8 - PROBABILIDADE

8.1 - Introdução

No capítulo anterior foi utilizado um raciocínio predominantemente indutivo: os dados eram coletados, e através da sua organização em distribuições de freqüências era possível caracterizar a variabilidade do fenômeno observado, e elaborar hipóteses ou conjecturas a respeito.

Suponha que se está estudando o percentual de meninos e meninas nascidos em um estado brasileiro. Consultando dados do IBGE, provenientes de censos e levantamentos anteriores (portanto distribuições de freqüências da variável qualitativa sexo dos recém-nascidos) há interesse em *prever* qual será o percentual de nascimentos no ano de 2020: em suma será usado um raciocínio dedutivo, a partir de algumas suposições sobre o problema (a definição dos resultados possíveis, os percentuais registrados em anos anteriores) tenta-se obter novos valores.

Se o percentual de meninos no passado foi de 49% a pergunta é: qual será o percentual de meninos nascidos no ano de 2020? É possível que seja um valor próximo de 49%, talvez um pouco acima ou um pouco abaixo, mas não há como responder com certeza absoluta, pela simples razão que o fenômeno ainda não ocorreu, e que sua natureza é ALEATÓRIA: ou seja é possível identificar quais serão os resultados possíveis (menino ou menina), e há uma certa regularidade nos percentuais de nascimentos (verificados anteriormente), mas não é possível responder qual será o resultado exato ANTES do fenômeno ocorrer. A regularidade citada acima (que foi observada para um **grande** número de nascimentos) permite que seja calculado o grau de certeza, ou confiabilidade, da previsão feita, que recebe o nome de PROBABILIDADE. Haverá uma grande probabilidade de que realmente o percentual de meninos nascidos em 1998 seja de 49%, mas NADA IMPEDE que um valor diferente venha a ocorrer.

Marcelo Menezes Reis

Sem saber montamos um **MODELO PROBABILÍSTICO** para o problema em questão:
- foram definidos todos os RESULTADOS POSSÍVEIS para o fenômeno (experimento);
- definiu-se uma REGRA que permite dizer quão provável será cada resultado ou grupo de resultados.

O Modelo Probabilístico permite expressar o grau de incertezas através de probabilidades.

A regra citada acima foi definida a partir de observações anteriores do fenômeno, mas também poderia ser formulada com base em considerações teóricas. Por exemplo, se há interesse em estudar as proporções de ocorrências das faces de um dado, e se este dado não é viciado espera-se que cada face ocorra em 1/6 do total de lançamentos: se o dado for lançado um grande número de vezes isso provavelmente ocorrerá, mas um resultado diferente poderia ser obtido sem significar que o dado está viciado, principalmente se forem feitos pouco lançamentos[1].

Neste ponto é importante ressaltar que os modelos probabilísticos não têm razão de ser para fenômenos (experimentos) NÃO ALEATÓRIOS: aqueles em que usando teorias e fórmulas apropriadas pode-se prever exatamente qual será o seu resultado **antes** do fenômeno ocorrer: por exemplo, o lançamento de uma pedra de 5 kg de uma altura de 10 metros, havendo interesse em cronometrar o tempo para que ela atinja o chão. Conhecendo o peso da pedra, a altura do

[1] Para construir ou utilizar modelos probabilísticos é necessário que haja um grande número de realizações do fenômeno (experimento) para que uma regularidade possa ser verificada: é a Lei dos Grandes Números. No início do século XX o estatístico inglês Karl Pearson lançou uma moeda não viciada 24000 vezes (!) para verificar a validade dessa lei: obteve 12012 caras, praticamente o valor esperado (12000, 50%).

lançamento, a aceleração da gravidade e as leis da física é perfeitamente possível calcular o tempo de queda, não há necessidade sequer de realizar o experimento.

Vamos passar a algumas definições importantes para estudar os modelos probabilísticos.

8.1.1 - Experimento Aleatório

Experimento Aleatório é um processo de obtenção de um resultado ou medida que apresenta as seguintes características:
- não se pode afirmar, ANTES de realizar o experimento, qual será o resultado de uma realização, mas é possível determinar o conjunto de resultados possíveis.
- quando é realizado um grande número de vezes (replicado) apresentará uma REGULARIDADE que permitirá construir um **modelo probabilístico** para analisar o experimento.

São experimentos aleatórios: lançamento de um dado não viciado e observação da face voltada para cima; cruzar espécimes de ervilha e observar os fenótipos dos descendentes.

8.1.2 - Espaço Amostral (Ω)

Espaço Amostral é o conjunto de **TODOS** os resultados possíveis de um experimento aleatório. “PARA CADA EXPERIMENTO ALEATÓRIO HAVERÁ UM ESPAÇO AMOSTRAL ÚNICO Ω ASSOCIADO A ELE “.

Exemplo 8.1 - Definir os espaços amostrais dos experimentos abaixo:
a- Lançar uma moeda e observar a face voltada para cima.
Os dois únicos resultados possíveis são cara e coroa: Ω = {*Cara, Coroa*}.

b- Resultado do cruzamento de 2 indivíduos heterozigotos.
Um indivíduo heterozigoto possui genótipo Aa, se dois indivíduos heterozigotos forem cruzados há 3 resultados possíveis: Ω = {*AA, Aa, aa*}.

c- Altura de homens adultos.
De uma forma genérica poderíamos definir indivíduo adulto como tendo mais de 1,40m de altura:
Ω = {*Altura > 1,40m*}

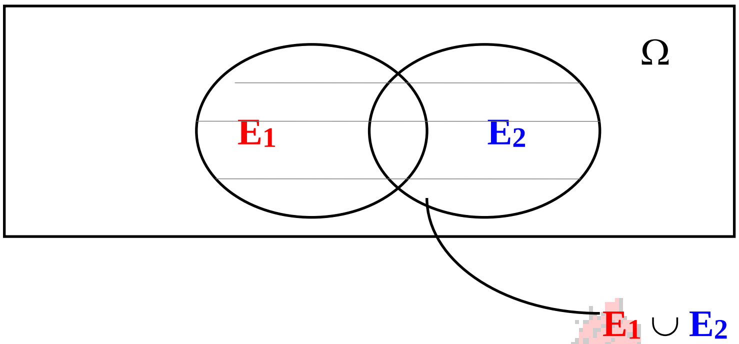
d - Observar o número de meninos em famílias de 5 filhos.
Cada família pode ter no mínimo 0 e no máximo 5 meninos: $\Omega = \{0, 1, 2, 3, 4, 5\}$

8.1.3 - Evento

Evento é QUALQUER subconjunto do espaço amostral. Um evento pode conter um ou mais resultados, se pelo menos um dos resultados ocorrer o evento ocorre! Geralmente há interesse em calcular a probabilidade de que um determinado evento venha a ocorrer, e este evento pode ser definido de forma <u>verbal</u>, precisando ser “traduzido” para as definições da Teoria de Conjuntos, que veremos a seguir.

