se, deste modo, uma segunda tabela, que é já uma *tabela de frequências*. O universo U é, então, substituído pelo universo U^* das classes e a tabela de frequência define, manifestamente, uma *distribuição de frequência* sobre U^* (recorde-se o que, a este respeito, se disse no n.º 8). Para ilustrar estas considerações, citaremos um exemplo extraído do trabalho de Daniel Nagore, *Biometria, nociones sobre este método de investigacion en genetica*, Ministerio de Agricultura, Madrid, 1941. O exemplo refere-se a uma sub-variedade de trigo, denominado vulgarmente “Catalão compacto”, classificado cientificamente entre os *Triticum vulgare crythrospermum* (Percival) e obtido em campos de estudo por selecção genealógica. Desse trigo foi constituída uma população de 400 espigas, na qual se determinou directamente, em cada indivíduo, *o número de grãos e o comprimento da espiga* (em milímetros). Destes caracteres biométricos, que designaremos, respectivamente por g e c , se deduziram os valores duma outra grandeza: *o número de grãos por 10 cm de comprimento*. Esta grandeza, denominada *densidade de espiga*, é definida em função das primeiras pela fórmula

$$d = \frac{100 g}{c} .$$

Estes dados foram registados numa tabela (que não reproduzimos), na qual, à direita do número da ordem de cada espiga se indica, numa coluna, o valor de g , noutra, o valor de c , e na última, o valor de d , calculado este a partir dos primeiros a menos de 0,01. Em seguida, escolheram-se os intervalos de classe, e daquela primeira tabela deduziu-se a seguinte tabela de frequência:

TABELA N.º 3

Densidade	Frequência absoluta	Densidade	Frequência absoluta
$14,5 \leq d < 15$	1	$22,5 \leq d < 23$	52
$15 \leq d < 15,5$	1	$23 \leq d < 23,5$	41
$15,5 \leq d < 16$	1	$23,5 \leq d < 24$	19
$16 \leq d < 16,5$	1	$24 \leq d < 24,5$	14
$16,5 \leq d < 17$	1	$24,5 \leq d < 25$	6
$17 \leq d < 17,5$	4	$25 \leq d < 25,5$	12
$17,5 \leq d < 18$	5	$25,5 \leq d < 26$	2
$18 \leq d < 18,5$	5	$26 \leq d < 26,5$	2
$18,5 \leq d < 19$	6	$26,5 \leq d < 27$	1
$19 \leq d < 19,5$	14	$27 \leq d < 27,5$	1
$19,5 \leq d < 20$	6	$27,5 \leq d < 28$	1
$20 \leq d < 20,5$	44	$28 \leq d < 28,5$	0
$20,5 \leq d < 21$	31	$28,5 \leq d < 29$	1
$21 \leq d < 21,5$	34	$29 \leq d < 29,5$	0
$21,5 \leq d < 22$	50	$29,5 \leq d < 30$	1
$22 \leq d < 22,5$	43		

Ao olhar para esta tábua pode haver um equívoco, que é preciso evitar. Por exemplo, à direita da função proposicional $19 \leq d < 19,5$ escreveu-se 14. Ora, isto não significa que, para todo o valor de d pertencente àquele intervalo, a frequência é 14, mas, sim, que é esta a frequência do atributo “*densidade maior ou igual a 19 e menor que 19,5*” na população considerada. O número 14 corresponde, pois, ao intervalo $[19; 19, 5[$ e não a cada ponto deste intervalo. Por outras palavras: é uma função *de conjunto* – uma *distribuição* – e não uma função de *ponto* (isto é, uma função da variável d). Estas observações vêm, assim, esclarecer o conceito de distribuição introduzido no n.º 9.

No trabalho a que nos referimos faz-se, depois, para determinado efeito, uma condensação dos dados, substituindo os intervalos de comprimento 0,5 por intervalos de comprimento 1 e aplicando a propriedade aditiva das distribuições. Obtém-se, deste modo, a seguinte tabela:

TABELA N.º 4

Densidade	Frequência absoluta	Densidade	Frequência absoluta
$14 \leq d < 15$	1	$22 \leq d < 23$	95
$15 \leq d < 16$	2	$23 \leq d < 24$	60
$16 \leq d < 17$	2	$24 \leq d < 25$	20
$17 \leq d < 18$	9	$25 \leq d < 26$	14
$18 \leq d < 19$	11	$26 \leq d < 27$	3
$19 \leq d < 20$	20	$27 \leq d < 28$	2
$20 \leq d < 21$	75	$28 \leq d < 29$	1
$21 \leq d < 22$	84	$29 \leq d < 30$	1

Note-se que muitas vezes, para se designar cada intervalo de classe, se indica o ponto médio desse intervalo. Assim, na tábua anterior, o intervalo $[14, 15[$ seria designado como a classe 14,5, etc. Nestes casos, convém escolher os intervalos de modo que os pontos médios sejam números inteiros de unidades ou de submúltiplos decimais da unidade. Recordemos que é este o critério usado no arredondamento das médias de notas de alunos; assim, quando se diz que a média é de 12 valores, está-se a indicar, resumidamente, que a média é um número x tal que $11,5 \leq x < 12,5$; analogamente, dizendo que a média é de 14,3, pretende-se dizer que a média verifica a condição $14,35 \leq x < 14,36$, etc.

O mesmo critério é ainda usado pelas companhias de seguros na atribuição da idade para o pagamento do prémio, em seguros de vida.

12. Representação gráfica das distribuições: histogramas e polígonos de frequência

Para ter uma visão mais directa e sugestiva duma distribuição de frequência, procede-se à sua representação gráfica, começando por fixar no plano um sistema de eixos coordenados rectangulares, com escalas geralmente diferentes. Uma vez marcados no eixo das abcissas os sucessivos intervalos de classe (não sendo indispensável que o ponto de abcissa 0 coincida com o cruzamento dos eixos), a representação gráfica pode, na prática, ser feita de dois modos diversos:

1.º processo – Sobre o segmento representativo de cada classe, constrói-se um rectângulo de altura (positiva) correspondente à frequência dessa classe. A reunião dos rectângulos assim obtidos é chamada o *histograma* ou *diagrama de colunas* de distribuição considerada. (Na Fig. 1 é dado o histograma correspondente à Tabela n.º 4).

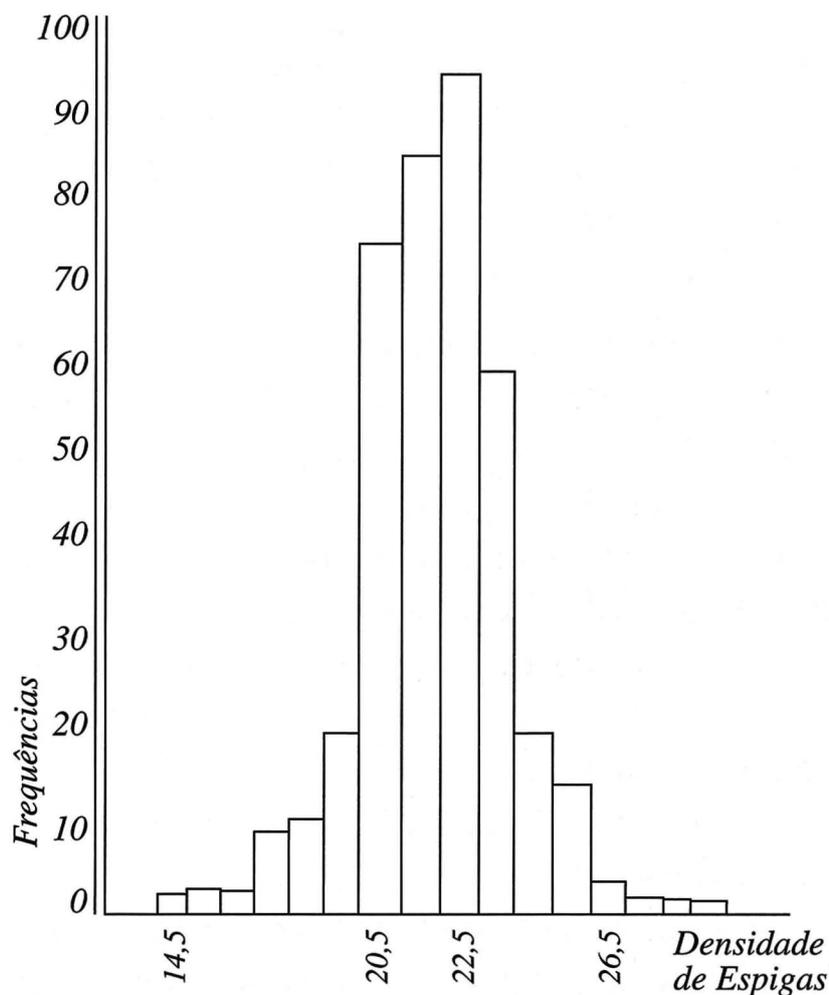


Fig. 1

É claro que as áreas dos diferentes rectângulos são proporcionais às frequências das respectivas classes, de modo que a área do histograma representa, na mesma escala, o número total de indivíduos. Em particular, o quociente da área dum rectângulo (ou duma reunião de rectângulos) pela área do histograma dá a frequência relativa da respectiva classe (ou da respectiva reunião de classes).

2.º processo – Pelo ponto médio de cada intervalo de classe conduz-se uma perpendicular ao eixo das abcissas, sobre a qual se marca o ponto de ordenada igual à frequência dessa classe. Traçam-se, depois, segmentos de recta, unindo entre si os pontos consecutivos assim obtidos e unindo o primeiro e o último destes pontos, respectivamente, com o primeiro e o último extremo dos intervalos marcados. Obtém-se, deste modo, uma linha poligonal, à qual se dá o nome de *polígono de frequência* ou *polígono de Johanssen* da distribuição considerada. Na Fig. 2 são dados os polígonos de frequência correspondentes à Tabela n.º 3 (a traço fino) e à Tabela n.º 4 (a traço cheio).

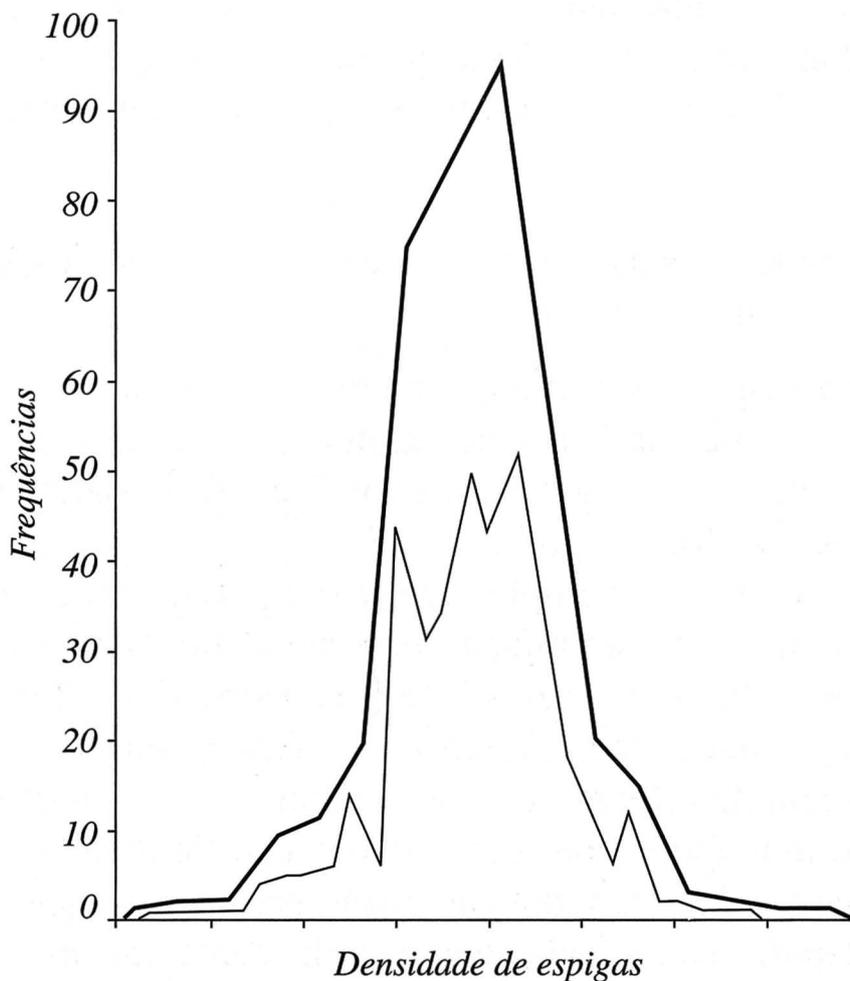


Fig. 2

Observe-se que a primeira poligonal é mais sinuosa, mais irregular do que a segunda.

Dum modo geral, não convém que os intervalos de classe sejam demasiado grandes, pois que, então, se obtém uma informação grosseira da distribuição. Mas também não convém que os intervalos sejam demasiado pequenos, porque, então, se tornam mais sensíveis às influências individuais, rompendo-se aquela *regularidade estatística*, aquele possível esboço duma lei simples que se colhe da visão dum conjunto, mas que se desvanece ao aproximarmos dos indivíduos, dos casos isolados.

É, ainda, manifesto que, uma tal regularidade estatística será tanto maior quanto maior for o número de indivíduos, podendo, então, reduzir-se cada vez mais a amplitude dos intervalos. Deste modo, a poligonal acabará por se confundir, sensivelmente, com uma curva. E, assim, a intuição nos vai aproximando do conceito de distribuição duma variável contínua, que será estudado em rigor a partir do capítulo seguinte. Só depois disso poderemos tirar conclusões dos exemplos agora citados.

13. Independência e associação de atributos. Distribuições de duas ou mais variáveis

Quando se pretende averiguar em que medida dois atributos ou dois fenómenos andam ligados, numa população ou num tipo de experiências, o que há a fazer é um inquérito estatístico de que vamos indicar as formas preliminares.

Suponhamos, por exemplo, que se trata de saber se, entre seres humanos, o atributo “ser míope” está, de qualquer modo, associado ao atributo “ter olhos azuis”. O que, ocorre desde logo, é indagar, numa população tão numerosa e variada quanto possível, qual a percentagem de míopes: 1.º – entre indivíduos com olhos azuis; 2.º – entre indivíduos com olhos não azuis. Se as duas percentagens são sensivelmente iguais, há razão para pensar que os atributos considerados são *independentes*; se as percentagens se afastam consideravelmente uma da outra, seremos inclinados a admitir que os dois atributos estão *associados* ou *correlacionados*: em sentido *positivo*, se a primeira percentagem é maior que a segunda, em sentido *negativo*, no caso oposto.

Analogamente se procederia para averiguar, por exemplo, se o fumar está ou não ligado com doenças de estômago, se um insecti-

cida é ou não eficaz no tratamento de certas plantas, etc., etc. Nestes exemplos, começa a desenhar-se a hipótese duma relação de causalidade entre dois fenómenos.

Estas considerações podem ser teorizadas do seguinte modo. Sejam α , β dois atributos definidos num dado universo U , composto de N indivíduos. Para indicar as frequências destes atributos e dos seus contrários, pode fazer-se uso duma tabela de duas entradas, do seguinte tipo:

	β	$\tilde{\beta}$	Total
α	$(\alpha\beta)$	$(\alpha\tilde{\beta})$	(α)
$\tilde{\alpha}$	$(\tilde{\alpha}\beta)$	$(\tilde{\alpha}\tilde{\beta})$	$(\tilde{\alpha})$
Total	(β)	$(\tilde{\beta})$	N

Para brevidade de linguagem diremos “os α ” em vez de “os indivíduos com o atributo α ”, e, analogamente, para β .

A proporção dos α em U é $(\alpha)/N$ (frequência relativa de α). A proporção dos α entre os β será $(\alpha\beta)/(\beta)$; chamar-lhe-emos *frequência condicional de α relativa a β* e representá-la-emos por $\text{fr}(\alpha|\beta)$. Ter-se-á, pois, por definição:

$$(13.1) \quad \text{fr}(\alpha|\beta) = \frac{(\alpha\beta)}{(\beta)} = \frac{\text{fr}(\alpha\beta)}{\text{fr}(\beta)} .$$

O atributo α dir-se-á *independente de β em U* se a proporção dos α entre os β é igual à proporção dos α entre os não β , isto é, se

$$\text{fr}(\alpha|\beta) = \text{fr}(\alpha|\tilde{\beta}), \text{ ou seja, se } \frac{(\alpha\beta)}{(\beta)} = \frac{(\alpha\tilde{\beta})}{(\tilde{\beta})} .$$

Mas, recordemos a seguinte propriedade das fracções:

Se $\frac{a}{b} = \frac{c}{d}$, então $\frac{a+c}{b+d} = \frac{a}{b}$, e reciprocamente.⁽¹⁾

Ter-se-á, portanto:

Se $\frac{(\alpha\beta)}{(\tilde{\beta})} = \frac{(\alpha\tilde{\beta})}{\beta}$, então $\frac{(\alpha\beta) + (\alpha\tilde{\beta})}{(\beta) + (\tilde{\beta})} = \frac{(\alpha\beta)}{(\beta)}$, e reciprocamente.

Mas $(\alpha\beta) + (\alpha\tilde{\beta}) = (\alpha)$ e $(\beta) + (\tilde{\beta}) = N$. Portanto, dizer que α é independente de β equivale a dizer que

$$(13.2) \quad \frac{(\alpha\beta)}{(\beta)} = \frac{(\alpha)}{N},$$

isto é, que a proporção dos α entre os β é igual à proporção dos α na população total. A fórmula precedente, pode ainda, escrever-se

$$\text{fr}(\alpha|\beta) = \text{fr}(\alpha).$$

Logo, dizer que α é independente de β equivale a dizer que a frequência condicional de α relativa a β é igual à frequência relativa de α (em U).

Uma outra maneira de escrever (13.2) é a seguinte:

$$(13.3) \quad (\alpha\beta) = \frac{(\alpha)(\beta)}{N},$$

fórmula esta que, pela sua simetria em α e β nos mostra imediatamente que:

(1) – Com efeito, se pusermos $\frac{a}{b} = \frac{c}{d} = k$, virá $a = bk$, $c = dk$, $a + c = (b + d)k$ e, portanto,

$$\frac{a+c}{b+d} = k = \frac{a}{b}.$$

Se α é independente de β , também β é independente de α .

