Estatística Aplicada à Medicina – 2020.1

Tópico 3 – Probabilidades (1) – conceitos iniciais e exemplos

1 – Introdução: por que estudar *Teoria das Probabilidades*?

- para dar a base matemática ao estudo da *Inferência Estatística*
- para servir de *apoio à decisão* (quantificação da incerteza)

1.1. Probabilidade como base da "Inferência Estatística"

- não é possível tirar conclusões, a partir de um único caso, já que a *variabilidade* entre as pessoas é muito grande; por outro lado, não é possível examinar toda a população
- grande parte da pesquisa em Medicina é feita utilizando-se *amostras*
- qualquer resultado obtido numa amostra é *incerto*; o que podemos fazer é buscar resultados *altamente prováveis*.
- como tirar conclusões a partir de uma amostra ? Usando a teoria de *Inferência Estatística*, que se baseia na *Teoria das Probabilidades.*

1.2. Probabilidade como apoio à decisão

O resultado de quase qualquer ação em Medicina (teste, exame, intervenção, procedimento, etc.) é *incerto*; nunca podemos ter certeza do resultado. O que podemos é analisar as probabilidades e escolher as ações que tem mais probabilidade de terem bons resultados.

Ex. de incerteza:

> Um estudo registrou que a probabilidade da mamografia como teste de triagem para detecção de câncer de mama dar positivo, para pacientes sabidamente com câncer, é de 0,85, enquanto que a probabilidade de dar negativo para pacientes sabidamente sem câncer é de 0,80...

2. Probabilidades – Definições e conceitos iniciais

A Teoria das Probabilidades é a parte da Matemática que estuda os *experimentos aleatórios*.

2.1. Experimento aleatório

- Experimento que, se repetido várias vezes sob as mesmas condições, dá a cada vez resultados definidos mas imprevisíveis

Exemplos de experimentos: lançar uma moeda e observar se dá cara ou coroa; lançar um dado não viciado e observar o resultado da face superior, realizar uma cirurgia com uma determinada técnica e observar se o paciente fica livre dos sintomas após um mês.

Exemplos nas ciências exatas:

- erro de medição (erros aleatórios e sistemáticos)

Exemplos nas ciências da vida:

- resultados do cruzamento de ervilhas de Mendel
- propriedades de um exame laboratorial (sens., espec., VPP, VPN)
- desfecho de uma cirurgia com determinada técnica
- inferência (a aleatoriedade é consequência do sorteio da amostra)

2.2. Porque estudar experimentos aleatórios?

- Porque existe uma regularidade ou padrão, que só se manifesta quando o experimento é repetido um grande número de vezes (“regularidade estatística”).

- A probabilidade nos permite fazer previsões aproximadas sobre o resultado de um grande número de repetições de um experimento; mas não diz muito sobre o que acontece com cada repetição individual, ou com uma pequena quantidade de repetições.

2.3. Conceito inicial: o que é probabilidade?

- A *probabilidade* é um número entre 0 e 1 que mede quão incerta é a ocorrência de um evento.

  $P = 1$ $\rightarrow$ a *ocorrência* é certa

  $P = 0$ $\rightarrow$ a *não-ocorrência* é certa.

  $P = 0.5$ $\rightarrow$ a *ocorrência* e a *não-ocorrência* têm a mesma probabilidade

- *Definição clássica*:

$$P = \frac{\text{número de casos favoráveis}}{\text{número de casos equiprováveis}}$$

Exercícios

1) Determine as probabilidades de se obter:
   a) no lançamento de um dado:
   - a face 6
   - uma face com número par de pontos

   b) na retirada aleatória (aleatório significa "ao acaso") de uma carta de um baralho completo:
   - um rei
   - uma figura qualquer
   - uma carta qualquer do naipe de paus

3. Cálculo de probabilidades: cálculo por enumeração

3.1. Os problemas de probabilidades: o que exigem

- calcular a probabilidade de eventos complexos, criados pela combinação de vários eventos elementares de probabilidade conhecida.

