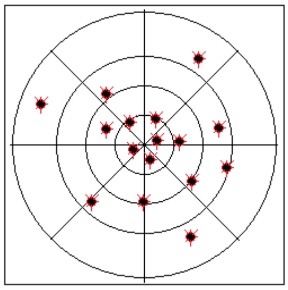
Introdução aos Modelos Probabilísticos

Prof. Sebastião de Amorim

UNICAMP - 2012

Capítulo 1 – Fundamentos

As leis da probabilidade interferem em todos os fenômenos naturais, sociais, econômicos e pessoais. Em diversas situações eles são determinantes, podendo-se dizer que literalmente governam os resultados – tal é o caso no ato corriqueiro do arremesso de uma moeda ou de um dado; em outras, a contribuição é parcial, mas fortemente influente nos resultados, como nos "experimentos" de tiro ao alvo. Neste caso, o nível de perícia do atirador delimita o grau de influência da componente probabilística no resultado final. Em muitas outras situações, a componente probabilística dos resultados é tão pequena comparada às componentes determinísticas que o fenômeno é visto em geral como puramente determinístico. Nos jogos de azar, como as loterias de números – a sena é um caso bem conhecido – as leis da probabilidade governam os resultados de forma completa. Em outros jogos, como a loteria esportiva, o nível de conhecimento do jogador pode exercer algum nível de controle sobre seu resultado, mas ele é ainda fundamentalmente comandado por leis probabilísticas.



É interessante observar que sob certos aspectos, quanto maior a extensão e complexidade – sob uma ótica determinística – de uma situação prática, tanto maior a sua simplicidade do ponto de vista de uma abordagem probabilística. Isto ressalta o valor do tratamento probabilístico de problemas reais, complementando a abordagem determinística usual. Nas situações em que a abordagem determinística é impotente, a probabilística brilha. Por exemplo, ao arremessar uma moeda 10 vezes, podemos afirmar – conforme veremos adiante – que, com probabilidade 0,89, a freqüência relativa de caras ficará entre 30 e 70%. Temos, portanto, sobre o resultado do experimento, uma previsibilidade não nula, mas limitada. Com um número muito maior de arremessos o problema ganha, do ponto de vista probabilístico, surpreendente nitidez. Com 10 mil arremessos, numa situação aparentemente muito mais complexa, podemos afirmar que, com probabilidade 0,99, a fração de caras ficará entre 48,7 e 51,3%. Aumente o número de arremessos para um milhão e teremos quase uma situação de nitidez absoluta: com probabilidade 0,99 a fração de caras estará entre 49,9 e 50,1%.

Numa outra direção: é muito mais fácil e preciso prever quanta chuva cairá em Campinas durante o ano de 2012 do que durante o dia 18 de janeiro de 2012.

Neste curso alcançaremos um domínio conceitual e operacional de um conjunto surpreendentemente poderoso, abrangente e eficaz de ideias e ferramentas básicas da Teoria da Probabilidade e da Estatística. Estes ganhos ampliarão a nossa compreensão do funcionamento dos fenômenos reais a nossa volta.

1.1 – Desenvolvimento formal do conceito básico de Probabilidade

O conceito de probabilidade pode ser definido matematicamente, de uma forma abstrata e completamente desvinculada de qualquer contexto material. Este desenvolvimento é útil por fornecer uma infra-estrutura simples e rigorosa, sobre a qual as propriedades matemáticas inerentes do conceito podem ser deduzidas. A estrutura matemática produzida, com a teoria matemática associada – a Teoria da Probabilidade – poderá então ser aplicada na modelação de problemas reais de natureza probabilística.

Construiremos aqui, de uma forma muito simplificada, a estrutura formal básica que será utilizada em seguida na dedução de diversas propriedades fundamentais úteis, e na modelação de diversas situações concretas interessantes.

Espaço de Probabilidades
Construiremos aqui a estrutura básica sobre a qual construiremos os fundamentos da Teoria da Amostragem. Os conceitos definidos a seguir são simples.

Experimento aleatório – também denominado experimento probabilístico ou, ainda, experimento estocástico, é qualquer ação cujo resultado não pode ser previsto, senão em termos probabilísticos. O lançamento de uma moeda ou de um dado são exemplos corriqueiros. Um experimento aleatório é dito binário se seu o conjunto de todos os seus resultados possíveis tem apenas dois elementos. Tal é o caso do lançamento de uma moeda: os resultados possíveis são Cara (C) e Coroa (c); o conjunto dos resultados possíveis é então {C, c}. Na sua forma genérica, os resultados de um experimento aleatório binário são Sucesso (S) e Fracasso (F). Experimentos aleatórios binários são também denominados de Bernouli, ou bernoulianos.

Como o termo “experimento” sugere intencionalidade, empregaremos também a expressão Fenômenos Aleatórios para os casos em que a ação humana – intencional ou acidental – não intenvem.

Uma partida de futebol, por exemplo, pode ser considerada um experimento aleatório, uma vez que o resultado final depende, em grande parte, de fatores aleatórios, daí a impossibilidade de se antecipar, na maioria dos casos, o resultado de uma partida. E aqui temos um exemplo de como a definição do espaço amostral associado a um experimento aleatório é sujeita, frequentemente, a convenções arbitrárias: sob certo ponto de vista o espaço amostral pode conter 3 resultados possíveis apenas, como {Vitória do time A, Empate, Vitória do time B}. Numa outra visão, mais detalhista, o placar é importante, e o espaço amostral, a rigor, infinitos elementos, como {... 3x0, 3x1, 3x2, 3x3, 2x3, 1x3, 0x3, 2x0, 2x1, 2,x2, 1x2, 0x2, 1x0, 1x1, 0x1, 0x0}

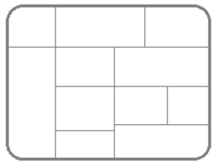
Espaço amostral – associado a um experimento aleatório, é o conjunto de todos os resultados possíveis deste experimento. Genericamente representado por Ω, no caso do arremesso de um dado Ω={1, 2, 3, 4, 5, 6}. Na conceituação teórica abstrata, pode-se prescindir da idéia de experimento aleatório, sendo Ω um conjunto qualquer. O espaço amostral pode também ser referido como Conjunto Universo ou, simplesmente, Universo.

- A cardinalidade de um conjunto é o número de elementos deste conjunto. Um conjunto é finito se sua cardinalidade é finita. O conjunto Ω em (D2) é finito e sua cardinalidade, que se representa por #(Ω), é 6.
- No caso do arremesso de uma moeda, Ω={C, c} e #(Ω)=2. O espaço amostral correspondente a um experimento aleatório binário genérico é Ω={S, F}. Outras representações genéricas, como por exemplo, {0, 1}, {Positivo, Negativo}, são também utilizadas.
- Um conjunto é enumerável se ele é finito, ou se infinito, mas seus elementos podem ser postos em relação unívoca com os naturais, isto é, se eles podem ser contados. Todo conjunto finito é enumerável. São exemplos de conjuntos infinitos enumeráveis:
    - {x inteiro e múltiplo de 3}
    - {x natural e primo}
    - {x é racional}
- Se um experimento consiste em arremessar uma moeda até a obtenção da primeira cara e o resultado do experimento for a sequência de resultados elementares obtidos, então o espaço amostral deste experimento é Ω={C, cC, ccC, cccC, ccccC, ... }, é infinito, portanto... mas enumerável.
- Um conjunto infinito que não pode ser contado (isto é, ter seus elementos associados biunivocamente aos de N) é dito não enumerável. São exemplos de conjuntos não enumeráveis:

- O conjunto R dos reais
- O conjunto dos reais no intervalo (0, 1)
- O conjunto de todos os irracionais no intervalo (-0,0001, 0,0001)

Evento – é qualquer subconjunto de Ω.

No caso do arremesso de um dado, {1}, {2, 4, 6} e {5, 6} são exemplo de eventos. Quantos eventos diferentes existem associados a um espaço amostral de cardinalidade 6?



Partição – é qualquer classe P de subconjuntos não vazios e disjuntos de Ω cuja união é Ω. A figura ao lado representa graficamente a ideia básica de partição, que pode ser, naturalmente, aplicada a qualquer conjunto. Denominamos classe (geralmente representadas por maiúsculas manuscritas, como $\mathscr{P}$) a qualquer conjunto cujos elementos são conjuntos.

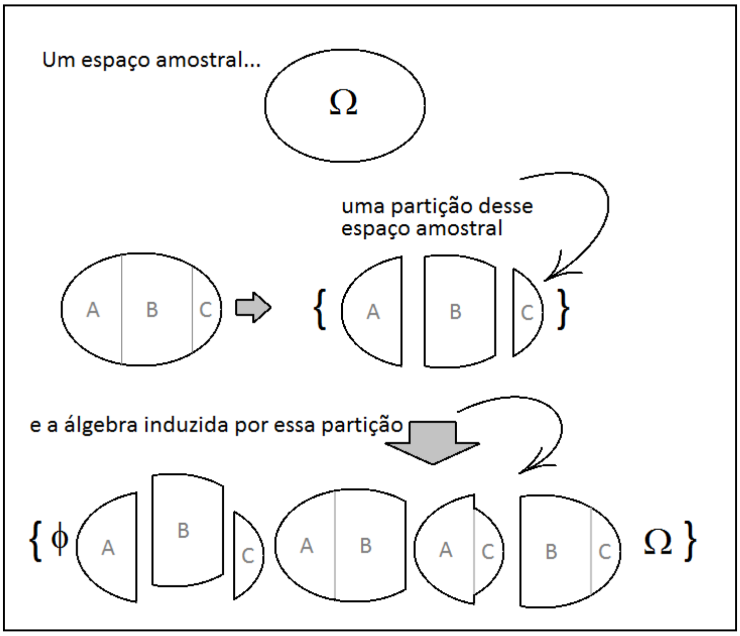
A classe { {1, 2}, {3, 4}, {5, 6} } é uma partição de Ω={1, 2, 3, 4, 5, 6}, assim como { {1, 2, 3, 4}, {5, 6} }. Uma partição com dois elementos apenas, como a anterior, é dita binária.

Quantas partições diferentes existem do conjunto Ω definido acima, isto é, de quantas maneiras diferentes ele pode ser particionado? E um conjunto de cardinalidade n?

Álgebra – é qualquer classe não vazia $\mathscr{F}$ de subconjuntos de Ω, com as seguintes propriedades de fechamento:

a. Se A pertence à classe $\mathscr{F}$, então seu complemento A^c também pertence: $A\in \mathscr{F} \Rightarrow A^c\in \mathscr{F}$
b. Se A e B pertencem a $\mathscr{F}$ então sua união também pertence: $A\in \mathscr{F}$ e $B\in \mathscr{F} \Rightarrow A\cup B\in \mathscr{F}$

Seja, por exemplo, Ω={1, 2, 3, 4, 5, 6}. Definindo os eventos A={1, 2}, B={3, 4, 5, 6}, C={1, 2, 3}, D={4, 5} e E={6} vemos que $\mathscr{P}_1$={A, B} é uma partição binária de Ω; $\mathscr{P}_2$={C, D, E} é outra partição de Ω, desta vez em três partes; por outro lado, a classe $\mathscr{N}$={A, D, E} não é uma partição de Ω, uma vez que a união de seus elementos não forma Ω.



A classe $\mathcal{F}_1=\{\phi, A, B, \Omega\}$ é uma álgebra de subconjuntos de Ω (verifique); $\mathcal{F}_2=\{\phi, C, D, E, C\cup D, C\cup E, D\cup E, \Omega\}$ também é uma álgebra de subconjuntos de Ω. Toda álgebra tem um conjunto de blocos básicos, formadores. Estes blocos básicos compõesm uma partição de Ω.

São consequências imediatas da definição de álgebra. Como exercício, prove esses resultados:

- O eventos ϕ e Ω sempre pertencem à álgebra.
- Se $A_1, A_2, \dots, A_n$ pertencem à álgebra, então a união $\bigcup_{i=1}^{n} A_i$ também pertence
- Se $A_1, A_2, \dots, A_n$ pertencem à álgebra, então a intersecção $\bigcap_{i=1}^{n} A_i$ também pertence

Note aqui que as propriedades definidoras de uma álgebra implicam que se $A\in\mathcal{F}$ e $B\in\mathcal{F}$, então $A\cap B\in\mathcal{F}$. Prova: Pelas leis de DeMoivre, $(A\cap B)^c=A^c\cup B^c$, isto é, o complemento da intersecção é a união dos complementos, logo $A\cap B=(A^c\cup B^c)^C$, que, já sabemos, pertence a $\mathcal{F}$, pelas suas propriedades definidoras de fechamento em relação a união e complementação. Uma álgebra é, portanto, também fechada em relação a operações de complementação.

Geratriz da Álgebra – É uma partição $\mathcal{P}$ de Ω, tal que:

$A\in P \Rightarrow A\in\mathcal{F}$

B é um subconjunto próprio de algum elemento de $\mathcal{P}$, então $B\notin\mathcal{F}$.

Toda álgebra tem uma partição geratriz, e ela é única; toda partição de Ω induz uma álgebra de subconjuntos de Ω : toda partição tem uma álgebra associada, e vice-versa.

Nota : Dado um conjunto não vazio qualquer, A, uma maneira construir uma álgebra de subconjuntos de A consiste em, primeiro, particionar A em n subconjuntos e, no segundo passo, compor a classe formada pelo vazio e todas as uniões, 1 a 1, 2 a 2, 3 a 3, até n a n (isto é, o próprio A), dos elementos da partição. Assim, de um conjunto A, pode-se formar tantas álgebras de subconjuntos de A quantas forem as partições diferentes possíveis de A.

Átomos de uma álgebra – são os elementos da partição geratriz da álgebra. Um elemento de uma álgebra é um átomo se nenhum subconjunto próprio pertence à álgebra. Neste sentido, os átomos de uma álgebra são os blocos básicos, os menores elemetos, indivisíveis, da mesma. Eles são os blocos básicos indivizíveis, formadores dos elementos da mesma; assim, cada elemento da álgebra, ou é o vazio, ou um átomo, ou a união de dois ou mais daqueles átomos.

Estamos prontos, agora para definir Probabilidade, formalmente :

Probabilidade – é qualquer função real definida em $\mathscr{F}$ e tomando valores no intervalo fechado [0, 1], com as seguintes propriedades:

1. $P(\phi)=0$
2. Se $A\in \mathscr{F}$, então $P(A)=1-P(A^c)$
3. Se $A\in \mathscr{F}$ e $B\in \mathscr{F}$ são disjuntos (isto é, $A\cap B=\phi$) então $P(A\cup B)=P(A)+P(B)$.

Uma função de probabilidade fica completamente definida com a definição de seu valor para cada átomo da geratriz de $\mathscr{F}$.

No caso dos experimentos binários, $\Omega=\{S, F\}$, e a partição $\mathscr{P}=\{ \{S\} , \{F\} \}$ define uma álgebra com 4 elementos, a única relevante no caso; a outra é $\mathscr{F}_0=\{\phi, \Omega\}$. A função de probabilidades é definida por $P\{S\}=p$, com $0<p<1$. Consequentemente, $P\{F\} = 1 - P\{S\} = 1 - p = q$, com $p+q=1$. Nestas notas designaremos por E_p ao experimento aleatório binário com $P\{S\}=p$.

Exemplo 1.1.1 – Seja $\Omega=\{1, 2, 3, 4, 5, 6\}$ e a partição $\mathscr{P}=\{ \{1, 2, 3\}, \{4, 5\}, \{6\} \}$.

a - Construa, elemento a elemento, a álgebra $\mathscr{F}$ de subconjuntos de Ω gerada por $\mathscr{P}$.

Para simplificar, vamos dar nomes a cada elemento da partição: $A=\{1, 2, 3\}$, $B=\{4, 5\}$, $C=\{6\}$; eles serão os átomos, ou blocos básicos, da álgebra. Então,

$\mathscr{F}=\{ \phi, A, B, C, A\cup B, A\cup C, B\cup C, A\cup B\cup C\}$ ou, $\mathscr{F}=\{ \phi, A, B, C, A\cup B, A\cup C, B\cup C, \Omega\}$

b – Defina arbitrariamente uma função de probabilidades P, escolhento valores consistentes para P dos átomos da árgebra, e use as propriedades definidoras de uma função de probabilidades para determinar P de forma completa, isto é, para cada elemento de $\mathscr{F}$.

Arbitremos : $P(A)=0,5$, $P(B)=0,4$ e $P(C)=0,1$. Note que, como $A\cup B\cup C=\Omega$, então, para consistência, é necessário que $P(A)+P(B)+P(C)=1$. Em seguida, obedecendo aos axiomas básicos : $P(\phi)=0$, $P(A\cup B)=0,9$, $P(A\cup C)=0,6$, $P(B\cup C)=0,3$ e, naturalmente, $P(A\cup B\cup C)=P(\Omega)=1$. A função $P : \mathscr{F}\rightarrow R$, assim definida, é uma *função de probabilidades no espaço (Ω, $\mathfrak{F}$). Nota : é possível que você esteja associando esse espaço amostral ao experimento de arremessar um dado. Neste caso, você acha que P(A)=0,5, P(B)=1/3 e P(C)=1/6 seria a definição correta para P, e que a definição proposta não faz sentido. Quanto a isto, alguns comentários são apropriados: (1)Como não estamos modelando nenhum experimento aleatório em particular, mas apenas construindo um modelo abstrato de probabilidades, a definição dada é correta, uma vez que cumpre todas as condições de uma função de probabilidade ; (2)a segunda definição também é correta, desde que os valores de P para os demais eventos em A sejam calculados segundo as leis básicas ; e, (3)para modelar apropriadamente o experimento concreto do lançamento de um dado normal, a segunda é, certamente, a proposta adequada.*

Espaço de Probabilidades – É a tríade composta por um espaço amostral, uma álgebra de subconjuntos deste espaço amostral, e uma função de probabilidades definida sobre esta álgebra: (Ω, $\mathfrak{F}$, P)

Evento Mensurável – Um evento A (isto é, A$\subset\Omega$) é dito mensurável se ele pertence à álgebra, ou seja, se P(A) for bem definida.

No Exemplo 1.1.1, o evento {1, 2, 3, 6} é mensurável mas {1, 2, 4}, não. Dizer que um evento é mensurável equivale a dizer que ele tem um valor bem definido de probabilidade no espaço métrico construído.

1.2 – Probabilidade Condicional, Independência e o Teorema de Bayes, com aplicações

O conceito de probabilidade condicional tem um papel central na teoria do conhecimento. Mais recentemente ele tem ganhado grande visibilidade pelo seu papel central em estratégias de busca automática na internet.

1.2.1 - Fundamentos

Probabilidade Condicional – Seja (Ω, $\mathfrak{F}$, P) um espaço de probabilidade e B um evento mensurável (isto é, B$\in\mathfrak{F}$) de probabilidade não nula. A função P_B: $\mathfrak{F}\rightarrow$R representada P[{A}\{B}] (lê, se: probabilidade do evento A, dada a ocorrência do evento B, ou, mais simplesmente, probabilidade de A dado B), é definida como

$$\mathbf{P}[\mathbf{A}\backslash\mathbf{B}] = \frac{\mathbf{P}(\mathbf{A}\cap\mathbf{B})}{\mathbf{P}(\mathbf{B})}$$

Tambem representada como $P_B[A]$, a probabilidade condicional é uma função probabilidade no espaço métrico (Ω, $\mathfrak{F}$, P).

Exemplos:

1. Seja o experimento D_5^2, conforme definido, cujo espaço amostral, Ω={11 12 … 54 55} tem cardinalidade 25. Definindo F como a álgebra completa, e P a função natural de probabilidades para o experimento, considerando-se dados equilibrados, dá probabilidade 0,04 para cada um dos eventos unitários em F. Desta forma, definindo:

   A={o maior resultado é 3}={13 23 31 32}

   e B={a soma dos resultados é 4}={13 22 31}

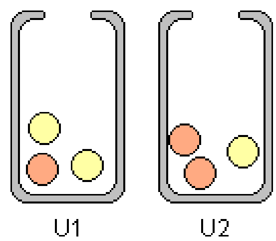
   temos que P[A]=0,16, P[B]=0,12, e, pela definição de probabilidade condicional:

   $P[A\backslash B] = P[A\cap B]/P[B] = P\{13\ 31\}/P[B] = 0{,}08/0{,}12 = 2/3$

   E $P[B\backslash A] = P[A\cap B]/P[A] = P\{13\ 31\}/P[A] = 0{,}08/0{,}16 = 1/2$

   Interpretando os resultados acima, temos que a probabilidade do maior resultado ser 3 dado que a soma dos dois resultados foi 4 é igual a 2/3; já a probabilidade da soma dos resultados ser quatro, dado que o maior dos dois resultados foi três, é igual a 0,50.

2. Uma urna contém 3 bolas, sendo uma vermelha e duas amarelas. Uma segunda urna contém também três bolas, sendo duas vermelhas e uma amarela. Um experimento é realizado em dois estágios. No primeiro, uma das urnas é selecionada ao acaso, ao arremesso de um dado : {1, 2}$\rightarrow u_1$, {3, 4, 5}$\rightarrow u_2$; no segundo, dela é sorteada uma bola e sua cor anotada...

   Qual a probabilidade de que a segunda bola seja amarela dado que a primeira foi vermelha?

   O espaço amostral melhor adequado a este experimento é

   $\Omega=\{u_1a,\ u_1v,\ u_2a,\ u_2v\}$

3.

D2. Eventos Independentes – Se dois eventos mensuráveis A e B são tais que P_B é bem definida e $P_B(A) = P(A)$, então dizemos que A e B são independentes. Neste caso, $P[A \cap B]=P[A] \times P[B]$.

Exemplo 1.3.1 **–** *Um experimento aleatório consiste de 2 arremessos sucessivos de um dado equilibrado. O espaço amostral associado a este experimento é*

$$\begin{aligned}\Omega = \{&(1\ 1),(1\ 2),(1\ 3),(1\ 4),(1\ 5),(1\ 6),\\ &(2\ 1),(2\ 2),(2\ 2),(2\ 4),(2\ 5),(2\ 6),\\ &(3\ 1),(3\ 2),(3\ 3),(3\ 4),(3\ 5),(3\ 6),\\ &(4\ 1),(4\ 2),(4\ 3),(4\ 4),(4\ 5),(4\ 6),\\ &(5\ 1),(5\ 2),(5\ 3),(5\ 4),(5\ 5),(5\ 6),\\ &(6\ 1),(6\ 2),(6\ 3),(6\ 4),(6\ 5),(6\ 6)\}\end{aligned}$$

Seja $\mathfrak{F}$ a classe de todos os subconjuntos de Ω; os átomos de $\mathfrak{F}$ são, portanto, os 36 subconjuntos unitários de Ω, e a partição geratriz de $\mathfrak{F}$ é { {(1 1)}, {(2 6)}, …, {(6 6)} }.