Diremos, então, simplesmente, que os atributos α , β são *independentes* (em U).

Mas a fórmula (13.3) ainda se pode escrever com o aspecto

$$\frac{(\alpha\beta)}{N} = \frac{(\alpha)}{N} \cdot \frac{(\beta)}{N}, \text{ ou seja, } \text{fr}(\alpha\beta) = \text{fr}(\alpha) \cdot \text{fr}(\beta),$$

donde o

TEOREMA DO PRODUTO. *Se dois atributos α , β são independentes, a frequência relativa de $\alpha\beta$ é igual ao produto das frequências relativas de α e de β . Reciprocamente, se esta condição se verifica, os dois atributos são independentes.*

Mais geralmente, da definição (13.1) de frequência condicional, deduz-se

$$\text{fr}(\alpha\beta) = \text{fr}(\beta) \cdot \text{fr}(\alpha|\beta)$$

ou, ainda, trocando os papéis de α e de β

$$\text{fr}(\alpha\beta) = \text{fr}(\alpha) \cdot \text{fr}(\beta|\alpha),$$

o que se exprime dizendo que:

A frequência relativa de $\alpha\beta$ é igual ao produto da frequência relativa de α pela frequência condicional de β relativa a α ou (vice-versa).

É claro que o teorema do produto se obtém como caso particular desta última proposição, pois que se tem $\text{fr}(\beta|\alpha) = \text{fr}(\beta)$, se e só se, α e β são independentes.

Suponhamos, agora, que α e β não são independentes em U . Então, de duas uma:

- 1) ou $\frac{(\alpha\beta)}{(\beta)} > \frac{(\alpha)}{N}$ (isto é, a proporção dos α é maior entre os β que no universo inteiro),

2) ou $\frac{(\alpha\beta)}{(\beta)} < \frac{(\alpha)}{N}$ (isto é, a proporção dos α é menor entre os β que no universo inteiro).

No primeiro caso, tem-se, mais simetricamente,

$$(13.4) \quad (\alpha\beta) > \frac{(\alpha)(\beta)}{N} \text{ ou } \text{fr}(\alpha\beta) > \text{fr}(\alpha) \text{fr}(\beta)$$

e diz-se que os atributos α , β estão *associados positivamente*, ou, simplesmente, *associados* (em U).

No segundo caso, também se pode escrever

$$(13.5) \quad (\alpha\beta) < \frac{(\alpha)(\beta)}{N} \text{ ou } \text{fr}(\alpha\beta) < \text{fr}(\alpha) \text{fr}(\beta),$$

e diz-se que α , β estão *associados negativamente ou desassociados* (em U).

Os casos extremos são aqueles em que a associação ou a desassociação é *completa*. Diremos que os atributos α , β estão *completamente associados* (em U) quando um deles, pelo menos, implica o outro, isto é, quando se tem:

$$\text{ou } \alpha \subset \beta \text{ (todo o } \alpha \text{ é um } \beta) \text{ ou } \beta \subset \alpha \text{ (todo o } \beta \text{ é um } \alpha),$$

podendo, em particular, verificar-se as duas hipóteses (será, então, α equivalente a β). É claro, que se tem $\alpha \subset \beta$ ou $\beta \subset \alpha$ conforme for $(\alpha) = (\alpha\beta)$ ou $(\beta) = (\alpha\beta)$. Portanto, dizer que os atributos α , β são completamente associados equivale a dizer que $(\alpha\beta)$ é igual ao menor dos números (α) , (β) .

Os atributos α , β dizem-se *completamente desassociados* em U quando se verifica uma, pelo menos, das seguintes condições:

$$\alpha\beta = \emptyset, \quad \tilde{\alpha}\tilde{\beta} = \emptyset,$$

isto, é quando são incompatíveis os atributos α , β (nenhum α é β) ou os seus contrários (todo o $\tilde{\alpha}$ é β).

Deste modo, é natural considerar a diferença entre $(\alpha\beta)$ e o valor $(\alpha)(\beta)/N$ (chamado *valor de independência*) como um índice da

intensidade da associação ou desassociação. É costume designar por δ esta diferença e por $(\alpha\beta)_0$ o valor de independência. Tem-se, pois:

$$(\alpha\beta)_0 = \frac{(\alpha)(\beta)}{N}, \quad \delta = (\alpha\beta) - (\alpha\beta)_0.$$

No entanto, para ter uma ideia comparativa dos graus de associação, é necessário substituir δ por um índice *relativo*. Entre os vários índices que têm sido propostos para este efeito, o mais simples é o seguinte, chamado *coeficiente de associação*:

$$Q = \frac{(\alpha\beta)(\tilde{\alpha}\tilde{\beta}) - (\alpha\tilde{\beta})(\tilde{\alpha}\beta)}{(\alpha\beta)(\tilde{\alpha}\tilde{\beta}) + (\alpha\tilde{\beta})(\tilde{\alpha}\beta)} = \frac{N\delta}{(\alpha\beta)(\tilde{\alpha}\tilde{\beta}) + (\alpha\tilde{\beta})(\tilde{\alpha}\beta)}.$$

É claro que se tem $Q=0$ se e só se os atributos são independentes (o que equivale a ser $\delta=0$). Os valores extremos de Q são $+1$ e -1 :

Tem-se $Q=+1$, se e só se os atributos estão completamente associados, pois que, só nesse caso um dos factores $(\alpha\tilde{\beta})$, $(\tilde{\alpha}\beta)$ se anula.

Tem-se $Q=-1$, se e só se os atributos estão completamente desassociados, pois que, só então, se tem $(\alpha\beta)=0$ ou $(\tilde{\alpha}\tilde{\beta})=0$.

Nos dois números seguintes, estudaremos um outro índice de associação, de grande interesse estatístico.

É fácil ver, ainda, que a condição (13.4) de associação positiva (simétrica em α e β) é equivalente a qualquer das seguintes (assimétricas):

$$(13.6) \quad \frac{(\alpha\beta)}{(\beta)} > \frac{(\alpha\tilde{\beta})}{(\tilde{\beta})}$$

$$(13.7) \quad \frac{(\alpha\beta)}{(\alpha)} > \frac{(\tilde{\alpha}\beta)}{(\tilde{\alpha})}$$

e, analogamente, para a associação negativa. Ora, na prática, qualquer destas comparações é, geralmente, mais elucidativa do que a

primeira para dar uma ideia do tipo de associação. Segundo a disposição adoptada na tabela-tipo atrás considerada, a fórmula (13.6) corresponde a uma *comparação por colunas* e a fórmula (13.7) a uma *comparação por linhas*. Qual das duas deve ser preferida? Isso depende da hipótese que se tiver em mente ao pôr o problema. Se a relação suposta entre os atributos é assimétrica, próxima duma das implicações $\alpha \subset \beta$ ou $\beta \subset \alpha$, deve preferir-se a comparação por linhas ou por colunas, conforme for α ou β que se presume no papel de *antecedente ou causa*. Se, porém, a relação suposta for de reciprocidade, vizinha da equivalência $\alpha \equiv \beta$, é preferível a comparação entre $(\alpha\beta)$ e $(\alpha\beta)_0$, como se faz nas fórmulas (13.4) e (13.5).

Consideremos o seguinte exemplo indicado por YULE e KENDALL (obra citada, p. 40). Trata-se de investigar a *associação entre inoculação contra a cólera e a isenção do ataque*; obtiveram-se os seguintes resultados:

TABELA N.º 5

	Não atacados	Atacados	Total
Inoculados	276	3	279
Não inoculados	473	66	539
Total	749	69	818

Pressupõe-se, é claro, uma relação assimétrica, com α no papel de causa. Convirá, pois, fazer uma comparação por linhas:

1) – Percentagem de inoculados que não foram atacados:

$$\frac{(\alpha\beta)}{(\alpha)} = \frac{276}{279} = 98,9\%$$

2) – Percentagem de não inoculados que não foram atacados:

$$\frac{(\tilde{\alpha}\beta)}{(\tilde{\alpha})} = \frac{473}{539} = 87,9\%$$

São, talvez, mais sugestivos (embora equivalentes) as proporções complementares:

1') – Percentagem de inoculados que foram atacados: 1,1

2') – Percentagem de não inoculados que foram atacados: 12,2

Em qualquer das formas é visível a associação positiva entre a *inoculação* e a *isenção do ataque* (ou associação negativa entre *inoculação* e *ataque*).

A comparação por colunas indica, igualmente, uma associação apreciável, mas é claro que não responde directamente à questão implícita no inquérito:

1'') – Percentagem de não atacados que foram inoculados: 36,8

2'') – Percentagem de atacados que foram inoculados: 4,3

14. Associação e independência de partições múltiplas. Tábuas de contingência

Consideremos, agora, mais geralmente, duas partições em atributos

$$\{\alpha_1, \alpha_2, \dots, \alpha_p\} \{\beta_1, \beta_2, \dots, \beta_q\}$$

num mesmo universo U com N indivíduos. Seja α a variável que toma por valores $\alpha_1, \alpha_2, \dots, \alpha_p$ e seja β a variável que toma por valores $\beta_1, \beta_2, \dots, \beta_q$. As duas partições, cruzadas sobre U , determinam uma nova partição, mais fina, constituída pelos $p \cdot q$ atributos $\alpha_i \beta_k$ ($i = 1, 2, \dots, p; k = 1, 2, \dots, q$); em particular, alguns destes atributos, podem ser impossíveis, o que reduz o seu número efectivo. As frequências dos atributos α_i, β_k podem ser indicadas numa tabela de duas entradas do seguinte tipo:

$\alpha \backslash \beta$	β_1	β_2	β_q	Totais
α_1	$(\alpha_1 \beta_1)$	$(\alpha_1 \beta_2)$	$(\alpha_1 \beta_q)$	(α_1)
α_2	$(\alpha_2 \beta_1)$	$(\alpha_2 \beta_2)$	$(\alpha_2 \beta_q)$	(α_2)
.....
α_p	$(\alpha_p \beta_1)$	$(\alpha_p \beta_2)$	$(\alpha_p \beta_p)$	(α_p)
Totais	(β_1)	(β_2)	(β_q)	N

Trata-se, manifestamente, duma generalização do tipo de tábuas consideradas no número precedente, para o caso de duas partições dicotómicas ($p = q = 0$). Dá-se-lhe o nome de *tábuas de contingência*. Como exemplo, indicaremos, a seguinte tabela apresentado por YULE e KENDALL (obra citada, p. 66), relativa às 2 variáveis “cor dos olhos” e “cor dos cabelos”, numa população constituída por 6.800 habitantes masculinos de Baden.

TABELA N.º 6

Cor dos olhos \ Cor dos cabelos	Cor dos cabelos				Total
	Louro	Castanho	Preto	Ruivo	
Azul	1.768	807	189	47	2.811
Cinzentos ou verde	946	1.387	746	53	3.132
Castanho	115	438	288	16	857
Total	2.829	2.632	1.223	116	6.800

Aqui, como se vê, a variável “cor dos olhos” é susceptível de três valores (“azul”, “cinzentos ou verde” e “castanho”), enquanto a variável “cor dos cabelos” é susceptível de quatro valores (“louro”, “castanho”, “preto” e “ruivo”).

Dum modo geral, diremos que a variável α é *independente* da variável β (em U), quando cada um dos valores de α é independente de cada um dos valores de β , no sentido precisado no número precedente; isto é, quando se tiver

$$(\alpha_i \beta_k) = \frac{(\alpha_i) (\beta_k)}{N}, \text{ ou seja, } fr(\alpha_i \beta_k) = fr(\alpha_i) fr(\beta_k)$$

para todos os valores possíveis de i e de k . É claro, que se α é independente de β , também β é independente de α : diremos, então, simplesmente, que *as variáveis, α , β são independentes em U* .

Se tal não acontecer, poremos, ainda,

$$(\alpha_i \beta_k)_0 = \frac{(\alpha_i) (\beta_k)}{N}, \text{ para } i = 1, 2, \dots, p, \quad k = 1, 2, \dots, q,$$

e chamaremos aos números $(\alpha_i \beta_k)_0$ *valores de independência*. Por sua vez, chamaremos *discrepâncias* às diferenças entre os valores observados e os valores de independência, isto é, aos números

$$\delta_{ik} = (\alpha_i \beta_k) - (\alpha_i \beta_k)_0, \text{ para } i = 1, 2, \dots, p; k = 1, 2, \dots, q.$$

Desde já, convém salientar que *as discrepâncias não são algebricamente independentes*. Tem-se, com efeito, para cada valor de i :

$$\begin{aligned} \delta_{i1} + \delta_{i2} + \dots + \delta_{iq} &= \sum_{k=1}^q \left[(\alpha_i \beta_k) - \frac{(\alpha_i) (\beta_k)}{N} \right] \\ &= \sum_{k=1}^q (\alpha_i \beta_k) - \sum_{k=1}^q \frac{(\alpha_i) (\beta_k)}{N}. \end{aligned}$$

Mas

$$\sum_{k=1}^q (\alpha_i \beta_k) = (\alpha_i \beta_1) + (\alpha_i \beta_2) + \dots + (\alpha_i \beta_q) = (\alpha_i)$$

e

$$\sum_{k=1}^q \frac{(\alpha_i) (\beta_k)}{N} = \frac{(\alpha_i)}{N} \sum_{k=1}^q (\beta_k) = \frac{(\alpha_i)}{N} \cdot N = (\alpha_i).$$

Logo

$$\delta_{i1} + \delta_{i2} + \dots + \delta_{iq} = 0, \text{ para } i = 1, 2, \dots, p;$$

isto é, *a soma das discrepâncias em cada linha é igual a zero*.

Analogamente se conclui que *é nula a soma das discrepâncias em cada coluna*, isto é, que:

$$\delta_{1k} + \delta_{2k} + \dots + \delta_{pk} = 0, \text{ para } k = 1, 2, \dots, q.$$

O número total dos δ_{ik} é pq . Mas é claro que basta conhecer os valores das discrepâncias em $p-1$ linhas e em $q-1$ colunas, pois

os valores correspondentes à linha e à coluna restantes se determinam por meio das relações estabelecidas. É fácil ver, agora, que não existem outras relações obrigatórias entre as discrepâncias; o número total dos δ_{ik} independentes é, pois, $(p-1)(q-1)$. Dá-se-lhe o nome de *número de graus de liberdade* da tábua em questão.

O conjunto das discrepâncias dá uma ideia do grau de associação entre os diferentes valores de α e de β . Mas convém resumir esses dados num índice único, de maneira tão simples e elucidativa quanto possível. A soma dos δ_{ik} não serve, visto ser nula, como vimos. Um índice independente dos sinais dos δ_{ik} é o que se obtém *dividindo o quadrado de cada discrepância pelo correspondente valor de independência e somando os resultados obtidos*. Este índice que se representa por χ^2 (*qui quadrado*) e se chama *contingência quadrática*, apresenta, como veremos, um grande interesse em ciências experimentais. Será, pois, por definição,

$$\chi^2 = \sum \frac{\delta_{ik}^2}{(\alpha_i \beta_k)_0}$$

(Segundo um hábito corrente em Estatística, omitiremos as indicações relativas aos índices nos somatórios, quando não houver perigo de confusão. Note-se que muitos autores usam a letra S em vez de Σ como símbolo de somatório).

Na prática, procede-se do seguinte modo para a determinação dos δ_{ik} e do χ^2 . Começa-se por deduzir da tábua inicial uma segunda tábua com os valores de independência nos lugares dos valores observados. Em seguida, constroi-se a tabela das discrepâncias por simples subtração entre os valores da primeira e os correspondentes da segunda. Finalmente, calcula-se o χ^2 segundo a definição deste índice.

No cálculo dos valores de independência convém, ainda, observar certos preceitos práticos. Visto que se tem

$$(\alpha_i \beta_k)_0 = \frac{(\alpha_i)(\beta_k)}{N} = (\alpha_i) \text{ fr}(\beta_k) = (\beta_k) \text{ fr}(\alpha_i),$$

pode-se começar por calcular, por exemplo, a frequência relativa de cada β e multiplicá-la, depois, pelas frequências absolutas de todos os α (ou vice-versa); sem esquecer que a soma das frequências relativas de todos os β ou de todos os α é 1, isto é:

$$\text{fr}(\alpha_1) + \text{fr}(\alpha_2) + \dots + \text{fr}(\alpha_p) = 1,$$

$$\text{fr}(\beta_1) + \text{fr}(\beta_2) + \dots + \text{fr}(\beta_q) = 1;$$

portanto, uma vez calculadas todas as frequências relativas menos uma, a restante obtém-se por subtração da unidade. Daremos exemplos de cálculo do χ^2 no n.º 16, em que abordaremos o estudo interpretativo deste índice.

15. Associações parciais de atributos. Independência de atributos no caso em que o seu número é superior a dois

Suponhamos agora, mais geralmente, que no universo U são dadas várias partições em atributos. Podemos limitar-nos ao caso em que as partições são dicotómicas. Consideremos, pois, as partições determinadas em U por vários atributos

$$\alpha, \beta, \gamma, \delta, \dots$$

e pelos seus contrários (em número finito).

Além da frequência condicional de α a respeito de β , podemos, agora, considerar:

1) – A frequência condicional de γ a respeito de α e β :

$$\text{fr}(\gamma|\alpha\beta) = \frac{(\alpha\beta\gamma)}{(\alpha\beta)} = \frac{\text{fr}(\alpha\beta\gamma)}{\text{fr}(\alpha\beta)};$$

2) – A frequência condicional de δ a respeito de α , β e γ :

$$\text{fr}(\delta|\alpha\beta\gamma) = \frac{(\alpha\beta\gamma\delta)}{(\alpha\beta\gamma)} = \frac{\text{fr}(\alpha\beta\gamma\delta)}{\text{fr}(\alpha\beta\gamma)};$$

etc., etc.

Destas definições resulta, desde logo, que se tem, em geral, a *fórmula do produto*:

$$(15.1) \quad \text{fr}(\alpha\beta\gamma\delta\dots) = \text{fr}(\alpha) \text{fr}(\beta|\alpha) \text{fr}(\gamma|\alpha\beta) \text{fr}(\delta|\alpha\beta\gamma)\dots,$$

fórmula esta que continua a ser válida permutando entre si, de todos os modos possíveis, as letras α , β , γ , δ , ...