Seja o Experimento Aleatório lançamento de um dado não viciado e observação da face voltada para cima: o seu espaço amostral será $\Omega = \{1, 2, 3, 4, 5, 6\}$. Definindo três eventos: **E_1** = {2, 4, 6}, **E_2** = {3, 4, 5, 6} e **E_3** = {1, 3} serão apresentadas as definições de Evento União, Evento Intersecção, Eventos Mutuamente Exclusivos e Evento Complementar.

Evento **União** de **E_1** com **E_2** (**$E_1 \cup E_2$**): evento que ocorre se E_1 **OU** E_2 **OU** ambos ocorrem.

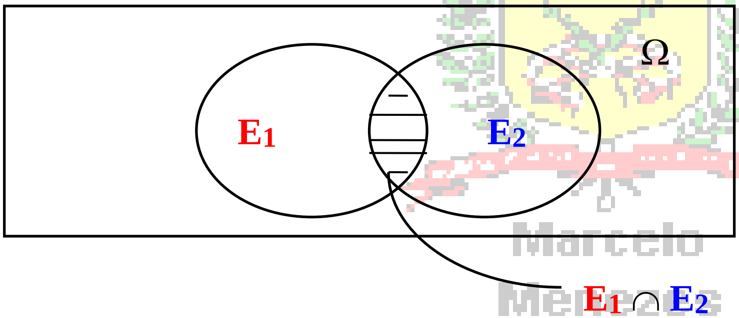


$E_1 \cup E_2$ = {2, 3, 4, 5, 6}

Composto por todos os resultados que pertencem a um **ou** ao outro, **ou** a ambos.

Figura 1 - Evento união

Evento **Intersecção** de **E_1** com **E_2** (**$E_1 \cap E_2$**) : evento que ocorre se E_1 **E** E_2 ocorrem SIMULTANEAMENTE.



$E_1 \cap E_2$ = {4, 6}

Composto por todos os resultados que pertencem a ambos.

Figura 2 - Evento intersecção

Eventos **Mutuamente Exclusivos** (M.E.): são eventos que NÃO PODEM OCORRER SIMULTANEAMENTE, não apresentando elementos em comum (sua intersecção é o conjunto vazio).

Dentre os três eventos definidos acima, observamos que os eventos **E_1** e **E_3** não têm elementos em comum:

E_3 = {1, 3} **E_1** = {2, 4, 6} **$E_1 \cap E_3$** = $\varnothing$ => **E_1** e **E_3** são mutuamente exclusivos

Evento **Complementar** de um evento qualquer é formado por todos os resultados do espaço amostral que NÃO PERTENCEM ao evento. A união de um evento e seu complementar formará o próprio Espaço Amostral, e a intersecção de um evento e seu complementar é o conjunto vazio.

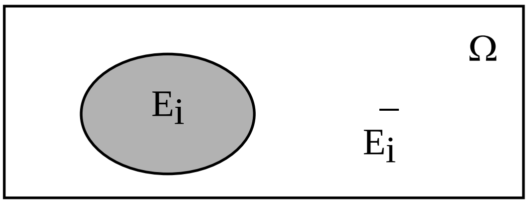


Figura 3 - Evento complementar

$E_i \cup \bar{E}_i = \Omega$ $\qquad$ $E_i \cap \bar{E}_i = \varnothing$

E_1 = {2, 4, 6} $\qquad$ **$\bar{E}_1$** = {1, 3, 5}

E_2 = {3, 4, 5, 6} $\qquad$ **$\bar{E}_2$** = {1, 2}

Marcelo Menezes Reis

8.2 - Definições de Probabilidade

A repetição de um experimento, mesmo sob condições semelhantes, poderá levar a resultados (eventos) diferentes. Mas se o experimento for repetido um número "suficientemente grande" de vezes haverá uma regularidade nestes resultados que permitirá calcular a sua probabilidade de ocorrência. Há três definições de probabilidade, que se complementam.

8.2.1 - Definição Clássica

Se um experimento aleatório puder resultar em **n** diferentes e igualmente prováveis resultados, e $\mathbf{n_{Ei}}$ destes resultados referem-se ao evento **Ei**, então a probabilidade do evento **Ei** ocorrer será: $P(Ei) = \frac{n_{Ei}}{n}$

O problema reside em calcular o número total de resultados possíveis e o número de resultados associados ao evento de interesse. Isso pode ser feito usando técnicas de análise combinatória (que serão vistas posteriormente) ou por considerações teóricas ("bom senso").

Exemplo 8.2 - Seja o seguinte Experimento Aleatório: lançamento de um dado não viciado e observação da face voltada para cima. Calcular as probabilidades de ocorrência dos seguintes eventos:
a) Face 1.
b) Face par.
c) Face menor ou igual a 2.

Marcelo
Menezes
Reis

*O Espaço Amostral deste experimento será: $\Omega = \{1, 2, 3, 4, 5, 6\}$. Sendo assim há um total de 6 resultados possíveis, resultando em **n** = 6. Basta então definir quantos resultados estão associados a cada evento para que seja possível calcular suas probabilidades pela definição clássica.*

O evento "face 1" tem apenas um resultado associado: { 1 }. Então $\mathbf{n_{Ei}} = 1$, e a probabilidade de ocorrer a face 1 será: $P(Ei) = \frac{n_{Ei}}{n} = \frac{1}{6}$

O evento "face par" tem três resultados associados: {2, 4, 6}. Então $\mathbf{n_{Ei}} = 3$, e a probabilidade de ocorrer face par será: $P(Ei) = \frac{n_{Ei}}{n} = \frac{3}{6} = \frac{1}{2}$

O evento "face menor ou igual a 2" tem dois resultados associados: {1, 2}. Então $\mathbf{n_{Ei}} = 2$, e a probabilidade de ocorrência de face menor ou igual a 2 será: $P(Ei) = \frac{n_{Ei}}{n} = \frac{2}{6} = \frac{1}{3}$

8.2.2 - Definição Axiomática

Seja um experimento aleatório e um espaço amostral Ω associado a ele. A cada evento **Ei** associaremos um número real denominado **P(Ei)** que deve satisfazer as seguintes condições:

a) $0 \leq$ **P(Ei)** $\leq 1{,}0$
A probabilidade de ocorrência de um evento SEMPRE é um número real entre 0 e 1 (0% e 100%)

b) **P (Ω)** = 1,0
A probabilidade de ocorrência do Espaço Amostral é igual a 1 (100%) pois pelo menos um dos resultados do Espaço Amostral ocorrerá. Por isso o Espaço Amostral é chamado de Evento Certo.