Ex: No lançamento de uma moeda, P(cara) = 1/2
No lançamento de duas moedas, P(2 caras) = ?

No lançamento de um dado, P(6) = 1/6
No lançamento de dois dados, P(soma = 7) =?

Na retirada de uma carta, P(K) = 1/13
Na retirada de quatro cartas, P(4 reis) = ?

3.2. Maneiras de resolver problemas de probabilidades:

- por enumeração dos casos possíveis (se equiprováveis)
- pelo cálculo do número de casos possíveis, usando análise combinatória
- por operações sobre conjuntos

Exercícios (cont.)

2) Uma mulher tem duas crianças. Qual a probabilidade de que:
   a) (exatamente) uma delas seja menina?
   b) sejam duas meninas?
   c) as duas crianças sejam do mesmo sexo?

3) Três moedas são lançadas simultaneamente.
   a) Qual é a probabilidade de que sejam obtidas 3 caras?
   b) Qual é a probabilidade de que sejam obtidas (exatamente) 2 caras?
   c) Qual é a probabilidade de que as três moedas mostrem a mesma face?

4) Dois dados comuns são lançados. Qual a probabilidade de a soma dos números mostrados ser
   a) igual a 7 ?
   b) igual ou maior que 10 ?

5) Dois dados comuns são lançados. Qual a probabilidade de os dois números mostrados serem iguais ?

6) Dez moedas são lançadas simultaneamente.
   a) Qual é a probabilidade de que sejam obtidas 10 caras?
   b) Qual é a probabilidade de as dez moedas mostrarem a mesma face?

7) Três dados são lançados simultaneamente.
   a) Qual é probabilidade de dois deles mostrem a face "6" ?
   b) Qual é a probabilidade de que todos eles mostrem a face "6" ?

Observação: Definições de "Probabilidade"

Definir o que é "probabilidade" não é fácil, apesar de o conceito ser intuitivamente simples de se entender. As definições têm mudado ao longo do tempo, refletindo o aumento da complexidade e da abrangência da teoria.

Os iniciadores do estudo de Probabilidades, Pascal e Fermat, usaram implicitamente um conceito que foi mais tarde explicitado por Laplace no primeiro livro escrito sobre o assunto (em 1812):

Probabilidade é o quociente do número de casos favoráveis sobre o número de casos igualmente possíveis.

Esta definição é útil para calcular as probabilidades básicas em alguns problemas elementares, mas tem vários defeitos. Como saber, por exemplo, se os casos são "igualmente possíveis"? Quando lançamos um par de dados e somamos os pontos mostrados obtemos um valor entre 2 e 12. Será que podemos considerar que todos os valores neste intervalo são "igualmente possíveis"?

Além disso, a definição não explica para que servem ou o que significam estes valores. Quando digo que a probabilidade de nascer um *menino* é aproximadamente igual a de nascer uma *menina*, o que isto implica?

Há outra definição que mostra mais claramente a utilidade do conceito de probabilidade, e se baseia na *freqüência relativa* de um evento:

Se após *n* repetições de um experimento (*n* suficientemente grande), se observam *h* repetições de um determinado evento, então a probabilidade do evento é *h/n*.

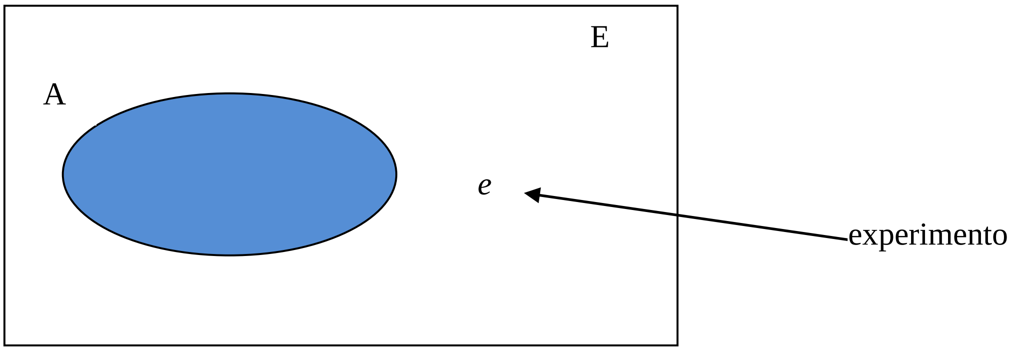
Essa probabilidade (criada no final do século XIX) é chamada de *probabilidade empírica*, e deixa clara a utilidade do número calculado: se a probabilidade de nascer um menino é aproximadamente igual a de nascer uma menina, isto quer dizer que, num grande número de partos, posso prever que o número de nascimentos de *meninas* será aproximadamente igual ao de *meninos*.