Os atributos β , γ dizem-se *parcialmente independentes a respeito de α* , se

$$\text{fr}(\gamma|\alpha\beta) = \text{fr}(\gamma|\alpha)$$

isto é, se

$$\frac{(\alpha\beta\gamma)}{(\alpha\beta)} = \frac{(\alpha\gamma)}{(\alpha)},$$

ou, ainda, o que é equivalente, se:

$$(\alpha\beta\gamma) = \frac{(\alpha\beta) (\alpha\gamma)}{(\alpha)}.$$

Caso contrário, os atributos β , γ dir-se-ão *parcialmente associados a respeito de α* – *positivamente ou negativamente*, conforme for⁽¹⁾

$$(\alpha\beta\gamma) \gtrless \frac{(\alpha\beta) (\alpha\gamma)}{(\alpha)}.$$

O caso da independência parcial também se pode conceber como associação parcial nula.

Analogamente, se define associação parcial de 3 ou mais atributos a respeito de um ou mais atributos.

(1) – O conceito de associação parcial a respeito de α corresponde, assim, a substituir o universo inicial U pelo universo dos α_i , isto é, pelo conjunto dos indivíduos com o atributo α .

Note-se que, para 3 atributos α , β , γ , se têm ao todo 3 associações totais (de α com β , de α com γ e de β com γ) e 6 associações parciais (de α com β a respeito de γ , de α com γ a respeito de β , etc.).

Os atributos α , β , γ , δ , ..., dizem-se *independentes* em U , se são nulas todas as possíveis associações (totais e parciais) entre estes atributos e seus contrários. Da fórmula (15.1) deduz-se a seguinte generalização do

TEOREMA DO PRODUTO. *Se os atributos α , β , γ , ..., são independentes, tem-se*

$$(15.2) \quad \text{fr}(\alpha\beta\gamma \dots) = \text{fr}(\alpha) \text{fr}(\beta) \text{fr}(\gamma) \dots$$

Todavia, pode reconhecer-se com exemplos que a recíproca não é, agora, verdadeira.

Também importa salientar que o facto de os atributos serem independentes dois a dois não implica a fórmula (15.2), ao contrário do que poderia pensar-se à primeira vista.

É, porém, verdadeiro o seguinte teorema, que contém o anterior como caso particular:

TEOREMA. *Para que os atributos α , β , γ , ... sejam independentes, é necessário e suficiente que sejam válidas, não só a fórmula*

$$\text{fr}(\alpha\beta\gamma \dots) = \text{fr}(\alpha) \text{fr}(\beta) \text{fr}(\gamma) \dots$$

como todas as que se deduzem desta, substituindo um ou mais dos atributos α , β , γ , ..., pelos seus contrários.

A condição é, evidentemente, necessária: se os atributos dados são independentes, a frequência condicional de cada atributo (ou do seu contrário) a respeito de um ou mais dos restantes atributos (ou dos seus contrários) é igual à frequência relativa do primeiro no universo. Então, a fórmula (15.1) simplifica-se tomando o aspecto (15.2), pois que

$$\text{fr}(\beta|\alpha) = \text{fr}(\beta), \quad \text{fr}(\gamma|\alpha\beta) = \text{fr}(\gamma|\alpha) = \text{fr}(\gamma), \text{ etc.};$$

e o mesmo se pode dizer da fórmula que se deduz de (15.1), substituindo um ou mais dos atributos pelos seus contrários.

Demonstremos agora que a condição é suficiente. Limitar-nos-emos ao caso de 3 atributos α , β , γ , pois que, no caso geral, a demonstração é análoga. Por se ter

$$\alpha\beta = \alpha\beta(\gamma + \tilde{\gamma}) = \alpha\beta\gamma + \alpha\beta\tilde{\gamma},$$

será

$$\text{fr}(\alpha\beta) = \text{fr}(\alpha\beta\gamma) + \text{fr}(\alpha\beta\tilde{\gamma}).$$

Então, se forem verdadeiras as igualdades

$$\text{fr}(\alpha\beta\gamma) = \text{fr}(\alpha) \cdot \text{fr}(\beta) \cdot \text{fr}(\gamma), \quad \text{fr}(\alpha\beta\tilde{\gamma}) = \text{fr}(\alpha) \cdot \text{fr}(\beta) \cdot \text{fr}(\tilde{\gamma}),$$

virá

$$\text{fr}(\alpha\beta) = \text{fr}(\alpha) \cdot \text{fr}(\beta) \cdot \text{fr}(\gamma) + \text{fr}(\alpha) \cdot \text{fr}(\beta) \cdot \text{fr}(\tilde{\gamma}) = \text{fr}(\alpha) \cdot \text{fr}(\beta),$$

donde

$$\text{fr}(\alpha|\beta) = \frac{\text{fr}(\alpha\beta)}{\text{fr}(\alpha)} = \text{fr}(\beta),$$

$$\text{fr}(\gamma|\alpha\beta) = \frac{\text{fr}(\alpha\beta\gamma)}{\text{fr}(\alpha\beta)} = \text{fr}(\gamma),$$

e, analogamente, para os restantes casos.

Para complementos e exemplos sobre este ponto, veja-se a citada obra de YULE e KENDALL.

Note-se, ainda, que todas as considerações deste número e dos dois precedentes se estendem, mutatis mutandis, ao caso dos acontecimentos.

16. Interpretação duma tábua de contingência. Testes de significância

Vamos, agora, abordar o problema da interpretação duma tábua de contingência, aproximando-nos do conceito de probabilidade.

Como primeiro exemplo, consideremos o seguinte, extraído da obra de FINNEY citada na advertência prévia:

Trata-se duma experiência hipotética relativa à protecção de animais contra determinada doença. Supõe-se que o experimentador dispõe de 300 animais, e que decide submeter 100 destes a um certo tratamento, reservando os outros 200 para controle. Supõe-se, além disso, que são os seguintes os resultados obtidos:

TABELA N.º 7

	Sobreviventes	Mortos	Total	Mortos %
Não tratados	152	48	200	24
Tratados	88	12	100	12
Total	240	60	300	20

É claro que se pretende averiguar em que medida estes resultados autorizam a admitir uma real influência do tratamento na doença.

Para isso, começa-se por adoptar uma atitude psicológica, que faz lembrar o método de demonstração por redução ao absurdo usado em Matemática, e que consiste em admitir precisamente o contrário do que se pretende estabelecer: *suponhamos que o tratamento não influi na doença nem num sentido nem noutro*. Nisto consiste a chamada *hipótese nula*. Admitida esta hipótese, seria natural esperar que os resultados obtidos fossem aqueles a que, no n.º 13, chamámos valores de independência e que constam da seguinte tabela:

TABELA N.º 8

	Sobreviventes	Mortos	Total	Mortos %
Não tratados	160	40	200	20
Tratados	80	20	100	20
Total	240	60	300	20

Observe-se como foi construída esta tabela:

De acordo com as instruções práticas do n.º 14, calculou-se a frequência relativa dos mortos, ou seja,

$$\frac{60}{300} = 0,2 = 20\%;$$

multiplicou-se depois, este resultado, pelos totais de animais não tratados e de animais tratados, obtendo-se

$$0,2 \times 200 = 40, \quad 0,2 \times 100 = 20.$$

As frequências dos sobreviventes entre animais não tratados e animais tratados são, respectivamente,

$$160 = 200 - 40, \quad 80 = 100 - 20.$$

Posto isto, subtraindo os valores da tabela n.º 8 (*valores de independência* ou *valores esperados* de acordo com a hipótese nula) dos valores correspondentes da tabela n.º 7 (*valores observados*), obtém-se a tabela das *discrepâncias*.

TABELA N.º 9

	Sobreviventes	Mortos	Total
Não tratados	- 8	+ 8	0
Tratados	+ 8	- 8	0
Total	0	0	0

Verificam-se, é claro, as relações algébricas entre os δ_{ik} indicadas no n.º 14 (o número de graus de liberdade é aqui igual a 1).

Mas pergunta-se agora:

Estão estes resultados em contradição com a hipótese nula? A resposta exige reflexão.

É claro que o facto de o tratamento não influir na doença não obriga, de nenhum modo, a uma rígida confirmação dos valores de independência: *pode haver discrepâncias não nulas devidas ao acaso; só se as discrepâncias excederem um “certo limite” haverá motivo para refutar a hipótese nula*. Tudo está, portanto, em avaliar esse limite para além do qual a hipótese nula é insustentável.

Note-se que uma questão análoga se nos apresenta, quando se lança ao ar uma moeda, várias vezes seguidas, e se regista de cada vez o lado que fica para cima: *coroa* ou *face*. Se a moeda estiver bem *balançada* (isto é, se os dois lados forem sensivelmente iguais do ponto de vista da gravidade) é de *esperar* que, por exemplo, em 100 lançamentos sucessivos, se apresente coroa 50 vezes, isto é, que seja $1/2$ a frequência relativa deste acontecimento. Mas é bem possível (e até mais *provável*) que, nos 100 lançamentos, se apresente coroa mais ou menos de 50 vezes: *simplesmente, a frequência absoluta do acontecimento será tanto menos de esperar (tanto menos provável), quanto mais se afastar do valor esperado, isto é, quanto maior for a discrepância*.

Este exemplo sugere o seguinte procedimento para averiguar em que medida as discrepâncias observadas no caso em estudo são atribuíveis ao acaso:

Tomem-se 300 bocados de cartão, sensivelmente iguais em substância, forma e dimensões, e distingam-se 100 desses cartões com a cor vermelha (representativos dos 100 animais tratados) e os 200 restantes com a cor branca (representativos dos animais não tratados). Execute-se, depois, *um grande número de vezes seguidas*, a prova \mathcal{P} que consiste nas seguintes operações:

- 1.º – *Fechar todos os cartões numa mesma caixa.*
- 2.º – *Agitar esta várias vezes.*
- 3.º – *Tirar da caixa ao acaso 60 cartões (representativos dos animais mortos).*

4.º – *Registrar o número de cartões vermelhos existentes nessa amostra de 60 cartões.*

É claro que a repetição \mathcal{P} implica a reposição dos cartões retirados na prova anterior.

A cada realização da prova corresponderá, assim, uma tábua de contingência de valores observados e, portanto, uma outra de discrepâncias. O que está agora naturalmente indicado é *verificar a frequência relativa com que se registam discrepâncias iguais ou superiores em valor absoluto às da tabela n.º 9, isto é, iguais ou superiores a 8 (em valor absoluto).*

Ora, quem tiver a paciência de efectuar um número muito grande de tais provas (pelo menos 1.000), verificará que *só em cerca de 2 por cento das provas se apresentam discrepâncias iguais ou superiores a 8, em valor absoluto.* A raridade de tais discrepâncias, embora não autorize, em absoluto, a rejeitar a hipótese nula, torna já bastante plausível a hipótese da eficácia do tratamento.

Convém, desde já, registrar este facto: a frequência relativa com que, numa série de realizações da prova \mathcal{P} , se apresentam discrepâncias cujo valor absoluto é igual ou superior a um dado limite L , tende a estabilizar-se num valor determinado, à medida que o número de realizações aumenta. Ao tratar do conceito de probabilidade, veremos, mais precisamente, em que consiste essa estabilização. Entretanto, podemos já dizer que o referido valor se chama *probabilidade dum discrepância igual ou superior a L (em valor absoluto)*⁽¹⁾.

Ora, como teremos ocasião de ver, tal valor pode ser calculado por dedução puramente matemática, que dispensa as longas e fastidiosas séries de provas com cartões a que fizemos referência. Por este simples exemplo se pode já fazer uma ideia da utilidade e do alcance do Cálculo das probabilidades.

O critério de interpretação atrás descrito pertence à categoria dos chamados *testes de significância*. Para averiguar se dado tratamento é eficaz, começa-se por formular a *hipótese nula*, que consiste em

(1) – Nesta antecipação, que nos parece útil do ponto de vista didáctico (fazendo preceder a teoria de exemplos concretos) seguimos a orientação da citada obra de FINNEY.

supor o contrário, isto é, que o tratamento não tem efeito nenhum. Da tabela dos *valores observados*, deduz-se, então, a dos *valores esperados* (de acordo com a hipótese nula) e, em seguida, a das discrepâncias. Posto isto, investiga-se a frequência relativa com a qual, sendo válida a hipótese nula, se apresentariam discrepâncias tão grandes ou maiores (em valor absoluto) do que os desvios observados. Tal investigação poderia ser feita directamente, pelo processo empírico, laborioso, de amostragem casual de cartões, segundo o esquema atrás descrito; mas a Matemática permite evitar esse trabalho, por meio do *Cálculo das probabilidades*. Se a frequência relativa (ou melhor, a *probabilidade*) de tais desvios for *muito pequena*, então, sim, haverá razão para rejeitar a hipótese nula, admitindo como real a influência do tratamento. Os desvios observados dizem-se, então, *estatisticamente significantes* ou apenas *significantes*. De contrário, continuará de pé a hipótese nula, enquanto novas experiências não vierem enriquecer o material de dados, para que o investigador se possa pronunciar num ou noutro sentido.

Mas o que se entende aqui por frequência relativa “muito pequena” e por frequência relativa “não muito pequena”? Haverá, necessariamente, uma extensa margem de subjectividade no emprego destes termos. O critério a adoptar no seu uso depende não só da experiência realizada, como, ainda, da intuição do experimentador, que neste campo dispõe de ampla liberdade.

Em investigações agronómicas é usual considerar já como “muito pequena”, em casos tais, a frequência relativa de 1 por 20 (0,05), embora, por vezes, se prefira considerar como “muito pequena” a frequência de 1 por 100 (0,01), sem esquecer que se trata aqui das frequências respeitantes a um grande número de provas (ou melhor, de *probabilidades*). Para distinguir aqueles dois graus de significância (ou *níveis de significância*, como usa dizer-se nestes casos), convencionou-se chamar ao primeiro, “significante” (sem qualquer restritivo) e ao segundo, “altamente significativo”. Convém salientar que se trata aqui apenas duma convenção de linguagem, reconhecida como apropriada por estatísticos e experimentadores. De resto, não são estes os únicos níveis de significância empregados: casos há em que uma probabilidade de 1 por 10 é já considerada signifi-

cante (especialmente em investigações médicas, em que o tratamento consista numa ligeira alteração de regime, pouco dispendiosa e sem risco para o doente); e casos há, pelo contrário, em que se requer um nível de 1 por 1.000 (quando se trate de alterações de regime que sejam profundas e dispendiosas).

Em síntese, podemos dizer que a aplicação dum teste de significância comporta os seguintes passos:

1) – *Formular a hipótese nula, contraditória do facto experimental a estabelecer.*

2) – *Escolher o nível de probabilidade que pareça mais adequado à natureza da experiência e ao tipo de decisão a tomar.*

3) – *Empregando o cálculo das probabilidades ou efectuando uma longa série de provas com cartões no género das que foram descritas, avaliar a probabilidade de que, sendo válida a hipótese nula, as discrepâncias casuais sejam pelo menos tão grandes (em módulo) como aquelas observadas na experiência.*

4) – *Se a probabilidade obtida em 3) é inferior à probabilidade escolhida em 2), rejeitar a hipótese nula; de contrário, deixar suspensa a conclusão.*

É, agora, o momento de mostrar o partido que se pode tirar da função χ^2 dos desvios, como teste de significância aplicado a uma experiência, cujos resultados estejam inscritos numa tábua de contingência. Mostra a análise matemática o seguinte: a probabilidade de que, sendo válida a hipótese nula, o valor de χ^2 exceda um dado limite, pode, sob certas restrições, ser calculada com suficiente aproximação, sem fazer intervir os números de indivíduos observados, mas, unicamente, o número n de graus de liberdade da tábua em questão. Deste modo, será possível calcular, uma vez por todas, para cada valor de n , as probabilidades correspondentes a diversos valores de χ^2 e registar os resultados numa tábua de duas entradas, que pode ser usada pelo investigador, com enorme economia de esforço e de tempo⁽¹⁾.

(1) – Só ao tratar das distribuições de variáveis contínuas, podemos referir-nos aos fundamentos teóricos deste método.

Assim, por exemplo, para $n = 2$, sendo válida a hipótese nula, está calculado que se tem:

- $\chi^2 \geq 0,5$ em cerca de 50 % dos casos,
- $\chi^2 \geq 2,7$ em cerca de 10 % dos casos,
- $\chi^2 \geq 3,8$ em cerca de 5 % dos casos,
- $\chi^2 \geq 6,6$ em cerca de 1% dos casos,
- $\chi^2 \geq 10,8$ em cerca de 1% dos casos.

Por outras palavras, a *probabilidade* de se ter $\chi^2 \geq 0,5$ é aproximadamente igual a 0,5, o que também se exprime escrevendo

$$\Pr(\chi^2 \geq 0,5) = 0,5$$

e, analogamente: $\Pr(\chi^2 \geq 2,7) = 0,1$; $\Pr(\chi^2 \geq 3,8) = 0,05$, etc.

Estes resultados podem ser confirmados, experimentalmente pelo processo das provas repetidas atrás descrito, calculando o valor de χ^2 correspondente aos resultados de cada prova e determinando, no final, a frequência relativa dos valores $\geq 0,5$, a dos valores $\geq 2,7$, etc.

Ora, tornando ao exemplo das tabelas n.ºs 7, 8 e 9, verifica-se que é:

$$\chi^2 = \frac{(-8)^2}{160} + \frac{8^2}{40} + \frac{8^2}{80} + \frac{(-8)^2}{20} = 6,00.$$

Segundo a doutrina exposta, a probabilidade de que seja $\chi^2 \geq 6,6$ deve ser um pouco superior a 1% (mais precisamente 0,0143), o que concorda com a frequência de cerca de 2% que se obteria numa longa série de provas com cartões.

Mas é preciso notar que *o uso do χ^2 e dos respectivos níveis de probabilidade envolve aproximações que só são aceitáveis quando os números de indivíduos são grandes*. Uma regra prática que costuma ser recomendada é a de que nenhum dos valores esperados (ver tabela n.º 8) deve ser inferior a 5. Contudo, a aproximação pode ser melhorada consideravelmente, em qualquer caso, por um simples artifício, que consiste em diminuir 0,5 ao módulo de cada discrepância, antes de calcular o χ^2 (*correção de YATES*). Assim, no caso em estudo, viria

$$\chi^2 = \frac{(-7,5)^2}{160} + \frac{(7,5)^2}{40} + \frac{(7,5)^2}{80} + \frac{(-7,5)^2}{20} = 5,27.$$

O valor da $\Pr(\chi^2 \geq 5,27)$, no caso ideal a que se refere a teoria matemática, é 0,0217, em vez de 0,0143 – bastante mais próximo do verdadeiro valor que é, no caso considerado, 0,0210 (a menos de 0,0001).