c) **P ($\varnothing$)** = 0
A probabilidade de ocorrência do conjunto vazio é NULA (igual a zero), uma vez que não há resultados no conjunto vazio. Por isso o conjunto vazio é chamado de Evento Impossível.

d) Σ **P(Ei)** = 1,0
Se a probabilidade de ocorrência do Espaço Amostral é igual a 1 (100%) ao somar as probabilidades de todos os eventos que compõem o Espaço Amostral o resultado deverá ser igual a 1 (100%).

e) **P(Ei)** = 1 - **P($\overline{\textbf{Ei}}$)**
A probabilidade de ocorrência de um evento qualquer será igual a probabilidade do Espaço Amostral (1 ou 100%) menos a probabilidade de seu evento complementar (a soma das probabilidades de todos os outros eventos do Espaço Amostral).

e) Sejam **Ei** e **Ej** dois eventos quaisquer: P (**Ei** $\cup$ **Ej**) = **P(Ei)** + **P(Ej)** - P(**Ei** $\cap$ **Ej**)
A probabilidade de ocorrência do evento União de dois outros eventos será igual a soma das probabilidades de cada evento menos a probabilidade de ocorrência do evento Intersecção dos mesmos dois eventos. Esta propriedade também é chamada de REGRA DA ADIÇÃO.

Exemplo 8.3 - Seja o Experimento Aleatório lançamento de um dado não viciado e observação da face voltada para cima definido no Exemplo 8.2: o seu espaço amostral será $\Omega = \{1, 2, 3, 4, 5, 6\}$.
Definindo três eventos: $\mathbf{E_1}$ = face 1 = {1}, $\mathbf{E_2}$ = face par = {2, 4, 6} e $\mathbf{E_3}$ = face $\leq$ 2 {1, 2}, cujas probabilidades já foram calculadas.
Calcular a probabilidade de ocorrência dos seguintes eventos:
a) Complementar de $\mathbf{E_1}$.
b) Complementar de $\mathbf{E_2}$.
c) União de $\mathbf{E_2}$ e $\mathbf{E_3}$.
d) União de $\mathbf{E_1}$ e $\mathbf{E_2}$.
No Exemplo 8.2 obteve-se $P(\mathbf{E_1}) = 1/6$, $P(\mathbf{E_2}) = 3/6$ *e* $P(\mathbf{E_3}) = 2/6$.
Usando os axiomas:
$\mathbf{P(E_1)} = 1 - \mathbf{P(\overline{E_1})}$ *então* $\mathbf{P(\overline{E_1})} = 1 - \mathbf{P(E_1)} = 1 - 1/6 = 5/6$ $\quad \overline{\mathbf{E_1}} = \{2, 3, 4, 5, 6\}$
$\mathbf{P(E_2)} = 1 - \mathbf{P(\overline{E_2})}$ então $\mathbf{P(\overline{E_2})} = 1 - \mathbf{P(E_2)} = 1 - 3/6 = 3/6$ $\quad \overline{\mathbf{E_2}} = \{1, 3, 5\}$
$\mathbf{P(E_2 \cup E_3) = P(E_2) + P(E_3) - P(E_2 \cap E_3)}$ *Observe que há apenas um elemento em comum entre os eventos* $\mathbf{E_2}$ *e* $\mathbf{E_3}$: *apenas um resultado associado =>* $\mathbf{P(E_2 \cap E_3)} = 1/6$

$\mathbf{P(E_2 \cup E_3)} = 3/6 + 2/6 - 1/6 = 4/6$

$\mathbf{P(E_1 \cup E_2) = P(E_1) + P(E_2) - P(E_1 \cap E_2)}$ *Não há elementos em comum entre os eventos* $\mathbf{E_1}$ *e* $\mathbf{E_2}$: *eles são mutuamente exclusivos, sua intersecção é o conjunto vazio, e a probabilidade de ocorrência do conjunto vazio é nula.* $\mathbf{P(E_1 \cup E_2)} = 1/6 + 3/6 - 0 = 4/6$

Marcelo Menezes Reis

8.2.3 - Definição Experimental

Seja um experimento aleatório que é repetido **n** vezes, e **Ei** um evento associado.
A freqüência relativa do evento **Ei**: $f_{REi} = \frac{n_{Ei}}{n} = \frac{\text{no vezes que Ei ocorreu}}{\text{total de tentativas}}$

Quando o número de repetições tende ao infinito (ou a um número suficientemente grande) $\mathbf{f_{REi}}$ tende a um limite: a probabilidade de ocorrência do evento **Ei**. A probabilidade do evento pode ser estimada através da freqüência relativa.

Quando não há outra maneira de obter as probabilidades dos eventos é necessário realizar o experimento várias vezes (replicá-lo) para que seja possível obter um número tal de tentativas que permita que as freqüências relativas estimem as probabilidades, para que seja possível construir um modelo probabilístico para o experimento.

8.3 - Probabilidade Condicional

Muitas vezes há interesse de calcular a probabilidade de ocorrência de um evento A qualquer, sabendo (ou supondo) que um outro evento B ocorreu previamente. Em outras palavras queremos calcular a probabilidade de ocorrência de A CONDICIONADA à ocorrência prévia de B, simbolizada por P(A | B) - lê-se probabilidade de A dado B - e a sua expressão será:

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} \quad \text{para} \quad P(B) > 0$$

A probabilidade de ocorrência de A condicionada à ocorrência de B será igual à probabilidade da intersecção entre A e B, dividida pela probabilidade de ocorrência de B (o evento que já ocorreu)[2].

Se houvesse interesse no oposto, probabilidade de ocorrência de B condicionada à ocorrência prévia de A:

$$P(B \mid A) = \frac{P(B \cap A)}{P(A)} \quad \text{para} \quad P(A) > 0$$

Neste caso o valor no denominador seria a probabilidade de A uma vez que este evento ocorreu previamente[3]. É importante ressaltar que a operação de intersecção é **comutativa**, implicando em:

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

Exemplo 8.4 - Seja o lançamento de 2 não viciados, um após o outro, e a observação das faces voltadas para cima. Calcular as probabilidades:
a) de que as faces sejam iguais supondo-se que sua soma é menor ou igual a 5.
b) de que a soma das faces seja menor ou igual a 5, supondo-se que as faces são iguais.
Observe que há interesse em calcular a probabilidade de eventos, supondo que outro evento ocorreu previamente.
Como todo problema de probabilidade é preciso montar o Espaço Amostral. Neste caso serão os

[2] No denominador da expressão é colocada **sempre** a probabilidade do evento que JÁ OCORREU.
[3] Tal como B na outra expressão.