Uma terceira definição (criado no século XX) considera a probabilidade como a medida da intensidade de uma crença pessoal:

A probabilidade é um número entre 0 e 1 que expressa minha crença na ocorrência de um evento. Se digo que há 0,90 de probabilidade de meu time ganhar o campeonato, isto quer dizer que estou *quase* certo de que ele vai ganhar.

Esta definição é chamada de definição *bayesiana*, em homenagem a Bayes, que primeiro a utilizou.

Existe também uma definição de base *geométrica*, baseada numa analogia entre a área de uma superfície e a probabilidade. Se a cada realização de um experimento obtemos um ponto *e* (evento simples), podemos considerar o conjunto de pontos que formam o retângulo abaixo como E, sendo os pontos uniformemente distribuídos na superfície de E, cuja área é igual a 1. O evento A (uma partição de E) representa o evento composto de todos os eventos simples que compõem a superfície marcada. Como P(E) = 1, a probabilidade de A é igual a área da superfície A. No desenho, o evento elementar *e* não atenderia ao evento A.



A maior parte dos livros prefere empregar atualmente uma definição puramente matemática, a *definição axiomática*:

Seja um E um experimento, e S o espaço amostral a ele associado. A cada evento A associaremos um número real representado por $P(A)$ e denominado *probabilidade de A*, que satisfaça às seguintes propriedades:

1) $0 < P(A) < 1$
2) $P(S) = 1$
3) Se A e B forem eventos mutuamente excludentes,
$$P(A \cup B) = P(A) + P(B)$$
4) Se $A_1, A_2, A_3, \ldots A_n$ forem, dois a dois, eventos mutuamente excludentes, então
$$P(A_1 \cup A_2 \cup A_3 \cup \ldots \cup A_n) = P(A_1) + P(A_2) + P(A_3) + \ldots + P(A_n)$$

Resumindo as definições de probabilidades:

- Laplace
- Empírica ou frequência relativa
- Bayesiana ou subjetiva
- Geométrica
- Axiomática

4. Cálculo de probabilidades usando teoria de conjuntos

4.1. *Termos usados*

- experimento / ponto amostral / espaço amostral / evento

4.2. *Relações entre dois conjuntos: complemento, união, e interseção*

Existem quatro relações possíveis entre 2 conjuntos:

Complemento:	A' ou $\overline{A}$ (= conjunto dos pts que **não** pertencem a A)
União:	$A \cup B$ (= pts que pertencem a A, **ou** B, **ou** a ambos)
Exclusão:	A-B (= pts que pertencem a A mas **não** a B)
Interseção:	$A \cap B$, ou então, AB (= pts que pertencem a A **e** B)

Correspondência

Lógica	Álgebra de conjuntos
E	interseção
OU	união
não	complemento

4.3. *Probabilidade da união de dois conjuntos*

- Se os dois eventos são *mutuamente exclusivos* (não podem ocorrer juntos): Para calcular a probabilidade de que ocorra *ou o evento A ou o evento B*, somamos a probabilidade de A com a de B.