Da mesma obra extraímos um segundo exemplo, relativo a uma tábua de contingência com 3 graus de liberdade. Trata-se de colheitas de cevada em 260 campos, feitas em 1942. No inverno e primavera precedentes, tinha-se avaliado em cada campo a população da larva dum coleóptero, comum em Inglaterra, com o nome de “wireworm” (alfinete). Segundo determinado critério, classificou-se a infestação dos diferentes campos em “baixa”, “moderada”, “alta” e “muito alta”. Por sua vez, o resultado das colheitas foi classificado em “satisfatório” e “não satisfatório”. Os resultados são os que constam da seguinte tabela:

TABELA N.º 10

Resultados das colheitas	Infestação				Total
	Baixa	Moderada	Alta	Muito alta	
Satisfatórios	94	62	31	15	202
Não satisfatórios	15	15	17	11	58
Total	109	77	48	26	260

Daqui se deduziu uma tabela de valores de independência e uma outra de discrepâncias (tabelas que não reproduzimos), achando-se o seguinte valor:

$$\begin{aligned} \chi^2 = & \frac{(9,3)^2}{84,7} + \frac{(2,2)^2}{59,9} + \frac{(-6,3)^2}{37,3} + \frac{(-5,2)^2}{20,2} + \\ & + \frac{(-9,3)^2}{24,3} + \frac{(-2,2)^2}{17,2} + \frac{(6,3)^2}{10,7} + \frac{(5,2)^2}{5,8} = 15,7. \end{aligned}$$

A tabela n.º 11, em que são indicados alguns níveis de significância para χ^2 , mostra-nos que o valor 15,7 calculado é sensivelmente superior ao valor 11,3, que, para 3 graus de liberdade, marca o nível “altamente significativo”. Há, pois, razão suficiente para excluir a hipótese nula.

TABELA N.º 11

Graus de liberdade	Frequências relativas		(Probabilidades)
	0,1	0,05	
1	2,7	3,8	6,6
2	4,6	6,0	9,2
3	6,3	7,8	11,3
4	7,8	9,5	13,3
6	10,6	12,6	16,8
8	13,4	15,5	20,1
10	16,0	18,3	23,2
15	22,3	25,0	30,6
20	28,4	31,4	37,6
25	34,4	37,7	44,3
30	40,3	43,8	50,9

Convem notar que:

1) *Dum modo geral, o teste do χ^2 , como foi descrito, não deve aplicar-se quando algum dos valores esperados for demasiado pequeno, exigindo-se, nesse caso, uma análise estatística especial.*

2) *A correcção de Yates é somente aplicável nas tábuas de 2×2 (com 1 grau de liberdade).*

Há, ainda, certos pormenores, relativos à realização da experiência, que precisam de ser observados, para que o teste de significância não resulte ilusório. Tornando ao primeiro exemplo citado neste número, é óbvio, que, se os animais tratados tiverem uma proveniência diferente da dos animais não tratados, as diferenças registadas podem muito bem não ser devidas ao tratamento. Para atenuar o mais possível os factores estranhos ao tratamento, o que há a fazer é destacar os 300 animais duma população tão *homogénea* quanto

possível, e, entre esses, escolher, completamente *ao acaso*, os 100 animais a serem tratados, dando a todos os 300 *igual probabilidade* de serem escolhidos. Esta operação, chamada *casualização*, tem de ser efectuada estritamente por um processo objectivo, em que não intervenha qualquer factor pessoal na escolha. Um tal processo pode consistir no seguinte: numerar os animais de 1 a 300, deitar numa caixa 300 cartões iguais, numerados de 1 a 300, e tirar 100 à sorte; os números saídos serão os dos animais a tratar.

Nunca é demais encarecer a importância da casualização, nestes e noutros tipos de experiência, como base duma investigação estatística bem orientada. Não podemos, contudo, indicar aqui os pormenores da técnica, a que por vezes tem de obedecer uma tal operação.

Note-se que o papel do estatístico não se limita à interpretação dos resultados experimentais: ele pode contribuir, com grande vantagem, para o *planeamento da experiência*. Não pertence ao âmbito do nosso curso o estudo deste assunto de alto interesse. Limitar-nos-emos, por isso, a breves indicações sobre a natureza da questão, no caso particular do tratamento hipotético atrás considerado. Suponhamos, por exemplo, que se tinham escolhido, apenas, 2 ou 3 animais para serem tratados, reservando os restantes para controle (ou vice-versa); é manifesto, mesmo *a priori*, que os resultados não autorizariam um juízo seguro. Deve, portanto, haver uma proporção *ótima* para revelar a eficácia do tratamento, caso este tenha, de facto, efeito. Porém, essa proporção não é, como poderia parecer à primeira vista, a de 50% para os dois grupos. Na Tabela n.º 12, estão indicados os valores do χ^2 para diferentes proporções de animais tratados, entre os 300, na hipótese de a percentagem de mortos entre animais tratados ser em todos os casos a mesma que se observa na Tabela n.º 7 (12%). Vê-se que o máximo valor do χ^2 é atingido na proporção de 175 animais tratados para 125 não tratados. Será, portanto, essa proporção ideal para a experiência.

É claro que, quanto maior for o número total de casos individuais estudados, melhor informação fornecem os resultados obtidos; mas esse número é forçosamente limitado por razões de ordem económica, e, por isso, mais necessário se torna tirar o maior rendimento possível do material de que se dispõe.

TABELA N° 12

N° de animais tratados	χ^2 para 12% de mortalidade
25	1,2
50	2,8
100	5,3
125	6,1
150	6,5
175	6,6
200	6,3
250	4,0
275	2,0

Muito mais haveria a dizer sobre testes de significância, mas, embora o assunto seja de grande interesse para agrónomos, não pode ser desenvolvido nesta cadeira. E, se nos alongámos um pouco a respeito do teste do χ^2 , foi sobretudo para dar uma ideia de como o Cálculo das Probabilidades pode intervir nas aplicações de carácter agronómico.

B – Probabilidades

1. Lógica indutiva

Suponhamos que, em n realizações, dum mesma prova \mathcal{P} , um dado acontecimento α se realizou n vezes: então, a frequência relativa do acontecimento α , na referida série de provas, será $f = n/n = 1$. Como já tivemos ocasião de observar, este facto não habilita a concluir que α seja um acontecimento certo. No entanto, se o número n de provas realizadas for *muito grande*, é-se levado a admitir que o acontecimento é *praticamente certo*, embora não se possa concluir que seja *certo* (*em absoluto*). Nisto consiste, esquematicamente, a *indução* ou *raciocínio indutivo*, que se encontra na base de toda a ciência experimental. É bem sabido que as *leis físicas* ou, melhor, as *leis da Natureza*, têm carácter *contingente*: não se pode garantir que sejam absolutamente infalíveis.

Seja, por exemplo, a seguinte experiência: aproximar uma chama dum frasco, com a boca para baixo, no qual se tenha introduzido

uma mistura de hidrogénio e oxigénio, em proporções convenientes. É, então, praticamente certo que se verifica o fenómeno “explosão”.

Trata-se, aqui, duma lei química qualitativa, de cuja contingência nos apercebemos, sobretudo, se não tiverem sido definidas, com certa precisão, as condições em que a experiência deve ser realizada. Muitos outros exemplos poderíamos citar, de leis qualitativas, com o mesmo carácter contingente.

As leis quantitativas correspondem a um grau mais elevado de conhecimento, a uma fase mais avançada da ciência: mas nelas subsiste o carácter de contingência. Primeiro que tudo, devemos lembrar-nos de que as medidas das grandezas físicas *nunca* podem ser exactas (não faz, mesmo, sentido falar de medida exacta duma grandeza empírica – seja comprimento, seja velocidade, seja carga eléctrica, seja um pH, seja qualquer outra). Já daqui resulta uma certa margem de incerteza, isto é, de contingência.

Consideremos, por exemplo, esta experiência: aquecer um pedaço de chumbo a uma temperatura de cerca de 335 graus centígrados, em condições normais. É praticamente certo que se verifica o fenómeno *fusão*. Mas no próprio carácter aproximativo da temperatura indicada, bem como no significado da expressão “condições normais”, há uma origem de incerteza, que pode ser reduzida (indicando, por exemplo, uma vizinhança de 335 na qual se prevê o acontecimento), mas nunca eliminada.

Recordemos, ainda, leis físicas, de nível mais elevado: a lei da gravitação, as leis da termodinâmica, as leis do electro-magnetismo, etc. Como é sabido, nenhuma destas leis se adapta exactamente à realidade: assentam todas em hipóteses simplificadoras, que só aproximadamente se podem realizar. Essas leis são, pois, apenas esquematizações duma realidade que é sempre demasiado complexa, demasiado mutável, para se deixar traduzir fielmente na simplicidade dos nossos símbolos.

Assim, a lei dos gases perfeitos deve o seu nome ao facto de se ter convencido chamar “gases perfeitos” aos gases que a seguem rigorosamente, o que imprime a essa lei um certo carácter de definição ou de postulado. A verdade, porém, é que não há gases perfeitos: há, apenas, gases que se aproximam, mais ou menos, dessa forma

ideal. Uma lei que substitui aquela, dando uma melhor aproximação da realidade, é a de Van der Waals; mas nem mesmo essa poderá ser exacta. Porque é lei metafísica das leis físicas, o serem contingentes.

O método experimental, e com ele o método matemático, tem-se estendido progressivamente, do âmbito restrito das ciências físico-químicas, ao das ciências biológicas, ao das ciências sociais, etc. Nestes novos domínios, mais acentuado se torna o carácter contingente das leis naturais. No entanto, algumas se apresentam com visos de certeza inabalável, como aquela que tem servido de premissa a um exemplo clássico de silogismo:

“Todos os homens são mortais”.

Há aqui, de certo modo, uma lei biológica qualitativa. Mas se tentarmos precisá-la quantitativamente, afirmando, por exemplo:

“Todos os homens morrem antes dos 300 anos de idade”,

já não sentiremos o mesmo grau de segurança. Conhecemos nós, suficientemente, o passado da espécie humana? E que sabemos nós sobre o futuro?

O que pode dizer-se é que, nas condições actuais, *é extremamente improvável, praticamente impossível* que um ser humano atinja a idade de 300 anos. Um outro exemplo análogo é o que se refere a alturas: é praticamente impossível que um ser humano cresça até atingir a altura de 4 metros.

Se formos baixando estes limites, o grau de incerteza aumentará – e entraremos, abertamente, no campo das probabilidades. É de salientar que o Cálculo das Probabilidades e a Estatística se têm desenvolvido principalmente no sector das ciências biológicas e das ciências sociais. Mas certo é, também, que, por um movimento de retrocesso, acabaram por invadir o campo das ciências físicas, principalmente no que se refere ao estudo do átomo. A física moderna tem carácter probabilista.

NOTA. Uma fonte primária de contingência está na impossibilidade que há em definir exactamente os conceitos relativos ao mundo empírico, – em delimitar com precisão os atributos dos entes naturais. Esta impossibilidade tem sido origem constante de especulações filosóficas através dos séculos. Platão resolvia a dificuldade, distinguindo duas formas de existência: a *realidade sensível* ou *mundo dos fenómenos*, que conhecemos por meio dos sentidos, e a *realidade inteligível* ou *mundo das Ideias*, que conhecemos por meio da razão e da qual a primeira é, apenas, imitação grosseira. Recorde-se, ainda, a querela que, na Idade Média, dividiu os filósofos em *realistas* e *nominalistas*, os primeiros proclamando a *realidade* dos universais (isto é, das classes, dos entes abstractos) e a sua supremacia sobre o contingente e o transitório; os segundos afirmando que os conceitos abstractos são, apenas, *nomes*, ficções cómodas, e que só os indivíduos existem.

No fundo, trata-se de duas tendências complementares, inerentes ao nosso espírito, com predomínio duma ou doutra conforme as pessoas e as situações. Assim, o matemático assume geralmente a atitude platónica (realista), reportando ao mundo dos seres ideais, de maneira mais ou menos consciente, os conceitos abstractos que são objecto do seu pensamento. Por sua vez, o experimentador tende para a atitude nominalista (ou empirista). Do equilíbrio das duas tendências, segundo o bom-senso, é que pode resultar o êxito da actividade intelectual: toda a teoria deve ser controlada pela experiência, assim como a prática deve sempre ser guiada pela teoria.

2. Lógica dedutiva

Ao raciocínio indutivo das ciências experimentais contrapõe a Matemática o raciocínio dedutivo. À contingência *física* das leis físicas opõe-se a certeza *matemática* das deduções matemáticas. Considerem-se, por exemplo, os seguintes teoremas:

“O quadrado dum número ímpar é sempre um número ímpar”.

“Toda a equação de coeficientes complexos admite pelo menos uma solução no campo complexo”.

“Toda a série de potências de x é derivável, termo a termo (em ordem a x), no seu intervalo de convergência”.

Encontramos aqui um carácter de certeza e precisão que contrasta vivamente com a natureza dúbia e aproximativa das leis naturais. Contudo, já se nota diferença, quando se passa da matemática pura para as matemáticas aplicadas. A própria Geometria é, sob certo aspecto, um ramo da Física. Consideremos, por exemplo, o bem conhecido teorema da geometria euclideana:

“A soma dos ângulos dum triângulo (plano) é sempre igual a um ângulo raso”.

Dois métodos diferentes se podem seguir para estabelecer a veracidade desta proposição:

1.º – *Método indutivo ou experimental* – Medem-se os ângulos internos de vários triângulos e verifica-se que a soma dos ângulos de cada triângulo se *aproxima bastante* de 180° , tanto mais quanto mais perfeito for o traçado dos triângulos e mais precisa for a medição. É este o método seguido na primeira fase do ensino da Matemática no liceu.

2.º – *Método dedutivo ou racional* – Demonstra-se a proposição, *deduzindo-a logicamente* de proposições mais simples, cuja veracidade se considera *evidente*. Estas proposições evidentes são os chamados *axiomas* ou *postulados* da geometria euclideana. No teorema em questão, o postulado que intervém de maneira essencial é precisamente o postulado de Euclides:

“Dados um ponto e uma recta, existe sempre uma recta, e uma só, que passa pelo ponto dado e é paralela à recta dada”.

Mas o declarar *evidentes* os postulados é uma atitude cómoda, que exige reflexão. Têm eles, de facto, carácter de certeza absoluta?

Um dos grandes progressos da história do pensamento, realizado no século passado, foi precisamente o de reconhecer que os postulados da geometria euclideana têm a natureza contingente e aproximativa das leis naturais. Nem leis se podem dizer, propriamente, mas sim hipóteses, cuja legitimidade se avalia indirectamente pelas suas consequências lógicas. Insinuam-se no nosso espírito, por um longo processo indutivo, dificilmente controlável, que se desenvolve na experiência quotidiana, no exercício constante da nossa actividade neuro-muscular. É esta uma forma de indução que se confunde com aquele processo directo de conhecimento a que se dá o nome de *intuição*.

Mas a nossa experiência quotidiana refere-se a uma região limitada do espaço. A teoria da relatividade veio mostrar, precisamente, que, nos espaços astronómicos, não é a geometria euclideana a que mais se aproxima da realidade: aí, a soma dos ângulos internos dum triângulo pode tornar-se sensivelmente superior a 180° .

O que se disse a respeito do teorema citado aplica-se a qualquer outro. A contingência dos postulados transmite-se a todos os teoremas. O que é matematicamente certo não é o teorema, mas, sim, o facto de o teorema ser consequência lógica dos postulados.

Entretanto registem-se as vantagens do método racional. Enquanto o método empírico exige um grande número de verificações fastidiosas, o método racional, com elegante simplicidade e perfeito rigor, reduz a veracidade do teorema à de proposições já reconhecidas como verdadeiras. Estas vantagens patenteiam-se, em particular, no Cálculo das Probabilidades.

3. Conceito natural de probabilidade

As considerações precedentes mostram que os termos “verdadeiro” e “falso”, aplicados a proposições, se tornam insuficientes, quando, da realidade inteligível da Matemática, se passa à realidade sensível dos fenómenos. Os conceitos de “verdadeiro” e “falso” cedem, então, o lugar ao conceito de “probabilidade”, correlativo do de “incerteza” ou “contingência”: um facto dir-se-á tanto mais *provável* quanto menos contingente for.

O conceito de probabilidade, como todos os conceitos relativos ao mundo empírico, não é susceptível de definição lógica: gera-se no nosso espírito por um processo indutivo. Mais até, faz parte intrínseca do próprio mecanismo da indução. É o que vamos ver, tentando esclarecer este conceito.

Seja α um dos acontecimentos a prever numa certa experiência \mathcal{P} e suponhamos que esta experiência foi efectuada em várias séries de provas, todas em *grande número*:

Série 1 – $\mathcal{P}_{1,1}, \mathcal{P}_{1,2}, \dots, \mathcal{P}_{1,n_1}$ (n_1 provas)
 Série 2 – $\mathcal{P}_{2,1}, \mathcal{P}_{2,2}, \dots, \mathcal{P}_{2,n_2}$ (n_2 provas)

 Série m – $\mathcal{P}_{m,1}, \mathcal{P}_{m,2}, \dots, \mathcal{P}_{m,n_m}$ (n_m provas).

Se forem, respectivamente, v_1, v_2, \dots, v_m as frequências absolutas do acontecimento α nestas séries de provas, serão

$$\frac{v_1}{n_1}, \frac{v_2}{n_2}, \dots, \frac{v_m}{n_m},$$

as correspondentes frequências relativas. Suponhamos, ainda, que se verificou uma sensível concordância entre estas frequências, isto é, que todas se localizaram num pequeno intervalo $]f_0 - \varepsilon, f_0 + \varepsilon[$:

$$f_0 - \varepsilon < \frac{v_i}{n_i} < f_0 + \varepsilon, \quad \text{para } i = 1, 2, \dots, m.$$

Aplicando o raciocínio indutivo, exactamente como se faz ao estabelecer as leis naturais, seremos levados a admitir, como praticamente certo, que:

Em toda a série formada por muitas realizações de \mathcal{P} (em número não inferior ao maior dos números n_1, n_2, \dots, n_m) a frequência relativa de α ficará situada entre $f_0 - \varepsilon$ e $f_0 + \varepsilon$.