Marcelo Menezes Reis

pares de faces dos dados, e como os dados são lançados um após o outro a ordem das faces é importante:

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

Há um total de 36 resultados possíveis: **n** *= 36.*

Agora é preciso definir os eventos de interesse:

a) "Faces iguais sabendo-se que sua soma é menor ou igual a 5" significa dizer probabilidade de ocorrência de faces iguais supondo-se que ***já ocorreram*** *faces cuja soma é menor ou igual a 5; chamando o evento faces iguais de* $\mathbf{E_1}$ *e o evento soma das faces menor ou igual a 5 de* $\mathbf{E_2}$ *estamos procurando* $P(\mathbf{E_1} \mid \mathbf{E_2})$*, probabilidade de ocorrência de* E_1 *condicionada à ocorrência PRÉVIA de* $\mathbf{E_2}$.

Usando a fórmula:

$$P(E_1 \mid E_2) = \frac{P(E_1 \cap E_2)}{P(E_2)}$$ *é preciso encontrar os valores das probabilidades.*

Primeiramente definir o número de resultados do Espaço Amostral que pertencem aos eventos de interesse, para que seja possível calcular a sua probabilidade usando a definição clássica de probabilidade:

$\mathbf{E_1}$ = **{**(1,1) (2,2) (3,3) (4,4) (5,5) (6,6) **}** - *faces iguais, 6 resultados,* $\mathbf{n_{E1}}$ = 6.

$\mathbf{E_2}$ = **{**(1,1) (1,2) (1,3) (1,4) (2,1) (2,2) (2,3) (3,1) (3,2) (4,1)**}** - *soma das faces* ≤ 5*, 10 resultados,* $\mathbf{n_{E2}}$ = 10.

Os elementos em comum formarão o evento intersecção: $\mathbf{E_1 \cap E_2}$ = {(1,1) (2,2)} - *faces iguais e soma das faces* ≤ 5*, 2 resultados,* $\mathbf{n_{E1 \cap E2}}$ = 2.

$\mathbf{P(E_2) = n_{E2} / n = 10/36}$ $\mathbf{P(E_1 \cap E_2) = n_{E1\cap E2} / n = 2/36}$

Tendo as probabilidades acima é possível calcular a probabilidade condicional:

$$P(E_1 \mid E_2) = \frac{P(E_1 \cap E_2)}{P(E_2)} = \frac{2/36}{10/36} = \frac{2}{10} = 0{,}2\,(20\%)$$

Então a probabilidade de que as faces são iguais sabendo-se que sua soma é menor ou igual a 5 é de 20%.

Este resultado poderia ser obtido de outra forma. Se a soma das faces é menor ou igual a 5, o evento $\mathbf{E_2}$ *já ocorreu previamente, então o Espaço Amostral* ***modificou-se****, passando a ser o conjunto de resultados do evento* $\mathbf{E_2}$:

novo Ω = **{** (1,1) (1,2) (1,3) (1,4) (2,1) (2,2) (2,3) (3,1) (3,2) (4,1)**}**

O novo Espaço Amostral tem 10 resultados, novo **n** = 10.

O número de resultados do evento faces iguais ($\mathbf{E_1}$) *no novo Espaço Amostral é igual a 2, novo* $\mathbf{n_{E1}}$ = 2 *(há apenas dois pares no novo Espaço Amostral, de soma das faces menor ou igual a 5, em que as faces são iguais).*

Então a probabilidade de ocorrer o evento $\mathbf{E_1}$ *no novo Espaço Amostral, ou seja a probabilidade de ocorrência do evento* $\mathbf{E_1}$ ***condicionada*** *à ocorrência prévia do evento* $\mathbf{E_2}$, $P(\mathbf{E_1} \mid \mathbf{E_2})$*, será:*

$P(\mathbf{E_1}|\ \mathbf{E_2})$ = novo $\mathbf{n_{E1}}$/ novo $\mathbf{n}$ = 2/10 = 0,2 (20%) *o mesmo resultado obtido anteriormente.*

b) *"Soma das faces menor ou igual a 5 sabendo-se que as faces são iguais" significa dizer probabilidade de ocorrência de faces cuja soma é menor ou igual a 5 supondo-se que **já ocorreram** faces que são iguais*[4]*; chamando o evento faces iguais de* $\mathbf{E_1}$ *e o evento soma das faces menor ou igual a 5 de* $\mathbf{E_2}$ *estamos procurando* $P(\mathbf{E_2} \mid \mathbf{E_1})$*, probabilidade de ocorrência de* $\mathbf{E_2}$ *condicionada à ocorrência PRÉVIA de* $\mathbf{E_1}$.

Usando a fórmula: $P(E_2 \mid E_1) = \dfrac{P(E_2 \cap E_1)}{P(E_1)}$ *todos os valores já foram obtidos no item a.*

$$P(E_2 \mid E_1) = \frac{P(E_2 \cap E_1)}{P(E_1)} = \frac{2/36}{6/36} = \frac{2}{6} = 0{,}33\,(33\%)$$

Então a probabilidade de que as faces tenham soma menor ou igual a 5 sabendo-se que são iguais é de 33%.

Da mesma forma que no item a o resultado poderia ser obtido se outra forma. Se as faces são iguais, o evento $\mathbf{E_1}$ *já ocorreu previamente, então o Espaço Amostral **modificou-se**, passando a ser o conjunto de resultados do evento* $\mathbf{E_1}$:

novo $\Omega = \{$ (1,1) (2,2) (3,3) (4,4) (5,5) (6,6) $\}$

O novo Espaço Amostral tem 6 resultados, novo $\mathbf{n}$ = 6.
O número de resultados do evento soma das faces menor ou igual a 5 ($\mathbf{E_2}$*) no novo Espaço Amostral é igual a 2, novo* $\mathbf{n_{E2}}$ = 2 *(há apenas dois pares no novo Espaço Amostral, de faces iguais, em que a soma das faces é menor ou igual a 5).*

Marcelo Menezes Reis

Então a probabilidade de ocorrer o evento $\mathbf{E_2}$ *no novo Espaço Amostral, ou seja a probabilidade de ocorrência do evento* $\mathbf{E_2}$ ***condicionada** à ocorrência prévia do evento* $\mathbf{E_1}$, $P(\mathbf{E_2}|\ \mathbf{E_1})$, *será*:

$P(\mathbf{E_2}|\ \mathbf{E_1})$ = novo $\mathbf{n_{E2}}$/ novo $\mathbf{n}$ = 2/6 = 0,33 (33%) *o mesmo resultado obtido anteriormente.*

É EXTREMAMENTE IMPORTANTE LEMBRAR QUE, CONCEITUALMENTE $P(A|B) \neq P(B|A)$[5]

8.3.1 - Regra do Produto

Uma das conseqüências da expressão da probabilidade condicional é a regra do produto, isolando a probabilidade da intersecção:

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} \quad => \quad P(A \cap B) = P(B) \times P(A \mid B)$$

Neste caso o evento B ocorreu previamente, e o segundo valor é a probabilidade de ocorrência de A dado que B ocorreu.

$$P(B \mid A) = \frac{P(A \cap B)}{P(A)} \quad => \quad P(A \cap B) = P(A) \times P(B \mid A)$$

Neste caso o evento A ocorreu previamente, e o segundo valor é a probabilidade de ocorrência de B dado que A ocorreu[6].