Sabemos que a função de probabilidades fica completamente definida quando definimos seus valores para cada um dos átomos da álgebra considerada. Sabemos que

$$P\{1 \text{ no primeiro lançamento}\} = P\{(1\ 1),(1\ 2),(1\ 3),(1\ 4),(1\ 5),(1\ 6)\}= 1/6$$

Da mesma forma como

$$P\{3 \text{ no segundo lançamento}\} = P\{(1\ 3),(2\ 3),(3\ 3),(4\ 3),(5\ 3),(6\ 3)\} = 1/6,$$

e assim por diante. Ora, no caso concreto de dois arremessos sucessivos de um dado, os dois eventos listados são independentes, logo

$P(\{1 \text{ no primeiro lançamento}\} \cap \{3 \text{ no segundo lançamento}\}) = P\{13\} = (1/6)x(1/6) = 1/36$

Podemos fazer o mesmo para cada um dos 35 demais eventos unitários, obtendo 1/36 todas as vezes. Assim, a função de probabilidade que associa a cada subconjunto unitário de Ω o valor 1/36, é fiel às características do experimento em questão. A função P, assim definida, é a função de probabilidade natural neste experimento em que o dado é equilibrado e os dois arremessos são feitos de forma independente. Assim, se $A \subset \Omega$ então, neste caso, por hipótese, $A \in \mathfrak{F}$, (isto é, A é mensurável) e $P(A)=\#(A)/36$. A tríade $(\Omega, \mathfrak{F}, P)$ é um espaço de probabilidades. Por exemplo, o evento A, definido por

A = {os dois resultados são iguais} = {(1 1) ,(2 2) ,(3 3) , (4 4) , (5 5) , (6 6)}

é um exemplo de evento mensurável. Neste caso, #(A)=6 então P(A)=6/36=1/6.
Seja agora outro evento mensurável que denominamos B:

B = { a soma dos dois resultados é maior que 8 }
= { (3 6) , (4 5) , (4 6) , (5 4) , (5 5) , (5 6) , (6 3) , (6 4) , (6 5) , (6 6) }

Sua probabilidade é 10/36. Como P(B)>0, podemos definir a função de probabilidade $P_B: \mathscr{F} \to R$ *como* $P_B(A)=P(A \cap B)/P(B)$*, a probabilidade condicional de A, dado B. Assim, se A é como definido acima, temos que* $A \cap B$ *= {(4 4) , (5 5) , (6 6)} e* $P(A \cap B)=1/12$*, isto é, a probabilidade de se ter dois resultados iguais e a soma dos mesmos maior que 7.*

$$P_B(A) = \mathrm{P(A/B)} = \frac{P(A \cap B)}{P(B)} = \frac{\frac{3}{36}}{\frac{10}{36}} = \frac{3}{10} = 0{,}3$$

Podemos dizer, então, que, em dois arremessos sucessivos de um dado, a probabilidade de se ter dois resultados iguais, dado que a soma dos dois resultados é superior a 8, é 0,3. Podemos também calcular a probabilidade de se ter um total superior a 8 dado que os dois resultados foram iguais. Dá 0,333 ; confira.

Da mesma forma como a probabilidade condicional de A, dado B, foi calculada, também se pode calcular a probabilidade condicional em B, de qualquer elemento de $\mathscr{F}$*. Por exemplo, se C = {o primeiro resultado é 6}, então C= {6 1 , 6 2 , 6 3 , 6 4 , 6 5 , 6 6}, e*

$C \cap B$={6 3 , 6 4 , 6 5 , 6 6} e $P_B[C]$=4/10 = 0,4.

Exemplo 1.3.2 – *Considerando o mesmo espaço métrico do exemplo anterior, seja agora o evento B definido como B={O segundo resultado parcial é seis}={1 6, 2 6, 3 6, 4 6, 5 6, 6 6}, com P[B]=1/6. A função de probabilidade* P_B*:* $\mathscr{F} \to \mathbf{R}$ *definida por* $P_B[A]=P[A \cap B]/P[B]$ *para qualquer* $A \in \mathscr{F}$ *é a probabilidade condicional de A, dado B. Esta função fica completamente definida quando definimos seu valor para cada um dos átomos de* $\mathscr{F}$*:*
Como {ij}$\cap${16, 26, 36, 46, 56, 66} é igual a ϕ se $j \neq 6$ e a {ij} se j=6

$$P_B\{ij\} = \frac{P[\{ij\} \cap B]}{P[B]} = \begin{cases} \dfrac{P[\phi]}{P[B]} = 0 \quad se\ j \neq 6 \\ \dfrac{P\{ij\}}{P[B]} = \dfrac{\frac{1}{36}}{\frac{1}{6}} - \dfrac{1}{6} \quad se\ j = 6 \end{cases}$$

Em outras palavras, os átomos {1 6}, {2 6}, {3 6}, {4 6}, {5 6} e {6 6}, que têm intersecção não vazia com B têm, cada um, P_B igual a 1/6; os demais 30 átomos de $\mathscr{F}$ têm P_B zero.

O valor de PB para qualquer dos demais elementos – não "atômicos" – de $\mathscr{F}$ fica perfeitamente determinado a partir das propriedades fundamentais das funções de probabilidade. Por exemplo, PB {1 2, 2 3, 3 4, 4 5, 5 6} = 1/6.

Exemplo 1.3.3 **–** Ainda no mesmo contexto do Exemplo 4, seja o evento:

C={o primeiro resultado parcial é impar}

Então, P[C]=½, já que C contem metade dos elementos de ⍰, portanto #(C)=18. Como
$B \cap C$={1 6, 3 6, 5 6}, temos que $P[B \cap C]$=3/36=1/12. Logo

$$P_B[C] = P[C / B] = \frac{P[B \cap C]}{P[B]} = \frac{\frac{1}{12}}{\frac{1}{6}} = \frac{1}{2}$$

Como $P_B[C]=P[C]$, concluímos que, neste espaço de probabilidades, os eventos

B={O segundo resultado parcial é 6} e C={O primeiro resultado parcial é ímpar}

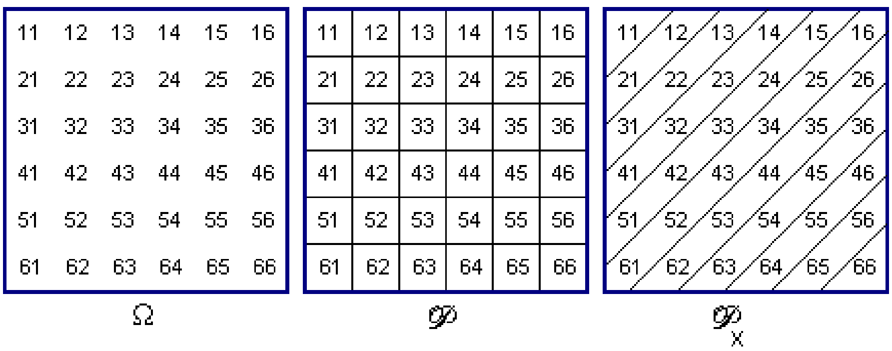
são independentes, um resultado que pode ser facilmente interpretado no contexto físico do exemplo.

Exemplo 1.3.4 **–** Continuemos no espaço de probabilidades do Exemplo 4. Seja a função $X:\Omega \rightarrow R$, definida por X(ij)=i+j, a soma dos dois resultados parciais. Note que X assume valores no conjunto {2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12} e define uma partição própria de Ω em 11 átomos, $\mathscr{P}_X=\{A_2, A_3, A_4, \ldots, A_{12}\}$ com $A_i=\{(\omega_1+\omega_2)\in\Omega$, tais que $\omega_1+\omega_2 = i\}$, para i=2, 3, 4, ..., 12. Ilustrando

A_2={11} , A_3={12 , 21}, A_4={13, 22, 31}, A_5={14, 23, 32, 41},

e assim por diante... Como X é constante dentro dos átomos de A, concluímos que X é uma variável aleatória. Na verdade, como adotamos neste exemplo a partição máxima de Ω, na qual cada átomo é constituído de um único elemento de Ω, qualquer função real definida em Ω será uma variável aleatória.

Seja $\mathscr{F}_X$ *a álgebra gerada por* $\mathscr{P}_X$. *Como cada átomo de* $\mathscr{P}_X$ *é formado pela união de átomos de* $\mathscr{P}$, *dizemos que esta partição é um refinamento daquela. Como a partição* $\mathscr{P}_X$ *é mais grossa que* $\mathscr{P}$, *a álgebra* $\mathscr{F}_X$ *define sobre* Ω *uma estrutura métrica de menor resolução – mais grosseira – que aquela definida por* $\mathscr{F}$. *Por outro lado, como qualquer átomo de* $\mathscr{F}_X$ *é a união de átomos de* $\mathscr{F}$, *concluímos que* $\mathscr{F}_X \subset \mathscr{F}$. *Na figura abaixo temos uma representação gráfica do espaço amostral* Ω *e das suas duas partições consideradas neste exemplo.*



A função de probabilidade P pode ser reduzida a uma versão definida sobre $\mathcal{F}_X$, denominada P_X, de forma que $(\Omega, \mathcal{F}_X, P_X)$ é menos refinado que $(\Omega, \mathcal{F}, P)$, mas é suficiente para estudar a variável aleatória X. Lembrando que $\mathcal{F}_X \subset \mathcal{F}$, qualquer elemento A de $\mathcal{F}_X$ é também elemento de $\mathcal{F}$, e $P_X[A]=P[A]$. Por outro lado, existem elementos de $\mathcal{F}$ que não são elementos de $\mathcal{F}_X$. É o caso, por exemplo, do evento D={O primeiro resultado parcial é 2} = {(2 1), (2 2), (2 3), (2 4), (2 5), (2 6)}. Os eventos desta classe são mensuráveis no espaço $(\Omega, \mathcal{F}_X, P)$ mas não o são no espaço $(\Omega, \mathcal{F}_X, P_X)$. Isto significa que eventos como D têm valor bem definido para P, mas não para P_X. No caso ilustrado temos: P[D]=1/6 e $P_X[D]$="indefinido". Finalmente, a álgebra $\mathcal{F}_X$ é a mais simples – mais grossa – que ainda permite a mensurabilidade de X. Qualquer redução da mesma e a função X perde a condição de variável aleatória.

A variável aleatória X assume valores no conjunto {2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12}. A função de distribuição de probabilidades de X é dada na tabela abaixo.

x	2	3	4	5	6	7	8	9	10	11	12
P[X=x]	0,0278	0,0556	0,0833	0,1111	0,1389	0,1667	0,1389	0,1111	0,0833	0,0556	0,0278

Exercícios 1.3. : [Todos os exercícios desta seqüência se referem ao espaço de probabilidades e ao evento B do exemplo anterior]

1. Determine explicitamente e calcule a probabilidade condicional P_B dos seguintes eventos:
    a. ϕ
    b. Ω
    c. {O segundo resultado é seis}
    d. {A soma dos dois resultados dá um número primo}
    e. {O segundo resultado é maior que o primeiro}
    f. {O resultado parcial "1" não ocorre}
    g. {O resultado parcial "1" ocorre pelo menos uma vez}
2. Encontre um evento que seja independente de B.
3. Seja o evento D={O primeiro resultado é maior que o segundo}. Defina a função de probabilidade condicional P_D, e calcule a P_D para todos os eventos citados no exemplo anterior e no exercício 12.
    a. Seja a função X: $\Omega \rightarrow R$ definida como a soma dos dois resultados parciais.
    b. Mostre que X é uma variável aleatória no espaço de probabilidades $(\Omega, \mathcal{F}, P)$.
    c. Determine $\mathcal{P}_X$, a partição de Ω induzida pela variável aleatória X.
    d. Mostre que se $\mathcal{F}_X$ é a álgebra gerada por $\mathcal{P}_X$, então $\mathcal{F}_X \subset \mathcal{F}$.

e. Determine a função de distribuição de probabilidades de X e trace seu gráfico.

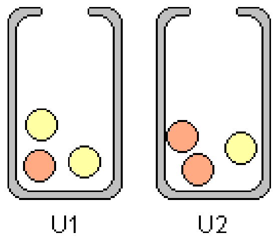
f. Calcule E(X) e V(X)

g. Determine explicitamente o evento $S_6=\{X\geq 6\}$

h. Determine a distribuição condicional de X dado S_6.

4. Determine $E(X/S_6)$ e $V(X/S_6)$ a esperança e variância condicionais de X, dado S_6.
5. Seja A_1={O primeiro resultado é 6} e A_2={O segundo resultado é 3}. Verifique que A_1 e A_2 são independentes.
6. Sejam X_1 e X_2 definidas como o valor do primeiro e o segundo resultado parcial, respectivamente. Determine $\mathscr{F}_{X_1}$ e $\mathscr{F}_{X_2}$.
7.

Exemplo 1.3.5 – *Uma urna* contém 3 bolas, sendo uma vermelha e duas amarelas. Uma segunda urna contém também três bolas, sendo duas vermelhas e uma amarela. Um experimento é realizado em dois estágios. No primeiro, uma das urnas é selecionada ao acaso e com chances iguais, no segundo, dela é sorteada uma bola e sua cor anotada...



Antes de repor a bola, uma segunda bola será sorteada da mesma urna. Qual a probabilidade de que a segunda bola seja amarela dado que a primeira foi vermelha?

O espaço amostral melhor adequado a este experimento é

$$\Omega=\{u_1aa,\ u_1av,\ u_1va,\ u_2av,\ u_2va,\ u_2vv\}$$

Note que os resultados u_1vv e u_2aa não foram incluidos em Ω, pois não podem ocorrer. Numa outra versao de Ω esses dois resultados poderiam ser incluídos, desde que, na definição de P, se cuidasse de fazer $P\{u_1vv\}=P\{u_2aa\}=0$.

Seja $\mathscr{F}$ a classe de todos os subconjuntos de Ω; para a definição da função de probabilidade P associada naturalmente ao experimento vamos inicialmente dar nome a alguns eventos interessantes:

U_1={ A urna 1 foi selecionada } = $\{u_1aa, u_1av, u_1va\}$ e $U_2=U_1^c$

A_1={A primeira bola selecionada é amarela} = $\{u_1aa, u_1va, u_2va\}$ e $V_1=A_1^c$

A_2={A segunda bola selecionada é amarela}=$\{u_1aa, u_1va, u_2va\}$ e $V_2=A_2^c$

Pelas hipótesers do experimento, sabemos que

$P[U_1] = 0,5$

$P[A_1 \backslash U_1] = 2/3$, $P[V_1 \backslash U_1] = 1/3$, $P[A_1 \backslash U_2] = 1/3$ e $P[V_1 \backslash U_2] = 2/3$

$P[A_2 \backslash (U_1\cap A_1)]= P[V_2 \backslash (U_1\cap A_1)] = 1/2$, $P[A_2 \backslash (U_2\cap A_1)] = 0$ e $P[V_2 \backslash (U_1\cap A_1)] =1/2$

$P[A_2 \backslash (U_1\cap V_1)] = 1$

$P[V_2 \backslash (U_1\cap V_1)] = 0$

$P[A_1 \backslash U_2] = 1/3$ e $P[V_1 \backslash U_2] = 2/3$

Estamos agora equipados para responder à pergunta: Qual a probabilidade de que a segunda bola seja amarela dado que a primeira foi vermelha ?

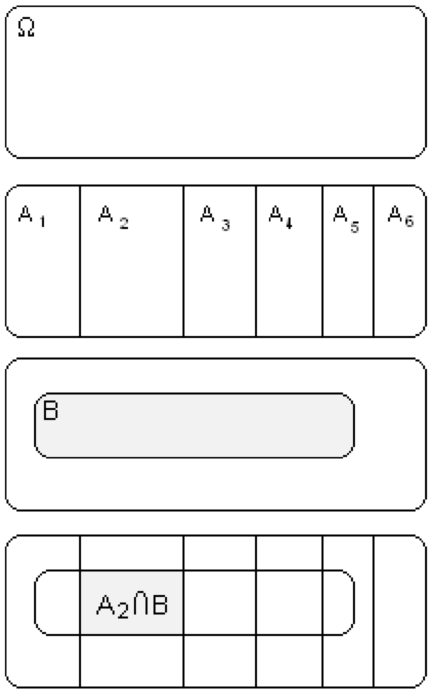
$$P[A_2 \setminus V_1] = \frac{P[A_2 \cap V_1]}{P[V_1]} = \frac{P\{u_1va, u_2va\}}{P\{u_1va, u_2va, u_2vv\}} = \frac{\frac{1}{2}\cdot\frac{1}{3}\cdot 1 + \frac{1}{2}\cdot\frac{2}{3}\cdot\frac{1}{2}}{\frac{1}{2}\cdot\frac{1}{3}\cdot 1 + \frac{1}{2}\cdot\frac{2}{3}\cdot\frac{1}{2} + \frac{1}{2}\cdot\frac{2}{3}\cdot\frac{1}{2}} = \frac{2}{3}$$

□

Exercícios 1.3. :

8. Modifiquemos um pouco o experimento original no exemplo acima: o segundo sorteio é feito na outra urna. Calcule $P[A_2 \setminus V_1]$.
9. Ainda no experimento modificado, calcule a probabilidade de que a urna sorteada tenha sido a primeira, dado que a primeira bola foi amarela: $P[U_1 \setminus A_1]$?
10. Em cada uma das duas variações do experimento das duas urnas, calcule $P[V_1 \setminus A_2]$
11. No caso do arremesso de dois dados, tratado anteriormente, calcule a probabilidade de que o primeiro resultado tenha sido par dado que a soma dos dois resultados foi maior que 8.

1.3.2 – Regra do Valor Total e Teorema de Bayes



Seja um espaço de probabilidades qualquer, $(\Omega, \mathscr{F}, P)$, uma partição enumerável e mensurável de Ω: $\mathscr{A}=\{A_1, A_2, A_3, \ldots\}$ e um evento mensurável qualquer B. Neste caso, $\mathscr{A}_B=\{A_1\cap B, A_2\cap B, A_3\cap B, \ldots\}$ forma uma partição disjunta e mensurável do evento B, e portanto

$$\mathrm{P(B)} = \sum_{\mathrm{i}=1}^{\infty} \mathrm{P(A_i \cap B)}$$

Mas $P[A_i \cap B]=P[B/A_i]\times P[A_i]$, logo

$$\mathrm{P(B)} = \sum_{\mathrm{i}=1}^{\infty} \mathrm{P(A_i) \times P(B \setminus A_i)}$$

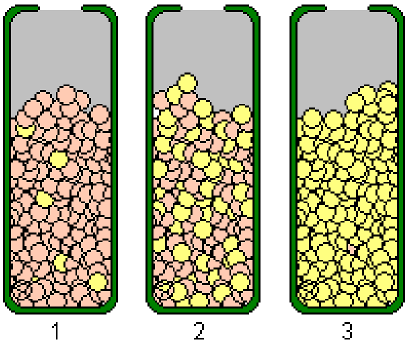
Vejamos um exemplo simples de aplicação deste resultado.

***Exemplo 1.3.6 –** Três urnas contêm 100 bolinhas cada, todas idênticas, exceto na cor. Na primeira, 90 bolinhas são vermelhas, na segunda 50 e na terceira, uma. As restantes são amarelas. Uma urna é escolhida segundo* *um sorteio com probabilidades 0,10, 0,50 e 0,40 para as urnas 1, 2 e 3, respectivamente. Da urna sorteada, uma bolinha é selecionada ao acaso, e sua cor é registrada.*

Podemos montar o espaço de probabilidades para este caso, sem dificuldade:

$\Omega=\{ u_1a, u_1v, u_2a, u_2v, u_3a, u_3v\}$

$\mathscr{F}$ *é a classe de todos os subconjuntos de* Ω



Outras definições operacionais úteis:

U_1*={a urna selecionada foi a 1}=*$\{u_1a, u_1v\}$

A = {a bola selecionada é amarela} = $\{u_1a, u_2a, u_3a\}$

e assim por diante.

A função P fica completamente definida quando definimos seu valor para cada um dos eventos unitários como:

$P\{u_1a\} = P(U_1 \cap A) = P\{U_1\} \times P\{A/U_1\} = 0{,}10 x 0{,}10 = 0{,}01$,

De maneira análoga os demais valores são calculados:

$P\{u_1v\}= 0{,}09$, $P\{u_2a\}= 0{,}25$, $P\{u_2v\}= 0{,}25$, $P\{u_3a\}= 0{,}396$, $P\{u_3v\}= 0{,}004$

A probabilidade de V={a bolinha sorteada é vermelha} pode ser calculada pela regra do valor total, já que $\{U_1, U_2, U_3\}$ *é uma partição disjunta e mensurável de* Ω *e o conjunto V={a bolinha sorteada é vermelha} é um evento mensurável. Assim*

$$P(V) = P(V \setminus U_1) \times P(U_1) + P(V \setminus U_2) \times P(U_2) + P(V \setminus U_3) \times P(U_3)$$

$$=0{,}9 \times 0{,}10 + 0{,}5 \times 0{,}50 + 0{,}01 \times 0{,}4 = 0{,}090 + 0{,}250 + 0{,}004 = 0{,}344$$

Como A é o complemento de V, P(A) = 1-P(V) = 0,656.

Da estrutura do problema, conhecemos a probabilidade condicional de se obter a cor vermelha, dado que a urna selecionada foi a número 1. Invertendo o ponto de vista, podemos agora responder a perguntas como "qual a probabilidade de que a urna selecionada tenha sido a número 1, dado que a bolinha selecionada foi vermelha?", ou seja, $P(U_1 \setminus V)$. *Vejamos*

$$P(U_1 \setminus V) = \frac{P(U_1 \cap V)}{P(V)} = \frac{P(U_1) \times P(V \setminus U_1)}{P(V)} = \left(\frac{P(U_1)}{P(V)}\right) \times P(V \setminus U_1)$$

Observe a elegância como a expressão acima promove a inversão de ponto de vista $P(V \setminus U_1) \Leftrightarrow P(U_1 \setminus V)$.

$$P(U_1 \setminus V) = \left(\frac{P(U_1)}{P(V)}\right) \times P(V \setminus U_1) = \frac{0{,}10}{0{,}344} \times 0{,}90 \;=\; 0{,}2907 \times 0{,}90 \;=\; 0{,}2616$$

Ou ainda, por um outro ângulo, agrupando os termos de outra forma, a maneira como a probabilidade a priori da urna é transformada na probabilidade condicional da urna, dada a cor da bola sorteada:

$$P(U_1 \setminus V) = \left(\frac{P(V \setminus U_1)}{P(V)}\right) \times P(U_1) = \frac{0,90}{0,344} \times 0,10 \ = 2,6163 \times 0,10 \ = \ 0,2616$$

A probabilidade condicional é, por si própria, uma medida de probabilidade: $P_V(U_1)=P(U_1\backslash V)$. Logo, já que $\{U_1, U_2, U_3\}$ é uma partição disjunta de Ω, a soma $P_V(U_1) + P_V(U_2) + P_V(U_3)$ é igual a 1. De fato, como podemos ver:

$$P(U_2 \setminus V) = \left(\frac{P(U_2)}{P(V)}\right) \times P(V \setminus U_2) = \frac{0,50}{0,344} \times 0,50 \ = 1,4535 \times 0,50 \ = \ 0,7267$$

e

$$P(U_3 \setminus V) = \left(\frac{P(U_3)}{P(V)}\right) \times P(V \setminus U_3) = \frac{0,40}{0,344} \times 0,01 \ = 1,1628 \times 0,01 \ = \ 0,0116$$

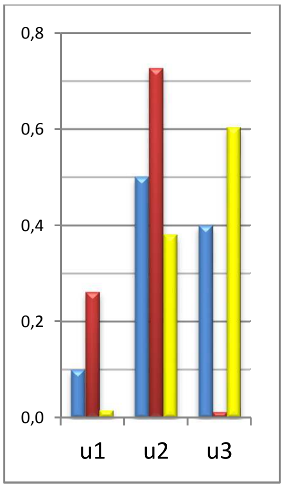
Da mesma forma, conhecendo $P(A)=1-P(V)=0,654$, determinamos $P(U_i\backslash A)$

$$P(U_1 \setminus A) = \left(\frac{P(U_1)}{P(A)}\right) \times P(A \setminus U_1) = \frac{0,10}{0,656} \times 0,10 \ = 0,1524 \times 0,10 = 0,0152$$

$$P(U_2 \setminus A) = \left(\frac{P(U_2)}{P(A)}\right) \times P(A \setminus U_2) = \frac{0,50}{0,656} \times 0,50 \ = 0,7622 \times 0,50 \ = \ 0,3811$$

$$P(U_3 \setminus A) = \left(\frac{P(U_3)}{P(A)}\right) \times P(A \setminus U_3) = \frac{0,40}{0,656} \times 0,99 \ = 0,6098 \times 0,99 = 0,6037$$

A tabela abaixo e a figura ao lado comparam as 3 distribuição de probabilidades entre as 3 urnas: a distribuição a priori, distribuições condicionais, dado que a bola selecionada é vermelha e dado que é amarela. Note como, dado que a bola selecionada foi vermelha, a urna 3 – onde esta cor é rara – se torna relativamente implausível, com a probabilidade a priori de 40% caindo para apenas 1,16%. Efeito semelhante se verifica na urna 1, quando a bola selecionada é amarela.



i	P(Ui)	P(Ui \ V)	P(Ui \ A)
1	0,1	0,2616	0,0152
2	0,5	0,7267	0,3811
3	0,4	0,0116	0,6037

É interessante notar

como, neste exemplo, pudemos determinar P(U_1 / V) a partir de P(V / U_i) e de P(U_i), para todo i. Este é um recurso muito útil na análise de diversas classes de problemas probabilísticos interessantes, conforme veremos mais adiante.