É, então, natural dizer que f_0 é um *valor aproximado da probabilidade de α , com erro inferior a ε (ou a menos de ε)*.

Por exemplo, se em tais séries de provas se registou sempre uma frequência relativa de α entre 7% e 11% (ou seja, entre $0,09 - 0,02$ e $0,09 + 0,02$), dir-se-á que 0,09 é um valor aproximado, a menos de 0,02, da probabilidade de α .

Suponhamos, agora, que se efectuaram novas séries de realizações de \mathcal{P} , em *números bastante maiores* que os anteriores, e que passou a registar-se uma frequência relativa de α , entre

$$f_1 - \frac{\varepsilon}{10} \text{ e } f_1 + \frac{\varepsilon}{10};$$

dir-se-á, então, que f_1 é valor aproximado da probabilidade de α , a menos de $\varepsilon/10$. Analogamente, poderíamos imaginar valores f_2, f_3, \dots , aproximados da probabilidade de α a menos de $\varepsilon/10^2$, de $\varepsilon/10^3$, etc.

A probabilidade aparece-nos, assim, como *grandeza que se mede*, tal como se fosse uma grandeza física (uma velocidade, um ponto

de fusão, uma densidade, etc.), com o carácter contingente e aproximativo que revestem todas as medições empíricas. Já as expressões “grandes números”, “números bastante maiores”, atrás usadas, envolvem a aplicação de critérios subjectivos, humanos, a que falta rigor matemático, mas que são inevitáveis.

Tal como sucede com as grandezas físicas, é-se levado a idealizar, para cada acontecimento α (em certos tipos de experiências), um valor *exacto*, p , da probabilidade de α , que seria, por assim dizer, o *limite* da sucessão dos valores aproximados f_0, f_1, f_2, \dots . Mas é claro que um tal valor exacto *não existe nos casos concretos* – assim como não existe o valor exacto duma grandeza física: trata-se apenas de convenções cómodas, que o nosso espírito transporta para o mundo platónico dos entes ideais, para sobre as mesmas poder construir uma teoria racional – *uma teoria matemática*⁽¹⁾.

Note-se, ainda, que a unidade adoptada para medir a probabilidade do acontecimento α é a probabilidade do acontecimento certo – isto é, a *certeza*⁽²⁾. Por sua vez, ao acontecimento impossível é atribuída a probabilidade 0. Deste modo, a probabilidade p de α deverá ser um número tal que

$$0 \leq p \leq 1.$$

Se p for aproximadamente igual a 1, dir-se-á que α é *praticamente certo*. Se p for aproximadamente igual a 0, dir-se-á que α é *praticamente impossível*. Mas o grau de aproximação a partir do qual se dirá uma ou outra coisa é questão manifestamente subjectiva. Registe-se, entretanto, que, para um acontecimento ser considerado praticamente certo, não é necessário que se realize em *todas* as provas que tenham sido efectuadas: admite-se a possibilidade de excepções, contanto que estas sejam extremamente raras.

E, analogamente, no caso oposto.

(1)–Esta concepção da probabilidade como constante física dos acontecimentos é devida ao matemático francês MAURICE FRÉCHET.

(2)–Recorde-se que, na medida dos arcos, se pode tomar unidade a circunferência. Uma das maneiras sugestivas de indicar graficamente as probabilidades (ou as frequências relativas) de vários acontecimentos incompatíveis, consiste em representá-los por meio de sectores dum círculo, de amplitudes proporcionais a essas probabilidades (ou frequências).

O termo “probabilidade”, tal como acabamos de o precisar, é usado na prática com maior ou menor propriedade. Recordemos o exemplo das tábuas de mortalidade, sobre as quais os actuários das companhias de seguros baseiam o cálculo dos prémios a pagar em seguros de vida. Essas tábuas são, no fundo, estatísticas, isto é, tabelas de frequência, relativas a populações numerosas. Assim, por exemplo, a tábua de mortalidade alemã para homens, correspondente ao decénio 1901-1910, indica que, entre 10.000 crianças com menos de 1 ano, 3.608 chegaram à idade de 65 anos. Deste modo, a probabilidade de que um recém-nascido venha a atingir os 65 anos será, aproximadamente:

$$\frac{3.608}{10.000} \approx 36\%.$$

Por sua vez, a mesma tábua mostra que, entre 10.000 crianças com menos de 1 ano, 7.065 chegam aos 20 anos. Daqui se deduz que a probabilidade de um rapaz de 20 anos chegar aos 65 anos é

$$\frac{3.608}{7.065} \approx 51\%,$$

maior que a precedente.

A mesma tábua dá para crianças de menos de 1 ano, a probabilidade 0,68% de chegar aos 90 anos, a probabilidade 0,0038% de atingir os 100 anos, etc.

Mas note-se que as tábuas de mortalidade têm de ser revistas de tempos a tempos e não devem ser as mesmas para todas as regiões, para todas as raças, para os dois sexos, etc.

A probabilidade dum acontecimento pode, assim, estar sujeita a uma evolução imprevisível no tempo e no espaço. É este um facto que, pelo menos aparentemente, está em contradição com o conceito de probabilidade, tal como foi atrás explanado. O mesmo é dizer que as leis físicas evoluem de maneira imprevisível no tempo e no espaço – *deixando, portanto, de ser leis*. A explicação do facto está, ainda, no carácter contingente do raciocínio indutivo, que impõe sempre a seguinte norma prudencial ao experimentador: *as induções*

só merecem confiança em regiões limitadas do espaço e do tempo; não podem ser extrapoladas muito para além do domínio das experiências efectuadas.

Portanto, só quando a evolução for lenta será lícito falar de probabilidades.

Recordemos, ainda, o exemplo dos desastres de avião: é hoje menor do que há vinte anos a frequência relativa dos desastres de aviação em percursos iguais. Neste caso, embora seja mais rápida a evolução, ainda, de certo modo, é lícito falar de probabilidade no sentido atrás considerado⁽¹⁾.

Casos há, porém, em que o emprego desta palavra escapa, de todo, a critérios quantitativos, que lhe confirmam interesse científico.

4. Axiomatização do conceito de probabilidade

Viu-se no número anterior que, para fundar uma teoria matemática das probabilidades, se idealiza para cada acontecimento, em certos tipos de experiências, um valor exacto da sua probabilidade. Contudo, seja esse valor uma convenção ou uma realidade, o que interessa para o desenvolvimento lógico da teoria não é, propriamente, o conceito em si, mas, antes, o conjunto das suas propriedades formais que intervêm nos cálculos e nos raciocínios. O que há a fazer, então, é escolher, entre essas propriedades, algumas mais simples, das quais se possam deduzir logicamente todas as outras. As propriedades escolhidas dir-se-ão *propriedades fundamentais* ou *axiomas* e o seu conjunto terá o nome de *axiomática*. Já em Matemáticas Gerais foi estudada uma axiomática dos grupos contínuos de grandezas. Trata-se, agora, de estabelecer uma axiomática do conceito de probabilidade, no caso dos corpos finitos de acontecimentos, considerando aquele conceito como *primitivo*, isto é, não definível logicamente à custa de outros.

Seja, pois, \mathcal{R} um corpo finito de eventualidades a considerar numa dada prova \mathcal{P} , e suponhamos que a cada eventualidade $\alpha \in \mathcal{R}$, corresponde uma determinada probabilidade, que designaremos pelo

(1) – Nos seguros relativos a acidentes de aviação, as companhias americanas fazem hoje o cálculo dos prémios como se houvesse morte de um passageiro em 20.000 viagens de ida e volta.

símbolo $\Pr(\alpha)$. Tomaremos como fundamentais as propriedades seguintes⁽¹⁾:

AXIOMA 1. *A probabilidade de α é sempre um número real não negativo, isto é:*

$$\Pr(\alpha) \geq 0, \text{ qualquer que seja } \alpha \in \mathcal{R}.$$

AXIOMA 2. *Se α e β são acontecimentos incompatíveis de \mathcal{R} , tem-se:*

$$\Pr(\alpha + \beta) = \Pr(\alpha) + \Pr(\beta).$$

(Convém aqui rever, cuidadosamente, as noções do n.º 2 e do n.º 3 relativas às álgebras de atributos e de acontecimentos).

AXIOMA 3. *A probabilidade do acontecimento certo é 1; isto é, tem-se:*

$$\Pr(\alpha) = 1, \text{ se } \alpha \text{ é certo.}$$

Estas propriedades concordam com o conceito empírico de probabilidade, tal como foi atrás introduzido. Sendo as probabilidades, por assim dizer, limites de frequências relativas, é natural que se conservem as propriedades das frequências relativas que não mudam na passagem ao limite.

Os axiomas 1 e 2 dizem-nos, simplesmente, que a função $\Pr(\alpha)$ é uma distribuição definida no corpo \mathcal{R} . O axioma 3 acrescenta que essa distribuição é relativa. Diremos, então, que se trata duma *distribuição de probabilidade* definida em \mathcal{R} . Mas só nas aplicações concretas será possível distinguir uma distribuição de probabilidade duma distribuição de frequência, pois que a anterior axiomática se limita a dar os caracteres duma distribuição abstracta.

(1) – Vários autores consideram, ainda, como axioma a propriedade relativa à probabilidade do produto. Mas essa propriedade pode considerar-se como definição ou como consequência de definição de probabilidade condicional, como veremos.

Da anterior axiomática deduzem-se, desde logo, as seguintes consequências (que se estendem a qualquer distribuição):

TEOREMA 1. *Se α implica β , então $\Pr(\alpha) \leq \Pr(\beta)$.*

Com efeito, se $\alpha \subset \beta$, tem-se $\beta = \alpha + \tilde{\alpha}\beta$ e, como α e $\tilde{\alpha}\beta$ são incompatíveis, vem, pelo axioma 2,

$$\Pr(\beta) = \Pr(\alpha) + \Pr(\tilde{\alpha}\beta)$$

donde se conclui, pelo axioma 1, que $\Pr(\beta) \geq \Pr(\alpha)$.

COROLÁRIO. *Qualquer que seja $\alpha \in \mathcal{R}$, tem-se:*

$$\Pr(\alpha) \leq 1.$$

Com efeito, todo o acontecimento α de \mathcal{R} implica o acontecimento certo: isto é, em fórmula:

$$\alpha \subset I,$$

donde, pelo teorema 1 e pelo axioma 3:

$$\Pr(\alpha) \leq 1.$$

TEOREMA 2⁽¹⁾. *Se $\alpha_1, \alpha_2, \dots, \alpha_n$ são acontecimentos (de \mathcal{R}) incompatíveis dois a dois, a probabilidade de que se realize um dos acontecimentos $\alpha_1, \alpha_2, \dots, \alpha_n$ é a soma das probabilidades destes acontecimentos, isto é, tem-se:*

$$\Pr(\sum \alpha_i) = \sum \Pr(\alpha_i), \text{ se } \alpha_i \alpha_k = \emptyset \text{ para } i \neq k.$$

Basta aplicar, repetidamente, o axioma 2, observando que é $\alpha_1 + \alpha_2 + \alpha_3 = (\alpha_1 + \alpha_2) + \alpha_3$, $\alpha_1 + \alpha_2 + \alpha_3 + \alpha_4 = (\alpha_1 + \alpha_2 + \alpha_3) + \alpha_4$, etc.,

(1) – Também conhecido por *princípio das probabilidades totais*.

e que, se $\alpha_1, \alpha_2, \dots, \alpha_n$ são incompatíveis dois a dois, também α_3 é incompatível com $\alpha_1 + \alpha_2$, α_4 com $\alpha_1 + \alpha_2 + \alpha_3$, etc.

TEOREMA 3. *Quaisquer que sejam $\alpha, \beta \in \mathcal{R}$, tem-se*

$$\Pr(\alpha + \beta) = \Pr(\alpha) + \Pr(\beta) - \Pr(\alpha\beta).$$

Basta notar que $\alpha + \beta = \alpha\tilde{\beta} + \tilde{\alpha}\beta + \alpha\beta$ e aplicar o teorema 2, notando que $\Pr(\alpha) = \Pr(\alpha\tilde{\beta}) + \Pr(\alpha\beta)$, $\Pr(\beta) = \Pr(\tilde{\alpha}\beta) + \Pr(\alpha\beta)$. O teorema pode generalizar-se ao caso de vários acontecimentos, obtendo-se a fórmula de DANIEL DA SILVA (cf. A-10).

TEOREMA 4. *A probabilidade do acontecimento contrário de α é a diferença para 1 da probabilidade de α ; isto é:*

$$\Pr(\tilde{\alpha}) = 1 - \Pr(\alpha).$$

Consequência imediata dos axiomas 2, 3 e da definição de contrário ($\alpha + \tilde{\alpha} = I$, $\alpha\tilde{\alpha} = \emptyset$), conforme já se viu para as distribuições em geral.

COROLÁRIO. *Se α é acontecimento impossível, tem-se*

$$\Pr(\alpha) = 0.$$

Basta lembrar que o acontecimento impossível é contrário do acontecimento certo e aplicar o teorema 4 com o axioma 3.

É claro que, em virtude do teorema 2, uma distribuição de probabilidade num corpo finito \mathcal{R} fica determinada, logo que se conheçam as probabilidades das células de \mathcal{R} : a probabilidade dum acontecimento $\alpha \in \mathcal{R}$ será a soma das probabilidades dos acontecimentos celulares de que α é a soma. Tudo está, pois, em determinar as probabilidades das células.

5. Alguns exemplos de cálculo de probabilidades *a priori*

Por vezes, a distribuição de probabilidade num corpo finito pode ser determinada *a priori* (isto é, antes de qualquer experiência deliberada) por simples considerações ditadas pelo *senso-comum*.

Um dos exemplos clássicos, já atrás invocados, é o que se refere à extracção casual de bolas ou de cartões duma caixa. Consideremos o caso duma urna com N bolas, que designaremos por x_1, x_2, \dots, x_N . Seja U o conjunto destas bolas e seja \mathcal{P} a experiência que consiste em tirar *ao acaso* uma bola da urna.

Os acontecimentos elementares que se podem verificar nesta experiência são: “sair a bola x_1 ”, “sair a bola x_2 ”, ..., “sair a bola x_N ”; ou, em notação proposicional:

$$x = x_1, x = x_2, \dots, x = x_N.$$

(Como se disse atrás, x é neste caso uma *variável casual*, que toma, em cada prova, um e um só dos valores x_1, x_2, \dots, x_N).

Estes acontecimentos elementares geram um corpo de que são as células. Os acontecimentos que formam o corpo serão, além das células e do acontecimento impossível, todos aqueles que se exprimem como somas lógicas de células⁽¹⁾, por exemplo:

$$(x = x_1) + (x = x_2) \text{ (sair } x_1 \text{ ou } x_2)$$

$$(x = x_1) + (x = x_3) + (x = x_5) \text{ (sair } x_1 \text{ ou } x_3 \text{ ou } x_5), \text{ etc.}$$

Entre estes está incluído o acontecimento certo:

$$\sum_{i=1}^p (x = x_i) \text{ (sair uma das bolas } x_1, x_2, \dots, x_N).$$

É claro, agora, que se conhecermos as probabilidades dos acontecimentos elementares, estamos aptos a conhecer a probabilidade de qualquer outro acontecimento do corpo, por simples aplicação do teorema 2. Designemos por p_1, p_2, \dots, p_N , respectivamente, as probabilidades de “sair x_1 ”, de “sair x_2 ”, ..., de “sair x_N ”, isto é, ponhamos

$$p_i = \Pr(x = x_i), \text{ para } i = 1, 2, \dots, N.$$

(1) – É evidente que os acontecimentos considerados são incompatíveis dois a dois, uma vez estabelecido que se tira *uma só* bola de cada vez.

Então, a probabilidade de “sair x_1 ou x_2 ” será

$$\Pr[(x = x_1) + (x = x_2)] = p_1 + p_2,$$

a probabilidade de “sair x_1 ou x_3 ou x_5 ” será $p_1 + p_3 + p_5$.

Como a probabilidade do acontecimento certo é igual à soma das probabilidades de todas as células, deverá ter-se

$$p_1 + p_2 + \dots + p_N = 1.$$

Suponhamos que as bolas x_1, x_2, \dots, x_v são vermelhas e as restantes brancas. Designando por V o conjunto das bolas vermelhas e por B o conjunto das bolas brancas, a probabilidade de sair bola vermelha será

$$\Pr(x \in V) = p_1 + p_2 + \dots + p_v$$

e a de sair bola branca

$$\Pr(x \in B) = p_{v+1} + p_{v+2} + \dots + p_N = 1 - \Pr(x \in V)$$

Mas como determinar as probabilidades elementares p_1, p_2, \dots, p_N , se é que existem?

Para isso, há que introduzir hipóteses suplementares. *Suponhamos, por exemplo, que as bolas são sensivelmente iguais, em substância, forma e dimensões.* Então, para dar a todas as bolas igual probabilidade de serem extraídas (*casualização*), ocorre a ideia de fechar a urna e agitá-la várias vezes antes de tirar a bola.

Esta ideia, ou melhor, esta intuição, é confirmada pela experiência: efectuando *muitas vezes* a prova nas condições indicadas (com reposição da bola retirada), serão sensivelmente iguais as frequências relativas com que aparecem as diferentes bolas⁽¹⁾.

(1) – Donde nos vem esta intuição? É, na verdade, anterior a qualquer experiência? Não parece que o seja. Será, antes, o produto de inúmeras experiências que fazemos, sem dar por isso, no decurso da nossa existência. Qualquer coisa de semelhante à intuição que nos leva a admitir os postulados da geometria euclídeana.

Uma vez admitido que todas as bolas têm a mesma probabilidade de saída, isto é, que:

$$p_1 = p_2 = \dots = p_N,$$

como se tem $p_1 + p_2 + \dots + p_N = 1$, será

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

a probabilidade de saída de qualquer bola. Então, a probabilidade de sair bola vermelha, será

$$\Pr(x \in V) = \frac{1}{N} + \frac{1}{N} + \dots + \frac{1}{N} \quad (v \text{ vezes}) = \frac{v}{N}.$$

Este último resultado corresponde directamente à definição clássica de probabilidade:

“A probabilidade dum acontecimento é o quociente do número de casos favoráveis ao acontecimento ⁽¹⁾ pelo número total de casos possíveis, supondo que estes são todos igualmente prováveis”.