---

[4] Houve uma mudança no evento que ocorreu previamente.
[5] Pois os eventos que ocorreram previamente são DIFERENTES.
[6] Não se esqueça que a intersecção é COMUTATIVA.

É importante que seja observada com cuidado a seqüência dos eventos para montar as expressões acima: analisar corretamente que evento já ocorreu.

8.3.2 - Eventos Independentes

Dois ou mais eventos são independentes quando a ocorrência de um dos eventos não influencia a probabilidade de ocorrência dos outros. Se dois eventos A e B são independentes então a probabilidade de A ocorrer dado que B ocorreu é igual à própria probabilidade de ocorrência de A, e a probabilidade de B ocorrer dado que B ocorreu é igual à própria probabilidade de ocorrência de B.

Se A e B são independentes então:

$$P(A \mid B) = P(A) \quad e \quad P(B \mid A) = P(B)$$

$$P(A \cap B) = P(A) \times P(B \mid A) = P(A) \times P(B)$$

$$P(A \cap B) = P(B) \times P(A \mid B) = P(B) \times P(A)$$

AS EXPRESSÕES ACIMA SÃO VÁLIDAS SE E SOMENTE SE OS EVENTOS **A** E **B** FOREM <u>INDEPENDENTES</u>!

Em situações práticas dois eventos são independentes quando a ocorrência de um deles não modifica, ou modifica muito pouco, o Espaço Amostral do Experimento Aleatório.

Exemplo 8.5 - Uma urna contém 2 bolas brancas e 3 vermelhas. Retiram-se 2 bolas ao acaso, uma após a outra. Resolva os itens abaixo:
a) Se a retirada for feita SEM REPOSIÇÃO.
   a.1- Qual é a probabilidade de que as 2 bolas retiradas sejam da mesma cor?
   a.2- Qual é a probabilidade de que as 2 bolas retiradas sejam vermelhas, supondo-se que são da mesma cor?
b) Se a retirada for feita COM REPOSIÇÃO.
   b.1- Qual é a probabilidade de que as 2 bolas retiradas sejam da mesma cor?
   b.2- Qual é a probabilidade de que as 2 bolas retiradas sejam vermelhas, supondo-se que são da mesma cor?

Como em todos os problemas de probabilidade primeiramente é preciso definir o Espaço Amostral. Há 2 cores e 2 retiradas, então podemos ter:
- *a 1ª e a 2ª bolas brancas (2 bolas da mesma cor)- evento* $E1 = B1 \cap B2$;
- *a 1ª bola branca e a 2ª bola vermelha - evento* $E2 = B1 \cap V2$;
- *a 1ª bola vermelha e a 2ª bola branca - evento* $E3 = V1 \cap B2$;
- *a 1ª bola vermelha e a 2ª bola vermelha (2 bolas da mesma cor) - evento* $E4 = V1 \cap V2$.

Então o Espaço Amostral será:

$\Omega = \{ B1 \cap B2, B1 \cap V2, V1 \cap B2, V1 \cap V2\}$

*Todos os quatro eventos acima são **mutuamente exclusivos**: quando as bolas forem retiradas apenas um, e somente um, dos eventos acima pode ocorrer.*

*Qual o significado das retiradas SEM REPOSIÇÃO e COM REPOSIÇÃO? Se a retirada for feita SEM REPOSIÇÃO as retiradas serão **dependentes**, pois o Espaço Amostral será modificado: a cada retirada as probabilidades de ocorrência são modificadas porque as bolas não são repostas.*

Se a retirada for feita COM REPOSIÇÃO as retiradas são ***independentes****, pois o Espaço Amostral não será mudado porque as bolas retiradas são repostas antes da próxima extração.*

a) As retiradas são feitas SEM REPOSIÇÃO: a segunda retirada depende do resultado da primeira.

- A probabilidade de retirar bola branca na 1ª retirada é de 2/5(2 bolas brancas no total de 5), $P(B1) = 2/5$;

- A probabilidade de retirar bola vermelha na 1ª retirada é de 3/5 (3 bolas vermelhas em 5), $P(V1) = 3/5$.

Se a primeira bola retirada foi branca (o evento B1 ocorreu previamente), restaram 4 bolas, 1 branca e 3 vermelhas:

- *a probabilidade de retirar uma bola branca na 2ª retirada se na 1ª foi extraída uma branca é de 1/4 (1 bola branca em 4[7]),* $P(B2 | B1) = 1/4$.
- *a probabilidade de retirar uma bola vermelha na 2ª retirada se na 1ª foi extraída uma branca é de 3/4 (3 bolas vermelhas em 4),* $P(V2 | B1) = 3/4$.

Se a primeira bola retirada foi vermelha (o evento V1 *ocorreu previamente), restaram 4 bolas, 2 brancas e 2 vermelhas:*

- *a probabilidade de retirar uma bola branca na 2ª retirada se na 1ª foi extraída uma vermelha é de 2/4 (2 bolas brancas em 4),* $P(B2 | V1) = 2/4$.
- *a probabilidade de retirar uma bola vermelha na 2ª retirada se na 1ª foi extraída uma vermelha é de 2/4 (2 bolas vermelhas em 4),* $P(V2 | V1) = 2/4$.

a.1

O evento que nos interessa: "bolas da mesma cor": brancas OU vermelhas, evento UNIÃO brancas-vermelhas. Chamando bolas da mesma cor de evento F: $F = [(B1 \cap B2) \cup (V1 \cap V2)]$

Usando as propriedades de probabilidade:

$$P(F) = P[(B1 \cap B2) \cup (V1 \cap V2)] = P(B1 \cap B2) + P(V1 \cap V2) - P(B1 \cap B2) \cup (V1 \cap V2)$$