O resultado implícito no desenvolvimento acima pode ser consolidado no Teorema de Bayes.

Sejam (Ω , $\mathscr{F}$, P) um espaço de probabilidades, $\mathscr{P}$ = {A_1, A_2, A_3, ... , A_n} uma partição enumerável, mensurável e disjunta de Ω, e B um evento mensurável qualquer com P[B]>0, então,

$$P(A_i \setminus B) = \frac{P(B \setminus A_i) \times P(A_i)}{\sum_{j=1}^{\infty} P(B \setminus A_j) \times P(A_j)}$$

Ora, o resultado acima, o Teorema de Bayes, já havia sido sugerida e utilizada na seção anterior. Ela é uma decorrência imediata da definição de probabilidade condicional e da regra do valor total:

$$P(A_i \setminus B) = \frac{P(B \cap A_i)}{P(B)} = \frac{P(B \setminus A_i) \times P(A_i)}{\sum_{j=1}^{\infty} P(B \cap A_j)} = \frac{P(B \setminus A_i) \times P(A_i)}{\sum_{j=1}^{\infty} P(B \setminus A_j) \times P(A_j)}$$

Na estrutura do experimento – voltamos aqui ao exemplo das três urnas, para ilustração – a determinação de P[B/A_i] é imediata, sendo decorrência direta da sua própria construção: a probabilidade de uma bola vermelha, dado que a urna selecionada foi a primeira, é 0,10; da mesma forma, P[V \ U_2)= 0,50 e P[V \ U_3)=0,01. A fórmula de Bayes permite uma reversão da ordem natural do experimento. Com ela podemos, por exemplo, calcular a probabilidade de U_1 dado V.

A fórmula de Bayes causa certo desconforto ao formalista mais rigoroso da Teoria da Probabilidade. Se, por um lado, é natural perguntar: "A urna u1 foi selecionada; qual a probabilidade, agora, de se sortear uma bola vermelha?", a pergunta reversa parece violar a ordem natural do experimento: "uma bola vermelha foi sorteada; qual a probabilidade de que a urna 1 tenha sido selecionada?".

A rigor, tendo uma bola vermelha sido selecionada, queda implícito que o estágio do experimento no qual a urna é selecionada já foi realizado. Neste caso, não haveria mais "probabilidade" envolvida, e a pergunta: *"qual a probabilidade de que a urna u1...",* seria uma impropriedade formal.

Analogamente, ao se preparar para arremessar uma moeda, você sabe que a probabilidade de Cara é ½. Estando já o arremesso no passado, mesmo que você ainda não tenha conhecimento do resultado, não seria próprio perguntar "qual a probabilidade de que o resultado tenha sido cara?", ou afirmar "a probabilidade de que o resultado tenha sido cara é ½". A idéia de probabilidade não se aplicaria a experimentos pretéritos.

Para contornar este desconforto formal, os estatísticos criaram o conceito de nível de confiança, e uma maneira objetiva de quantificá-lo. Se uma urna contém 100 bolas idênticas, exceto na cor – cinco são amarelas e 95 são vermelhas – sabemos que, ao sortear aleatoriamente uma bolinha desta urna, a probabilidade de sair uma bolinha amarela é 0,05. Digamos agora que eu sorteio aleatoriamente uma bolinha da urna e a mantenho fechada na minha mão. Você não sabe a cor da bolinha, mas não precisa dizer que a probabilidade de que ela seja vermelha é 0,95. Em vez disto, você pode dizer que *"como a bolinha foi selecionada aleatoriamente de uma urna com 100 bolinhas, das quais 95 são vermelhas, eu tenho 95% de confiança de que a bolinha na sua mão é vermelha."*

Viu? A questão foi tratada com elegância formal, sem usar probabilidade num experimento pretérito. E o conceito de **nível de confiança**, central em Estatística, foi introduzido. Mais adiante voltaremos a ele.

A Fórmula de Bayes fala de probabilidade, associada a um experimento pretérito, causando desconforto formal; e causará ainda mais desconforto quando sugerir o conceito de probabilidade subjetiva. Tanto que provocará a divisão dos estatísticos e probabilistas em dois campos mais ou menos antagônicos e mais ou menos irreconciliáveis, o dos **freqüentistas** e o dos **bayesianos**. Há ainda um campo intermediário, dos **pragmáticos**, que procuram utilizar o ferramental teórico e operacional dos dois lados, maximizando a eficácia no tratamento de problemas reais.

E os Bayesianos têm realmente ferramentas muito interessantes, conforme veremos adiante, num contexto bem básico.

> ***Exemplo 1.3.7****: Um pequeno restaurante em Paris é muito freqüentado por turistas de todo o mundo; em particular, de Portugal e do Brasil.*
>
> *Chega um grupo de turistas e o recepcionista percebe que eles falam Português, mas seu conhecimento do idioma de Camões é apenas vestigial, e ele não consegue distinguir, pelo sotaque, se são brasileiros ou portugueses. Acontece que, conforme se sabe no pequeno bistrô, historicamente, cerca de 80% dos turistas de língua portuguesa que aparecem por lá são brasileiros e os outros, portugueses (para simplicidade, e com todo o respeito, não consideramos aqui, os demais paises que compartilham conosco a "última flor do Lácio", e que contribuem para fazer do nosso o quarto idioma mais falado no mundo.). Há ainda outros fatos históricos, de conhecimento do perceptivo recepcionista. Observador atento de hábitos e costumes se sua clientela*

internacional, ele sabe, por exemplo, que 80% dos seus clientes brasileiros tomam cerveja e os demais tomam vinho. Entre os portugueses é diferente: apenas 10% tomam cerveja; 90% preferem vinho. Pois bem, um grupo de turistas chega, falando português. Ele sabe, então, ***a priori*** *que, com probabilidade 0,80, eles são brasileiros e com 0,20, portugueses.*

a – Qual é a probabilidade de que o grupo peça vinho?

Uma definição conveniente do espaço amostral para este "experimento" é Ω={bc, bv, pc, pv}; a álgebra é a completa (com todos os 16 subconjuntos Ω e a função própria de probabilidades neste caso é definida a partir das probabilidades dos átomos da álgebra máxima (que são, portanto, os subconjuntos unitários de Ω.

$P\{bc\}= P[B]\times P[C\backslash B]=0,8\times 0,8=0,64$

$P\{bv\}= P[B]\times P[V\backslash B]=0,8\times 0,2=0,16$

$P\{pc\}= P[P]\times P[C\backslash P]=0,2\times 0,1=0,02$

$P\{pv\}= P[P]\times P[V\backslash P]=0,2\times 0,9=0,18$

Podemos agora calcular $P[V]=P\{bv, pv\}=P\{bv\}+P(pv)=0,16+0,18=0,34$

b – Qual a probabilidade de que eles sejam portugueses, dado que pediram vinho?

A probabilidade a priori de que os clientes sejam portugueses é apenas 0,20. Dado que eles pediram vinho, uma ação fortemente associada aos portugueses, a probabilidade de serem portuguesas deve aumentar. Vejamos:

$$P_V\left[P\right]=P\left[P\setminus V\right]=\frac{P\left[P\cap V\right]}{P\left[V\right]}=\frac{P\left[V\setminus P\right]P\left[P\right]}{P\left[V\setminus B\right]P\left[B\right]+P\left[V\setminus P\right]P\left[P\right]}=\frac{0,90\times 0,20}{0,20\times 0,80+0,90\times 0,20}=\frac{0,18}{0,16+0,18}=\frac{18}{34}=0,529$$

$$P_V(P)=P(P\setminus V)=\frac{P(P\cap V)}{P(V)}=\frac{P(V\setminus P)\times P(P)}{P(V\setminus B)\times P(B)+P(V\setminus P)\times P(P)}$$

$$=\frac{0,90\times 0,20}{0,20\times 0,80+0,90\times 0,20}=\frac{0,18}{0,16+0,18}=\frac{18}{34}=0,529$$

Como P_V é uma função de probabilidades em (Ω, $\mathscr{F}$), logo $P_V(B)$ =$P_V(P^c)$=1-$P_V(P)$=0,471. Este resultado pode, naturalmente, ser confirmado pelo cálculo direto da expressão de $P_V[P]$.

c – Qual a probabilidade de que eles sejam brasileiros, dado que pediram cerveja?

Neste caso, a bebida corrobora a avaliação a priori de que eles seriam brasileiros. O nível de convicção na brasilidade dos turistas deve, então, aumentar, a partir do "resultado experimental": eles pediram cerveja.

$$P_CB=P(B\setminus C)=\frac{P(B\cap C)}{P(C)}=\frac{P(V\setminus B)\times P(B)}{P(V^C)}=\frac{0,80\times 0,80}{1-0,34}=\frac{64}{66}=0,970$$

E, portanto, $P_C[P]=0{,}030$. A priori, a probabilidade de o grupo ser brasileiro já é elevada; ao pedir cerveja o grupo reforça a evidência ao nível de quase certeza: eles são mesmo brasileiros.

d – Vamos agora admitir que uma distribuição a priori objetiva não exista para a nacionalidade – brasileira ou portuguesa – dos clientes que entram no bistrô falando português. Mas o recepcionista, pessoa observadora e inteligente, tem uma percepção subjetiva aguda. Por motivos que ele não consegue definir de forma nem clara nem objetiva, ele sente os recém chegados são brasileiros. Treinado a pensar estatisticamente, ele consegue, inclusive, quantificar – subjetivamente – o grau de convicção neste seu "feeling": j'ai 90% de convicción que ils sont brasilienes.

Pois bem, admitindo então a distribuição a priori de probabilidades, podemos calcular, à moda bayesiana, a distribuição a posteriori de probabilidades (brasileiros ou portugueses) relativa aos turistas. Novamente, se eles pedirem cerveja, a convicção do maïtre será reforçada:

$$P_C(B) = \frac{P\{bc\}}{P\{bc, bv\}} = \frac{0{,}64}{0{,}64 + 0{,}02} = 0{,}9697$$

As outras três probabilidades condicionais de interesse poderiam aqui ser calculadas de forma análoga. Deixo a tarefa ao leitor.

□

No exemplo anterior vimos como a Fórmula de Bayes permite conjugar informação a priori, muitas vezes de natureza subjetiva, com resultados experimentais objetivos, compondo uma distribuição a posteriori de convicções. A grande utilidade da abordagem decorre do fato que muitas vezes conhecimento não baseado em experimentos cuidadosamente planejados e executados, são valiosos por agegarem percepções e aprendizado adquirido de forma expontânea, como através do acúmulo gradual de experiência sobre fenômenos específicos.

Teremos adiante mais exposição ao método e pensamento bayesianos.

Exercícios 1.3. :

12. Determinada doença tem, numa dada população, prevalência igual a, em média, cinco casos para cada 100 mil indivíduos. Um exame clínico para diagnosticar esta doença tem especificidade 98% e sensibilidade 99%. Isto quer dizer que a probabilidade deste exame dar um falso positivo, ou seja $P_-\{+\}$, é igual a =0,02, enquanto que a probabilidade de um falso negativo, isto é $P_+\{-\}$, é 0,01. Um indivíduo é selecionado ao acaso desta população e examinado. Calcule :
    a. P{Resultado dá positivo}
    b. $P_+\{+\}$, a probabilidade que o paciente seja realmente doente, dado que o resultado deu positivo.

1.3.3 – Jogando dados e contando

O arremesso de um dado oferece um contexto fértil para o desenvolvimento e consolidadção de algumas idéias bastante sofisticadas em probabilidade; vamos elaborar mais sobre o D_5 e suas variantes. De quantas maneiras diferentes se pode obter 16 pontos em 4 arremessos de D_5? A resposta a esta pergunta é essencial para se determinar $P\{S_4=16\}$. A resposta é 35. E de quantas maneiras diferentes se pode obter 60 pontos em D_5^{20}, ou 20 arremessos de nosso numericamente amigável dado de 5 lados? A resposta agora é 5,966,636,799,745, fazendo $P\{S_{20}=60\}=S_{20}\times 0{,}2^{20}$ igual a 0.062564721 ou, aproximadamente, 6,26%. Como foram feitos esses cálculos?

s	#{S_n=s}							
	n=1	n=2	n=3	n=4	n=5	n=6	n=7	n=8
1	1							
2	1	1						
3	1	2	1					
4	1	3	3	1				
5	1	4	6	4	1			
6		5	10	10	5	1		
7		4	15	20	15	6	1	
8		3	18	35	35	21	7	1
9		2	19	52	70	56	28	8
10		1	18	68	121	126	84	36
11			15	80	185	246	210	120
12			10	85	255	426	455	330
13			6	80	320	666	875	784
14			3	68	365	951	1520	1652
15			1	52	381	1246	2415	3144
16				35	365	1506	3535	5475
17				20	320	1686	4795	8800
18				10	255	1751	6055	13140
19				4	185	1686	7140	18320
20				1	121	1506	7875	23940
21					70	1246	8135	29400
22					35	951	7875	34000
23					15	666	7140	37080
24					5	426	6055	38165
25					1	246	4795	37080
26						126	3535	34000
27						56	2415	29400
28						21	1520	23940
29						6	875	18320
30						1	455	13140
31							210	8800
32							84	5475
33							28	3144
34							7	1652
35							1	784
36								330
37								120
38								36
39								8
40								1

Seja D_5 o experimento aleatório que corresponde ao lançamento de um dado não viesado de cinco lados, e D_5^n o experimento composto por n repetições sucessivas e independentes de D_5. Seja X_i o resultado parcial do i-ésimo lançamento. Temos agora uma linguagem para este problema: queremos determinar $\#\{S_n=s\}$. Vamos desenvolver uma solução recursiva. Como $S_n=s$ pode ser alcançado a partir dos valores s-1, s-2, s-3, s-4 e s-5 para S_{n-1}, sempre de uma única maneira, com X_n igual a 1, 2, 3, 4 ou 5, respectivamente, conclui-se que

$$\#\{S_n = s\} = \sum_{i=1}^{5}\{S_{n-1} = s - i\}$$

Na tabela ao lado tem-se $\#\{S_n=s\}$ para n de 1 ao 8; seu formato sugere o esquema construtivo. As células sombreadas mostram como cada valor é obtido, a partir da primeira coluna, para n=1. Por exemplo:

$$\#\{S_8 = 21\} = 29.400 = \sum_{i=1}^{5}\{S_{20} = 21 - i\}$$

$$= 3535 + 4795 + 6055 + 7140 + 7875$$

Com uma planilha tipo Excel é simples; as colunas podem ir sendo construídas sucessivamente. Este método recursivo pode ser adaptado para cálculos análogos em diversas situações, como lançamentos de dados de qualquer número de faces, entre outros. Como os resultados individuais de D_5 são equiprováveis, as probabilidades de eventos como $\{S_n=s\}$ são determinadas diretamente: $P\{S_n=s\}=\#\{S_n=s\}\times 0{,}2^n$, podemos determinar $P\{S_8=21\}=0{,}075266$. Simples.

Queremos agora desenvolver uma abordagem para a resposta a questões como:

- ✓ Em D_5^4, dado que $S_4=16$, qual o valor mais provável de X_1? Posto numa linguagem coloquial: se, em quatro lançamentos de D_5, a pontuação acumulada foi 16, qual o resultado do primeiro lançamento é mais provável de ter acontecido?
- ✓ Generalizando: Em D_5^n, se $S_n=s$, qual a distribuição de X_m, para $m\leq n$? e para $m>n$?

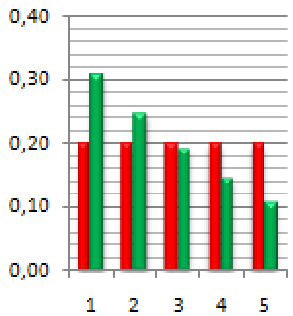
Exemplo

Seja o experimento aleatório D_5^{20}. Qual a distribuição de X_{15}, dado $S_{20}=50$?

Ora, podemos determinar a distribuição a priori (incondicional) de X_{15}: uniforme sobre o suporte {1, 2, 3, 4, 5}. Queremos saber como o conhecimento de $S_{20}=50$ distorce esta distribuição. Para isto, temos que determinar $P\{X_{15}=x / S_{20} = 50\}$, para todo x em $\Omega_x=\{1, 2, 3, 4, 5\}$. Primeiro, vamos determinar $P\{X_{15}=1 / S_{20} = 50\}$.

$$P\{X_{15}=1 \setminus S_{20}=50\}=\frac{P\{X_{15}=1\ e\ S_{20}=50\}}{P\{\ S_{20}=50\}}=\frac{P\{X_{15}=1\ e\ S_{19}=49\}}{P\{\ S_{20}=50\}}$$

$$P\{X_{15}=1\ \setminus S_{20}=50\}=\frac{\#\{X_{15}=1\ \}\times\#\{S_{19}=49\}\times 0{,}2^{20}}{\#\{\ S_{20}=50\}\times 0{,}2^{20}}=\frac{1\times 538.320.708.340}{1.751.059.016.758}=0{,}307426$$



De forma análoga determinamos $P\{X_{15}=x\}$, para x=2, 3, 4 e 5, obtendo, respectivamente, 0.246527, 0.192365, 0.145987 e 0.107695. A figura ao lado mostra as distribuições a priori, em vermelho, e a posteriori, em verde, de X_{15}. Vemos como a informação de que $S_{20}=50$ distorce a distribuição de probabilidades de X_{15}, aumentando a verossimilhança de valores baixos e diminuindo a de valores altos. Isto era esperado, uma vez que $S_{20}=50$ é um resultado baixo, inferior ao valor médio esperado, 60. Valores ainda mais baixos diminuiriam ainda mais a plausibilidade de $\{X_{15}=5\}$. Para $S_{20}=40$, a probabilidade condicional de $\{X_{15}=5\}$ fica reduzida a 4,41% enquanto a de $\{X_{15}=1\}$ sobe para 45,2%. Todos estes cálculos foram feitos diretamente sobre a tabela estendida de $\#\{S_n=s\}$, da qual apresentamos um segmento acima.

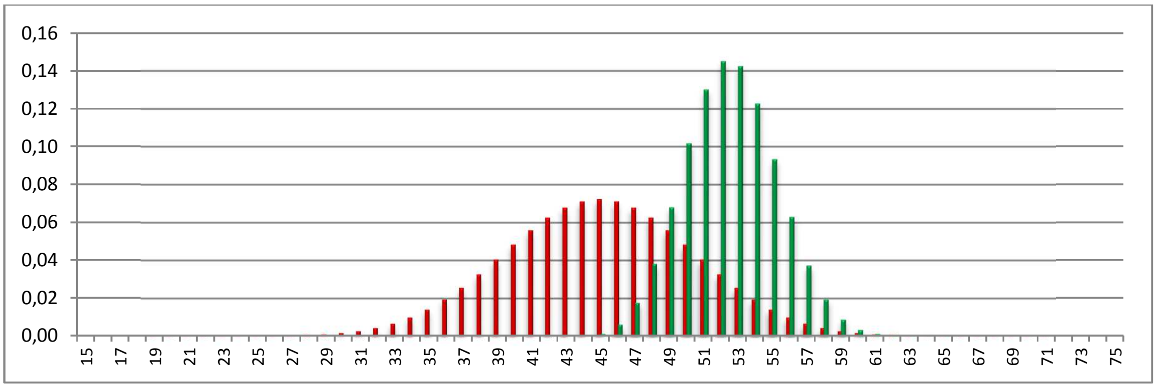
Vamos agora explorar uma outra questão. Já resolvemos o desafio numérico associado à determinação da distribuição exata a priori, ou incondicional, de S_n. Que tal agora enfrentar o seguinte desafio: Qual a distribuição de S_m, dado S_n, para m<n? E para m>n? O desenvolvimento segue a mesma linha:

$$P\{S_m=s\ \setminus S_n=s_n\}=\frac{P\{S_m=s\ e\ S_n=s_n\}}{P\{\ S_n=s_n\}}=\frac{P\{S_m=s\ e\ S_{n-m}=s_n-s_m\}}{P\{\ S_n=s_n\}}$$

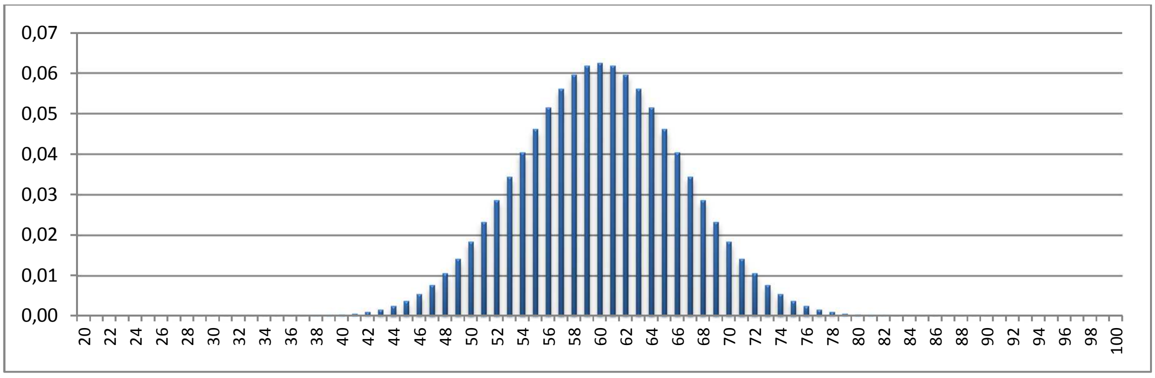
$$=\frac{P\{S_m=s\ \}\times P\{S_{n-m}=s_n-s_m\}}{P\{\ S_n=s_n\}}=\frac{\#\{S_m=s\ \}\times\#\{S_{n-m}=s_n-s_m\}}{\#\{\ S_n=s_n\}}=$$

Note que $P\{S_m = s\ e\ S_n = s_n\}$ é igual à probabilidade de S_m ser igual a m e os n-m arremessos seguintes resultarem em um total de s_n-s pontos. Os dois eventos, $\{S_m=s\}=\{X_1+X_2+\ldots+X_m=s\}$ e $\{S_n-S_m=s_n-s\}=\{X_{m+1}+X_{m+2}+\ldots+X_n=s_n-s\}$ são, obviamente, independentes, logo a probabilidade da intersecção deles é o produto de suas probabilidades. Alem disto, também obviamente, $P\{X_{m+1}+X_{m+2}+\ldots+X_n=s_n-s\}$ é igual a $P\{X_1+X_2+\ldots+X_{n-m}=s_n-s\}$, que é a probabilidade de se totalizar s_n-s pontos em n-m arremessos, portanto igual a $P\{S_{n-m}=s_n-s\}$. Assim explica-se a passagem anterior. As contagens são obtidas da tabela de $\#\{S_n=s\}$, da qual um segmento foi apresentado acima.