Segundo o ponto de vista adoptado nesta definição, considera-se como primitivo, não o conceito de probabilidade, mas, sim, o de “igualmente provável”.

É claro que a hipótese de as bolas serem iguais (em substância, forma e dimensões) nunca se realiza exactamente na prática: podem ser, apenas, *aproximadamente iguais* e, então, as respectivas probabilidades, p_1, p_2, \dots, p_N , serão aproximadamente iguais, com um grau de aproximação correspondente. Se as bolas diferem entre si de maneira sensível, as probabilidades elementares p_1, p_2, \dots, p_N serão, também, sensivelmente diferentes e só poderão ser avaliadas *a posteriori*, efectuando numerosas extracções com reposição e registando as frequências relativas dos acontecimentos $x = x_1, x = x_2, \dots, x = x_N$. Por comodidade, em tudo o que segue, ao tratar de problemas de bolas numa urna, supomos verificada a hipótese da igualdade.

(1) – Aqui a palavra “caso” é empregada na acepção de acontecimento celular. Chamam-se “casos favoráveis ao acontecimento” os casos que *implicam* o acontecimento.

Recordemos que a extracção casual de bolas com números constitui a base de certos jogos, como o loto ou a lotaria. O cálculo das probabilidades nasceu precisamente das reflexões de certos matemáticos, nomeadamente PASCAL, sobre questões relativas a *jogos de azar*⁽¹⁾.

Um jogo de azar conhecido desde a antiguidade é o dos dados. Em princípio, um dado deve ser um cubo formado de substância homogénea (*dado perfeito*), mas é claro que, na prática, estas condições só aproximadamente se realizam. As faces do cubo estão numeradas de 1 a 6. Admite-se que, lançando o dado, depois de o agitar dentro dum copo (*casualização*), todas as faces têm igual probabilidade de se apresentar em cima e será, pois, $1/6$ a probabilidade correspondente a cada face. Esta hipótese é, geralmente, confirmada pela experiência: quando não houver confirmação, a experiência vem, apenas, revelar imperfeições que tenham passado despercebidas.

Recordemos, ainda, os jogos de cartas, alguns dos quais são puramente de azar. Nestes jogos, a casualização consiste em *baralhar* as cartas.

É, ainda, de citar o jogo da roleta – círculo dividido em sectores numerados, que roda em torno dum eixo que passa pelo centro. A casualização consiste em imprimir à roleta um impulso que a faça dar várias voltas: se os sectores foram iguais, a probabilidade de parar num dado ponto é a mesma para todos. Se os sectores forem diferentes, as probabilidades serão proporcionais às respectivas amplitudes, de modo que, se foram $\Theta_1, \Theta_2, \dots, \Theta_N$, essas amplitudes em radianos, serão

$$p_1 = \frac{\Theta_1}{2\pi}, p_2 = \frac{\Theta_2}{2\pi}, \dots, p_N = \frac{\Theta_N}{2\pi},$$

as respectivas probabilidades, pois que se deve ter

$$p_1 + p_2 + \dots + p_N = 1.$$

(Este exemplo conduz, naturalmente, à consideração de probabilidade no contínuo).

(1) – Aqui “azar” não é usado como sinónimo de “pouca sorte”, mas, simplesmente, como sinónimo de “acaso” (tal como o francês “hasard”).

Recordemos, por último, o mais simples de todos os jogos de azar: o lançamento dum moeda ao ar. Se a moeda for bem *balançada*, as duas eventualidades “sair coroa” e “sair face” terão igual probabilidade, ou seja, $1/2$. Se não, haverá uma diferença de probabilidade que a experiência acusará.

Para ilustrar as considerações precedentes, convém apresentar aqui alguns exemplos numéricos, que, no fundo, se reduzem geralmente a problemas de Cálculo Combinatório.

Exemplo 1 – Determinar a probabilidade de que, no lançamento dum dado perfeito, se obtenha um múltiplo de 3.

São 2 os casos favoráveis ao acontecimento: $x = 3$, $x = 6$. Visto haver ao todo 6 casos possíveis (igualmente prováveis), a probabilidade pedida será

$$\frac{2}{6} = \frac{1}{3}.$$

Exemplo 2 – Determinar a probabilidade de que, em dois lançamentos sucessivos dum dado perfeito, se obtenham 2 números pares.

As células, neste caso, podem ser representadas pelos pares ordenados (x, y) , em que x designa o número saído no 1.º lançamento e y o número saído no 2.º lançamento. Estes pares são em número de $6^2 = 36$ e todos igualmente prováveis, como é evidente. Os casos que implicam o acontecimento “saída de dois números pares” são: $(2,2)$, $(2,4)$, $(2,6)$, $(4,2)$, $(4,4)$, $(4,6)$, $(6,2)$, $(6,4)$, $(6,6)$, em número de 9. A probabilidade pedida será, pois,

$$\frac{9}{36} = \frac{1}{4}.$$

NOTA. Como se disse em Matemáticas Gerais, chama-se produto cartesiano dum conjunto A por um conjunto B , e designa-se por $A \times B$, o conjunto de todos os pares ordenados (a, b) que se obtém tomando um elemento a em A e um elemento b em B ; o número de elementos de $A \times B$ será, então, o produto dos números de elementos de A e de B . Analogamente, chama-se produto cartesiano de 3

conjuntos A , B , C , na ordem por que estão escritos, e designa-se por $A \times B \times C$, o conjunto de todos os ternos ordenados (a, b, c) com $a \in A$, $b \in B$, $c \in C$; o número destes ternos é igual ao produto dos números de elementos de A , B , e C . E assim por diante.

Também convencionámos escrever A^2 como abreviatura de $A \times A$, A^3 como abreviatura de $A \times A \times A$, etc.

Deste modo, os casos possíveis, no exemplo 2, serão os elementos do produto cartesiano

$$\{1, 2, 3, 4, 5, 6\} \times \{1, 2, 3, 4, 5, 6\} = \{1, 2, 3, 4, 5, 6\}^2$$

e os casos que implicam o acontecimento considerado serão os elementos do produto cartesiano $\{2, 4, 6\}^2$.

Note-se que os pares ordenados, os ternos ordenados, etc., também se chamam *arranjos com repetição* (*dois a dois, três a três*, etc.).

Exemplo 3 – Determinar a probabilidade de que, em dois lançamentos sucessivos dum dado perfeito, saia pelo menos uma vez número ímpar.

O acontecimento “sair pelo menos uma vez número ímpar” é o contrário do acontecimento “sair duas vezes número par” cuja probabilidade, há pouco calculada, é $1/4$. A probabilidade pedida será, pois,

$$1 - \frac{1}{4} = \frac{3}{4}.$$

Exemplo 4 – Determinar a probabilidade de que, em dois lançamentos sucessivos dum dado perfeito, saia primeiro um número par e depois um múltiplo de 3.

Os casos possíveis (células) são os mesmos que no exemplo 2 e no exemplo 3. Os casos favoráveis ao acontecimento considerado obtêm-se desenvolvendo o produto cartesiano

$$\{2, 4, 6\} \times \{3, 6\},$$

cujos elementos são em número de $3 \times 2 = 6$. A probabilidade pedida será, pois, $6/36 = 1/6$.

Exemplo 5 – *Determinemos a probabilidade de que, em dois lançamentos sucessivos dum dado perfeito, saia uma vez um número par e outra vez um múltiplo de 3.*

O acontecimento considerado é a soma lógica dos dois seguintes: “sair primeiro número par e depois um múltiplo de 3” e “sair primeiro um múltiplo de 3 e depois um número par”. Estes dois acontecimentos, que designaremos por α_1 e α_2 , têm ambos a probabilidade $1/6$ (exemplo 4), *mas não são incompatíveis*. O acontecimento $\alpha_1\alpha_2$ equivale a “sair 6 duas vezes”; como se trata de um só caso entre 36 possíveis, a sua probabilidade é de $1/36$. A probabilidade pedida será, pois (teorema 3):

$$\Pr(\alpha_1) + \Pr(\alpha_2) - \Pr(\alpha_1\alpha_2) = 2 \cdot \frac{1}{6} - \frac{1}{36} = \frac{11}{36}.$$

NOTA. Nos exemplos anteriores, excepto o primeiro, as considerações ficam essencialmente as mesmas, se substituirmos a expressão “em dois lançamentos sucessivos dum dado perfeito” pela expressão “num lançamento simultâneo de dois dados perfeitos”.

Exemplo 6 – *Duma urna que contem 12 bolas, das quais 4 são brancas e 8 pretas, tiram-se duas bolas à sorte, uma após a outra, sem reposição. Calcular a probabilidade de que: a) sejam ambas brancas; b) sejam ambas pretas; c) sejam da mesma cor; d) sejam de cores diferentes; e) sejam a 1.ª branca e a 2.ª preta; f) sejam a 2.ª preta e a 1.ª branca.*

Suponhamos as bolas numeradas de 1 a 12, sendo as 4 primeiras brancas; designemos por x o primeiro número saído e por y o segundo. Neste caso, os acontecimentos celulares são representados pelos pares ordenados, (x, y) , com $x \neq y$; trata-se, portanto, de arranjos (sem repetição!) de 12 elementos 2 a 2; o número desses arranjos é $12 \times 11 = 132$.

O acontecimento considerado em a) corresponde aos pares (x, y) , em que x e y variam de 1 a 4, mas sendo $x \neq y$. Trata-se, portanto, de arranjos de 4 elementos 2 a 2, cujo número é $4 \times 3 = 12$. A probabilidade pedida será, pois, $12/132$.

Analogamente, se reconhece que a probabilidade pedida em b) é

$$\frac{8 \times 7}{132} = \frac{56}{132}.$$

O acontecimento considerado em c) é a soma lógica dos acontecimentos anteriores, que são incompatíveis. Portanto, a probabilidade pedida em c) será

$$\frac{12}{132} + \frac{56}{132} = \frac{68}{132}.$$

O acontecimento considerado em d) é o contrário do anterior. Logo, a sua probabilidade será

$$1 - \frac{68}{132} = \frac{64}{132}.$$

É fácil reconhecer, finalmente, que as probabilidades pedidas em e) e f) são ambas iguais a

$$\frac{8 \times 4}{132} = \frac{32}{132};$$

a sua soma dá a anterior, como era de esperar.

Exemplo 7 – Duma urna que contem N bolas, sendo a brancas e b pretas, extrai-se, ao acaso, uma amostra de n bolas (tiradas, por exemplo, uma após outra, sem reposição). Achar a probabilidade de que, na amostra obtida, haja v bolas brancas e $n-v$ bolas pretas.

Começemos por notar que a ordem pela qual são tiradas as bolas não interessa à questão. Os casos possíveis (todos igualmente prováveis) serão, pois, as combinações das N bolas tomadas n a n , cujo número é $\binom{N}{n}$. Por sua vez, as amostras com v bolas brancas e $n-v$ bolas pretas podem obter-se, sem omissão nem repetição, do seguinte modo:

1) – Formando, por um lado, as combinações das a bolas brancas tomadas v a v_1 cujo número é $\binom{a}{v}$.

2) – Formando, por outro lado, as combinações das b bolas pretas tomadas $n-v$ a $n-v$, cujo número é $\binom{b}{n-v}$.

3) – Arranjando todos os possíveis pares $A_v B_{n-v}$ constituídos por uma combinação A_v obtida em 1) e uma combinação B_{n-v} obtida em 2). Como o número destes pares é $\binom{a}{v}\binom{b}{n-v}$, a probabilidade pedida será

$$\frac{\binom{a}{v}\binom{b}{n-v}}{\binom{N}{n}}.$$

Esta distribuição de variável casual v é chamada *distribuição hipergeométrica*.

Exemplo 8 – Duma urna que contem N bolas, sendo a brancas e b pretas, tiram-se, ao acaso, sucessivamente, n bolas, repondo a bola retirada após cada extracção. Achar a probabilidade de que saia v vezes bola branca e $n-v$ vezes bola preta.

Discorrendo como há pouco, seríamos levados a considerar, agora, os casos possíveis sob a forma de *combinações com repetição*. Porém, esses casos *já não são igualmente prováveis, como é fácil ver*. Suponhamos, por exemplo, $n = 3$; neste caso, a combinação $x_1 x_2 x_3$ será mais provável que a combinação $x_1 x_1 x_3$, pois que a primeira se pode apresentar com as seguintes ordens de saída de bolas:

$$x_1 x_2 x_3, x_1 x_3 x_2, x_2 x_1 x_3, x_2 x_3 x_1, x_3 x_1 x_2, x_3 x_2 x_1,$$

ao passo que a segunda se pode apresentar só dos seguintes modos:

$$x_1 x_1 x_3, x_1 x_3 x_1, x_3 x_1 x_1.$$

Este exemplo mostra, bem, o cuidado que é necessário ter na escolha dos acontecimentos celulares, para que sejam igualmente prováveis. Neste caso, há que tomar para células os arranjos com repetição das N bolas n a n (isto é, sistema de n bolas).

Veremos, mais adiante, como se resolve este problema, considerado sob um aspecto mais geral (distribuição de BERNOULLI).

Exemplo 9 – *Nas condições do exemplo 7, achar a probabilidade de que, numa amostra, o número v de bolas brancas: a) não seja inferior a um dado número L_1 nem superior a um dado número L_2 ; b) seja inferior a L_1 ou superior a L_2 .*

Em a) trata-se de achar a probabilidade do acontecimento

$$L_1 \leq v \leq L_2$$

o qual é a soma lógica dos acontecimentos

$$v = L_1, v = L_1 + 1, \dots, v = L_2 - 1, v = L_2,$$

incompatíveis dois a dois. Ter-se-á, portanto,

$$\Pr(L_1 \leq v \leq L_2) = \Pr(v = L_1) + \Pr(v = L_1 + 1) + \dots + \Pr(v = L_2),$$

ou seja, atendendo ao resultado do exemplo 7:

$$\Pr(L_1 \leq v \leq L_2) = \sum_{v=L_1}^{L_2} \frac{\binom{a}{v} \binom{b}{n-v}}{\binom{N}{n}}.$$

Em b) pede-se a probabilidade do acontecimento

$$(v < L_1) + (L_2 < v) \quad (v \text{ inferior a } L_1 \text{ ou } L_2 \text{ inferior a } v)$$

que é o contrário do acontecimento $L_1 \leq v \leq L_2$. Será, pois,

$$\Pr[(v < L_1) + (v > L_2)] = 1 - \Pr(L_1 \leq v \leq L_2).$$

Eis como se poderiam calcular exactamente as probabilidades que fomos levados a considerar em A-16, a propósito da experiência hipotética sobre animais. Porém, estas fórmulas, embora simples na aparência, exigem cálculos muito laboriosos. É possível substituí-las por outras que, com muito menos trabalho, fornecem boas aproximações quando o número n excede um certo limite.

NOTA IMPORTANTE RELATIVA ÀS NOTAÇÕES E TERMINOLOGIA

Convencionámos no n.º 4 designar por $\Pr(\alpha)$ a probabilidade do acontecimento α . Nesta ordem de ideias, representámos, atrás, por $\Pr(x=x_1)$ a probabilidade de “sair a bola x_1 ”, por $\Pr(x \in V)$ a probabilidade de “sair bola vermelha”, etc. Mas poderíamos, igualmente, para abreviar, designar por $\Pr(x_1)$, $\Pr(V)$, etc., aquelas probabilidades, dizendo “a probabilidade do elemento x_1 ”, a “probabilidade do conjunto V ”, etc. Assim, a distribuição de probabilidade passa a conceber-se como distribuição definida, não num corpo de acontecimentos, mas, sim, num corpo de conjuntos: todos os possíveis conjuntos de bolas da urna (cf. A-4). Poderíamos, também, conceber esta distribuição como função $\Pr(x)$ da variável casual x , tendo-se, na hipótese de as bolas serem iguais,

$$\Pr(x) = \frac{1}{N}, \text{ para todo o valor de } x.$$

Estas considerações estendem-se, *mutatis mutandis*, a qualquer distribuição de probabilidade.

6. Independência e associação de acontecimentos

Sendo as probabilidades valores ideais de frequências relativas, as definições de independência e associação que demos para o caso das frequências traduzem-se, imediatamente, em termos de probabilidade. Seja \mathcal{R} um corpo de eventualidades a considerar numa dada prova \mathcal{P} , com determinadas probabilidades. Dados dois acontecimentos α , β , chama-se *probabilidade condicional de β em relação a α* e designa-se por $\Pr(\beta|\alpha)$ ao número dado pela fórmula

$$(6.1) \quad \Pr(\beta|\alpha) = \frac{\Pr(\alpha\beta)}{\Pr(\alpha)}.$$

Para avaliar $\Pr(\beta|\alpha)$ empiricamente, o processo a seguir consistiria em efectuar a prova \mathcal{P} um grande número de vezes e registar: 1) a frequência absoluta (α) de α ; 2) a frequência absoluta ($\alpha\beta$) de $\alpha\beta$. O quociente de ($\alpha\beta$) por (α) dar-nos-ia, então, um valor aproximado de $\Pr(\beta|\alpha)$.

É claro que, da definição (6.1), resulta, logo, a *fórmula do produto*:

$$(6.2) \quad \Pr(\alpha\beta) = \Pr(\alpha) \cdot \Pr(\beta|\alpha) = \Pr(\beta) \cdot \Pr(\alpha|\beta),$$

a qual nos diz que: *a probabilidade de se verificarem ao mesmo tempo os acontecimentos α e β é o produto da probabilidade de α pela probabilidade condicional de β na hipótese de se verificar α (ou vice-versa).*

Os acontecimentos α , β dizem-se *estocasticamente independentes* ou, apenas, *independentes*, quando $\Pr(\beta|\alpha) = \Pr(\beta)$ ou, o que é equivalente, quando $\Pr(\alpha|\beta) = \Pr(\alpha)$; no caso contrário, dizem-se *associados*. Desta definição e da fórmula do produto deduz-se, logo, o

TEOREMA DO PRODUTO. *Se os acontecimentos α , β são independentes (e só neste caso), a probabilidade de realização simultânea de α e β é igual ao produto das probabilidades de α e de β .*

Suponhamos, por exemplo, que uma urna contém 8 bolas brancas e 24 pretas, estando 6 bolas brancas marcadas com o sinal +, 18 bolas pretas com este mesmo sinal, e todas as restantes com o sinal –. Então, como existem na urna 32 bolas, ao todo, sendo 24 marcadas com o sinal +, a probabilidade (incondicional) de aparecer o sinal + é $24/32 = 3/4$. Por sua vez, a probabilidade de aparecer o sinal + em bola branca é $6/8 = 3/4$, igual à primeira. Os acontecimentos “sair bola branca” e “sair sinal +” são, pois, independentes; deste modo, a probabilidade de “sair bola branca e sinal +” é:

$$\frac{8}{32} \cdot \frac{3}{4} = \frac{1}{4} \cdot \frac{3}{4} = \frac{3}{16}$$

(probabilidade de sair bola branca vezes probabilidade de sair sinal +).