Os eventos $(B1 \cap B2)$ *e* $(V1 \cap V2)$ *são* ***mutuamente exclusivos****, se as bolas são da mesma cor ou são brancas ou são vermelhas, então a intersecção entre eles é o conjunto vazio, e a probabilidade do conjunto vazio ocorrer é igual a zero (ver definição axiomática de probabilidade), então simplesmente:*

$$P(F) = P[(B1 \cap B2) \cup (V1 \cap V2)] = P(B1 \cap B2) + P(V1 \cap V2)$$

Usando a regra do produto:

$P(B1 \cap B2) = P(B1) \times P(B2 | B1) = (2/5) \times (1/4) = 2/20 = 1/10$

$P(V1 \cap V2) = P(V1) \times P(V2 | V1) = (3/5) \times (2/4) = 6/20 = 3/10$

Substituindo na expressão:

$P(F) = P[(B1 \cap B2) \cup (V1 \cap V2)] = P(B1 \cap B2) + P(V1 \cap V2) = 1/10 + 3/10 = 4/10 = 0{,}4\ (40\%)$

Então se as retiradas forem feitas sem reposição a probabilidade de que as 2 bolas sejam da mesma cor será igual a 0,4 (40%).

a.2 - Neste caso sabe-se que as 2 bolas são da mesma cor (o evento F *acima JÁ OCORREU) e há interesse em saber a probabilidade de que as duas bolas sejam vermelhas:*

Marcelo Menezes Reis

---

[7] Repare que o número de bolas, número de resultados, diminuiu de 5 para 4 porque as retiradas são feitas SEM REPOSIÇÃO.

$$\mathbf{P[(V1 \cap V2)|\ F] = P\{(V1 \cap V2)\ |\ [(B1 \cap B2) \cup (V1 \cap V2)]\}}$$

Usando a expressão de probabilidade condicional:

$$\mathbf{P\{(V1 \cap V2)\ |\ [(B1 \cap B2) \cup (V1 \cap V2)]\}} = \frac{P\{(V1 \cap V2) \cap [(B1 \cap B2) \cup (V1 \cap V2)]\}}{P[(B1 \cap B2) \cup (V1 \cap V2)]}$$

A probabilidade do denominador já é conhecida do item a.1. E a do numerador pode ser obtida facilmente.
Repare: o que há em comum entre o evento $(V1 \cap V2)$ *e o evento* $\mathbf{[(B1 \cap B2) \cup (V1 \cap V2)]}$, *em suma qual será o evento intersecção? O que há em comum entre 2 bolas vermelhas e 2 bolas da mesma cor?* *<u>O próprio evento 2 bolas vermelhas</u>* $(V1 \cap V2)$, *então*:
$(V1 \cap V2) \cap \mathbf{[(B1 \cap B2) \cup (V1 \cap V2)]} = (V1 \cap V2)$;
$P\{(V1 \cap V2) \cap \mathbf{[(B1 \cap B2) \cup (V1 \cap V2)]\}} = P\ (V1 \cap V2) = 3/10$.

Sabendo que $P\{(V1 \cap V2)\ |\ \mathbf{[(B1 \cap B2) \cup (V1 \cap V2)]}\} = 4/10$ *(do item a.1) e substituindo os valores na fórmula:*

$$P\{(V1 \cap V2)\ |\ \mathbf{[(B1 \cap B2) \cup (V1 \cap V2)]}\} = \frac{P(V1 \cap V2)}{P[(B1 \cap B2) \cup (V1 \cap V2)]} = \frac{3/10}{4/10} = \frac{3}{4}$$

$P\{(V1 \cap V2)\ |\ \mathbf{[(B1 \cap B2) \cup (V1 \cap V2)]}\} = 0{,}75\ (75\%)$
Então se as retiradas forem feitas sem reposição, e as duas bolas forem da mesma cor, a probabilidade de que sejam vermelhas será igual a 0,75 (75%).

As retiradas e as probabilidades podem ser representadas através de um diagrama chamado de "Árvore de Probabilidades":