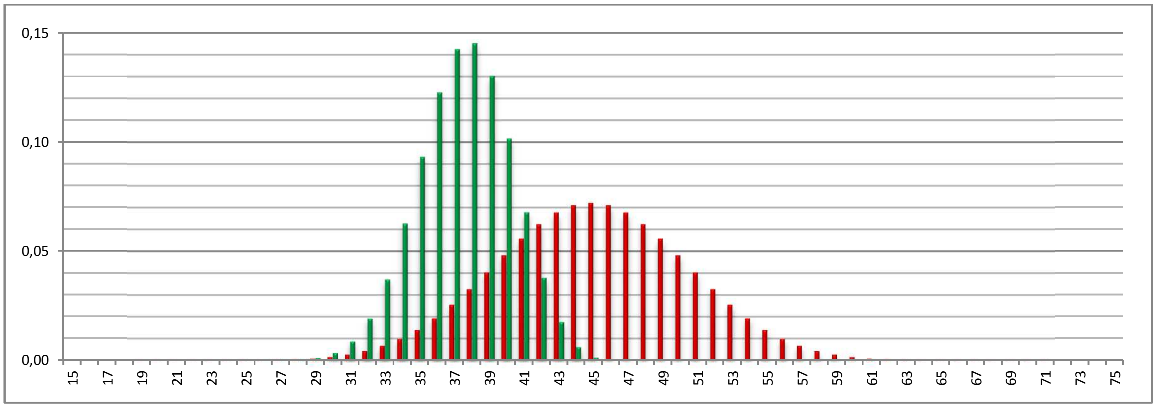
Na figura abaixo temos a distribuição a priori de S_{15} (em vermelho) e a diatribuição a priori de S_{15}, dado S_{20}=70. Note que um resultado muito grande de S_{20}, bem acima do valor médio esperado, $E(S_{20})$=60, tornam improváveis valores pequenos de S_{15}; como nos do 15º ao 20º a máxima pontuação possível é 25, valores de S_{15} menores que 45, dado que S_{20}=70, são impossíveis; pelo mesmo motivo, também são impossíveis valores de S_{15} superiores a 65.

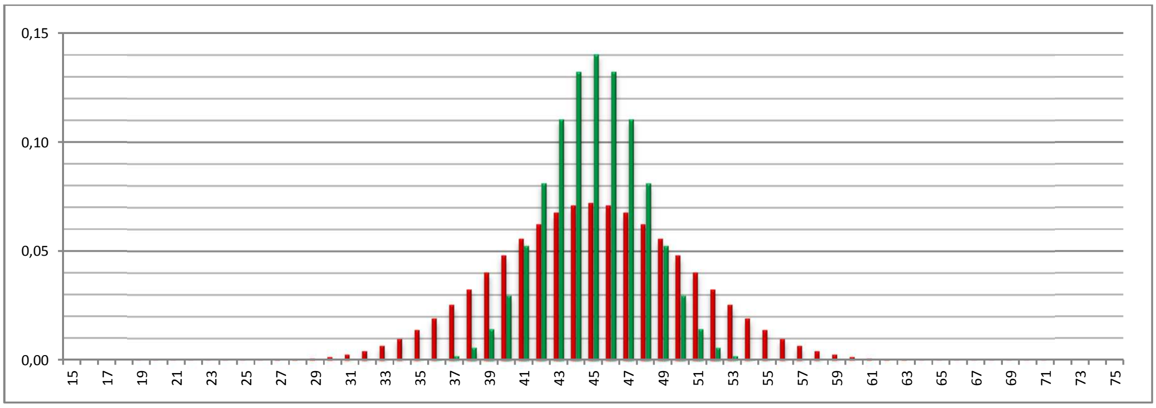


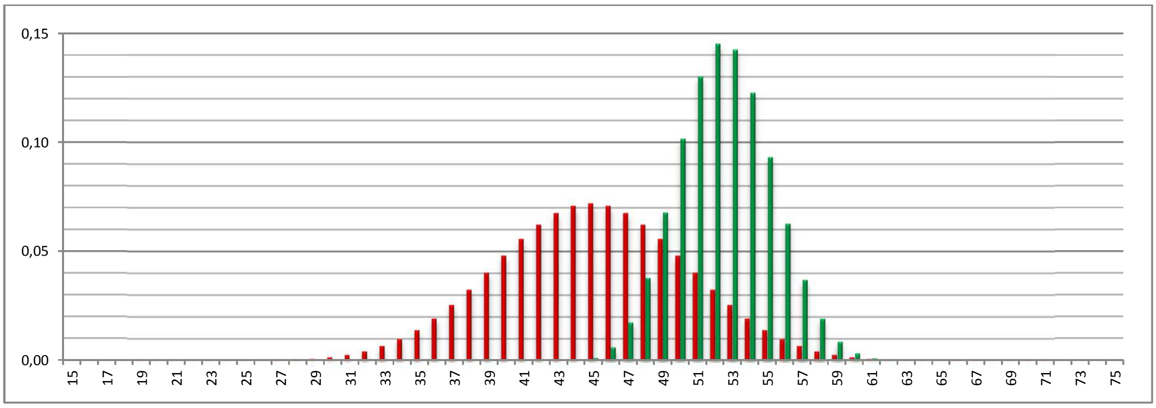
A distribuição de S_{20} está dada na figura abaixo. Note que o resultado 70 é realmente grande, o que induziu o deslocamento da distribuição de S_{15} para a direita.



As figuras abaixo mostram, sobre a distribuição a priori de S_{15}, em vermelho, as distribuições contidicionais de S_{15}, condicionais a S_{20} igual a, respectivamente, 50, 60 e 70.







Capítulo 2

Variáveis Aleatórias

D3. Variável Aleatória – Dado um espaço de probabilidades (Ω, $\mathcal{F}$, P), variável aleatória é qualquer função real $X:\Omega\rightarrow R$ que é constante dentro dos átomos de $\mathcal{F}$, isto é, se ω_1 e ω_2 são dois elementos de Ω pertencentes a um mesmo átomo de $\mathcal{F}$, então $X(\omega_1)=X(\omega_2)$.

Continuando o Exemplo 1.1.1: Na tabela ao lado, X_1 não é uma variável aleatória ; X_2 e X_3 são. E X_4 ?

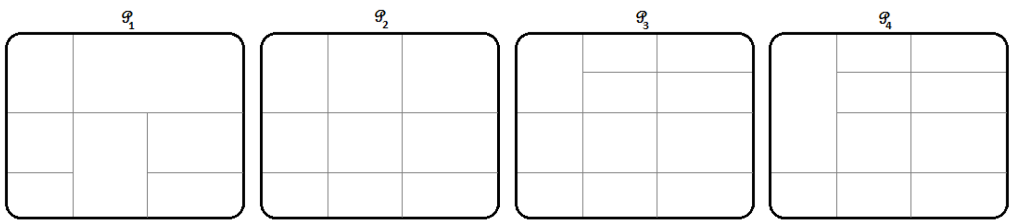
X	ω					
	1	2	3	4	5	6
$X_1(\omega)$	1	4	9	16	25	36
$X_2(\omega)$	1	1	1	2	2	3
$X_3(\omega)$	-1	-1	-1	1	1	1
$X_4(\omega)$	-1	-1	0	0	1	1

D4. Partição de Ω induzida por uma Variável Aleatória X – Se X é uma variável aleatória no espaço de probabilidades (Ω, $\mathcal{F}$, P), então $\mathcal{P}_X$ é a partição que divide Ω em regiões de X constantes: isto é, se ω_1 e ω_2 são dois elementos de Ω, então $X(\omega_1)=X(\omega_2)$ se ω_1 e ω_2 pertencem a um mesmo elemento de $\mathcal{P}_X$, e $X(\omega_1)\neq X(\omega_2)$ no caso contrário.

Exemplo 1.1.2 – Um experimento consiste em arremessar uma moeda três vezes, anotando os resultados. O espaço amostral, neste caso, é Ω={CCC, CCc, CcC, cCC, Ccc, cCc, ccC, ccc}. Seja $\mathcal{P}$ a partição de Ω em oito elementos unitários, e $\mathcal{F}$ a álgebra induzida por esta partição – simplesmente a classe, de cardinalidade 256, de todos os subconjuntos de Ω. A função de probabilidade que adotaremos é a naturalmente associada ao experimento – assumindo moeda não viciada, isto é, P{C}=P{c}=0,5 – então, P{CCC}=1/8, assim como a de qualquer outro evento unitário. Vamos agora definir uma variável aleatória $X : \Omega\rightarrow R$, como $X(\omega)$=número de caras em ω. A variável aleatória X induz uma partição de Ω menos fina que $\mathcal{P}$:

$$\mathcal{P}_X = \{ X^{-1}(0), X^{-1}(1), X^{-1}(2), X^{-1}(3)\} = \{ \{ccc\}, \{ccC, cCc, Ccc\}, \{cCC, CcC, CCc\}, \{CCC\} \}$$

Note que X naturalmente particiona Ω segundo $\mathcal{P}_X$. A induzida por X é mais grosseira que $\mathcal{P}$ uma vez que dois de



seus quatro átomos são iguais a átomos de $\mathcal{P}$, mas os outros dois são uniões de átomos de $\mathcal{P}$. Na figura acima, $\mathcal{P}_2$ é um refinamento de $\mathcal{P}_1$ e é refinada por $\mathcal{P}_3$; $\mathcal{P}_4$, todavia, guarda uma relação de ordem de refinamento com nenhuma das 3 anteriores.

Exercício 1.1.1 – Prove que se X é uma variável aleatória no espaço de probabilidades (Ω, $\mathcal{F}$, P), então $\mathcal{P}_X \subseteq \mathcal{P}$, onde P é a partição geratriz de $\mathcal{F}$.

D5. Álgebra de subconjuntos de Ω induzida por uma Variável Aleatória X – Se X é uma variável aleatória no espaço de probabilidades (Ω, $\mathcal{F}$, P), então $\mathcal{F}_X$ é a álgebra induzida pela partição $\mathcal{R}_X$.

Nota: Prove que $\mathcal{F}_X \subseteq \mathcal{F}$.

Exemplo 1.1.3 – Um experimento consiste em arremessar uma moeda três vezes, anotando os resultados. O espaço amostral, neste caso, é Ω={CCC, CCc, CcC, cCC, Ccc, cCc, ccC, ccc}. Seja $\mathcal{P}$ a partição de Ω em oito elementos unitários, e $\mathcal{F}$ a álgebra induzida por esta partição – simplesmente a classe, de cardinalidade 256, de todos os subconjuntos de Ω. A função de probabilidade que adotaremos é a naturalmente associada ao experimento – assumindo moeda não viciada, isto é, P{C}=P{c}=0,5 – então, P{CCC}=1/8, assim como a de qualquer outro evento unitário. Vamos agora definir uma função X : $\Omega \rightarrow$R, como X(ω)=número de caras em ω. A variável aleatória X induz uma partição, $\mathcal{P}_X$, de Ω menos fina que $\mathcal{P}$:

$$\mathcal{P}_X = \{ X^{-1}(0), X^{-1}(1), X^{-1}(2), X^{-1}(3)\} = \{ \{ccc\}, \{ccC, cCc, Ccc\}, \{cCC, CcC, CCc\}, \{CCC\} \}$$

Como cada elemento de $\mathcal{P}_X$ está contido em $\mathcal{F}$, então X é uma função mensurável no espaço (Ω, $\mathcal{F}$, P), portanto uma variável aleatória neste espaço. A álgebra $\mathcal{F}_X$ de subconjuntos de, induzida por $\mathcal{P}_X$, é a que contém os 4 elementos de $\mathcal{P}_X$ como blocos básicos, ou átomos ; ela contém apenas 2^4 elementos, e qualquer elemento seu é também elemento de $\mathcal{F}$, ou seja, $\mathcal{F}_X \subset \mathcal{F}$. Note que a função X é tambem mensurável no espaço de probabilidades (Ω, $\mathcal{F}_X$, P).

Vamos agora definir outra função, Y: $\Omega \rightarrow$R, correspondente à representação binária com C=1 e c=0, logo Y(ccc)=0, Y(ccC)=1, Y(cCc)=2, Y(cCC)=3, Y(Ccc)=4, Y(CcC)=5, Y(CCc)=6, Y(CCC)=7 (A função Y tambem poderia ser descrita como $Y=4x_1+2x_2+x_3$, com x_1, x_2 e x_3 igual ao número de caras no primeiro, segundo e terceiro lançamento, respectivamente). Note que, diferente de X, a partição $\mathcal{P}_Y$ induzida por Y em Ω é tão fina quanto $\mathcal{P}$; de fato, $\mathcal{P}_Y = \mathcal{P}$ e, portanto, $\mathcal{F}_Y = \mathcal{F}$.. Desta forma, a álgebra que induz é a própria $\mathcal{F}$. Portanto, para ser mensurável, a função Y é mais exigente que X em termos de espaço métrico. Verifique que Y é mensurável, portanto uma variável aleatória, em (Ω, $\mathcal{F}$, P), mas não é em (Ω, $\mathcal{F}_X$, P).

Exercício 1.1.2 – Seja D_5 um dado não viesado de 5 lados. Considere um experimento estocástico consistindo de 2 arremessos sucessivos e independentes deste dado.

a – Construa o espaço amostral Ω correspondente ;

b – Descreva a álgebra mais fina de subconjuntos de Ω, $\mathscr{F}$;

c – Defina de forma sucinta, mas completa, a função de probabilidades P :$\mathscr{F}\rightarrow$R, físicamente adequada ao experimento ;

d – Defina X : $\Omega \rightarrow$R como a soma dos pontos nos dois arremessos, descreva $\mathscr{P}_X$ de forma econômica e completa ;

e – Determine Ω_X, o conjunto de valores possíveis de X, e determine P{X=x}, para todo $x\in\Omega_X$;

f – descreva o espaço de probabilidades (Ω, $\mathscr{F}_x$, P), $\mathscr{F}_x$ a álgebra induzida por $\mathscr{P}_X$, e P consistente com o experimento. ;

g – Defina Y : $\Omega \rightarrow$R, como $Y(w_1, w_2)=5(w_1-1)+(w_2-1)$, onde w_1 e w_2 são os resultados do primeiro e do segundo arremesso, respectivamente ; descreva $\mathscr{P}_Y$, compare-a com $\mathscr{P}_X$;

h – Mostre que Y mensurável – portanto uma variável aleatória – em (Ω, $\mathscr{F}$, P), mas não em (Ω, $\mathscr{F}_X$, P).

I – Defina uma função Z : $\Omega \rightarrow$R, que induza uma partição $\mathscr{P}_Z$ de Ω, situada entre $\mathscr{P}_X$ e $\mathscr{P}_Y$ em termos de refinamento ; determine Ω_Z e P{Z=z}, para todo $z\in\Omega_z$.

D6. **Esperança de uma Variável Aleatória: E(X)** – Seja (Ω, $\mathscr{F}$, P) um espaço de probabilidade, $\mathscr{P}=\{A_1, A_2, A_3, \ldots\}$ a partição geratriz de $\mathscr{F}$, e $X(\omega)=x_i$ o valor de X para todo $\omega\in A_i$, então

$$E(X) = \sum_{i=1}^{\infty} x_i P(A_i)$$

E(X) pode ser definida de forma alternativa mas equivalente como

$$E(X) = \sum_{x\in\Omega_X} xP\{X = x\}$$

com Ω_X o conjunto de todos os valores possíveis de X, e {X=x} ao subconjunto de Ω dos elementos tais que $X(\omega)=x$.

Variância de uma Variável Aleatória: V(X) – Seja (Ω,$\mathscr{F}$, P) um espaço de probabilidade, $\mathscr{P}=\{A_1, A_2, A_3, \ldots\}$ a partição geratriz de $\mathscr{A}$, e $X(\omega)=x_i$ o valor de X para todo $\omega\in A_i$, então

$$\mathrm{V(X) = E\left[\left(X - E(X)\right)^2\right] = \sum_{i=1}^{\infty}[x_i - E(X)]^2 P(A_i)}$$

Ou, alternativamente,

$$\mathrm{V(X) = E(X) = \sum_{x\in\Omega_X}[x - E(X)]^2 P\{X = x\}}$$

Os conceitos de esperança e variância desempenham papel central em Probabilidade.

Propriedades básicas da Esperança e da Variância

Seja X uma variável aleatória no espaço (Ω, $\mathcal{F}$, P), e sejam a e b constantes reais, com $a\neq 0$. Então, Y=aX+b é uma variável aleatória.

Prova : Y : $\Omega\rightarrow R$ é tal que $Y(\omega)=aX(\omega)+b$, para todo $\omega\in\Omega$. Então, $\Omega_Y=\{y=ax+b\ ;\ x\in\Omega_X\}$, e $\{Y=y\} = \{\omega\in\Omega\ ;\ Y(\omega)=y\} = \{X=(y-b)/a\} = \{\omega\in\Omega\ ;\ X(\omega)=x\}$, logo {Y=y} é mensurável em (Ω, $\mathcal{F}$, P), uma vez que X sendo, por hipótese, uma variável aleatória, {X=(y-b)/a} é mensurável.

A variável aleatória Y, uma transformação linear de X, tem as seguintes propriedades :

i – E(Y) = a·E(X) + b

Partindo da definição de Esperança, e aplicando as propriedades da somatória, temos:

$$E(Y)=E(aX+b)=\sum_{i=1}^{\infty}(ax_i+b)P\{X-x_i\}=a\sum_{i=1}^{\infty}x_iP\{X-x_i\}+b\sum_{i=1}^{\infty}P\{X-x_i\}=aE(X)+b$$

ii – $V(Y) = a^2\cdot V(X)$

Partindo da definição de variância, e lembrando que E(X) é uma constante – na verdade um parâmetro associado à distribuição da variável aleatória X, e não mais uma variável aleatória – e, portanto, E[E(X)]=E(X), temos:

e

$$V(Y)=E\left[\left(Y-E(Y)\right)^2\right]=E[(aX+b-aE(X)-b)^2]=E\left[\left(aX-aE(X)\right)^2\right]=a^2E\left[\left(X-E(X)\right)^2\right]$$

$$V(Y)=a^2V(X)$$

iii – $V(X) = E(X^2) - E^2(X)$

Sendo X uma variável aleatória, X^2 também o é [prove]. Partindo da definição de variância e das propriedades do operador somatório, temos:

$$\mathrm{V(X)}=\sum_{i=1}^{\infty}\left[\left(x_i-E(X)\right)^2\right]P\{X=x_i\}=\sum_{i=1}^{\infty}\left(x_i^2-2E(X)x_i+E^2(X)\right)P\{X=x_i\}$$

$$\mathrm{V(X)}=\sum_{i=1}^{\infty}x_i^2P\{X=x_i\}-2E(X)\sum_{i=1}^{\infty}x_iP\{X=x_i\}+E^2(X)\sum_{i=1}^{\infty}P\{X=x_i\}$$

$$\mathrm{V(X)}=\mathrm{E(X)}-2E^2(\mathrm{X})+\mathrm{E}^2(\mathrm{X})$$

$$\mathrm{V(X)}=\mathrm{E(X^2)}-\mathrm{E}^2(\mathrm{X})$$

Exemplo 1.1.4 – *Seja o espaço amostral Ω={a, b, c, d, e} e os eventos A={a, b} e B={c, d, e}. A classe $\mathcal{F}$ = {ϕ, A, B, Ω} é uma álgebra de subconjuntos de Ω. A função P: $\mathcal{F} \to$R definida por:*

P(ϕ)=0 P(A)=0,2 P(B)=0,8 e P(Ω)=1

é uma função de probabilidade. Note que os eventos A e B formam uma partição disjunta de Ω. A tríade (Ω, $\mathcal{A}$, P) é um espaço de probabilidades.

Exemplo 1.1.5 – *Um modelo probabilístico para o arremesso de um dado pode ser construído como se segue:*

Ω={1, 2, 3, 4, 5, 6}

$\mathcal{F}$={todos os subconjuntos de Ω}, cuja geratriz é $\mathcal{P}$={{1}, {2}, {3}, {4}, {5}, {6}}

P definida por P{1} = P{2} = P{3} = P{4} = P{5} = P{6} = 1/6. Como cada átomo tem a mesma probabilidade 1/6, a probabilidade de qualquer evento não unitário é simplesmente o número de elementos do evento, vezes 1/6. Assim, por exemplo, P{1,3,5}=3x1/6=1/2.

Exemplo 1.1.6– *No espaço de probabilidades definido no exemplo acima, sejam as seguintes variáveis aleatórias e suas propriedades fundamentais:*

a – X{i} = i.

$$E(X) = \sum_{x=1}^{6} x \cdot P\{X = x\} = \frac{1}{6}\sum_{x=1}^{6} x = \frac{21}{6} = 3{,}5$$

$$V(X) = E(X^2) - E^2(X) = \sum_{x=1}^{6} x^2 \cdot P\{X = x\} - 3{,}5^2 = \frac{1}{6}\sum_{x=1}^{6} x^2 - 12{,}5 = \frac{91}{6} - 12{,}5 = \frac{16}{6} \approx 2{,}67$$

b – Y{i} = 2·X{i} – 1, com X{i}=i, como em (a)

Neste caso, E(Y)=2·E(X) – 1= 7 – 1 = 6

V(Y) = 2^2·V(X)=4G8/3 = 32/3 ≈ 10,67

c – Z{i} = X^2{i}, com X{i}=i, como em (a)

Como já visto em (a), E(Z)=E(X^2)=91/6. Para V(Z), calcularemos primeiro E(Z^2) que equivale a E(X^4)=(1^4+2^4+…+6^4)/6 = 2275/6. Logo V(X) = 2275/6 – $(91/6)^2$≈149,14

□

Exemplo 1.1.7 – *Um experimento consiste em arremessar um dado; o espaço amostral é Ω={1, 2, 3, 4, 5, 6}. Seja $\mathcal{F}$ a álgebra completa de subconjuntos de Ω; ela contém 2^6=64 elementos, sendo seus átomos os 6 subconjuntos unitários de Ω. Qualquer função de probabilidades em (Ω, $\mathcal{F}$) será completamente definida, pela definição de seus valores para esses átomos. Em particular, a função melhor associada às propriedades físicas do experimento, considerando-se que o dado é bem equilibrado, é a que dá valor 1/6 para cada átomo. Vamos agora definir uma variável aleatória X :Ω→R, de forma trivial : X(w)= w. Assim, X pode assumir valores inteiros de 1 a 6, com probabilidades iguais a 1/6. Vamos calcular E(X) e V(X)*

$$E(X) = \sum_{x=1}^{6} xP(X = x) = \sum_{x=1}^{6} x\frac{1}{6} = \frac{1}{6}\sum_{x=1}^{6} x = \frac{21}{6} = 3,5$$

Para o cálculo de V(X), lembramos que V(X) = E(X^2) – E^2(X). Mas X^2 é, por sua vez, uma variável aleatória, função da variável aleatória original, X ; poderíamos até dar-lhe um nome provisório, digamos, Y. Assim, Y=X^2, e pode assumir valores 1, 4, 9, 16, 25 e 36, com probabilidades implícitas pelas probabilidades correspondentes do X. Por exemplo, P{Y=4}=P{X=2}=1/6. Então

$$E(Y) = E(X^2) = \sum_{x=1}^{6} x^2P(X = x) = \sum_{x=1}^{6} x^2\frac{1}{6} = \frac{1}{6}\sum_{x=1}^{6} x^2 = \frac{91}{6}$$

Portanto,

$$V(X) = E(X^2) - E^2(X) = \frac{91}{6} - \frac{49}{4} = \frac{182 - 147}{12} = \frac{35}{12} \cong 2,91667$$

Seria natural o leitor, nesta altura, se perguntar o que os valores 3,5 e 2,916666…, respectivamente Esperança e Variância da v.a. X, tem a ver com o prosaico experimento estocástico do arremesso de um dado equilibrado.