$$P(A \cup B) = P(A) + P(B)$$

Exemplo 1 : No lançamento de 2 dados,

$$P(\text{soma} > 9) = P(10) + P(11) + P(12)$$

Exemplo 2 : Na retirada de uma carta de um baralho,

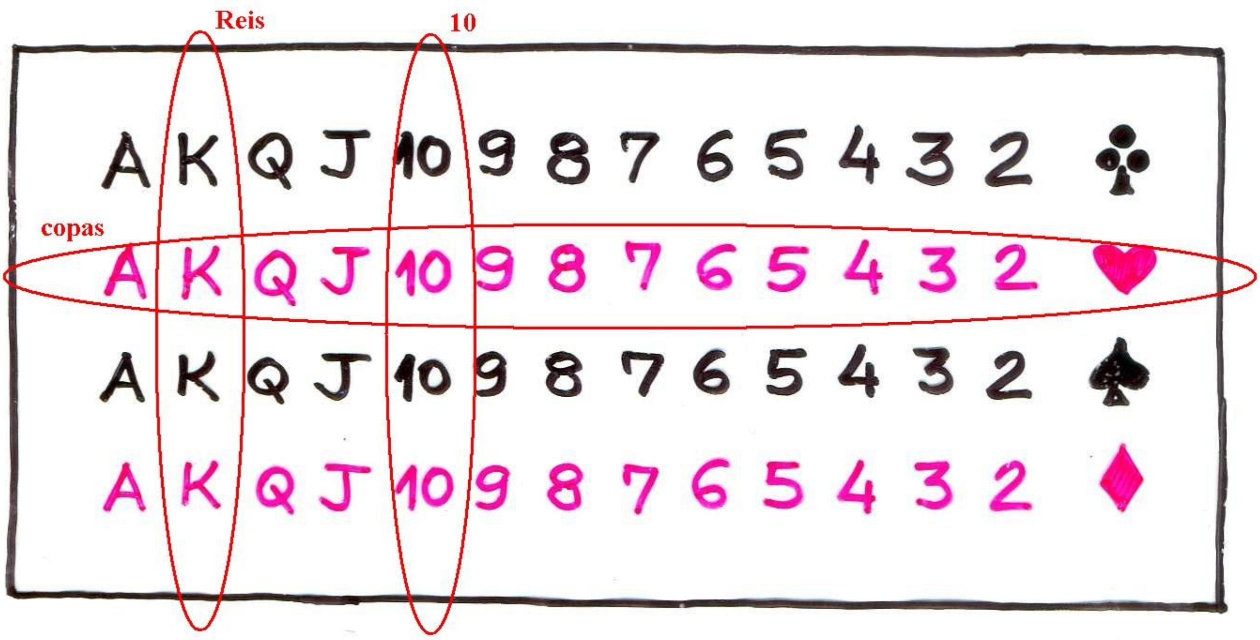
P(vermelha) = P(*ouros* ou *copas*) = P(*ouros*) + P(*copas*)

- Se dois eventos **não** *são mutuamente exclusivos* (ou seja, eles podem ocorrer juntos): Para calcular a probabilidade de que ocorra *ou o evento A, ou o evento evento B ou ambos juntos*, somamos a probabilidade de A com a de B e subtraimos a probabilidade da interseção ($A \cap B$) ou então simplesmente AB.

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

Exemplo 3 : Na retirada de uma carta de um baralho,

- P(*rei* ou *10*) = P(*rei*) + P(*10*)
- P(*rei* ou *copas*) = P(*rei*) + P(*copas*) - P(*rei* ∩ *copas*)



4.4. *Probabilidade condicional*

Exemplo 4 : A tabela abaixo mostra os resultados do levantamento do número de casos de daltonismo numa amostra. Se uma pessoa é escolhida aleatoriamente nesta amostra, qual é a probabilidade de que ela

1. seja uma mulher?
2. seja daltônica?
3. seja uma mulher daltônica?
4. seja um homem daltônico?
5. se uma mulher é escolhida aleatoriamente nesta amostra, qual é a probabilidade de que ela seja daltônica?
6. se um homem é escolhido aleatoriamente, qual é a probabilidade de que ele seja daltônico?