Um exemplo mais sugestivo, embora menos rigoroso, será o seguinte. Consideremos a prova que consiste em *semear uma certa quantidade de trigo num certo campo*. Seja p a probabilidade (incondicional) *de colher então uma quantidade de sementes superior a um dado limite L* e seja p' a probabilidade do mesmo acontecimento, *na hipótese de a altura pluviométrica, nos meses de outono e inverno, ser inferior a um dado limite λ* . Será, então, p' uma probabilidade condicional do primeiro acontecimento em relação ao segundo; e é natural que seja $p' \neq p$, isto é, que a eventualidade “colher trigo em quantidade superior a L ” depende da eventualidade “chover em quantidade inferior a λ ”.

Mas, geralmente, dados vários acontecimentos α , β , γ , ... do corpo \mathcal{R} tem-se, como é fácil ver,

$$(6.3) \quad \Pr(\alpha\beta\gamma \dots) = \Pr(\alpha) \Pr(\beta|\alpha) \Pr(\gamma|\alpha\beta) \dots$$

Os acontecimentos α , β , γ , ... são independentes, se for verdadeira, não só a igualdade

$$\Pr(\alpha\beta\gamma \dots) = \Pr(\alpha) \Pr(\beta) \Pr(\gamma) \dots$$

como todas as que resultam desta substituindo um ou mais dos acontecimentos α , β , γ , ... pelos seus contrários (cf. A-15)⁽¹⁾.

7. Sistema de duas experiências

Os conceitos de associação e independência de acontecimentos têm interesse, principalmente, quando se trata de várias experiências combinadas. Começemos por considerar o caso das duas experiências \mathcal{P} , \mathcal{P}' , iguais ou diferentes, e sejam \mathcal{R} e \mathcal{R}' dois corpos de eventualidades a considerar, respectivamente, nas provas \mathcal{P} e \mathcal{P}' . A realização destas duas provas, ao mesmo tempo ou uma após a outra, pode

(1) – Pode considerar-se esta proposição como definição de independência, no caso de vários acontecimentos.

conceber-se como *experiência* composta, única, que designaremos por $(\mathcal{P}, \mathcal{P}')$. Sejam $\alpha_1, \alpha_2, \dots, \alpha_m$ as células de \mathcal{R} e $\beta_1, \beta_2, \dots, \beta_n$ as células de \mathcal{R}' . Designaremos, então, por

$$(\alpha_i, \beta_k)$$

o acontecimento que consiste em realizar-se α_i na prova \mathcal{P} e β_k na prova \mathcal{P}' (*acontecimento composto de α_i e β_k*). Todos estes acontecimentos (α_i, β_k) são resultados possíveis da prova $(\mathcal{P}, \mathcal{P}')$, são as células dum novo corpo, que designaremos por $\mathcal{R} \times \mathcal{R}'$. Note-se que cada acontecimento α_i pode ser considerado como elemento do corpo $\mathcal{R} \times \mathcal{R}'$, pondo

$$\alpha_i = (\alpha_i, \beta_1) + (\alpha_i, \beta_2) + \dots + (\alpha_i, \beta_n),$$

visto que, $\beta_1 + \beta_2 + \dots + \beta_n = I$. Analogamente:

$$\beta_k = (\alpha_1, \beta_k) + (\alpha_2, \beta_k) + \dots + (\alpha_m, \beta_k).$$

Nesta ordem de ideias, podemos identificar o acontecimento composto (α_i, β_k) com o produto lógico $\alpha_i \beta_k$, desde que não haja perigo de confusão⁽¹⁾.

Suponhamos, agora, definido no corpo $\mathcal{R} \times \mathcal{R}'$ uma distribuição de probabilidade. Esta poderá indicar-se numa tábua do seguinte tipo (cf. tabela-tipo de A-14):

	β_1	β_2	β_n	Total
α_1	$\text{Pr}(\alpha_1 \beta_1)$	$\text{Pr}(\alpha_1 \beta_2)$	$\text{Pr}(\alpha_1 \beta_n)$	$\text{Pr}(\alpha_1)$
α_2	$\text{Pr}(\alpha_2 \beta_1)$	$\text{Pr}(\alpha_2 \beta_2)$	$\text{Pr}(\alpha_2 \beta_n)$	$\text{Pr}(\alpha_2)$
.....
α_m	$\text{Pr}(\alpha_m \beta_1)$	$\text{Pr}(\alpha_m \beta_2)$	$\text{Pr}(\alpha_m \beta_n)$	$\text{Pr}(\alpha_m)$
Total	$\text{Pr}(\beta_1)$	$\text{Pr}(\beta_2)$	$\text{Pr}(\beta_n)$	1

(1) – A confusão pode surgir quando \mathcal{P}' é apenas a repetição de \mathcal{P} . Neste caso, um acontecimento α deverá ser designado por dois símbolos diferentes, conforme se realize em \mathcal{P} ou em \mathcal{P}' . Um outro modo de evitar a confusão é chamar a (α_i, β_k) o *produto cartesiano de α_i por β_k* . Este produto, *que não é comutativo*, será designado por $\alpha_i \times \beta_k$ para o distinguir do produto lógico vulgar $\alpha_i \beta_k$.

Note-se que, para $i = 1, \dots, m$, $k = 1, 2, \dots, n$, se tem:

$$\Pr(\alpha_i) = \Pr(\alpha_i \beta_1) + \Pr(\alpha_i \beta_2) + \dots + \Pr(\alpha_i \beta_n),$$

$$\Pr(\beta_k) = \Pr(\alpha_1 \beta_k) + \Pr(\alpha_2 \beta_k) + \dots + \Pr(\alpha_m \beta_k).$$

Como estas probabilidades estão registadas à margem (sob a indicação “total”), dá-se-lhes, também, o nome de *probabilidades marginais* da distribuição considerada. Pode acontecer, em particular, que se tenha

$$\Pr(\alpha_i \beta_k) = \Pr(\alpha_i) \Pr(\beta_k),$$

quaisquer que sejam i, k . Então, é fácil ver que *a probabilidade de qualquer acontecimento $\alpha\beta$, composto dum acontecimento α de \mathcal{R} e dum acontecimento β de \mathcal{R}' será igual ao produto das probabilidades de α e de β ⁽¹⁾. Os acontecimentos de \mathcal{R} dir-se-ão, neste caso, *independentes* dos acontecimentos de \mathcal{R}' . De contrário, terá de usar-se a fórmula (6.2).*

Estas considerações tornam-se mais intuitivas quando se fala em termos de variáveis casuais. Consideremos um par (x, y) de variáveis casuais com uma determinada distribuição de probabilidade, $\Pr(x, y)$. Chama-se *probabilidade condicional dum valor x_i de x a respeito dum valor y_k de y* , e representa-se por $\Pr(x_i|y_k)$, o número dado por

$$\Pr(x_i|y_k) = \frac{\Pr(x_i, y_k)}{\Pr(y_k)}.$$

As variáveis x, y dizem-se *independentes (estocasticamente)*, se for $\Pr(x|y) = \Pr(x)$ para todos os valores de x e y . Tem-se, pois, nesta hipótese

$$\Pr(x, y) = \Pr(x) \Pr(y).$$

(1) – Esta regra, ou, mais geralmente, a fórmula (6.3), é conhecida tradicionalmente, como *princípio das probabilidades compostas*.

NOTA. Seria mais correcto designar por símbolos diferentes as distribuições de probabilidade de x e y : por exemplo, a primeira por $\text{Pr}_1(x)$ e a segunda por $\text{Pr}_2(y)$. Mas, para não sobrecarregar as notações, e não havendo perigo de confusão, usamos aqui, para ambas as distribuições, o mesmo símbolo.

Exemplos – Consideremos uma urna que contenha M bolas x_1, x_2, \dots, x_M , sendo as v primeiras vermelhas e as b últimas brancas. E consideremos uma outra urna, que contenha N bolas y_1, y_2, \dots, y_N , sendo as v' primeiras vermelhas e as b' últimas brancas. Sejam \mathcal{P} , \mathcal{P}' , respectivamente, as experiências que consistem em tirar à sorte uma bola da primeira urna e tirar à sorte uma bola da segunda urna. Neste caso, os resultados elementares da experiência composta $(\mathcal{P}, \mathcal{P}')$ serão expressos pelos diferentes valores (x_i, y_k) da variável casual (x, y) . Como estes valores são em número de MN e *todos igualmente prováveis*, a probabilidade de cada par (x_i, y_k) será

$$\frac{1}{MN} = \frac{1}{M} \cdot \frac{1}{N} = \text{Pr}(x_i) \text{Pr}(y_k),$$

o que significa que *as duas variáveis casuais x, y são independentes (estocasticamente)*. Os resultados da prova \mathcal{P}' são pois independentes dos resultados da prova \mathcal{P} . Por exemplo, a probabilidade do acontecimento *sair bola branca em \mathcal{P} e bola vermelha em \mathcal{P}'* , será

$$\frac{b}{M} \cdot \frac{v'}{N} = \frac{bv'}{MN}.$$

Casos análogos ao anterior serão, ainda, todos aqueles em que \mathcal{P}' é simplesmente a repetição de \mathcal{P} . Suponhamos, por exemplo, que se trata de duas extracções sucessivas duma bola da 1.^a urna, *com reposição*. A probabilidade de sair, em primeiro lugar, bola vermelha e, depois, bola branca será

$$\frac{v}{M} \cdot \frac{b}{M} = \frac{bv}{M^2}.$$

igual à probabilidade de sair, primeiro, branca e, depois, vermelha (será, portanto, $2bv/M^2$ a probabilidade do acontecimento “saírem cores diferentes”).

Mas suponhamos, agora, que \mathcal{P} consiste em tirar uma bola da 1.^a urna, e \mathcal{P}' em tirar uma segunda bola da mesma urna *sem repor a primeira bola tirada*. Os resultados elementares da prova composta (\mathcal{P} , \mathcal{P}') serão, ainda, os valores da variável casual (x, y) , *mas com a condição suplementar $x \neq y$* . Isto basta para ver que, neste caso, as variáveis casuais x, y não são independentes. Por exemplo, a probabilidade (condicional) de sair bola vermelha na 2.^a extracção, *tendo saído branca na 1.^a*, é

$$\frac{v}{M-1},$$

ao passo que a probabilidade (condicional) de sair vermelha na 2.^a, *tendo saído vermelha na 1.^a* é

$$\frac{v-1}{M-1}.$$

Assim, a probabilidade de *sair vermelha duas vezes* é

$$\frac{v}{M} \cdot \frac{v-1}{M-1} = \frac{v(v-1)}{M(M-1)}$$

e a de *sair primeiro branca e depois vermelha*:

$$\frac{b}{M} \cdot \frac{v}{M-1} = \frac{bv}{M(M-1)}.$$

Estes resultados podem recolher-se na seguinte tabela:

TABELA N.º 13

1. ^a \ 2. ^a	Vermelha	Branca	Total
Vermelha	$\frac{v(v-1)}{M(M-1)}$	$\frac{bv}{M(M-1)}$	$\frac{v}{M}$
Branca	$\frac{bv}{M(M-1)}$	$\frac{b(b-1)}{M(M-1)}$	$\frac{b}{M}$
Total	$\frac{v}{M}$	$\frac{b}{M}$	1

Na margem direita estão indicadas as probabilidades de sair bola vermelha e a de sair bola branca na 2.^a extracção, respectivamente, iguais às probabilidades de sair vermelha e de sair branca na 1.^a extracção (registadas na margem inferior). A tabela patenteia, assim, a associação dos acontecimentos considerados.

8. Sistema de várias experiências

Podemos, agora, conceber, mais geralmente, uma experiência $(\mathcal{P}, \mathcal{P}', \mathcal{P}'', \dots)$, composta de várias experiências $\mathcal{P}, \mathcal{P}', \mathcal{P}'', \dots$, (em número finito). Sendo $\alpha, \beta, \gamma, \dots$ eventualidades a considerar em cada uma destas experiências, designaremos por $(\alpha, \beta, \gamma, \dots)$ a eventualidade que consiste em acontecer α em \mathcal{P} , β em \mathcal{P}' , γ em \mathcal{P}'' , etc., e diremos que $(\alpha, \beta, \gamma, \dots)$ é o *acontecimento composto* de $\alpha, \beta, \gamma, \dots$. Porém, segundo o ponto de vista explanado no número anterior, podemos considerar este novo acontecimento como o produto lógico dos acontecimentos $\alpha, \beta, \gamma, \dots$, desde que se tomem as devidas precauções⁽¹⁾. Suponhamos que, a todo o acontecimento a considerar na prova $(\mathcal{P}, \mathcal{P}', \mathcal{P}'', \dots)$, corresponde uma determinada probabilidade. Então, o que se disse em B-6 é aqui aplicável: os acontecimentos $\alpha, \beta, \gamma, \dots$ dizem-se *independentes (estocasticamente)*, se for verdadeira não só a igualdade

$$(8.1) \quad \Pr(\alpha\beta\gamma \dots) = \Pr(\alpha) \Pr(\beta) \Pr(\gamma) \dots,$$

como todas as que se deduzem desta substituindo um ou mais dos acontecimentos $\alpha, \beta, \gamma, \dots$ pelos seus contrários. Não sendo assim, terá de usar-se a fórmula geral do produto, com as probabilidades condicionais.

Poderíamos, de novo, dar aqui exemplos de tiragens de bolas de uma ou várias urnas, com ou sem reposição, mas as considerações do número seguinte bastam para esclarecer as precedentes.

(1) – Isto é, se $\mathcal{P}', \mathcal{P}'', \dots$ consistem apenas na repetição de \mathcal{P} , um mesmo acontecimento deverá ser designado de modos diversos, conforme as provas em que se realiza. É o que faremos mais adiante.

9. Distribuição binomial ou de BERNOULLI

Seja α um acontecimento a esperar com determinada probabilidade numa certa prova \mathcal{P} e consideremos n realizações $\mathcal{P}_1, \mathcal{P}_2, \dots, \mathcal{P}_n$, da prova genérica $\mathcal{P}^{(1)}$. Estas provas particulares constituem uma experiência composta $(\mathcal{P}_1, \mathcal{P}_2, \dots, \mathcal{P}_n)$. Tendo em vista o conceito empírico de probabilidade, é fácil reconhecer que a probabilidade do acontecimento α deverá ser a mesma em cada uma das provas, quaisquer que sejam os resultados das restantes provas. Por outras palavras:

Os acontecimentos a esperar na repetição duma dada prova são independentes dos resultados das provas já efectuadas.

É claro que este facto não se deduz da anterior axiomática das probabilidades. Poderia, sim, assumir-se como novo axioma, porém com carácter bem diverso do dos primeiros.

Um exemplo típico é o das sucessivas extracções casuais de bolas de uma urna, com reposição da bola após cada extracção.

Pode, pois, aplicar-se nestes casos a fórmula (8.1).

Proponhamo-nos, então, resolver o seguinte problema:

Determinar a probabilidade de que, em n realizações da prova \mathcal{P} , um acontecimento α de probabilidade p se realize x vezes.

(Para concretizar, podemos supor que α consiste em tirar à sorte uma bola duma urna com bolas brancas e pretas, sendo α o acontecimento “sair bola branca”. Porém, as considerações que se seguem são de todo gerais).

Representemos por q a probabilidade de $\tilde{\alpha}$, isto é, ponhamos $q = 1 - p$. Para evitar confusões, designaremos por α_i o acontecimento particular que consiste em realizar-se α na prova \mathcal{P}_i ($i = 1, 2, \dots, n$); é claro que será sempre

$$\Pr(\alpha_i) = p, \Pr(\tilde{\alpha}_i) = q.$$

Suponhamos, por exemplo, $n = 3$ e $x = 2$. Três são os modos de α se realizar 2 vezes numa série de 3 provas: na 1.^a e na 2.^a, na 1.^a e na 3.^a ou na 2.^a e na 3.^a; isto é, em símbolos:

$$\alpha_1\alpha_2\tilde{\alpha}_3, \quad \alpha_1\tilde{\alpha}_2\alpha_3, \quad \tilde{\alpha}_1\alpha_2\alpha_3.$$

(1) – Dizendo que $\mathcal{P}_1, \mathcal{P}_2, \dots, \mathcal{P}_n$ são realizações da mesma prova \mathcal{P} , fica implícito que $\mathcal{P}_1, \mathcal{P}_2, \dots, \mathcal{P}_n$ se realizam em condições idênticas.

O acontecimento que consiste em α se realizar 2 vezes nas 3 provas, será, pois, a soma lógica

$$\alpha_1\alpha_2\tilde{\alpha}_3 + \alpha_1\tilde{\alpha}_2\alpha_3 + \tilde{\alpha}_1\alpha_2\alpha_3.$$

Ora, as três modalidades consideradas são incompatíveis duas a duas e têm todas as mesmas probabilidades; por exemplo:

$$\Pr(\alpha_1\tilde{\alpha}_2\alpha_3) = \Pr(\alpha_1) \Pr(\tilde{\alpha}_2) \Pr(\alpha_3) = pqp = p^2q.$$

A probabilidade pedida será, pois, neste caso, $3p^2q$.