Só para dar um sabor das coisas que virão nas próximas aulas, adianto que, por exemplo, em mil arremessos sucessivos de um dado equilibrado, a probabilidade do resultado médio, vamos chamá-lo $\bar{X}_{1000}$ (lê-se xis barra mil), cair entre 3,361 e 3,636 é igual a 0,99, ou quase certeza, e que estes valores foram obtidos de $E(X) - 2,58\sqrt{\frac{V(X)}{1000}}$ e $E(X) + 2,58\sqrt{\frac{V(X)}{1000}}$, respectivamente. Mais formalmente,

$$P\left\{E(X) - 2,58\sqrt{\frac{V(X)}{1000}} < \bar{X}_{1000} < E(X) + 2,58\sqrt{\frac{V(X)}{1000}}\right\} = 0,99$$

Como nesse caso, a exploração dos fundamentos da Probabilidade revelará um mundo inteiramente novo, onde o leitor assumirá o controle intelectual de fenômenos que julgava, até agora, completamente fora de qualquer possibilidade de entendimento, previsão e controle. A propósito, generalizando o resultado acima, arriscando um pouco, construa um intervalo de probabilidade 0,99, para $\bar{X}_{10000}$, o resultado médio de dez mil arremessos sucessivos de um dado.

Exemplo 1.1.8 : O dado de 5 faces*. Na construção dos fundamentos da Teoria da Probabilidade, usaremos com frequência exemplos simples, não tão simples, e até bastante complexos, combinando arremessos de moedas e dados. O dado normal, de 6 lados tem inconvenientes: a esperança e variância do resultado são números fracionários, sendo a variância uma dízima. Isto tira um pouco a simplicidade operacional dos exemplos. Para simplicidade empregaremos mais frequentemente, o dado não viezado de 5 lado, ao qual já nos referimos no Exercício 1.1.1, e representaremos por D_5. Os resultados possíveis são Ω={1, 2, 3, 4, 5} com chances iguais. Agora, fazendo (Ω, $\mathscr{F}$, P) um espaço de probabilidades, com P a função de probabilidades que representa as propriedades reais do experimento, isto é P{w}=1/5, para todo $\omega\in\Omega$, e definindo a variável aleatória X(ω)=w, teremos : E(X)=3 e V(X)=2. Confira.*

Exercício 1.1.3 – Calcule esperança e variância de todas as variáveis aleatórias definidas no exercício 1.1.1.

Exercícios 1.1

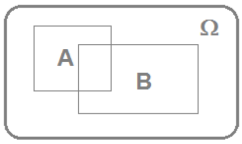
4. Mostre que se A, B e C pertencem a uma álgebra, então $A \cup B \cup C$ também pertence.

5. Mostre que uma álgebra é também fechada com relação a operações de intersecção.

6. Seja um espaço amostral Ω, de cardinalidade n. Qual é a menor álgebra de subconjuntos de Ω e quantos elementos possui a maior?

7. Definindo a álgebra gerada por uma partição de Ω como sendo a menor álgebra de subconjuntos de Ω que contém a partição, mostre como construir a álgebra gerada pela partição {A, B, C} de Ω.

8. Representações gráficas de conjuntos são úteis na visualização de relações diversas entre os mesmos. A figura ao lado representa um conjunto Ω com dois subconjuntos A e B representados. Dê nomes às entidades importantes para este exercício e ainda não denominadas, determine explicitamente a partição associada e construa a álgebra $\mathcal{A}$ induzida.



9. Construa uma função abstrata de probabilidades sobre $\mathcal{A}$ como definida em (5).

10. Seja $\mathcal{P}$ ={A_1, A_2, A_3, A_4} uma partição de Ω, de cardinalidade 4.

    a. Qual é a cardinalidade da álgebra gerada por $\mathcal{P}$?
    b. Represente-a de forma extensiva, listando todos os seus elementos?
    c. Defina uma função abstrata qualquer de probabilidades sobre a álgebra gerada.

11. Mostre que se $\mathcal{P}$ é uma partição disjunta finita de Ω tal que $\#(\mathcal{P})=m$, e $\mathcal{A}$ é a álgebra correspondente, então $\#(A)=2^m$.

12. Tratando agora de um problema concreto. Seja um experimento binário E_p, com espaço amostral Ω={S, F}, com P{S}=p. Um experimento composto corresponde a 4 repetições sucessivas de E_p.

a. Mostre Ω.
b. Construa uma partição de Ω que agrupe num mesmo átomo resultados com mesmo número de S. De nomes econômicos apropriados a esses átomos.
c. Descreva de forma sucinta mas suficiente a álgebra induzida por aquela partição.
d. Defina a função de probabilidades apropriada ao caso concreto em questão.

13. Seja o experimento aleatório que consiste de 10 realizações sucessivas de um mesmo experimento binário E_p, onde P(S) = p, com $0 \le p \le 1$ genérico.

a. Determine Ω de forma genérica (isto é, não tendo que listar todos os seus elementos), determine sua cardinalidade, e descreva sumariamente, mas de forma suficiente, a álgebra mais fina de subconjuntos de Ω.

b. Construa a função de probabilidades sobre a álgebra associada ao experimento físico concreto em questão.

14. Seja E_p um experimento aleatório binário, com Ω_o={S, F} e P{S}=p ; seja E um experimento aleatório composto por 3 repetições sucessivas e independentes de E_p.

a. Construa o espaço amostral Ω, associado a E.
b. Seja $X : \Omega \to R$ a função X. Determine $\mathscr{P}_X$, a partição de Ω induzida por X, e a álgebra $\mathscr{F}_X$, induzida por $\mathscr{P}_X$.
c. Complete o espaço métrico (Ω, $\mathscr{F}_X$) com uma função de probabilidade P consistente com o E.
d. Mostre que X é mensurável em no espaço métrico (Ω, $\mathscr{F}_X$, P), portanto é uma variável aleatória neste espaço métrico.
e. Calcule esperança e variância de X.

f. Defina $Y : \Omega \to R$ como $Y(\omega)=4\times w_1+2\times w_2+w_3$, onde w_i é o número de S na i-ésima repetição de E_p. Determine $\mathscr{P}_Y$ e a álgebra induzida $\mathscr{F}_Y$. Defina o espaço métrico (Ω, $\mathscr{F}_Y$, P), com P coerente com E.

g. Mostre que Y é uma variável aleatória em (Ω, $\mathscr{F}_Y$, P), mas não é em (Ω, $\mathscr{F}_X$, P), mas que X é mensurável em ambos os espaços.

h. Considerando (Ω, $\mathscr{F}_Y$, P), determine E(X) e V(X).

1.2 – A soma de duas ou mais variáveis aleatórias independentes

Sejam dois experimentos estocásticos elementares, E_1 e E_2, com espaços amostrais Ω_1 e Ω_2, e sejam (Ω_1, $\mathscr{F}_1$, P_1) e (Ω_2, $\mathscr{F}_2$, P_2) espaços de probabilidades apropriados respectivos. Suponha E_1 e E_2, realizados sucessiva e independentemente, isto é, o resultado de E_1 não influi no de E_2. Podemos pensar no experimento composto E, correspondente à realização sucessiva de E_1 e de E_2. O espaço amostral associado ao experimento composto E pode ser representado pelo produto cartesiano $\Omega = \Omega_1 \times \Omega_2$.

Da mesma forma, sejam $\mathscr{P}_1$ e $\mathscr{P}_2$ as partições geratrizes de $\mathscr{F}_1$ e de $\mathscr{F}_2$, respectivamente. Então $\mathscr{P}=\mathscr{P}_1\times\mathscr{P}_2$ é a geratriz da álgebra $\mathscr{F}$ de subconjuntos de Ω, adequada ao tratamento simultâneo dos dois experimentos básicos, agregados num único experimento composto, E. No caso em que $\mathscr{P}_1$ e $\mathscr{P}_2$ forem partições máximas (em peças unitários) de Ω_1 e Ω_2, respectivamente, $\mathscr{P}$ também será a partição máxima de Ω.

Exemplos ajudam a clarear a ideia: Considere dois experimentos estocásticos elementares E_1, o arremesso de uma moeda, e E_2, o de um D_5 (nosso numericamente conveniente dado de cinco lados). Temos então Ω_1={C, c} e Ω_2={1, 2, 3, 4, 5}. Considere o experimento composto: arremesso de uma moeda, seguido do arremesso de um D_5. O espaço amostral para o experimento composto será, então, $\Omega = \Omega_1\times\Omega_2$ = {(C,1), (C,2), (C, 3), (C, 4), (C, 5), (c, 1), (c, 2), (c, 3), (c, 4), (c, 5)}. [Nota : sempre que não houver possibilidade de confusão, adotaremos, para produtos cartesianos como este, a notação mais simples Ω={C1, C2, C3, C4, C5, c1, c2, c3, c4, c5}].

Num caso genérico, em que Ω_1={a, b, c, d, …} e Ω_1={1, 2, 3, 4, …}, o produto cartesiano será $\Omega = \Omega_1\times\Omega_2$ = {a1, a2, a3, a4, …, b1, b2, b3, b3, …}, e o evento {resultado a em E_1} é igual a {a}$\times\Omega_2$={a1, a2, a3, a4, …}, da mesma forma em que o evento {resultado 4 em E_2} é $\Omega_1\times${4}={a4, b4, c4, d4, …}. Admitindo-se que E_1 e E_2 são experimentos estocásticos independentes, a probabilidade de eventos unitários em Ω, como {a1}={a em E_1 e 1 em E_2} pode ser determinada por

$$P\{a1\} = P\{a1, a2, a3, a4, a5, \ldots\} \times P\{a1, b1, c1, d1, \ldots\} = P_1\{a\} \times P_2\{1\}$$

Numa ilustração ligeira, no caso dos arremessos de moeda e D_5, P{C2}=0,5×0,2=0,1

Uma vez definidos os valores de P para todos os eventos unitários, a função probabilidade fica completamente definida.

Exemplo

1.2.1 – Sejam dois experimentos aleatórios E_1 e E_2, respectivamente o arremesso de uma moeda e de um D_5. Seja E o experimento composto de E_1 e E_2, realizados sucessiva e independentemente. Aqui, Ω_1={C, c} e Ω_2{1, 2, 3, 4, 5}, logo $\Omega = \Omega_1 \times \Omega_2$={C1, C2, C3, C4, C5, c1, c2, c3, c4, c5}. Sejam $\mathfrak{F}_1$ e $\mathfrak{F}_2$ as álgebras completas de subconjuntos de Ω_1, Ω_2. A função $P_1:\mathfrak{F}_1 \rightarrow R$ é definida por $P_1\{C\}=P_1\{c\}=0{,}5$; a função $P_2:\mathfrak{F}_2 \rightarrow R$ é definida por $P_2\{1\} = P_2\{2\} = P_2\{3\} = P_2\{4\} = P_2\{5\} = 0{,}2$. Considerando agora o espaço amostral conjunto e a álgebra completa de subconjuntos de Ω, e a função de probabilidades derivada de P_1 e P_2, temos que a probabilidade de {C1},, bem como a de qualquer outro evento unitário, é igual a 0,5x0,2=0,1; a partir daí P fica completamente definida para todo elemento de $\mathfrak{F}$.

Sejam dois experimentos, E_1 e E_2, realizados independentemente, e sejam duas variáveis aleatórias, X e Y, associadas a E_1 e E_2, e definidas sobre os espaços de probabilidade (Ω_1, $\mathfrak{F}_1$, P_1) e (Ω_2, $\mathfrak{F}_2$, P_2), respectivamente. Vamos admitir que X assume valores em $\Omega_X=\{x_1, x_2, x_3, \ldots\}$ e Y, em $\Omega_Y=\{y_1, y_2, y_3, \ldots\}$. Por definição, temos :

$$E(X) = \sum_{i=1}^{\infty} x_i P_1(X = x_i) \quad e \quad E(Y) = \sum_{j=1}^{\infty} y_j P_2(Y = y_j)$$

Acomodando X e Y no espaço métrico produto (Ω, $\mathfrak{F}$, P), adequadamente definido, e definindo Z=X+Y, temos que Z é uma variável aleatória (os subconjuntos de Ω do tipo {Z=z} pertencem a $\mathfrak{F}$), e

$$E(Z) = \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} (x_i + y_j) P_1(X = x_i) P_2(Y = y_j)$$

$$= \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} x_i P_1(X = x_i) P_2(Y = y_j) + \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} y_j P_1(X = x_i) P_2(Y = y_j)$$

$$E(Z) = \sum_{i=1}^{\infty} x_i P_1(X = x_i) \sum_{j=1}^{\infty} P_2(Y = y_j) + \sum_{i=1}^{\infty} P_1(X = x_i) \sum_{j=1}^{\infty} y_j P_2(Y = y_j)$$

$$E(Z) = \sum_{i=1}^{\infty} x_i P_1(X = x_i) + \sum_{j=1}^{\infty} y_j P_2(Y = y_j) = E(X) + E(Y)$$

Para V(Z), lembramos que $V(Z)=E(Z^2)-E^2(Z)$. Sendo E(Z) já conhecida, resta-nos determinar $E(Z^2)$:

$$E(Z^2) = \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} (x_i + y_j)^2 P_1(X = x_i) P_2(Y = y_j)$$

$$= \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} x_i^2 P_1(X = x_i) P_2(Y = y_j) + \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} y_j^2 P_1(X = x_i) P_2(Y = y_j)$$

$$+ 2 \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} x_i y_j P_1(X = x_i) P_2(Y = y_j)$$

$$= E(X^2) + E(Y^2) + 2E(X)E(Y)$$

Portanto,

$$V(Z) = E(Z^2) + E^2(Z) = E(X^2) + E(Y^2) + 2E(X)E(Y) - [E(X) + E(Y)]^2$$
$$V(Z) = E(X^2) - E^2(X) + E(Y^2) - E^2(Y) = V(X) + V(Y)$$

Temos então, completos, dois resultados muito importantes relativos à soma de duas variáveis independentes: Se X e Y são duas variáveis aleatórias independentes, e Z=X+Y, então a esperança de Z é a soma das esperanças de X e de Y, e a variância de Z é a soma das variâncias de X e de Y.

Exemplos:

1.2.2 – Um experimento elementar D_5 consiste do arremesso de um dado de cinco lados. O experimento $(D_5)^2$ é composto por duas repetições sucessivas e independentes de D_5. Sejam X_1 e X_2 variáveis aleatórias independentes, correspondentes à pontuação em cada repetição de D_5, e seja $Y= X_1 + X_2$.

O espaço de probabilidades (Ω, $\mathfrak{F}$, P) melhor ajustado a este problema tem Ω = {1 1, 1 2, 1 3, 1 4, 1 5, 2 1, 2 2, … , 5 3, 5 4, 5 5}, com 25 elementos; $\mathfrak{F}$ é a álgebra completa de subconjuntos de Ω (tem, portanto, 2^{25} elementos… não vamos listá-los todos, portanto), e P é bem definida uma vez que qualquer evento unitário tem probabilidade $0{,}2^2$=0,04.

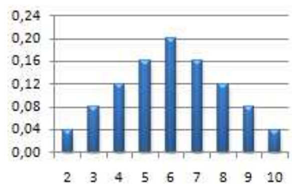
Já sabemos que $E(X_i)=3$ e $V(X_i)=2$, portanto E(Y)=6 e V(Y)=4. Podemos verificar esses resultados diretamente, determinando a distribuição de probabilidades de Y. Sabemos que Y assume valores em Ω_Y={2, 3, …, 10}; precisamos agora determinar a probabilidade de cada um desses 9 valores:

P{Y=2} = P{1 1} = 0,04 P{Y=3} = P{1 2, 2 1} = 2×0,04 = 0,08 P{Y=4} = P{1 3, 2 2, 3 1} = 0,12

P{Y=5} = P{1 4, 2 3, 3,2 4 1} = 0,16 P{Y=6} = P{1 5, 2 4, 3 3, 4 2, 5 1} = 0,20 P{Y=7} = P{2 5, 3 4, 4 3, 5 2} = 0,16

P{Y=8} = P{3 5, 4 4, 5 3} = 0,12 P{Y=9} = P{4 5, 5 4} = 0,08 P{Y=10} = P{5 5} = 0,04

Note que, enquanto a distribuição de probabilidades de X é uniforme sobre seus 5 valores possíveis, a de Y não, concentrando mais probabilidade nos valores centrais que nos extremos, como mostra a figura ao lado. Veremos mais adiante como esta é uma tendência geral.



O valor médio esperado de Y pode ser calculado diretamente como

$$\mathrm{E(Y)} = \sum_{y=2}^{10} \mathrm{yP\{Y = y\}}$$

$$= 2\times 0{,}04 + 3\times 0{,}08 + 4\times 0{,}12 + 5\times 0{,}16 + 6\times 0{,}20 + 7\times 0{,}16 + 8\times 0{,}12 + 9\times 0{,}08 + 10\times 0{,}04 = 6$$

Para calcular a variância de Y, primeiro vamos calcular $E(Y^2)$:

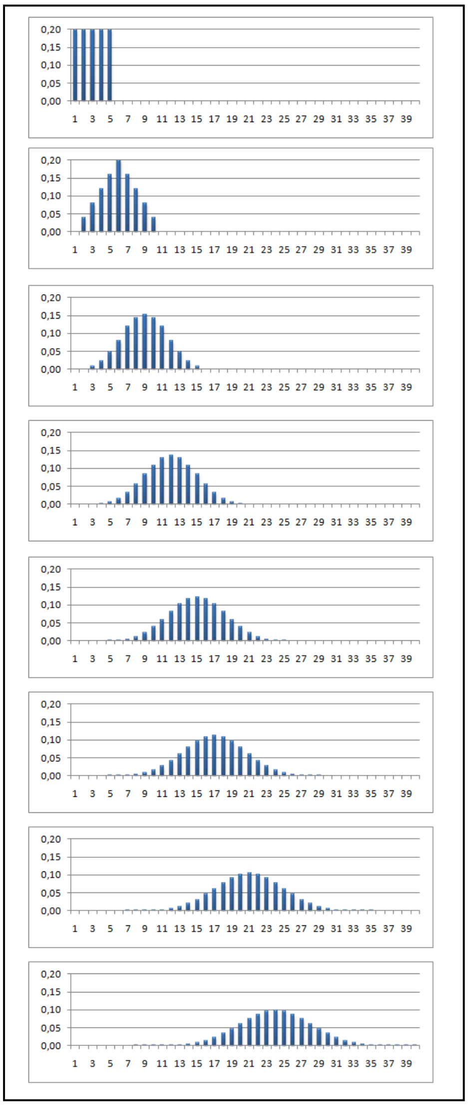
$$\mathrm{E(Y^2)} = \sum_{y=2}^{10} \mathrm{y^2P\{Y = y\}}$$

$$= 4\times 0{,}04 + 9\times 0{,}08 + 16\times 0{,}12 + 25\times 0{,}16 + 36\times 0{,}20 + 49\times 0{,}16 + 64\times 0{,}12 + 81\times 0{,}08 + 100\times 0{,}04$$

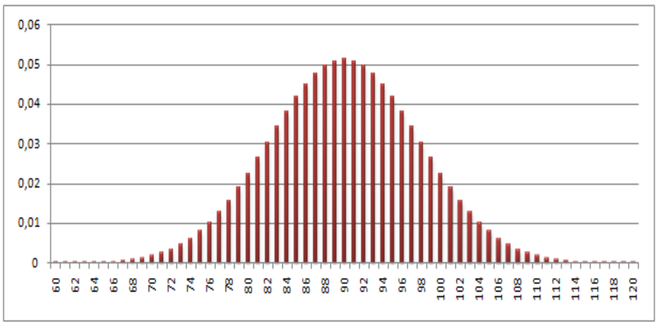
$$= 40$$

Portanto, $V(Y) = E(Y^2) - E^2(Y) = 40 - 6^2 = 4$, conforme já havíamos previsto pelas propriedades da soma de duas variáveis aleatórias independentes.

As propriedades da soma de duas variáveis aleatórias independentes, demonstradas acima (para variáveis discretas), pode ser facilmente generalizada para a soma de um número n qualquer de variáveis aleatórias independentes. Assim, se X_1, X_2, X_3, …, X_n são variáveis aleatórias independentes, tais que $E(X_i)=\mu_i$ e $V(X_i)=\sigma_i^2$, então $E(Y) = \sum_{i=1}^{n} \mu_i$ e $V(Y) = \sum_{i=1}^{n} \sigma_i^2$.



A figura ao lado mostra a distribuição de probabilidades de S_n, para n de 1 a 8; abaixo vê-se a distribuição de probabilidades de S_{30}, a soma dos resultados individuais de 30 arremessos sucessivos e independentes de D_5. Embora, teoricamente os valores possíveis são os inteiros de 30 a 150, os valores de 75 a 105 totalizam probabilidade superior a 0,95, enquanto que a probabilidade total acumulada dos valores inferiores a 60 é apenas 26,6 milionésimos, o mesmo valendo para valores acima de 120, já que a distribuição, naturalmente, é simétrica.



Exemplos

1.2.3 – Seja E_p o experimento binário elementar, com P{S}=p. Um experimento composto $(E_p)^n$ consiste em repetir E_p sucessiva e independentemente, n vezes. O espaço amostral para $(E_p)^n$ é o produto cartesiano de {S, F} por si mesmo, n vezes: Ω = {S, F} × {S, F} ×{S, F} ×{S, F} ×…×{S, F}. Os elementos de Ω são sequências de S's e F's de comprimento n, indo desde uma sequência de n S's até uma de n F's, passando por sequências com um

F e n-1 S's – existem n dessas – ; com dois F's e n-2 S's – das quais existem C_n^2 sequências distintas – e assim por diante. No total existem 2^n sequências diferentes em Ω. A probabilidade de cada um desses eventos unitários compostos depende de p e do número de S's. Assim, um evento unitário consistindo de uma sequência com x Sucessos e, consequentemente, n-x Fracassos, para qualquer x inteiro tal que $0 \leq x \leq n$, tem probabilidade igual a $p^x(1-p)^{n-x}$.