	Sexo		
Daltonismo	Masculino (M)	Feminino (F)	Total
Presente (D+)	423	65	488
Ausente (D-)	4.848	4.664	9.512
Total	5.271	4.729	10.000

A probabilidade mencionada no item (5) é uma *probabilidade condicional*, representada como P(D+|F)

A probabilidade pedida no item (6) é a *probabilidade condicional de a pessoa ser daltônica, dado que ela é do sexo masculino*, representada por:

$$P(D+|M) = P(D+|F')$$

Observe que, na amostra acima, estas duas probabilidades condicionais ao sexo são diferentes: $P(D+|M) \neq P(D+|F)$

Outro exemplo: suponha que desejamos verificar, a partir da tabela abaixo, se a probabilidade de ocorrência de *resfriado* é condicional ao *sexo* da pessoa

	Sexo		
Resfriado	M	F	Totais
Presente (R+)	8	2	10
Ausente (R-)	40	10	50
Totais	48	12	60

$$P(R+ \mid M) = P(R+ \mid F) = P(R+)$$

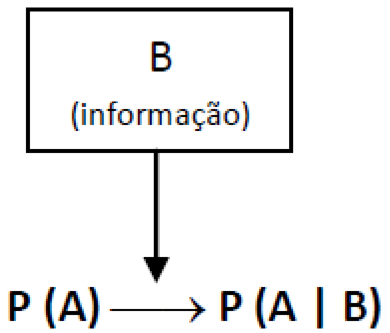
(R+: Resfriado presente; M: sexo masc.; F: sexo fem.)

4.5. *Probabilidade da interseção de eventos (quando há redução do espaço amostral)*

P(*A*|*B*) = Prob. condicional de ocorrer evento *A*, dado que evento *B* ocorreu
= Probabilidade de A, dado B

Por definição, temos que:

$$\mathbf{P(A|B) = \frac{P\,(A \cap B)}{P(B)}}$$



Se P(B|A) = P(B|A’) = P(B) → A e B são *independentes*
Se P(B|A) ≠ P(B|A’) → A e B são *dependentes probabilisticamente*

Dois eventos A e B são *independentes* se:

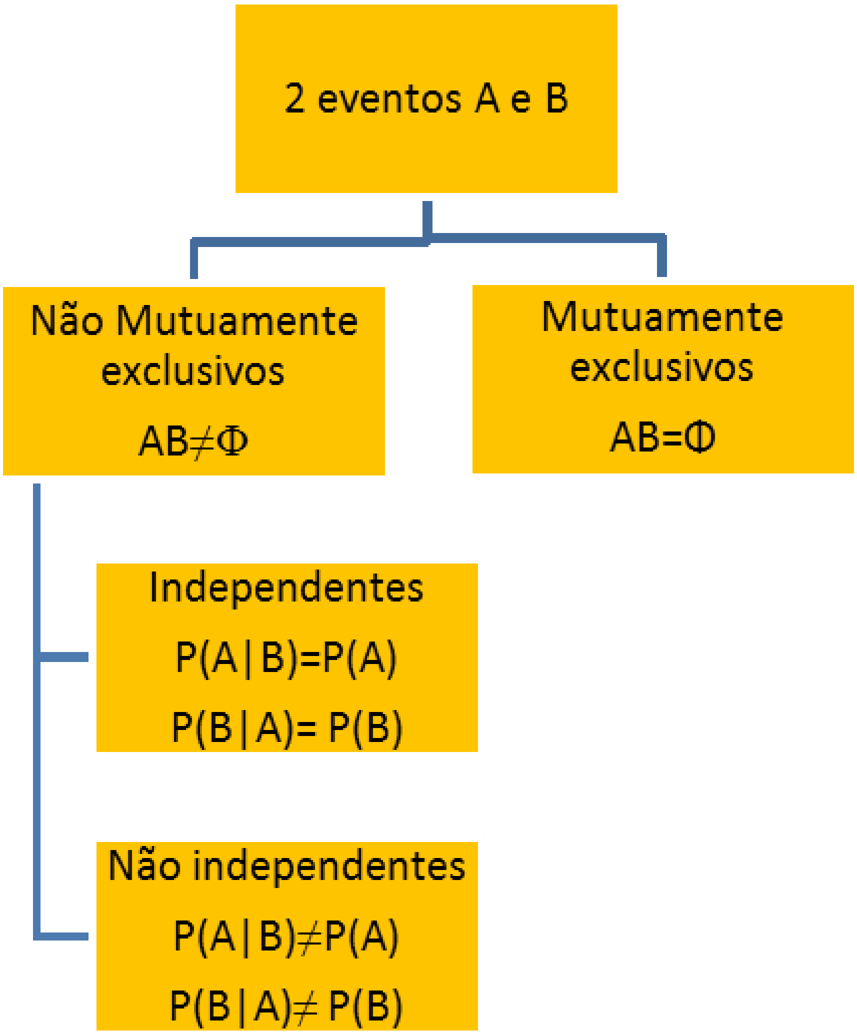
$$P(A \cap B) = P(A)\ .\ P(B)$$

Caso contrário, A e B são *dependentes probabilisticamente* se:

$$P(A \cap B) = P(A)\ .\ P(B|A)$$
$$P(A \cap B) = P(B)\ .\ P(A|B)$$

Ao utilizarmos cálculo de probabilidades devemos sempre considerar as seguintes relações entre dois eventos A e B quaisquer:

4.6. *Probabilidade do evento complementar e da exclusão*



a) se queremos que o evento A **não** aconteça, queremos A’

$$P(A') = 1 - P(A)$$

Os eventos A e A’ são ditos *complementares* se $A \cap A' = \phi$ e $A \cup A' = S$, onde S é o espaço amostral.

b) Se queremos que o evento A ocorra, mas B **não** ocorra, queremos o evento A–B

$$P(A\text{-}B) = P(A) - P(A \cap B)$$

Exemplo 5 : Supor que em um determinado país há 1% da população total infectada pelo VHC (prevalência de VHC) e que, pela literatura, sabemos que 50% dos usuários de drogas injetáveis são VHC+ (positivo para o vírus C da hepatite). Além disso, a literatura especializada informa que 1 em cada 10 pacientes VHC+ é usuário de droga injetável. Definir os diferentes eventos apresentados acima, suas respectivas probabilidades e probabilidades condicionais. Calcule então o percentual de usuários de droga injetável nesta população.

A: ser VHC+
B: ser usuário de drogas injetáveis
P (A) = 0,01
P(A|B) = 0,50
P(B|A) = 0,10
P(B) = ?
Portanto, P(B) × 0,50 = 0,01 ×0,10.
P(B) = 0,002 = 0,2%

Exemplo 6 : Suponhamos que um fundo de formatura faz uma rifa de uma moto, com 100 bilhetes e outra de uma televisão, com 50. Se uma pessoa compra um bilhete de cada rifa e definimos os eventos
M = {a pessoa ganha a moto}; T = {a pessoa ganha a televisão}

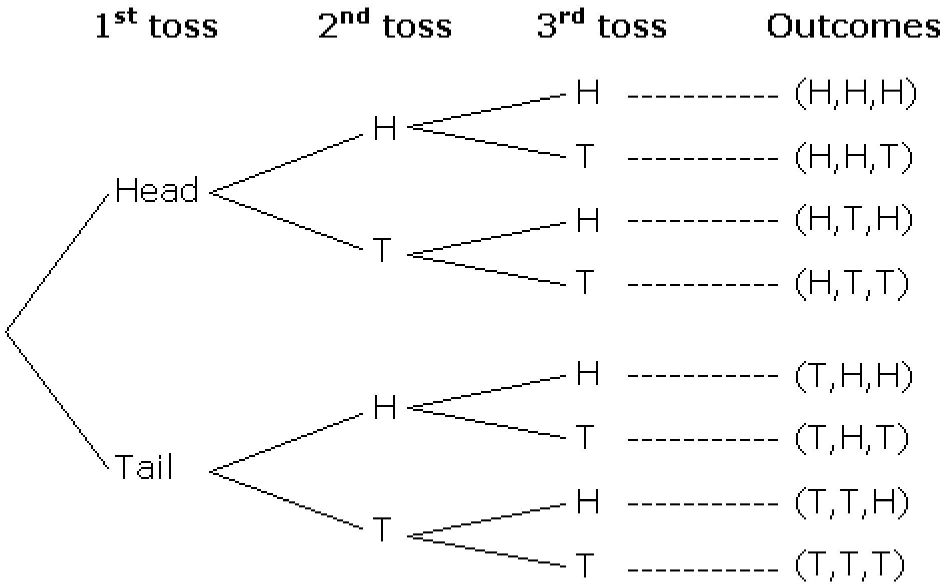
a) Podemos dizer que os eventos M e T são complementares?
b) Podemos dizer que os eventos M e T são independentes probabilisticamente?