Passemos, agora, ao caso geral. Um dos modos de α se realizar x vezes nas n provas será

$$\alpha_1\alpha_2 \dots \alpha_x\tilde{\alpha}_{x+1} \dots \tilde{\alpha}_n.$$

Qualquer outra modalidade se obtém escolhendo, entre as n provas, as x provas em que se realize α . As modalidades possíveis correspondem, pois, às combinações das n provas x a x . Como o número total destas combinações é $\binom{n}{x}$, o acontecimento que consiste em α se realizar x vezes nas n provas é a soma lógica de $\binom{n}{x}$ eventualidades, incompatíveis duas a duas e todas com a mesma probabilidade, que é

$$\Pr(\alpha_1) \Pr(\alpha_2) \dots \Pr(\alpha_x) \Pr(\tilde{\alpha}_{x+1}) \dots \Pr(\tilde{\alpha}_n) = p^x q^{n-x}.$$

A probabilidade pedida será, pois,

$$\Pr(x) = \binom{n}{x} p^x q^{n-x}$$

fórmula esta muito importante. Note-se que a variável casual x representa, aqui, a frequência absoluta de α numa série de n provas. A distribuição de probabilidade desta variável, dada pela fórmula anterior, tem o nome de *distribuição de BERNOULLI*. Também se

Ihe chama distribuição *binominal*, atendendo a que os diferentes valores de $\Pr(x)$ são os termos do desenvolvimento da potência n do binómio $p + q$:

$$(p + q)^n = \sum_{x=0}^n \binom{n}{x} p^x q^{n-x}.$$

É claro que $(p + q)^n = 1$. Procuremos, agora, a probabilidade de que a frequência absoluta de α seja inferior ou igual a um dado limite L (com $L \leq n$). Como o acontecimento $x \leq L$ é a soma lógica dos acontecimentos incompatíveis $x = 0, x = 1, \dots, x = L$, virá:

$$\Pr(x \leq L) = \Pr(x = 0) + \Pr(x = 1) + \dots + \Pr(x = L) = \sum_{x=0}^L \binom{n}{x} p^x q^{n-x}.$$

Analogamente, se reconhece que

$$\Pr(L_1 \leq x \leq L_2) = \sum_{x=L_1}^{L_2} \binom{n}{x} p^x q^{n-x}.$$

(Confrontar com os exemplos 7, 8 e 9 do n.º 5).

Se pusermos $\xi = x/n$ (frequência relativa de α nas n provas), a distribuição da variável ξ será, manifestamente:

$$\Pr(\xi) = \binom{n}{n\xi} p^{n\xi} q^{n(1-\xi)}.$$

A tabela n.º 14 representa a distribuição binominal para $p = 1/2$, $n = 10$. Por exemplo, a probabilidade de que em 10 lançamentos sucessivos duma moeda “correcta” se apresente 3 vezes coroa, é

$$\Pr(3) = \binom{10}{3} \left(\frac{1}{2}\right)^3 \left(\frac{1}{2}\right)^7 = \left(\frac{10 \cdot 9 \cdot 8}{3 \cdot 2 \cdot 1}\right) \cdot \left(\frac{1}{2^{10}}\right) = \frac{15}{128},$$

valor que concorda com o correspondente da tabela (a menos de 0,001).

TABELA N.º 14

x	$\text{Pr}(x)$	x	$\text{Pr}(x)$
0	0,001	6	0,205
1	0,010	7	0,117
2	0,044	8	0,044
3	0,117	9	0,010
4	0,205	10	0,001
5	0,246		

Na Fig. 3 é dado o histograma desta distribuição.

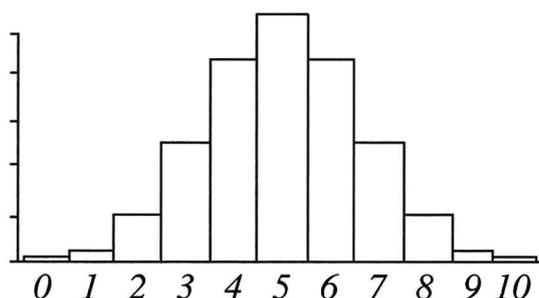


Fig. 3

Note-se que o histograma é *simétrico* a respeito da recta $x = 5$ e que a função $\text{Pr}(x)$ atinge um máximo no ponto 5.

Podemos ver um exemplo curioso desta distribuição no campo da genética. Sabe-se que, no cruzamento dum touro avermelhado do Shorthorn com uma vaca malhada da mesma raça, há a probabilidade de $1/2$ de se obter um vitelo avermelhado sem malhas⁽¹⁾. Então, a probabilidade de que, em 10 destes cruzamentos, se obtenham x vitelos avermelhados (sem malhas), é dada pela tabela n.º 14.

(1) – Traduz-se por “touro avermelhado” a expressão inglesa “red bull” e por “vaca malhada” a expressão “roan cow”. Neste caso, “malhado” significa “avermelhado, com malhas brancas ou cinzentas dispersas”.

Por sua vez, sabe-se que a probabilidade de que, no cruzamento dum touro malhado Shorthorn com uma vaca malhada da mesma raça se obtenha um vitelo avermelhado (sem malhas) é $1/4$. Então, a probabilidade de que, em 10 destes cruzamentos, se obtenham x vitelos avermelhados (sem malhas) será:

$$\Pr(x) = \binom{10}{x} \left(\frac{1}{4}\right)^x \left(\frac{3}{4}\right)^{10-x}.$$

Na Fig. 4 é dado o histograma desta distribuição, que já não apresenta simetria.

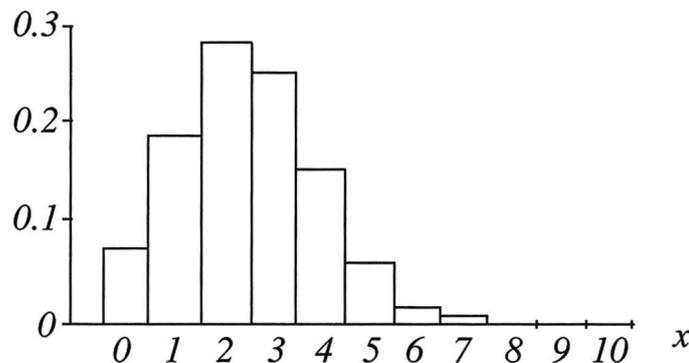


Fig. 4

Os dois casos extremos na distribuição binomial são os seguintes:

- 1) $x = n$ (o acontecimento α realiza-se nas n provas)
- 2) $x = 0$ (o acontecimento α não se realiza em nenhuma das provas).

A probabilidade do primeiro caso é (como se poderia reconhecer mesmo directamente):

$$\Pr(n) = p^n.$$

A probabilidade do segundo caso é:

$$\Pr(0) = q^n.$$

Note-se, porém, que a negação do acontecimento $x=0$ não é o acontecimentos $x=n$, mas sim o acontecimento que consiste em α se realizar pelo menos uma vez em n provas.

A probabilidade deste acontecimento será, pois,

$$\Pr(x \neq 0) = 1 - q^n = 1 - (1-p)^n = \sum_{x=1}^n (-1)^{x-1} \binom{n}{x} p^x.$$

Exemplo – Calcular a probabilidade de que, jogando 1.000 vezes num número duma lotaria com 30.000 números, se tenha pelo menos uma vez a sorte grande.

A probabilidade de, em cada extracção, se ter a sorte grande, é, manifestamente, $1/30.000$. Então, a probabilidade pedida será

$$1 - \left(1 - \frac{1}{30.000}\right)^{1.000} = \frac{1.000}{30.000} - \binom{1.000}{2} \frac{1}{(30.000)^2} + \dots \approx 0,0328$$

probabilidade esta ainda muito fraca, apesar do grande número de tentativas.

10. Conceito de moda. Caso da distribuição normal

Chama-se *moda* duma distribuição de probabilidade duma variável x (com um número finito de valores) todo o valor de x , cuja probabilidade seja superior ou igual à de qualquer outro valor de x . Há distribuições com uma só moda, distribuições com mais de uma moda e distribuições sem moda.

Vamos ver que a distribuição binominal

$$\Pr(x) = \frac{n!}{x! (n-x)!} p^x q^{n-x} \quad (\text{com } q = 1-p)$$

apresenta uma ou, quando muito, duas modas.

Considerando três valores consecutivos, $X-1$, X , $X+1$ da variável x , tem-se

$$\Pr(X-1) = \frac{n!}{(X-1)! (n-X+1)!} p^{X-1} q^{n-X+1}$$

$$\Pr(X) = \frac{n!}{X! (n-X)!} p^X q^{n-X}$$

$$\Pr(X+1) = \frac{n!}{(X+1)! (n-X-1)!} p^{X+1} q^{n-X-1}.$$

Então, virá, como é fácil verificar,

$$\frac{\Pr(X)}{\Pr(X-1)} = \frac{(n-X+1)p}{Xq},$$

$$\frac{\Pr(X)}{\Pr(X+1)} = \frac{(X+1)q}{(n-X)p}.$$

Por conseguinte, para que X seja uma moda, deve ser, simultaneamente,

$$\begin{aligned}(n-X+1)p &\geq Xq, \\ (X+1)q &\geq (n-X)p.\end{aligned}$$

Ora, esta dupla condição equivale à seguinte

$$np - q \leq X \leq np + p.$$

Como a diferença entre $np + p$ e $np - q$ é igual a $p + q = 1$, a anterior condição será verificada por um só valor inteiro de X (a moda), a não ser que $np + p$ e $np - q$ sejam números inteiros, que serão, nesse caso, as duas modas existentes.

Assim, a moda ou as modas da distribuição binominal são sempre números inteiros que diferem de np menos de uma unidade.

No primeiro exemplo atrás considerado ($p = 1/2$, $n = 10$), a moda é precisamente $np = 10 \cdot 1/2 = 5$. No segundo exemplo ($p = 1/4$, $n = 10$), a moda X deve verificar a condição

$$1,75 = 10 \cdot \frac{1}{4} - \frac{3}{4} \leq X \leq 10 \cdot \frac{1}{4} + \frac{1}{4} = 2,75;$$

só poderá, então, ser $X = 2$. Mas, se em vez de $n = 10$, tivéssemos tomado $n = 15$, com $p = 1/4$, já teríamos duas modas: $X_1 = 3$, $X_2 = 4$.

Temos aqui, pois, um exemplo duma distribuição de r variáveis casuais x_1, x_2, \dots, x_r . É claro que, por ser $x_1 + x_2 + \dots + x_r = n$, estas variáveis não são independentes, nem sequer algebricamente.

A referida distribuição diz-se *polinomial*, atendendo a que os valores de $\Pr(x_1, x_2, \dots, x_r)$ são os termos do desenvolvimento da potência n do polinómio $p_1 + p_2 + \dots + p_r$. Tem-se, com efeito, segundo a fórmula de LEIBNIZ:

$$(p_1 + p_2 + \dots + p_r)^n = \sum_{x_1 + \dots + x_r = n} \frac{n!}{x_1! x_2! \dots x_r!} p_1^{x_1} p_2^{x_2} \dots p_r^{x_r}.$$

Note-se que a distribuição binominal é um caso particular desta, correspondente a $r = 2$ (partição dicotómica). Tem-se, então, $x_2 = n - x_1$, donde

$$\frac{n!}{x_1! x_2!} = \frac{n!}{x_1! (n - x_1)!} = \binom{n}{x_1}.$$

Este resultado aplica-se em questões de *amostragem casual*. Chama-se *amostra casual* a todo o sistema (u_1, u_2, \dots, u_n) de n indivíduos tirados *ao acaso* duma dada população U .

Como já tivemos ocasião de verificar em exemplos, a amostragem casual pode fazer-se de dois modos: *com reposição ou sem reposição*. No primeiro caso, cada um dos indivíduos u_1, u_2, \dots, u_n é tirado e, em seguida, reposto na população, antes de se tirar o seguinte: *pode assim acontecer que um mesmo indivíduo apareça repetido na amostra*. No segundo caso, os indivíduos são tirados sucessivamente, sem regressarem à população após as tiragens.

Suponhamos dada no universo U uma partição em r atributos $\alpha_1, \alpha_2, \dots, \alpha_r$ e sejam a_1, a_2, \dots, a_r , respectivamente, as frequências absolutas de $\alpha_1, \alpha_2, \dots, \alpha_r$ em U . Suponhamos, ainda, que todos os indivíduos têm igual probabilidade de ser tirados. Então, a probabilidade de aparecer um indivíduo com o atributo α_i será, manifestamente,

$$p_i = \frac{a_i}{N}, \quad i = 1, 2, \dots, r,$$

em que $N = a_1 + a_2 + \dots + a_r$ (número de elementos de U).

Qual é, neste caso, a probabilidade de que, numa amostragem casual de n elementos, com reposição, as frequências absolutas de $\alpha_1, \alpha_2, \dots, \alpha_r$ sejam, respectivamente, x_1, x_2, \dots, x_r ?

A resposta é, segundo o que vimos,

$$\Pr(x_1, x_2, \dots, x_r) = \frac{n!}{x_1! x_2! \dots x_r!} \left(\frac{a_1}{N}\right)^{x_1} \left(\frac{a_2}{N}\right)^{x_2} \dots \left(\frac{a_r}{N}\right)^{x_r}.$$

Mas seja, agora, este outro problema:

Qual é a probabilidade de que, numa amostragem casual de n elementos, sem reposição, as frequências absolutas de $\alpha_1, \alpha_2, \dots, \alpha_r$, sejam, respectivamente, x_1, x_2, \dots, x_r ?

Raciocinando como no exemplo 7 do n.º 21, chega-se, agora, ao resultado

$$\Pr(x_1, x_2, \dots, x_r) = \frac{\binom{a_1}{x_1} \binom{a_2}{x_2} \dots \binom{a_r}{x_r}}{\binom{N}{n}}.$$

Eis aqui um novo exemplo de distribuição de r variáveis casuais, que, no caso particular $r = 2$, tem o nome de *distribuição hipergeométrica*.

É fácil ver que, se n é bastante pequeno em relação a N (isto é, se a razão n/N é bastante pequena), esta distribuição coincide, sensivelmente, com a anterior.

Note-se, ainda, que, na prática, as fórmulas obtidas exigem cálculos muito laboriosos quando n é grande. Têm de ser então substituídas por outras que, embora não sejam exactas, se adaptam muito melhor ao cálculo numérico, dando uma boa aproximação, para valores de n elevados.

BIBLIOGRAFIA

Além das obras indicadas na advertência prévia, recomendamos, ainda, as seguintes:

- A. C. AITKEN – *Statistical Mathematics*. University Mathematical Texts. Oliver and Boyd; Edimburg and London, 1944.
- A. HALD – *Statistical Theory with Engineering Application*, New York (John Wiley & Sons Inc.) and London (Chapman & Hall, Ld.), 1952.
- P. LEVY – *Calcul des probabilités*, Gauthiers – Villars, Paris.

ÍNDICE

CÁLCULO DAS PROBABILIDADES

ADVERTÊNCIA PRÉVIA	317
I.4.1 INTRODUÇÃO AO CÁLCULO DAS PROBABILIDADES: POPULAÇÕES FINITAS	319
A – Frequências	319
1. Primeiros exemplos	319
2. Populações. Álgebra dos atributos	322
3. Álgebra dos acontecimentos	325
4. Acontecimentos expressos em forma proposicional	326
5. Frequência dum atributo numa população	328
6. Frequência de um acontecimento numa série de provas	330
7. Partições	331
8. Corpos de conjuntos, corpos de atributos, corpos de acontecimentos	333
9. Distribuição em universos finitos	337
10. Soma de conjuntos não disjuntos (atributos ou acontecimentos compatíveis)	339
11. Atributos quantitativos	341
12. Representação gráfica das distribuições: histogramas e polígonos de frequência	345
13. Independência e associação de atributos. Distribuições de duas ou mais variáveis	348

14. Associação e independência de partições múltiplas. Tábuas de contingência	355
15. Associações parciais de atributos. Independência de atributos no caso em que o seu número é superior a dois	359
16. Interpretação de uma tábua de contingência. Testes de significância	363
B – Probabilidades	373
1. Lógica indutiva	373
2. Lógica dedutiva	376
3. Conceito natural de probabilidade	378
4. Axiomatização do conceito de probabilidade.	382
5. Alguns exemplos de cálculo de prodabilidades <i>a priori</i>	385
6. Independência e associação de acontecimentos	396
7. Sistema de duas experiências	398
8. Sistema de várias experiências	403
9. Distribuição binomial ou de Bernoulli	404
10. Conceito de moda. Caso da distribuição normal	409
11. Distribuição polinomial. Amostras casuais	411
BIBLIOGRAFIA	414

I.4.2 APONTAMENTOS DE CÁLCULO DAS PROBABILIDADES	415
A – Distribuições de uma variável contínua real	415
B – Valores médios para distribuições de uma variável real	425
C – Valores médios para distribuições de mais de uma variável real	438
D – Aplicação à distribuição binomial. Teorema de BERNOULLI	451
E – Distribuição normal	455
F – Convergência de distribuições. Relação entre as distribuições normal e binomial.	464
G – A distribuição de χ^2 de PEARSON	465
NOTA SOBRE A AVALIAÇÃO DA VARIÂNCIA	469

I.4.3 ADITAMENTO ÀS LIÇÕES DE CÁLCULO DE PROBABILIDADES	471
A – Regressões. Ajustamentos. Correlação	471
1. Formulação geral do problema	471
2. Ajustamentos pelo método dos mínimos quadrados	476
3. Outra forma das equações normais	479
4. Regressão polinomial	480
5. Regressão linear. Correlação	481
6. Segunda recta de regressão	485
7. Organização prática dos cálculos	487
8. Ajustamentos com mudanças não lineares de variáveis	490
9. Regressão múltipla. Índice de correlação em geral	493
10. Nota sobre as notações	497
B – Distribuições de STUDENT e de FISHER.	
Suas aplicações	497
1. A melhor estimativa do desvio padrão deduzida duma amostra	497
2. Distribuição t de STUDENT	499
3. A melhor estimativa de σ deduzida a partir de várias amostras	501
4. Distribuição da diferença entre duas médias	502
5. Prova do t (de significação)	503
6. Intervalos de tolerância e intervalos de confiança	506
7. Intervalos de confiança quando não é conhecido o σ da população	509
8. Aplicações agronómicas	510
9. Distribuição de F e z de FISHER	511
TÁBUA DA DISTRIBUIÇÃO NORMAL	512
TÁBUA DA DISTRIBUIÇÃO DE t DE STUDENT	513
TÁBUA DA DISTRIBUIÇÃO DE χ^2 DE PEARSON	514
INDICAÇÕES BIBLIOGRÁFICAS	515