Voltando a cada experimento elementar, com Ω_1={S, F} e $\mathfrak{F}$={ϕ, {S}, {F}, Ω}, e P definida por P{S}=p. seja X a variável aleatória correspondente ao número de Sucessos obtido no experimento elementar. X pode então assumir valor 0 ou 1, com P{X=1}=p; logo, E(X)=p e $E(X^2)$=p, portanto $V(X)= E(X^2)–E(X)=p–p^2=p(1–p)$.

Com as n repetições independentes de E_p, teremos n versões independentes e identicamente distribuídas da variável aleatória X, que denominaremos X_1, X_2, X_3, ..., X_n, todas com mesma esperança p e mesma variância, p(1-p). Definindo Y:$\Omega \rightarrow$R como $Y = \sum_{i=1}^{n} X_i$, temos uma variável aleatória, com esperança e variância conhecidas: E(Y)=np e V(Y)=np(1-p). Sabemos ainda que Y pode assumir valores inteiros, de 0 – no caso em que as n repetições resultaram em n fracassos – até n, no caso oposto, em que as n repetições resultaram em n sucesso. A probabilidade de cada um desses n+1 resultados possíveis de Y pode ser determinada, simplesmente contando o número de elementos em cada evento do tipo {Y=y}. Para y=0 já sabemos: só existem uma sequência – n fracassos consecutivos – isto é, a cardinalidade de {Y=0} é 1, e sua probabilidade, portanto, é cuja probabilidade é $(1-p)^n$. Já a cardinalidade de {Y=1} é $C_n^1 = n$, correspondente às n posições alternativas para o único S em meio aos (n-1) F's da sequência. Isoladamente, essas n sequências compõem n eventos unitários, todos com a mesma probabilidade $p(1-p)^{n-1}$, logo

$$P\{Y = 1\} = C_n^1 \times p \times (1-p)^{n-1} = n \times p \times (1-p)^{n-1}$$

Generalizando :

$$P\{Y = y\} = \#\{Y = y\} \times p^y \times (1-p)^{n-y} = C_n^y \times p^y \times (1-p)^{n-y}.$$

Na expressão anterior, #{Y=y} representa a cardinalidade – ou número de elementos – de {Y=y}.

Este resultado tem extraordinária importância e terá papel central neste curso. Voltaremos logo e frequentemente a ele.

Exercícios:

1.2.1 – Um experimento elementar D_5 consiste em arremessar um dado de cinco lados. O experimento $(D_5)^n$ é composto de n repetições sucessivas e independentes de D_5. Seja X o resultado uma realização de D_5 e Y a soma dos n resultados. Monte um espaço de probabilidades apropriado a este experimento composto, assumindo-se que o dado é não viesado. Descreva genericamente o espaço amostral Ω para $(D_5)^n$ – cardinalidade, cardinalidade do evento {Y=y}, etc. – procure desenvolver uma expressão para P{Y=y}, mas, mesmo que não consiga, determine E(Y) e V(Y).

1.2.2 – Sejam dois experimentos aleatórios independentes, E_1 e E_2; o primeiro consiste do arremesso de quatro moedas e o segundo o de dois dados D_5. Um experimento E é composto por E_1, seguido de E_2. Assumindo moedas e dados não viciados, construa o espaço de probabilidades adequado ao caso. Definindo X_1 como o número de Caras em E_1, e X_2 como a pontuação total em E_2 e $Y=X_1+X_2$, determine a esperança e a variância de cada uma dessas três variáveis aleatórias.

Um experimento elementar E_1 pode ser repetido sucessiva e independentemente, um número arbitrário n de vezes. Suponha uma variável aleatória X associada a este experimento elementar ; repetido n vezes, teremos $X_1, X_2, X_3, \ldots, X_n$, variáveis aleatórias independentes e identicamente distribuidas. Interessa-nos, em particular, o comportamento probabilístico da média aritmética desses n resultados individuais :

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$

Denominando $S_n = \sum_{i=1}^{n} X_i$, podemos afirmar, da generalização dos resultados para esperança e variância da soma de duas variáveis aleatórias independentes, para o caso da soma de n variáveis aleatórias independentes, que $E(S_n)=n\mu_X$ $V(S_n)= n\sigma_X^2$, onde μ_X e σ_X^2 são, respectivamente, a esperança e a variância de X. Como $\bar{X}_n = \frac{1}{n}S_n$ concluimos :

$$E(\bar{X}_n) = \frac{1}{n}E(S_n) = \mu_x \text{ e } V(X) = \frac{1}{n^2}V(S_n) = \frac{1}{n}\sigma_X^2$$

Notem que a média de repetições independentes de uma variável X, tem a mesma esperança que X, mas uma variância que cai com o número de observações, n. Este resultado é extraordinariamente importante, conforme veremos a adiante. Antes vamos provar um resultado básico fundamental, a Desigualdade de Tchebyshev.

Desigualdade de Tchebyshev : Seja uma variável aleatória X, com esperança μ_X e variância σ_X^2, e ε uma constante positiva arbitrariamente pequena ; então :

Pafnuti Lvovitch Tchebychev

P. L. Tchebychev

Nascimento	4 de maio (16 de maio no Calendário juliano) de 1821 Borovsk
Morte	26 de novembro/8 de dezembro de 1894 São Petersburgo
Nacionalidade	Russo
Alma mater	Universidade de São Petersburgo
Tese	1849: On Integration by Means of Logarithms
Orientador(es)	Nikolai Brashman
Orientado(s)	Andrei Andreyevich Markov, Aleksandr Lyapunov

$$P\{|X-\mu_X| > \varepsilon\} \le \frac{\sigma_X^2}{\varepsilon^2}$$

A ideia da prova deste resultado é simples e charmosa :

$$V(X) = E[(X-\mu_X)^2] = \sum_{x \in \Omega_X} (x-\mu_X)^2 P\{X=x\}$$

$$V(X) = \sum_{x \in \Omega_X; |x-\mu_X| < \varepsilon} (x-\mu_X)^2 P\{X=x\} + \sum_{x \in \Omega_X; |x-\mu_X| \ge \varepsilon} (x-\mu_X)^2 P\{X=x\}$$

$$V(X) \ge \sum_{x \in A; |x-\mu_X| < \varepsilon} (x-\mu_X)^2 P\{X=x\} + \varepsilon^2 \cdot \sum_{x \in A; |x-\mu_X| \ge \varepsilon} P\{X=x\}$$

$$V(X) \ge \sum_{x \in \Omega_X; |x-\mu_X| < \varepsilon} (x-\mu_X)^2 P\{X=x\} + \varepsilon^2 P\{|X-\mu_X| \ge \varepsilon\}$$

Logo, como a somatória no segundo membro da desigualdade acima é não negativa, concluímos que :

E, daí, o resultado :

$$P\{|X-\mu_X| \ge \varepsilon\} \le \frac{V(X)}{\varepsilon^2} = \frac{\sigma_X^2}{\varepsilon^2}$$

Conforme queríamos demonstrar.

Ao limite $\frac{V(X)}{\varepsilon^2}$ denominaremos limite superior de Tchebychev.

Pode não parecer, assim, à primeira vista, mas o resultado acima é extraordinário por estabelecer uma primeira relação de limitação de valores extremos de X, em função de sua variância. Vemos aqui que, quanto menor a variância de X, menor a probabilidade de X assumir valores muito afastados de sua esperança. Combinando este resultado, devido ao matemático russo P. L. Tchebychev (1821-1894), com o resultado anterior, para

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$

chegamos a conclusões notáveis, como, por exemplo, que, à medida que n cresce, a média $\bar{X}_n$ vai sendo espremida probabilisticamente para uma vizinhança cada vez mais estreita de sua esperança μ_X. Dizemos que $\bar{X}_n$ converge em probabilidade para μ_X. Este é um resultado glorioso, que permite um avanço considerável na compreensão profunda de como o Universo funciona; a sua descoberta é uma gigantesca conquista do intelecto humano.

Exemplo 1.2.3

Seja D_5 o experimento aleatório do arremesso de um dado não viciado de 5 lados ; $(D_5)^n$ é o experimento composto por n repetições independentes de D_5. Sejam X_1, X_2, ..., X_n os n resultados independentes, $S_n = \sum_{i=1}^{n} X_i$ é a pontuação total acumulada nas n repetições e $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ é a pontuação média por repetição. Podemos agora provar que, à medida em que n cresce, $\bar{X}_n$vai ficando probabilísticamente confinada em uma vizinhança cada vez mais estreita de 3, a esperança de X.

Ora, sabemos que $V(\bar{X}_n) = \frac{1}{n}V(X) = \frac{2}{n}$, logo vai tendendo a zero na mendida em que n cresce para ∞. Assim, não importa quão pequeno for o ε, podemos fazer $P\{3 - \varepsilon < \bar{X}_n < 3 + \varepsilon\}$ tão grande quanto desejarmos, bastando fazer n suficientemente grande.

Como ilustração, seja ε=0,01. Vamos determinar n para que possamos garantir, pela desigualdade de Tchebychev, a probabilidade de $\bar{X}_n$ cair entre 2,99 e 3,01 seja pelo menos 0,99. Por Tchebychev, temos :

$$P\{|\bar{X}_n - E(\bar{X}_n)| \geq \varepsilon\} \leq \frac{V(\bar{X}_n)}{\varepsilon^2}$$

Logo :

$$P\{3 - \varepsilon < \bar{X}_n < 3 + \varepsilon\} \geq 1 - \frac{V(\bar{X}_n)}{\varepsilon^2}$$

$$P\{3 - 0{,}01 < \bar{X}_n < 3 + 0{,}01\} \geq 1 - \frac{V(\bar{X}_n)}{0{,}01^2}$$

$$P\{2{,}99 < \bar{X}_n < 2{,}99\} \geq 1 - \frac{\frac{2}{n}}{0{,}01^2} = 0{,}99$$

Portanto :

$$1 - \frac{20.000}{n} = 0{,}99 \;\therefore n = \frac{20.000}{0{,}01} = 2.000.000$$

Concluimos, então, que após 2 milhões de arremessos do dado não viciado, de 5 lados, a média dos resultados estará, com probabilidade pelo menos 0,99, entre 2,99 e 3,01.

A desigualdade de Tchebychev, de valor teórico extraordinário, tem relativamente pouca utilidade no cálculo prático de probabilidades, uma vez que, sendo muito conservadora, fornece limites inferiores $\frac{1}{\varepsilon^2}\mathrm{V}(\overline{\mathrm{X}}_\mathrm{n})$ altos demais, como nesse exemplo, o que leva a exigências exageradas sobre n. Mais adiante introduziremos resultados teóricos mais poderosos. Com eles podemos determinar o valor preciso de n para se garantir $P\{2{,}99 < \overline{\mathrm{X}}_\mathrm{n} < 2{,}99\} = 0{,}99$. Passando um pouco o carro diante dos bois : são necessários 133.128 arremessos. Com 133.128 arremessos de D_5, a probabilidade da média dos 133.128 resultados individuais cair no intervalo (2,99, 3,01) é igual a 0,99. Mais adiante veremos como se chegou a este resultado.
Ainda ilustrando a baixa utilidade da desigualdade de Tchebychev, para cálculos prático : com n=200 e ε=0,1, tem-se $P\{|\overline{\mathrm{X}}_{200} - 3| \geq 0{,}1\} \leq \frac{\frac{2}{200}}{0{,}01} = 1$, completamente não informativa, como informar que a chance de sobrevivência de um paciente é menor ou igual a 100%. Tente n=100, ou ε=0,01.

Exercício 1.2.3 – Uma população de tamanho N=1 milhão de indivíduos, tem a altura A (em cm) distribuída segundo a tabela abaixo (a é a altura em cm e F(a) a frequência absoluta de a, ou seja, o número total de indivívuos na população com A=a)

a	f(a)	a	f(a)	a	f(a)	a	f(a)	a	f(a)
141	1	151	1.353	161	33.284	171	14.832	181	127
142	4	152	2.360	162	36.963	172	11.258	182	57
143	5	153	3.625	163	39.114	173	7.920	183	37
144	12	154	5.284	164	39.854	174	5.369	184	12
145	17	155	8.005	165	39.169	175	3.448	185	4
146	65	156	11.240	166	36.867	176	2.209	186	1
147	128	157	14.960	167	32.852	177	1.380	187	2
148	236	158	19.549	168	28.970	178	812	188	1
149	456	159	24.332	169	23.827	179	448		
150	773	160	29.005	170	19.551	180	222		

Um experimento E_1 consiste em sortear um desses indivíduos ao acaso e medir-lhe a altura. Seja o valor da altura o resultado do experimento.

a – Qual é o conjunto de todos os resultados possíveis deste experimento?

b – Construa, conceitualmente, a álgebra completa, $\mathfrak{F}$, de subconjuntos do espaço amostral, bem como uma função de probabilidade P consistente com o experimento.

c – Defina a função X :Ω→R por X(ω)=ω. Mostre que X é uma variável aleatória em (Ω, $\mathfrak{F}$, P). Determine sua esperança e sua variância. Mostre que $E(X) = \frac{1}{N}\sum_{i=1}^{N} A_i$.

d – Repita E_1, sucessiva e independentemente, n vezes. Para isto, a rigor, você deverá fazer sorteio com reposição : cada indivíduo sorteado é medido e devolvido à população, podendo inclusive vir a ser sorteado novamente. Seja $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$. Determine a esperança e a variância de $\bar{X}_n$.

e – Usando a desigualdade de Tchebychev, determine um limite superior para $P\{|\bar{X}_n - E(X)| > 1cm\}$, para n=1.000 e para n=10.000.

f – Usando-se $\bar{X}_n$ como estimativa de E(X) – a altura média da população – para n=1000, qual a probabilidade de se cometer um erro superior a 1cm ?

g – Qual deveria ser o número de repetições n de E para que o limite superior de Tchebychev para que $P\{|\bar{X}_n - E(X)| > 0,5cm\}$ seja 0,05.

Nota: a desigualdade de Tchebychev tem extraordinário valor teórico mas, como vimos, valor prático limitado. Seu limite superior é geralmente muito frouxo, frequentemente dando valores superiores a 1. Mais adiante desenvolveremos ferramental específico para tratar de problemas práticos como o sugerido no exercício anterior.

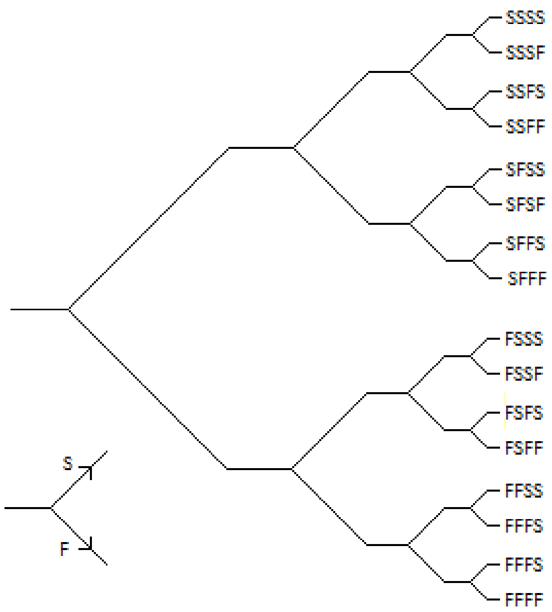
1.4 – Alguns Modelos Discretos Importantes

Desenvolveremos aqui alguns modelos probabilísticos discretos importantes, explorando suas propriedades e possibilidades de aplicação a problemas concretos. Começaremos com modelos derivados de experimentos binários.

1.4.1 – O Modelo Binomial

Seja um experimento binário com P{S} = p, que representamos por E_p. Um experimento composto consiste em repetir E_p, independentemente, n vezes. O espaço amostral associado a este experimento composto – vamos representá-lo por E_p^n – contém, portanto, 2^n elementos, Nos casos para n=3 e n=4, os espaços amostrais são: Ω_3 ={FFF, FFS, FSF, SFF, FSS, SFS, SSF, SSS}

Ω_4 ={ FFFF, FFFS, FFSF, FSFF, SFFF, FFSS, FSFS, FSSF, SFFS, SFSF, SSFF, FSSS, SFSS, SSFS, SSSF, SSSS }



Uma função X: $\Omega \to R$ interessante é aquela que conta os sucessos em cada $\omega \in \Omega$. Com n=4, por exemplo, X(SFSF)=2. Para estudar o comportamento probabilístico desta função, vamos definir $\mathfrak{F}$ como a álgebra máxima de subconjuntos de Ω – isto é, a classe de todos os subconjuntos de Ω. A função de probabilidade P: $\mathfrak{F} \to [0, 1]$ naturalmente associada às características concretas do experimento é tal que $P\{\omega\}=p^x(1-p)^{n-x}$, onde x é o número de S's em ω, para qualquer $\omega \in \Omega$

A função X, como definida, é uma variável aleatória no espaço de probabilidades $(\Omega, \mathfrak{F}, P)$[1], uma vez que a partição de Ω induzida por X está contida em $\mathfrak{F}$.

---

[1] A rigor, a álgebra máxima não é necessária para se acomodar X como uma variável aleatória. Para isto bastaria a álgebra gerada pela partição de Ω induzida por X, ou seja, $\{A_0, A_1, A_2, \ldots, A_n\}$, onde A_i é o conjunto de todos os elementos de $\omega \in \Omega$ com $X(\omega)$=i.

No caso n=4, a partição induzida por X tem 5 elementos:

A_0 = {X=0} = {FFFF}

A_1 = {X=1} = {FFFS, FFSF, FSFF, SFFF}

A_2 = {X=2} = {FFSS, FSFS, FSSF, SFFS, SFSF, SSFF}

A_3 = {X=3} = {FSSS, SFSS, SSFS, SSSF}

A_4 = {X=4} = {SSSS}

Uma descrição completa do comportamento probabilístico da variável aleatória X é dada pela sua função de distribuição de probabilidades. Esta função dá a probabilidade de X assumir cada um de seus valores possíveis. No caso dos experimentos binomiais, a função de distribuição de probabilidades de X é desenvolvida seguindo-se um roteiro lógico muito simples. Como sabemos, qualquer evento unitário {ω} tem probabilidade igual a $p^x(1-p)^{n-x}$, onde x é o número de S's em ω, logo, para o cálculo de P{X=x}, basta determinar a cardinalidade do evento {X=x}, e multiplicá-la por $p^x(1-p)^{n-x}$. Por exemplo, com n = 3 :

P{SSF} = P{SFS} = P{FSS} = p^2q

P{X=2} = P{SSF, SFS , FSS} = $C_3^2\, p^2q = 3p^2q$

Com n=4...

P{FFSS} = P{FSFS} = P{FSSF} = P{SFFS} = P{SFSF} = P{SSFF} = p^2q^2

P{X=2} = P{FFSS, FSFS, FSSF, SFFS, SFSF, SSFF} = $C_4^2\, p^2q^2 = 6p^2q$

Ora, o evento {X=x} é o conjunto de todos os elementos de Ω compostos de x sucessos e, consequentemente, (n-x) fracassos. Como a probabilidade associada a qualquer evento unitário com estas caracteríscicas é, simplesmente, p^xq^{n-x}, para calcular P{X=x} é suficiente determinar o número de maneiras diferentes que se pode compor uma seqüência de comprimento n, formada por de x sucessos e (n-x) fracassos e multiplicar este número por p^xq^{n-x}. Este número é C_n^x, ou seja:

$$\#\{X = x\} = C_n^x = \frac{n!}{x!\,(n-x)!}$$

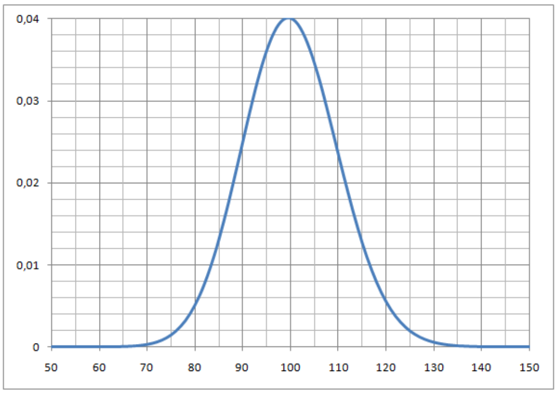
Portanto:

$$P\{X = x\} = C_n^x p^x (1-p)^{n-x}$$

Dizemos que uma variável aleatória como X, que representa o número de sucessos em n repetições independentes de um mesmo experimento binário E_p, tem distribuição binomial, com parâmetros n e p, e representamos por X~b(n, p). A denominação “binomial” vem da estreita associação formal da expressão algébrica de P{X=x} com o binômio de Newton. De fato, a expansão do binômio de Newton $(p+q)^n$ é algebricamente idêntica a soma de, justificando a denominação. Para 0<p<1 e fazendo q=1-p:

$$1 = 1^n = (p+q)^n = \sum_{x=0}^{n} C_n^x \cdot p^x \cdot q^{n-x} = \sum_{x=0}^{n} P\{X = x\} = 1$$

O cálculo de P{X=x} para X~b(n, p) pode ser feito sem grandes dificuldades – mesmo para valores grandes de n como por exemplo no caso X~b(1000, 0,5) Neste caso a fórmula recursiva $P\{X = x+1\} = \frac{n-x}{x-1} \cdot \frac{p}{q} \cdot P\{X = x\}$ pode ser útil na montagem de planilhas que calculem a função de distribuição de probabilidades para binomiais com n muito grande, mesmo com recursos de cálculo computacional de uso genérico, como o Excel.

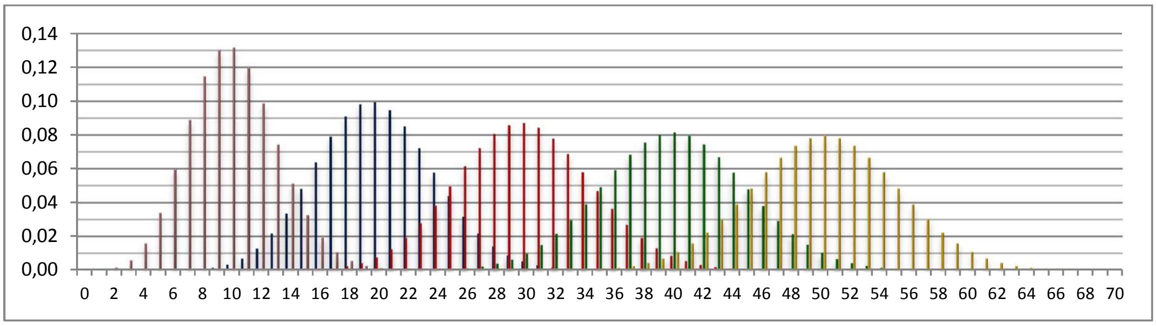


A figura ao lado mostra a curva envoltória da função de distribuição de probabilidades para a b(5000, 0,02), para valores de x variando de 60 a 140. Como se pode ver, além deste intervalo, abaixo ou acima, as probabilidades de ocorrência de X são desprezíveis.

De fato,

$$P\{60 \le X \le 140\} = \sum_{x=60}^{140} P\{X = x\} = \sum_{x=60}^{140} C_{5000}^x \cdot p^x \cdot (1-p)^{5000-x} = 0{,}999463 .$$

A figura abaixo, representa, de uma forma mais aproriada, o comportamento de binomiais muito menos extremas: X~b(100, p), para p=0,10, 0,20, 0,30, 0,40 e 0,50.