Como se representam, em notação de conjuntos, os eventos onde a pessoa:
a) ganha os dois prêmios?
b) ganha alguma coisa (um ou dois prêmios)?
c) não ganha nada?
d) ganha apenas a moto?
Calcule as probabilidades de ocorrência de cada um destes eventos.

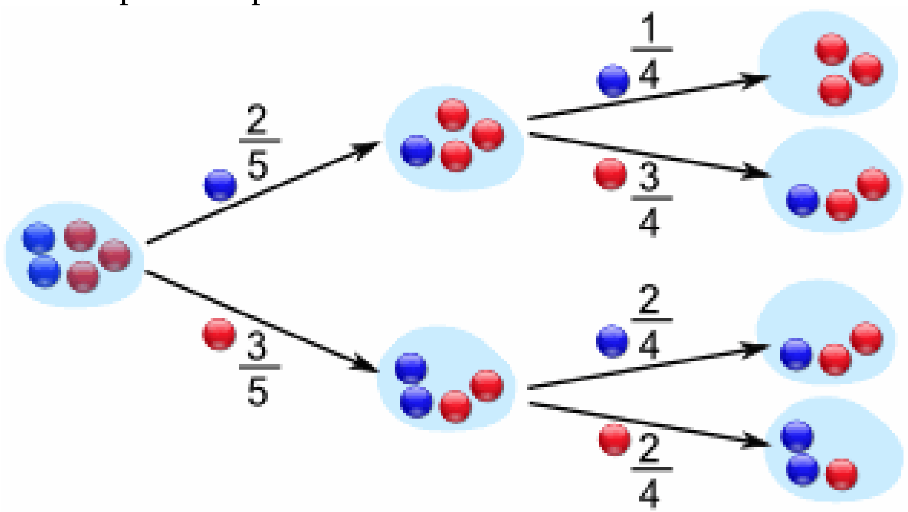
5. Cálculo de probabilidades: experimentos repetidos, diagr. de árvore

- Quando o diagrama de árvore que representa um problema tem ramos *não-equiprováveis*, é preciso calcular separadamente a probabilidade de cada ramo;
- A probabilidade de ramo é calculada pelo produto das probabilidades dos nós que o compõem.

Exemplo 7 : Lançamento de moedas: P(3 caras)



Exemplo 8: se temos dois eventos de interesse A: azul e B = A'=vermelho, podemos representar a retirada aleatória de 2 elementos (a sequência de duas retiradas sem reposição), onde nos "nós" foram representados os respectivos espaços amostrais e nos "ramos" foram representadas os eventos e as respectivas probabilidades:



Exemplo 9 : Lançamento de 3 dados (x: número mostrado por cada dado)

P(3 números pares)

P(666)

Exemplo 10 :Três cartas são retiradas aleatoriamente, sem reposição, de um baralho completo. Quais são as probabilidades de serem:

P(3 ases) ?

P(3 figuras) ?

P(apenas 2 ases) ?

Exemplo 11 : Um animal foi treinado para executar uma tarefa. A probabilidade de que ele falhe na sua primeira tentativa é de 0,40. Se o animal falhar, ele faz nova tentativa; o animal aprende com cada erro cometido, de modo que a probabilidade de um erro em cada tentativa é apenas metade da probabilidade de erro na tentativa fracassada anterior. Se é permitido ao animal fazer apenas 3 tentativas, qual é a probabilidade de que ele consiga realizar a tarefa?