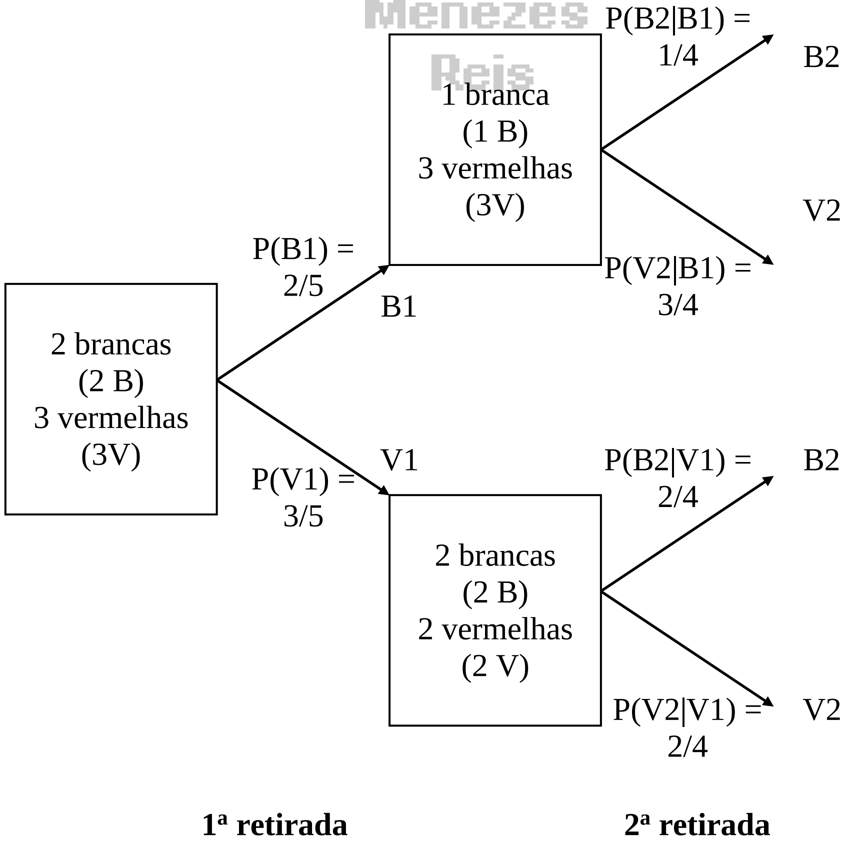


Figura 4 - Árvore de Probabilidades - Retiradas sem reposição

Observe que através da Árvore de Probabilidades podemos chegar aos mesmos resultados obtidos anteriormente. Partindo do Espaço Amostral original um dos ramos significa 1ª bola branca (B1) *e o outro 1ª bola vermelha* (V1)*. Dependendo do resultado da primeira retirada haverá um Espaço Amostral diferente: 1 bola branca e 3 vermelhas se na 1ª retirada obteve-se uma bola branca, ou 2 bolas brancas e 2 vermelhas se na 1ª retirada obteve-se uma bola vermelha. A partir dos novos Espaços Amostrais é possível calcular as probabilidades condicionais para cada caso, e depois substituí-las nas fórmulas adequadas. Contudo, a árvore será inútil se o evento para o qual se deseja calcular a probabilidade não for definido adequadamente: neste caso, no item a.1, bolas da mesma cor* $[(B1 \cap B2) \cup (V1 \cap V2)]$*, e no item a.2, bolas vermelhas sabendo que são da mesma cor* $\{(V1 \cap V2) \mid \mathbf{[(B1 \cap B2) \cup (V1 \cap V2)]}\}$*. A árvore será igualmente inútil se não forem usadas as definições de eventos dependentes (porque não há reposição) e de eventos mutuamente exclusivos (porque os eventos não podem ocorrer simultaneamente), e as expressões de probabilidade condicional e os axiomas de probabilidade.*

O grande inconveniente da Árvore de Probabilidades surge quando o número de "retiradas" aumenta e/ou o número de resultados possíveis para cada retirada é considerável: torna-se impraticável desenhar a Árvore, enumerando todos os resultados. Nestes casos usa-se Análise Combinatória, que veremos adiante. Vamos resolver agora o item b.

b) As retiradas são feitas COM REPOSIÇÃO: a segunda retirada ***independe*** *do resultado da primeira.*
- A probabilidade de retirar bola branca na 1ª ou na 2ª retirada é a mesma, 2/5 (2 bolas brancas no total de 5 sempre porque há reposição), $P(B1) = 2/5$, *e* $P(B2 \mid B1) = P(B2) = P(B1) = 2/5$, *e* $P(B2 \mid V1) = P(B2) = P(B1) = 2/5$.
- A probabilidade de retirar bola vermelha na 1ª ou na 2ª retirada é a mesma, 3/5 (3 bolas vermelhas no total de 5 sempre porque há reposição), $P(V1) = 3/5$, *e* $P(V2 \mid B1) = P(V2) = P(V1) = 3/5$, *e* $P(V2 \mid V1) = P(V2) = P(V1) = 3/5$.

b.1
O evento que nos interessa: "bolas da mesma cor": brancas OU vermelhas, evento UNIÃO brancas-vermelhas. Chamando bolas da mesma cor de evento F: $F = [(B1 \cap B2) \cup (V1 \cap V2)]$
Usando os axiomas de probabilidade:

$$P(F) = P\,[(B1 \cap B2) \cup (V1 \cap V2)] = P(B1 \cap B2) + P(V1 \cap V2) - P\,[(B1 \cap B2) \cap (V1 \cap V2)]$$

Os eventos $(B1 \cap B2)$ *e* $(V1 \cap V2)$ *são* ***mutuamente exclusivos****, se as bolas são da mesma cor ou são brancas ou são vermelhas, então a intersecção entre eles é o conjunto vazio, e a probabilidade do conjunto vazio ocorrer é igual a zero (ver definição axiomática de probabilidade), então simplesmente:*

$$P(F) = P\,[(B1 \cap B2) \cup (V1 \cap V2)] = P(B1 \cap B2) + P(V1 \cap V2)$$

Usando a regra do produto e lembrando que as retiradas são ***independentes****:*
$P(B1 \cap B2) = P(B1) \times P(B2 \mid B1) = P(B1) \times P(B2) = (2/5) \times (2/5) = 4/25$
$P(V1 \cap V2) = P(V1) \times P(V2 \mid V1) = P(V1) \times P(V2) = (3/5) \times (3/5) = 9/25$

Substituindo na expressão:
$P(F) = P\,[(B1 \cap B2) \cup (V1 \cap V2)] = P(B1 \cap B2) + P(V1 \cap V2) = 4/25 + 9/25 = 13/25 = 0{,}52\ (52\%)$

Então se as retiradas forem feitas com reposição a probabilidade de que as 2 bolas sejam da mesma cor será igual a 0,52 (52%)[8].

[8] Observe que o resultado é substancialmente DIFERENTE do caso em que as retiradas foram feitas sem reposição.

b.2 - Neste caso sabe-se que as 2 bolas são da mesma cor (o evento F *acima JÁ OCORREU) e há interesse em saber a probabilidade de que as duas bolas sejam vermelhas:*

$$P[(V1 \cap V2)|\,\mathbf{F}] = P\{(V1 \cap V2)\ |\ \mathbf{[(B1 \cap B2) \cup (V1 \cap V2)]}\}$$

Usando a expressão de probabilidade condicional:

$$P\{(V1 \cap V2)\ |\ \mathbf{[(B1 \cap B2) \cup (V1 \cap V2)]}\} = \frac{P\{(V1 \cap V2) \cap [(B1 \cap B2) \cup (V1 \cap V2)]\}}{P[(B1 \cap B2) \cup (V1 \cap V2)]}$$

A probabilidade do denominador já é conhecida do item b.1. E a do numerador pode ser obtida facilmente.

Repare: o que há em comum entre o evento $(V1 \cap V2)$ *e o evento* $\mathbf{[(B1 \cap B2) \cup (V1 \cap V2)]}$*, em suma qual será o evento intersecção? O que há em comum entre 2 bolas vermelhas e 2 bolas da mesma cor? O próprio evento 2 bolas vermelhas* $(V1 \cap V2)$*, então*:

$$(V1 \cap V2) \cap \mathbf{[(B1 \cap B2) \cup (V1 \cap V2)]} = (V1 \cap V2);$$

$$P\{(V1 \cap V2) \cap \mathbf{[(B1 \cap B2) \cup (V1 \cap V2)]}\} = P\,(V1 \cap V2) = 9/25.$$

Sabendo que $P\{(V1 \cap V2)\ |\ \mathbf{[(B1 \cap B2) \cup (V1 \cap V2)]}\} = 9/25$ *(do item b.1) e substituindo os valores na fórmula:*

$$P\{(V1 \cap V2)\ |\ \mathbf{[(B1 \cap B2) \cup (V1 \cap V2)]}\} = \frac{P(V1 \cap V2)}{P[(B1 \cap B2) \cup (V1 \cap V2)]} = \frac{9/25}{13/25} = \frac{9}{13}$$

$$P\{(V1 \cap V2)\ |\ \mathbf{[(B1 \cap B2) \cup (V1 \cap V2)]}\} = 0{,}69\ (69\%)$$

Então se as retiradas forem feitas com reposição, e as duas bolas forem da mesma cor, a probabilidade de que sejam vermelhas será igual a 0,69 (69%).