Exercícios 1.4. : (use computador quando necessário):

13. Seja X ~ b(10, 0,3). Calcule, utilizando as expressões e definições apropriadas:
    c. P{X=5]
    d. A tabela completa da função de distribuição de probabilidades de X.
    e. A esperança de X
    f. A variância de X
14. Um dado é arremessado duas vezes. Seja X igual ao maior dos resultados, e Y o menor. Determine a função de distribuição de probabilidades, a esperança e a variância, tanto de X como de Y. Mostre através de um argumento simples e direto que as duas variáveis aleatórias não são independentes.
15. Expanda a solução acima para 3 arremessos do dado, e depois generalize a solução acima para n de arremessos.
16. Um dado é arremessado 120 vezes. Seja X o número de vezes que se conseguiu o resultado máximo.
    g. Calcule P{X=18}.
    h. Calcule E(X) e V(X)
    i. Calcule P{X<10}
17. Uma moeda não tendenciosa é arremessada 1000 vezes. Seja X o número de caras obtidas. Calcule
    j. P{X=500}
    k. P{X<400}
    l. P{450≤X≤550}
    m. Determine o menor valor inteiro positivo **a**, tal que o intervalo de valores inteiros [500-**a**, 500+**a**] tenha probabilidade superior a 0,95.
18. Uma urna tem 100 bolinhas idênticas, menos na cor: 5 são brancas e as restantes são pretas. Uma bolinha é retirada por sorteio aleatório; sua cor é anotada e ela é devolvida à urna. Esta operação é

repetida 100 vezes. Qual a probabilidade de que, no final, uma bolinha branca tenha saído exatamente 5 vezes?

19. Os eleitores de uma cidade muito grande estão divididos: 30% pensam em votar no candidato A, os demais têm outras preferências, ou ainda não decidiram. Você sorteia 10 eleitores aleatoriamente. Qual a probabilidade de que, entre eles
    n. tenham exatamente 3 eleitores de A?
    o. não tenha nem um eleitor de A?
    p. sejam, todos, eleitores de A?
    q. O número de eleitores de A esteja no intervalo [2, 4]

20. Refaça o exercício anterior para, agora, uma amostra aleatória de n eleitores. Qual a probabilidade de que, entre eles
    r. tenham exatamente 30 eleitores de A?
    s. não tenha nem um eleitor de A?
    t. sejam, todos, eleitores de A?
    u. O número de eleitores de A esteja no intervalo [20 , 40].

21. No contexto dos 2 exercícios anteriores você vai, agora, sortear 1000 eleitores, aleatoriamente. E você usará a fração de eleitores de A na amostra (X/1000) como uma estimativa da fração de eleitores de A na cidade (admita que a população seja tão grande que p pode ser considerado constante ao longo do processo amostral).
    v. Qual a probabilidade de que o erro absoluto cometido |(X/1000-0,30)| seja superior a 0,05.?
    w. Qual a probabilidade de que sua estimativa seja inferior a 0,20. E superior a 0,40?
    x. Determine um intervalo em torno do valor verdadeiro, 0,30, no qual a probabilidade de sua estimativa cair seja igual ou superior a 0,95.
    y. Comente sobre o potencial deste procedimento para se estimar p em uma população muito grande, a partir de amostra de uma fração muito pequena da mesma.

22. Sejam n repetições de um mesmo experimento binário E_p. Seja X o número de S's nas n repetições, U_m o número de S's nas m primeiras e V_m o número de S's nas m últimas repetições de E_p, com $m \leq n$. Seja n=10 e p=0,40.

a) Calcule

a. $P[X>5 / U_6<3]$

b. $P[V_7<3 / U_5>2]$

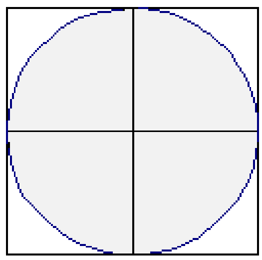
c. $P[U_1=0 / X=9]$

b) Determine a distribuição condicional de U_6, dado $V_7=3$

c) $E(U_6 / V_7=3)$ a esperança condicional de U_6, dado $V_7=3$.

d) Mostre que U_5 e V_5 são independentes

23. Um ponto (x, y) é escolhido aleatóriamente no plano cartesiano, uniformemente sobre o quadrado (-1, 1) x (-1, 1). Se o mesmo cai dentro do círculo de raio unitário, com centro em (0, 0), o experimento é dito ser um sucesso. Em 400 repetições independentes deste experimento, seja X o número total de sucessos obtidos.



z. Calcule E(X) e V(X)

aa. P{ X>300 }

Algumas propriedades básicas da Distribuição Binomial.

Seja E_p um experimento binomial com, com Ω={S, P}, $\mathfrak{F}$={ϕ, {S}, {F}, Ω}, e P definida por P{S}=p. O espaço de probabilidades (Ω, $\mathfrak{F}$, P) é uma estrutura extraordinariamente simples ; não se pode pensar em espaço métrico mais simples. Neste espaço vamos definir a função T : $\Omega \rightarrow$R, como T(S)=1 e T(F)=0. A função T é mensurável, logo é uma variável aleatória, com P{T=1}=1-P{T=0}=p.

Jakob Bernoulli

Nascimento	27 de Dezembro de 1654 Basileia
Morte	16 de Agosto de 1705 (50 anos) Basileia
Nacionalidade	Suíço
Campo(s)	Matemática
Instituições	Universidade de Basileia
Alma mater	Universidade de Basileia
Tese	1684: *Solutionem tergemini problematis arithmetici, geometrici et astronomici*
Orientador(es)	Gottfried Leibniz
Orientado(s)	Johann Bernoulli, Jakob Hermann, Nicolau I Bernoulli

Uma variável aleatória assim definida – binária, com P{T=1}=1-P{T=0}= p é dita ter distribuição de Bernoulli, com parâmetro p, o que se representa por T~B(p). Sua esperança é E(T)=0×(1-p)+1×p=p e sua variância é $V(T) = E(T^2) - E^2(T) = p - p^2 = p(1 - p) = pq$

Sumarizando : Se T~B(p), então E(T) = p e V(T) =p(1-p) = pq

Repetindo-se E_p sucessiva e independentemente, n vezes, teremos T_1, T_2, ..., T_n, n variáveis aleatórias independentes, com mesma distribuição B(p). Definindo $X = \sum_{i=1}^{n} T_i$ podemos dizer que X é o número de sucessos em n repetições independentes de um mesmo E_p, portanto X~b(n, p). Concluímos assim que uma binomial de parâmetros n e p é a soma de n Bernoullis independentes de parâmetro p. Podemos pois concluir : E(X) = nE(T) = np, e V(X) = nV(T) = np(1-p), dois resultados importantes no estudo da distribuição binomial. A seguir damos provas algébricas, bem mais trabalhosas, mas tecnicamente interessantes, desses mesmos dois resultados.

Teorema 1 - Uma variável aleatória X tem distribuição b(n, p); então, sua esperança é igual a np.

Prova:

$$E(X) = \sum_{x=0}^{n} x \cdot P[X = x] = \sum_{x=0}^{n} x \cdot C_n^x \cdot p^x \cdot q^{n-x} = \sum_{x=1}^{n} x \frac{n!}{x!(n-x)!} p^x \cdot q^{n-x} = \sum_{x=1}^{n} \frac{n!}{(x-1)!(n-x)!} p^x \cdot q^{n-x}$$

$$E(X) = np \cdot \sum_{x=1}^{n} \frac{(n-1)!}{(x-1)!(n-x)!} p^{x-1} \cdot q^{n-x} = np \cdot \sum_{x=1}^{n} C_{n-1}^{x-1} \cdot p^{x-1} \cdot q^{n-x}$$

Substituindo no somatório, y=x-1 e m=n-1, teremos

$$E(X)=np\cdot\sum_{y=0}^{m}C_m^y\cdot p^m\cdot q^{m-y}=np\cdot(p+q)^m=np$$

Quod erat demonstrandum.

Teorema 2 – Se X~b(n, p), então V(X)=npq.

Prova: Por definição, $V(X) = E[(X – E(X)]^2$, logo

$$V(X) = E[X^2 – 2XE(X) + E^2(X)] = E(X^2) – 2E[XE(X)] + E^2(X) = E(X^2) – 2E^2(X) + E^2(X)$$

$$V(X) = E(X^2) – E^2(X)$$

Como, no caso da distribuição binomial, E(X) já é conhecida, passemos à determinação de $E(X^2)$.

$$E\left(X^2\right)=\sum_{x=0}^{n}x^2P\{X=x\}=\sum_{x=1}^{n}x^2\cdot C_n^x\cdot p^x\cdot q^{n-x}=\sum_{x=1}^{n}x^2\frac{n!}{x!(n-x)!}\cdot p^x\cdot q^{n-x}=np\sum_{x=1}^{n}x\frac{(n-1)!}{(x-1)!(n-x)!}p^{x-1}q^{n-x}$$

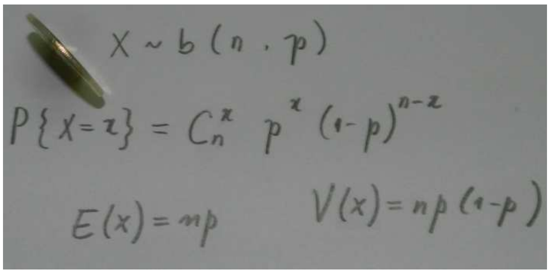
Substituindo na expressão acima, y=x-1 e m=n-1 e fazendo Y~b(m, p) temos.

$$E\left(X^2\right)=np\sum_{y=0}^{m}(y+1)\frac{m!}{y!(m-y)!}\cdot p^yq^{m-y}=np\left[\sum_{y=0}^{m}y\cdot P\{Y=y\}+\sum_{y=0}^{m}P\{Y=y\}\right]=np\left[mp+1\right]=np\left[(n-1)p+1\right]=(np)^2-np^2+np$$

$$E\left(X^2\right)=(np)^2+np\cdot(1-p)=(np)^2+npq$$

$$V(X)=E\left(X^2\right)-E^2(X)=(np)^2+npq-(np)^2=npq$$

Q.E.D.



Teorema 3 – Se X~b(n, p), então P[X=x] atinge o valor máximo para $x=x_{max}$ igual ao valor inteiro mais próximo de

Prova : Para X~b(n, p), o $P\{X=0\}=q^n$, e cresce à medida em que x cresce, enquanto P{X=x+1}/P{X=x}>1. Usaremos esta abordagem para determinar x_{max}, o valor de x em {0, 1, 2, ... , n} para o qual P{X=x} assume seu valor máximo é int(np-q)+1.

$$\frac{P\{X=x+1\}}{P\{X=x\}}=\frac{C_n^{x+1}}{C_n^x}\frac{p^{x+1}\cdot q^{n-x-1}}{p^x\cdot q^{n-x}}=\frac{\frac{n!}{(x+1)!\cdot(n-x-1)!}}{\frac{n!}{x!\cdot(n-x)!}}\frac{p}{q}=\frac{x!}{(x+1)!}\frac{(n-x)!}{(n-x-1)!}\frac{p}{q}=\frac{n-x}{x+1}\frac{p}{q}$$

Então

$$\frac{n-x}{x+1}\cdot\frac{p}{q}>1\Rightarrow np-q>x$$

Portanto, P{X=x} cresce com x, enquanto x<np-q, logo x_{max}=int(np-q)+1. Note que, se np-q for inteiro, então P{X=x} será máximo tanto para x=np-q como para x=np-q+1.

Q.E.D.

Exemplo 1.4.1 – O Brasil tem cerca de 140 milhões de eleitores. Numa grande eleição nacional, o candidato A tem uma fração p das intenções já definidas de votos, e o candidato B tem 1-p. Para se estimar p, a fração de eleitores de A, planeja-se uma pesquisa amostral, com uma amostra sorteada aleatoriamente, de mil eleitores. A fração de eleitores de A na amostra será considerada a estimativa da fração de eleitores de A no eleitorado todo. À primeira vista, parece uma amostra pequena demais para tão grande tarefa; afinal, são 140 milhões de elitores. Vejamos.

Suponha p=0,55. Qual a probabilidade de se cometer um erro inferior a, digamos, 3 pontos percentuais. Mais formalmente, seja X o número de respostas favoráveia a A. Como, a cada eleitor sorteado (chances iguais para todos os eleitores no universo), a probabilidade de ser sorteado um eleitor de A é p, sabemos que, a cada sorteio, o número de eleitores de A (0 ou 1) tem uma distribuição B(p), e os n resultados são completamente independentes entre si, concluimos que X~b(n, p) (Nota : se você estiver se sentindo desconfortável com o fato de que a composição do universo vai se alterando à medida que eleitores forem sendo sorteados, fazendo com que a fração de eleitores de A vá se alterando, suponha que o sorteio se dá com reposição, isto é, cada eleitor sorteado é restituído ao cadastro, podendo inclusive vir a ser sorteado novamente. Na prática, com uma amostra tão menor que o universo a perturbação daquela neste é desprezível, e tal preocupação é exagerada e desnecessária). Pois bem, queremos então determinar $\mathrm{P}\{0{,}52<\hat{\mathrm{p}}<0{,}58\}$, onde $\hat{\mathrm{p}}=\frac{x}{n}$ é a frando amostral de eleitores favoráveis a A. Ora,

$$\mathrm{P}\{0{,}52<\hat{\mathrm{p}}<0{,}58\}=\mathrm{P}\{520<X<580\}=\sum_{x=520}^{580}C_{1000}^{x}\times 0{,}55^{x}\times 0{,}45^{1000-x}$$

O cálculo exige computador ou uma boa calculadora científica. Usando o Excel, obtive

$$\mathrm{P}\{0{,}52<\hat{\mathrm{p}}<0{,}58\}=0{,}9475$$

Como se vê, a probabilidade de se acertar próximo do valor verdadeiro é grande, e não apenas para p=0,55, como você pode verificar, agora que já construiu um programa para fazer os cálculos.

Exemplo 1.4.2 – Uma grande montadora automobilística produz – e vende – cem mil unidades de um dado modelo, por mês. Nos três primeiros meses do ano, 38 veículos voltaram ao revendedor por apresentar um tipo de defeito considerado grave, desses que expõem o usuário a riscos de acidente sério, e a montadora a danos de reputação e legais. Ações são tomadas com o objetivo de, gradualmente, corrigir a fonte de defeitos. Uma comissão especial é formada dentro do poderoso Departamento de Melhoria Contínua da Qualidade e Produtividade da empresa. No mês seguinte, são registrados apenas 8 casos, e a empresa celebra a redução da taxa de ocorrência, como indicativo do progresso dos trabalhos da comissão. Ainda sem todo o ferramental necessário para uma abordagem formal completa da questão, vamos mostrar, apelando em parte para a intuição, mas já procurando lançar as bases para ideias que estão por vir, que a celebração é precipitada; os dados não só não dão suporte à tese de que melhoria tenha sido alcançada, como também não comprova que não tenha havido, de fato, deterioração.

A ocorrência de 38 casos em 300 mil unidades produzidas, sugere que a montadora produz itens defeituosos, aleatoriamente, a uma taxa p em torno de 0,0001267 (em situações como esta é conveniente referir a p com milhonésimos, ou partes por milhão ; assim: 126,7ppm). Contudo, valores vizinhos deste, abaixo ou acima, podem também ser probabilisticamente compatíveis com o resultado observado (38 em 300 mil).
É importante aqui é a percepção de que, como unidade de produção, a montadora é também um sistema que produz, aleatóriamente, de forma não intencional, certamente indesejada, e por causas desconhecidas, itens defeituosos segundo um processo binário com probabilidade p de sucesso (novamente o significado invertido do termo): a cada unidade produzida realiza-se um experimento aleatório binário; probabilidade p de unidade defeituosa e (1-p) de unidade conforme. Como em 300 mil repetições do experimento, produziu-se 38 itens defeituosos, a taxa 126,7ppm é uma estimativa natural para o verdadeiro p subjecente ao processo produtivo da montadora, o qual vamos admitir aqui – sem adequada fundamentação, mas para simplicidade conceitual – constante.

Mas o resultado observado não permite uma avaliação exata do p subjacente; existe uma faixa de possibilidades para este parâmetro. A largura desta região e de incerteza mede a incerteza estatística associada à nossa estimativa de p a partir dos resultados observados. Obviamente, valores muito distantes da frequencia observada de 126,7ppm, como 0,001 – ou 10.000ppm – são completamente incompatíveis, de um ponto de vista probabilistico, com o resultado observado e não merecem ser considerados. Por outro lado, valores próximos de 126,7ppm, (80, 100, 150 ppm ?) podem ser compatíveis com o resultado observado, não podendo, portanto, serem descartados como provável valor verdadeiro de p.

Um critério para avaliar o grau de compatibilidade de determinado candidato a p – digamos p_0 – é medir a probabilidade de se obter 38 casos positivos em 300 mil tentativas, sob a hipótese $p=p_0$, ou seja, computar $C^{38}_{300.000} \times p_0^{38} \times (1-p_0)^{299.962}$, e verificar se o valor é tal que não torna a hipótese completamente incompatível com o resultado observado de 38 casos em 300 mil tentativas. A propósito a probabilidade de se obter o resultado observado sob a hipótese de que p=38/300.000=0,0001267, é

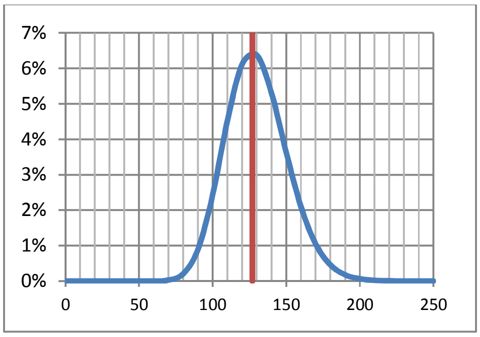
$$C^{38}_{300.000} \times 0{,}0001267^{38} \times (1-0{,}0001267)^{299.962} = 0{,}064579 \approx 6{,}46\%$$

Existe, portanto um bom nível de compatibilidade entre a hipótese p=0,0001267 e o resultado, digamos, X=38, o que já era esperado, uma vez que esta foi a taxa observada. Já a hipótese p=0,01 é, naturalmente, incompatível com o resultado observado. Se a taxa p de defeituosos produzidos pela montadora fosse 0,01, ou 1%, o resultado observado teria sido algo próximo de 3 mil, um desastre que provavelmente traria enormes danos de reputação e dificuldade econômica séria para montadora. Em todo o caso, qual a probabilidade, sob a hipótese p=0,01, de X=38 defeituosos em 300 mil unidades produzidas? Vejamos:

$$C^{38}_{300.000} \times 0{,}10^{38} \times (1-0{,}01)^{299.962} = 1{,}36556{\times}10^{-1222}$$

Realmente, um valor tão abismalmente pequeno revela a absoluta incompatibilidade entre a hipótese (p=0,01) e o resultado observado. Como não podemos descartar o resultado – fato empírico observado – descartamos, com absoluta segurança, a hipótese.

Mesmo uma hipótese aparentemente menos antagônica com os dados, p=0,001, não passa pelo teste de compatibilidade: a probabilidade de 38 carros defeituosos em 300 mil carros produzidas, sob a hipótese de que, a cada carro produzido, a probabilidade de ocorrência do defeito é 0,001, é de $1{,}1859{\times}10^{-81}$, revelando a total e absoluta incompatibilidade da hipótese com os fatos observados.

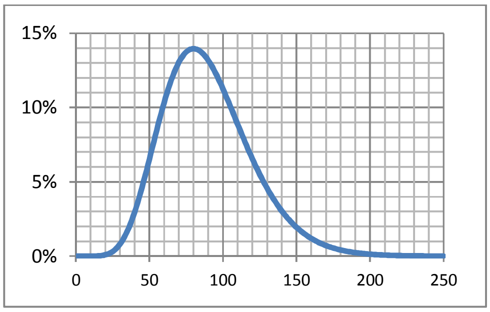


Vamos usar este critério – o da probabilidade do resultado observado sob cada hipótese alternativa sobre p – para testar a compatibilidade de valores de p numa vizinhança do candidato natural: X/n, ou 38/300000.

O resultado desta investigação é de uma elegância extraordinária, como ilustra a figura ao lado, o gráfico de P{X=38 em 300000}, em função de p (em ppm), na região em torno de p = 126,7ppm.

O máximo, de cerca de 6,4%, ocorre para p=0,0001267=126,7ppm, mas valores de p desde 80 até 180ppm são razoavelmente compatíveis com o resultado observado. Conclusão: o resultado observado nos primeiros 3 meses do ano permitem estimar o intervalo de 80 a 180ppm como faixa de valores prováveis para p, a taxa de falha implícita no processo produtivo da montadora. Valores centrais desse intervalo são mais prováveis.

Até este ponto, podemos dizer que ao longo dos três primeiros meses do ano, por causas desconhecidas, a montadora operou com uma taxa de falha (a probabilidade, suposta constante, de ocorrência do defeito em questão a cada unidade produzida) situada no intervalo de 80 a 180 ppm.



Uma comissão formada em regime de urgência, vem trabalhando no sentido de localizar e eliminar (ou, pelo menos, atenuar) as causas dos defeitos. O resultado do quarto mês (`apenas` 8 casos em cem mil unidades produzidas, ou 80ppm) levou a direção da montadora a concluir que melhoras substanciais já haviam sido alcançadas após um mês de trabalhos da comissão, provocando celebrações entusiasmadas. A figura co lado, contudo, recomenda cautela. O gráfico mostra P{X=8 em 100000}, em função de p: $C^{8}_{100.000} \times p^{38} \times (1-p)^{99.992}$. Vemos então que, agora, valores de 40 a 150 ppm são compatíveis com o resultado.

Comparando as duas curvas, de antes e depois do início dos trabalhos da comissão, vemos grande sobreposição das duas faixas de valores compatíveis para p. Nelas encontramos argumentos em favor da hipótese de que teria havido melhora, de que nada teria mudado, e de que teria havido piora. Uma possibilidade – perfeitamente compatível com os dados observados – é, por exemplo : p_{antes}=90ppm e p_{depois}=120ppm. Também pode ser: p_{antes}=150ppm e p_{depois}=60ppm. Os dados observados, embora sugiram melhora, não a comprovam estatisticamente.

Voltaremos mais tarde a essa classe de questão, com ferramental formal mais poderoso e completo. Por enquanto recomendo reflexões cuidadosas sobre as ideias envolvidas.

1.4.2 – O Modelo Geométrico

Outro modelo muito simples e interessante, derivado dos experimentos aleatórios binários, é o Modelo Geométrico. Considere um experimento bernoulliano E_p, e seja um experimento composto que consiste na repetição de E_p até a obtenção do primeiro Sucesso. O espaço amostral deste experimento é infinito enumerável : Ω={S, FS, FFS, FFFS, FFFFS, FFFFFS, FFFFFFA, …}. Seja a álgebra máxima $\mathfrak{F}$ – gerada pela partição de W em átomos unitários – e a função X : Ω→R, definida como o número total de fracassos antes do primeiro sucesso. Assim, X{S}=0, X(FS)=1, X{FFS}=2, e assim por diante. A função de probabilidades naturalmente associada a este experimento é definida a partir de P{S}=p, P{FS}=qp, P{FFS}=q^2p, … . A tríade (Ω, $\mathfrak{F}$, P) é um espaço de probabilidades ; a função X, definida acima é uma variável aleatória neste espaço, com P{X=x} = q^xp.