Conclusão: O diagrama de árvore serve para facilitar a enumeração de todos os eventos possíveis em uma sequência de eventos, independentes ou não.

6. Como obter as probabilidades elementares?

- pela definição clássica
- pela observação da *frequência relativa* num grande numero de repetições (definição *frequencista* ou *empírica*)

6.1. *Definição com base na frequência relativa (prob. empírica)*

- A probabilidade é o valor para o qual tende a frequência relativa de um evento A, quando o número *n* de repetições de um experimento tende para infinito.

$$P(A) = \lim_{n \to \infty} \frac{h_A}{n}$$

- A *probabilidade empírica* é estimada por h_A/n, para *n* suficientemente grande

Exemplo 12 : Entre os 499 registrados no arquivo para um dado ano da maternidade da USP, há 260 meninos e 235 meninas.

P(meninos) ≅ 260/499 = 0,52

Exemplo 13: De 4.065.014 nascimentos nos EUA em 1992, houve 2.081.287 meninos e 1.983.727 meninas (Pagano e Gauvreau, p.136).

P(meninos) ≅ 2081287/4065014 = 0,5120

Exemplo 14 : Uma mulher tem 3 crianças. Supondo P(menina)= 0.48

Qual é a probabilidade de que sejam todas meninas?

P = (0,48)^3 = 0,1106

Qual a probabilidade de que a mais velha seja um menino, e as outras duas meninas?

P(m,f,f) = 0,52 x 0,48 x 0,48 = 0,1198

7. O Teorema de Bayes para Probabilidade Condicional

Já conhecemos a definição de probabilidade condicional entre dois eventos A e B:

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(AB)}{P(B)} = \frac{P(BA)}{P(B)}$$

Logo, podemos reescrever $P(A|B)$ em função de $P(B|A)$:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \dots\dots\dots (1)$$

Sabemos que $B = \{A \cap B\} \cup \{\bar{A} \cap B\}$, ou seja, B pode ser representado pela união de dois eventos mutuamente exclusivos. Então, pelo Axioma 2, temos que:

$$P(B) = P(A \cap B) + P(\bar{A} \cap B)$$

As duas parcelas do lado direito da igualdade acima podem ser reescritas a partir da definição de probabilidade condicional:

$$P(B) = P(B|A)P(A) + P(B|\bar{A})P(\bar{A}) \dots\dots\dots (2)$$

Substituindo (2) em (1) temos:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|\bar{A})P(\bar{A})} \dots\dots\dots (3)$$

Em (3) temos o que é conhecida como fórmula de Bayes ou representação do Teorema de Bayes, que define uma probabilidade condicional em termos de outras probabilidades conhecidas.

Veremos a seguir que o Teorema de Bayes tem uma importância muito grande para a Medicina, em especial para avaliação de testes de triagem ("screening") e de testes e procedimentos diagnósticos.

Exemplo 15: Em estudo caso-controle sobre marcadores para diagnóstico de apendicite (D) foi avaliado, dentre outros marcadores, a "contagem de glóbulos brancos acima de 17^9 células por L" (B). Observou-se que dentre os que realmente apresentaram apendicite 15% tinham o valor do marcador alterado e, dentre aqueles que não apresentaram apendicite apenas 2% tinham o valor do marcador alterado. Outros estudos indicaram uma prevalência de apendicite de aproximadamente igual a 20% dentre aqueles que apresentavam sinais clínicos de acordo com o protocolo vigente. Qual a probabilidade de apendicite após um resultado alterado de contagem de glóbulos brancos em indivíduos que apresentem sinais clínicos de acordo com o protocolo vigente?

B: marcador alterado; D: apendicite
P(B|D) = 0,15; P(B|D') = 0,02; P(D) = 0,20; P(D|B) = ?

$$P(D|B) = \frac{0{,}15 \times 0{,}20}{0{,}15 \times 0{,}20 + 0{,}02 \times 0{,}80}$$

$$P(D|B) = 0{,}65$$

Qual a sua conclusão?