Exemplo 1.4.2 - Um experimento consiste em arremessar uma moeda sucessivamente, até a obtenção da primeira cara. Seja X o número de coroas obtidas no experimento, isto é, antes da primeira cara. Então,

a – P{X=x} = (½)^x

b – P{X=5} = (½)^5 = 0,03125

c – P{X<10}

Exercício 1.4.1 – Um experimento binário E_p será repetido até a obtenção do primeiro sucesso. Seja X o número de fracassos alcançados, antes do sucesso final. Seja p=0,1.

a – Calcule P{X>10}

b – Calcule a probabilidade condicional de pelo menos mais 10 tentativas sem sucesso, dado que já foram feitas 20 tentativas fracassadas.

(continua)

Exemplo 1.4.3 – Uma operação repetitiva em uma indústria metalúrgica envolve certo risco de acidente grave, com mutilação de mãos e braços. Especialistas em segurança do trabalho, contudo, avaliam o risco de acidente numa repetição da operação como de um milhonésimo, e a operação é então considerada segura o suficiente para não demandar maiores preocupações. Na linha de produção, um operário realiza a operação cerca de 100 vezes por hora. Sua jornada de trabalho é de 8 horas, e ele trabalha 220 dias por ano. Na fábrica, 120 trabalhadores realizam a mesma tarefa na linha de produção.

a) Qual a probabilidade de Severino perder a mão no próximo ano?

Em um ano ele realizará a operação 220×8×100 = 176.000 vezes. Seja X o número de repetições seguras realizadas antes do primeiro acidente. Queremos saber P{X<176.000} :

$$P\{X < 176.000\} = 1 - P\{X \geq 176.000\} = 1 - \sum_{x=176.000}^{\infty} P\{X = x\} = 1 - \sum_{x=176.000}^{\infty} p \times (1-p)^x$$

$$= 1 - p \sum_{x=176.000}^{\infty} (1-p)^x = 1 - p\frac{(1-p)^{176.000}}{1-(1-p)} = 1 - (1-p)^{176.000}$$

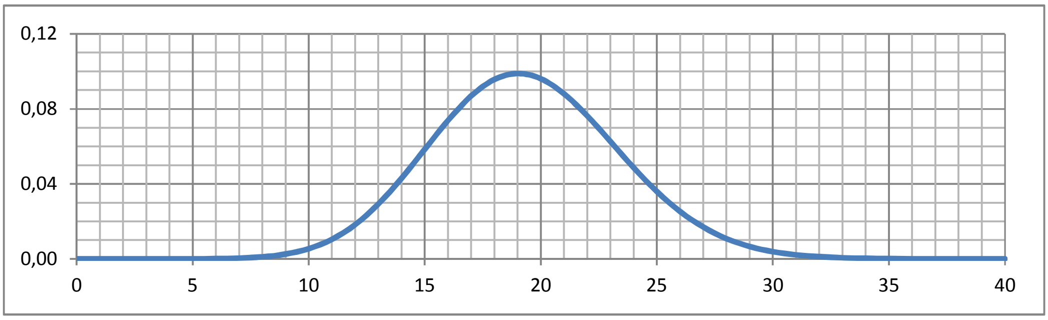
$$= 1 - (1 - 10^{-6})^{176.000} = 1 - 0{,}999999^{176.000} = 1 - 0{,}8386$$

$$= 0{,}1614$$

Severino tem 83,86% de chances de não perder uma mão no próximo ano, mas a perderá com probabilidade 0,1614. A probabilidade dele passar incólume pelos seus primeiros 5 anos na atividade é igual a $0{,}8386^5$=0,4148.

b) Quantos trabalhadores perderão a mão no próximo ano?

A probabilidade de um trabalhador perder a mão no ano é 0,1614. São 120 trabalhadores, portanto o número deles que perderá a mão antes do final do primeiro ano de trabalho é uma variável alatória, digamos, com distribuição binomial :Y~b(120, 0,1614). A distribuição de probabilidades de Y é dada por $P\{Y = y\} = C_{120}^{y} \times 0{,}1614^{y} \times 0{,}8386^{120-y}$, e E(Y)=19,4. A distribuição de probabilidades está representada na figura abaixo.



Como se pode ver ocorrerão, num ano, no minimo 10 acidentes. Com probabilidade 0,9542 o número de ocorrências no ano ficará entre 12 e 27, e se, por um lado, a chances de 30 acidentes ou mais no ano são praticamente nulas, o mesmo ocorre com a probabilidade de 10 ou menos. O procedimentos não são seguros.

1.4.3 – O Modelo de Poisson

Como espalhar pontos casualmente ao longo de uma linha?

Analisando os acidentes ocorridos em um ano ao longo de uma rodovia, buscando detectar os locais de mais alto risco, engenheiros de segurança de trânsito têm que decidir se a concentração relativamente elevada de acidentes num dado trecho é indicação segura de que aquele trecho é mais perigoso, e portanto demanda ações especiais, ou se a relativa concentração é mero produto de flutuações aleatórias naturais.

A representação gráfica do local exato de ocorrência de cada acidente ao longo da rodovia, normalmente revela trechos relativamente longos e vazios de pontos, sucedidos por outros onde estes pontos se concentram em número elevado. Os trechos de alta concentração imediatamente chamam a atenção, como os locais de alto risco. O tratamento especial desses trechos, a partir daí, pode ser um erro primário; pode-se estar focando a atenção – e as ações e os recursos econômicos e técnicos – num trecho da rodovia que não tem absolulamente nada de especial que o difira dos demais como de risco mais elevado. Mais ou menos da mesma forma que a pessoa que fez a maior pontuação numa loteria de números não tem necessariamente habilidades especiais para a escolha dos números "certos".

Siméon Denis Poisson

Nascimento	21 de Junho de 1781 Pithiviers
Morte	25 de Abril de 1840 (58 anos) Paris
Nacionalidade	Francês
Campo(s)	Matemática e física
Instituições	École Polytechnique, Bureau des Longitudes, Escola Militar Especial de Saint-Cyr
Alma mater	École Polytechnique
Tese	1800:
Orientador(es)	Joseph-Louis Lagrange e Pierre Simon Laplace
Orientado(s)	Michel Chasles, Johann Peter Gustav Lejeune Dirichlet, Joseph Liouville
Conhecido(a) por	Equação de Poisson, Parênteses de Poisson, Coeficiente de Poisson, Distribuição de Poisson, Regressão de Poisson

Antes de concluir que certos trechos são mais propensos a acidentes que outros, os engenheiros normalmente se perguntam se a distribuição longitudinal observada dos acidentes na rodovia não poderia ter sido o produto de uma distribuição puramente casual dos mesmos. Para responder a esta pergunta, devemos antes entender o comportamento de uma "distribuição puramente casual" de pontos ao longo de uma linha. Para isto vamos pensar na distribuição – por hipótese, puramente casual – de acidentes ao longo de uma extensão homogênea rodovia. A casualidade implica em algumas propriedades:

- ✓ Cada ponto da rodovia tem a mesma chance que qualquer outro de receber ou não um acidente;
- ✓ A probabilidade de mais que uma ocorrência no exato mesmo ponto, é nula;

- ✓ A probabilidade da ocorrência de um acidente em um ponto é independente do ocorrido em quaisquer outros pontos, vizinhos ou distantes.

Obedecendo a essas três condições básicas, vamos distribuir acidentes ao longo da rodovia, na concentração média de λ acidentes por quilômetro. Para focar as ideias, vamos admitir λ=5. Esta concentração média resultante, de 5/km, corresponde, numa distribuição casual, a cada trecho de 100 metros recebendo, com probabilidade 0,5, um acidente. Assim como Maxwell, lancemos mão aqui de um demônio. Com poderes atemporais, ele percorre todo o comprimento da rodovia, com sua roleta maligna, decidindo a cada 100 metros, “aqui ocorrerá um”, ou dependendo do resultado do giro da roleta – par ou impar – “aqui não ocorrerá nenhum”. Ao final dos, digamos, 100 km da rodovia, ele terá algo como a figura abaixo, onde cada segmento corresponde a 10 km. Para esta distribuição o demônio, num toque de sutil perversidade, decidiu posicionar os acidentes, sempre que fosse o caso, exatamente no centro do respectivo segmento de 100 m.

Como forma de produzir uma distribuição completamente casual de pontos sobre uma reta, o método adotado, contudo, é imperfeito. A configuração acima tem um aspecto claramente não natural, que sugere intervenção intencional, controle, propósito. O posicionamento de acidentes, em certos trechos, como pérolas num colar – nas sequências longas de 1’s, no sorteio 0 ou 1 do demônio – chama atenção como anti natural. Alem disto, a impossibilidade implicada pelo método adodado, de ocorrência de mais que um acidente num mesmo segmento de 100 metros, não corresponde à realidade de uma distribuição perfeitamente casual. Vamos tentar corrigir este defeito, escolhendo, sempre que for o caso, a posição do acidente, aleatoriamente dentro do segmento correspondente de 100 metros; não mais exatamente e sempre no centro do segmento. A figura abaixo lado mostra o resultado.

O efeito colar de pérolas foi, pelo menos parcialmente, removido, mas ainda sem corrigir o jeitão manipulado da distribuição. Há algo de errado nela; uma certa regularidade não natural, incompatível com a ausência absoluta de intenção, propósito, plano pré estabelecido. Pode-se, agora, ter dois pontos dentro de um mesmo segmento de 100 metros, mas nunca três.

O defeito da distribuição vem da divisão excessivamente grosseira da linha em segmentos de δ=0,1km, longos demais. Podemos corrigir este problema fazendo o demônio trabalhar mais. Em vez de um sorteio para cada 100 metros, com probabilidade de sucesso (eis um uso invertido da palavra) 0,50, por que não um sorteio a cada dez metros, com p=0,05. Teremos a mesma concentração média de 5/km, agora com maior semelhança de naturalidade.

Já nem tempo nem espaço é problema para o demônio, por que não fazê-lo tomar decisão a cada um metro (0,001 km), com p=0,005?

A distribuição acima, obtida efetuando-se um sorteio binário a cada um metro (c=0,001km) da rodovia, com p=0,005 (para garantir concentração média λ=p/δ=0,005/0,001km=5/km) já não apresenta os evidentes defeitos das anteriores, embora a perfeição exigiria levar δ ao infinitésimo, decrescendo p proporcionalmente, de forma a manter λ=p/δ constante e igual ao valor desejado; no caso do exemplo, 5.

A distribuição perfeitamente casual apresentará trechos vazios e outros com concentração muito elevada. Teremos que responder a perguntas como: qual a probabilidade de que, depois de feita a distribuição, um dado trecho extenso, de comprimento c, não receba nenhum acidente?

A resposta é surpreendentemente simples. Seja X o número de acidentes "alocados" nesse trecho durante a distribuição casual dos mesmos. Como a cada trecho elementar de comprimento δ é feito um sorteio com probabilidade de sucesso p=$\lambda\delta$, a probabilidade de que nenhum acidente ocorra num trecho de extensão c>>δ é igual à probabilidade de que o sorteio fatídico tenha dado negativo (negativo, aqui, é bom!) todas as c/δ vezes em que foi realizado ao longo do referido trecho, ou simplesmente $\mathrm{P}\{\mathrm{X}=0\}=\left(1-\frac{\lambda}{\delta}\right)^{\frac{c}{\delta}}$. Aqui assumimos, sem perda de generalidade, dado que δ é muito pequeno, tendendo ao infinitesimal, que c/δ é inteiro.

Podemos agora promover uma distribuição casual perfeita, fazendo $\delta\to 0$. Assim, empregando um resultado bem conhecido de limite*,

$$\mathrm{P}\{\mathrm{X}=0\}=\lim_{\delta\to 0}(1-\lambda\delta)^{\frac{c}{\delta}}=\mathrm{e}^{-\lambda c}$$

No caso do exemplo, com a concentração média de 5 acidentes por km por ano, com distribuição perfeitamente casual, a probabilidade de um segmento pré definido de um quilômetro de pista (digamos, por exemplo, aquele que começa na placa de 13km), não testemunhar, ao longo de todo o ano, nenhum acidente, é igual a $e^{-5/km\times 1km}$ = e^{-5} = 0,0067. É uma probabilidade baixa, mas, considerando-se os 100 segmentos de um quilômetro, a probabilidade de que pelo menos um deles passe o ano "em branco", já sabemos calcular, é igual a $[1-(1-0{,}0067)^{100}]$ = 0,4914, bastante considerável; com em média uma ocorrência a cada dois anos.

Definimos como X o número de ocorrências num dado trecho de comprimento c, e calculamos P{X=0}. Devemos agora calcular P{X=x}, para x=1, 2, 3, … Primeiro, P{X=1}. Ora, o número de sucessos em c/δ sorteios com probabilidade de sucesso igual a λ/δ em cada sorteio, segue uma distribuição binomial, com parâmetros (n, p), com n=c/δ e p=$\lambda\delta$, logo

$$P\{X=1\}=C_n^1 p(1-p)^{n-1}=\frac{c}{\delta}\times(\lambda\delta)\times(1-\lambda\delta)^{\frac{c}{\delta}-1}$$

Passando ao limite,

$$P\{X=1\}=\lim_{\delta\to 0}\left(\frac{c}{\delta}\times(\lambda\delta)\times(1-\lambda\delta)^{\frac{c}{\delta}-1}\right)=\lambda c\lim_{\delta\to 0}(1-\lambda\delta)^{\frac{c}{\delta}-1}$$

$$P\{X=1\}=\lambda c\lim_{\delta\to 0}\frac{(1-\lambda\delta)^{\frac{c}{\delta}}}{(1-\lambda\delta)}=\lambda c\frac{\lim_{\delta\to 0}(1-\lambda\delta)^{\frac{c}{\delta}}}{\lim_{\delta\to 0}(1-\lambda\delta)}=\lambda c\lim_{\delta\to 0}(1-\lambda\delta)^{\frac{c}{\delta}}=\lambda c\cdot e^{-c\lambda}$$

De forma análoga, podemos determinar

$$P\{X=2\}=\frac{(\lambda c)^2}{2}\cdot e^{-c\lambda}$$

$$P\{X=3\}=\frac{(\lambda c)^3}{3!}\cdot e^{-c\lambda}$$

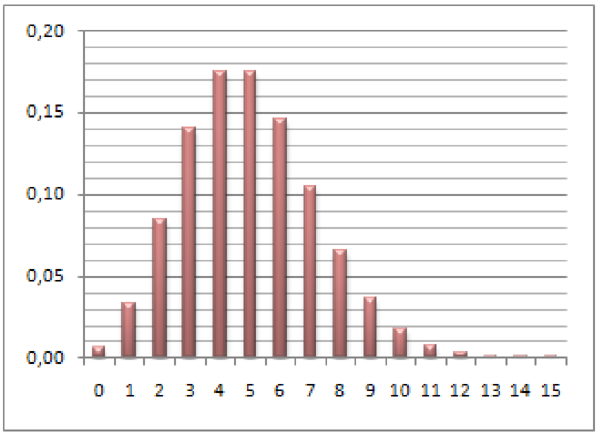
$$P\{X=4\}=\frac{(\lambda c)^4}{4!}\cdot e^{-c\lambda}$$

Generalizando,

$$P\{X=x\}=\frac{(\lambda c)^x}{x!}\cdot e^{-c\lambda}$$

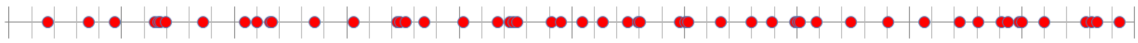
O desenvolvimento original do conceito apresentado se deve a S. D. Poisson e, em sua homenagem, uma variável aleatória com esta distribuição de probabilidade é dita ter distribuição de Poisson, com parâmetro cλ. Na forma padrão, para comprimento unitário c=1, temos $P\{X=x\}=\frac{\lambda^x}{x!}\cdot e^{-\lambda}$, e representamos por X~P(λ). Adiante veremos suas propriedades básicas, mas antes, vamos calcular P{X=x} para X~P(5).

x	P{X=x}	x	P{X=x}
0	0,0067	9	0,0363
1	0,0337	10	0,0181
2	0,0842	11	0,0082
3	0,1404	12	0,0034
4	0,1755	13	0,0013
5	0,1755	14	0,0005
6	0,1462	15	0,0002
7	0,1044	16	0,0000
8	0,0653		



Como se pode ver, numa rodovia de 100km onde se verificou uma média de 5 ocorrências por quilômetro no ano anterior, dividida em 100 trechos de 1km, pode-se ter trechos com apenas 1 acidente (em torno de 3 dos cem trechos), e outros com 10 acidentes (cerca de dois dos cem), mesmo que os acidentes tenham sido distribuidos de maneira completamente casual.

Da mesma forma como modelamos distribuições casuais de pontos longitudinalmente sobre uma linha, podemos também modelar ocorrências ao longo do tempo. Uma situação clássica é a ocorrência, ao longo do tempo, de fissões atômicas ocorrendo numa amostra de material radioativo. Um registro gráfico de ocorrências ao longo de um eixo do tempo, pode resultar em uma figura como a abaixo.



Novamente, sob a aparência de um regime caótico e irregular está, na verdade, um sistema governado por lei probabilística simples, estável e, sobretudo, completamente compreendida. Seus resultados são previsíveis em termos probabilísticos no curto prazo e, de forma progressivamente determinística, no longo prazo. (explicar melhor). Sabendo o número de átomos do elemento radioativo na amostra, podemos deduzir, a partir da análise dos dados acima, seu fator de decaimento e tempo de meia vida.

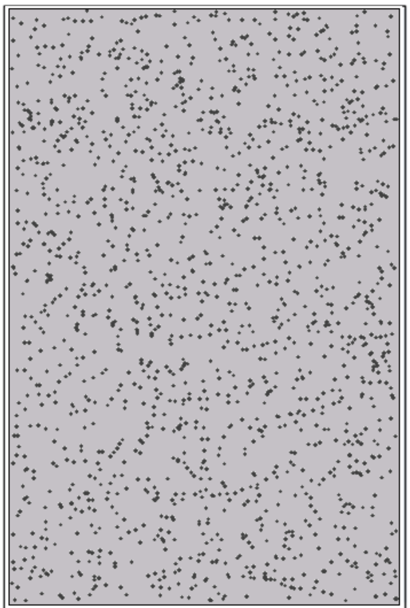
Exemplo 1.4.1 – Uma amostra contendo 1 mol de Carbono, extraída de um pedaço de madeira fossilizado é analisada num contador de emissões radioativas. Sabe-se que a concentração de ^{14}C em carbono atmosférico natural é de 1ppt (uma parte por trilhão), e que o tempo de meia vida deste isótopo radioativo é de 5730 anos. Quer-se estimar a idade do objeto fóssil do qual a amostras de C foi extraída. A amostra é mantida no contador por 30 horas, com registro do instante preciso (hh :mm :ss) de cada emissão. Ao final,

(continua)

Distribuições casuais no plano

Vimos como se distribui pontos, casualmente, ao longo de uma linha. As ideias podem ser generalizadas para duas ou mais dimensões.

Gotas de chuva no asfalto
Campos do Jordão
07 nov 2009

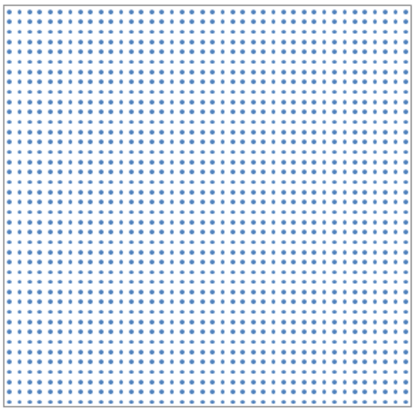


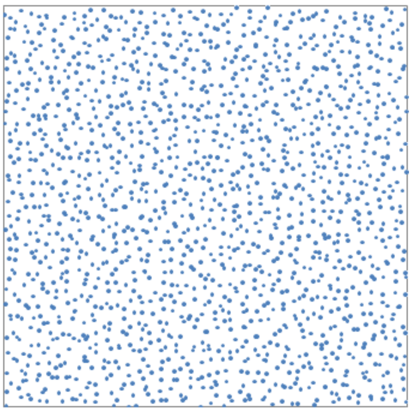
1600 pontos distribuídos ao acaso

A distribuição casual – não intencional, totalmente desprovida de propósitos ou preferências – cria padrões curiosos que, frequentemente, enganam o observador.

Nas duas figuras ao lado tem-se, à esquerda, o desenho feito no asfalto pelas primeiras gotas de uma chuva fina, mansa e súbita, num final de tarde em Campos do Jordão e, à direita, a distribuição, por sorteio aleatório, de 1600 pontos sobre um quadrado. Observe a semelhança.

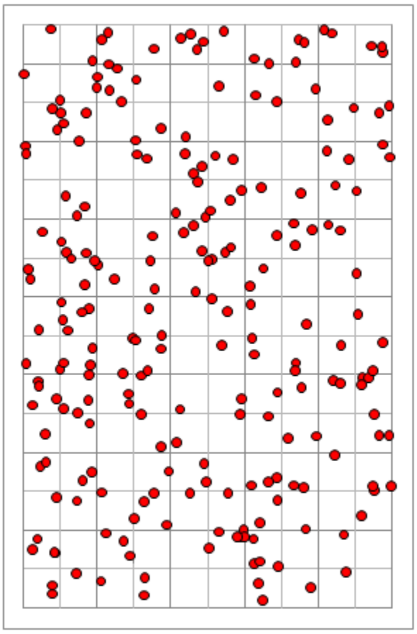
Para as duas figuras ao lado, o quadrado foi inicialmente dividido numa grade de 40x40, em 1600 células. Na da esquerda, os 1600 pontos são distribuídos de maneira totalmente intensional, no centro de cada uma das células da grade, numa estrutura cristalina absolutamente previsível, regular. Na da direita, um ponto foi colocado, não no centro, mas aleatoriamente, dentro de cada um dessas 1600 células. Decidir se uma distribuição linear, plana ou espacial de pontos segue uma distribuição perfeitamente casual, ou apresenta algum tipo de regularidade, de previsibilidade, de manifestação implicita de propósito ou





intensão, está no centro de muito processo de investigação científica. Tratamos aqui da questão de decidir se uma dada distribuição de pontos é casual, ou não.

Exemplo 1.4.2 – Um experimento científico busca estudar o efeito de diferentes fórmulas no estímulo ou inibição da formação de colônias de certa bactéria, em um tipo de cultura. Uma placa de 20cmX30cm é dividida, na forma de grade, em 150 células quadradas de 4cm^2 cada. A placa é coberta com uma película da cultura, e cada célula é impregnada de uma formulação diferente dos agentes inibidores/estimulantes. A placa é mantida por tempo pré deterrminado em atmosfera contaminante. Ao final, a placa é submetida a um banho quimico que revela os focos de colônias de bactérias formados. Os pesquisadores procurarão localizar as células mais contaminadas e associarão o tratamento correspondente a possível efeito estimulante à formação de culturas, e as células menos contaminadas a provável efeito inibidor das formulações correspondentes. A tabela ao lado dá o número de células por número de focos apresentados. Foram contadas, no total, 223 colônias espalhadas pela placa. Das 150 células, 35 não apresentaram nenhum foco, enquanto, no outro extremo, uma célula apresentaram 6 colônias, e outras duas, cinco colônias cada. Analise os dados e verifique se existe evidência em favor da existência de células mais hospitaleiras e células mais inóspitas às bactérias. Ou não.



Número de focos	Número de células
0	35
1	48
2	40
3	17
4	7
5	2
6	1