A Árvore de Probabilidades neste caso seria:

Marcelo Menezes Reis

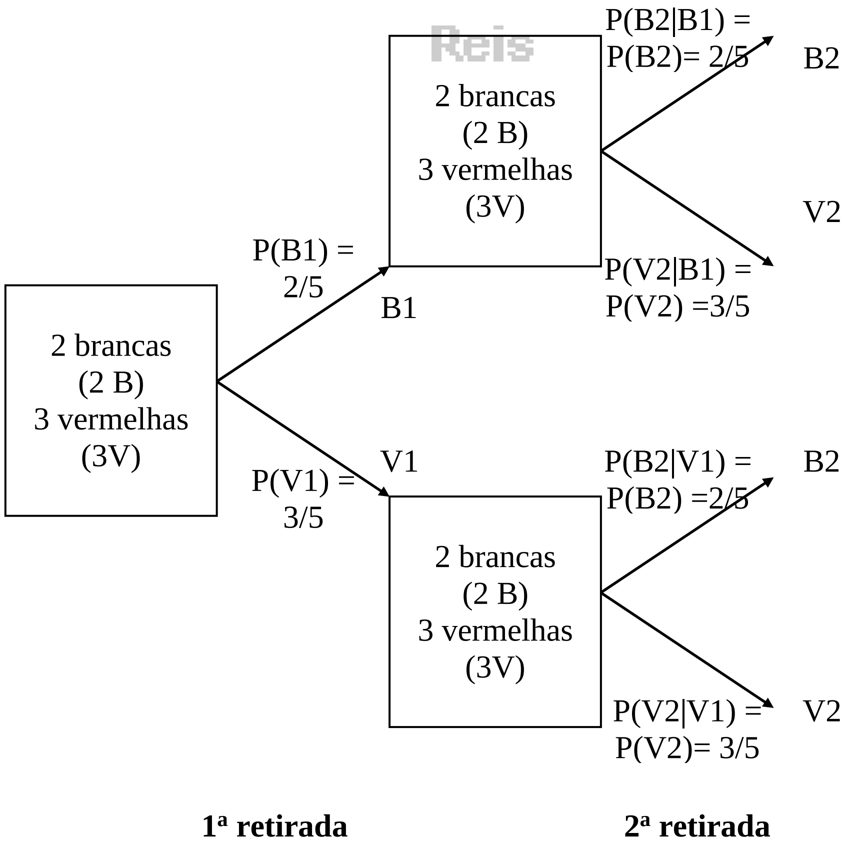


Figura 5 - Árvore de Probabilidades - Retiradas com reposição

Observe que como há reposição após cada retirada o Espaço Amostral não se altera, e todas as probabilidades permanecem constantes, independentemente dos resultados anteriores. O inconveniente da Árvore de probabilidades permanece porém.

8.4 - Probabilidade Combinatória

Como mencionado anteriormente, em muitos casos a resolução dos problemas de probabilidade enumerando todos os resultados possíveis torna-se extremamente difícil. Há uma forma mais rápida de enumerar os resultados: as técnicas de Análise Combinatória.

Relembrando a definição clássica de probabilidade que consistia em calcular o quociente entre o número de resultados associados ao evento e o número total de resultados possíveis. O cálculo desses números de resultados pode ser feito utilizando Análise Combinatória, tanto para os casos em que os eventos são dependentes quanto quando há independência.

As técnicas de Análise Combinatória buscam basicamente calcular o número de maneiras de dispor um certo número de "objetos" em um número limitado de "espaços" distintos (menor do que o número de objetos), sendo um objeto em cada espaço. Se o número de "objetos" é, teoricamente, infinito (ou ilimitado) temos a Análise Combinatória com Repetição Ilimitada (situação de **independência**): é o que ocorre nos casos em que há reposição. Se, porém, o número de "objetos" é limitado temos a Análise Combinatória sem Repetição (situação de **dependência**): casos em que não há reposição.

Marcelo Menezes Reis

8.4.1 - Análise Combinatória com Repetição Ilimitada

Há **n** objetos disponíveis em número ilimitado, em outras palavras há reposição, de quantas maneiras diferentes é possível preencher **k** espaços distintos com os objetos, cada espaço com um objeto?

Se temos um espaço e **n** objetos, há **n** maneiras de dispô-los no espaço. Se temos dois espaços, e os mesmos **n** objetos disponíveis para cada um haverá $\mathbf{n}^2$ maneiras: as **n** maneiras do primeiro espaço multiplicadas pelas **n** maneiras do segundo. Se houver três espaços, haverá $\mathbf{n}^3$ maneiras, e assim por diante.

Generalizando, se há **n** objetos disponíveis em número <u>ilimitado</u> para preencher **k** espaços distintos, cada espaço com um objeto, há $\mathbf{n}^{\mathbf{k}}$ maneiras de fazê-lo, e cada preenchimento é independente dos outros.

Exemplo 8.6 - Quantas palavras de 5 letras podem ser escritas com as 26 letras do alfabeto, sem se preocupar com o significado?
Primeiramente é preciso identificar os objetos e os espaços.
Os objetos neste caso são as letras do alfabeto, 26, então **n** = 26. *Como não há preocupação com o significado das palavras os objetos estão disponíveis em número <u>ilimitado</u>.*

Os espaços são as letras da palavra: cada palavra deve ter 5 letras, então **k = 5**.
Usando a expressão de Análise Combinatória com repetição ilimitada, o número de palavras será:

$$\mathbf{n}^{\mathbf{k}} = 26^5 = 11\ 881\ 376 \text{ palavras}$$

Exemplo 8.7 - Uma urna contém 2 bolas brancas e 3 vermelhas. Retiram-se ao acaso, uma após a outra, com reposição. Qual a probabilidade de que as 2 bolas sejam da mesma cor? Usar análise combinatória.

Este exemplo é uma repetição do item b.1 do Exemplo 8.5 visto anteriormente. Aqui chegaremos ao mesmo resultado usando Análise Combinatória.

O evento de nosso interesse: bolas da mesma cor = $F = [(B1 \cap B2) \cup (V1 \cap V2)]$. *Vimos que os eventos* $(B1 \cap B2)$ e $(V1 \cap V2)$ *são mutuamente exclusivos, então*:

$$P(F) = P[(B1 \cap B2) \cup (V1 \cap V2)] = P(B1 \cap B2) + P(V1 \cap V2)$$

Vamos calcular então as probabilidades necessárias.

$P(B1 \cap B2)$ = (N^o resultados para 2 bolas brancas) / (N^o total de resultados)

$P(V1 \cap V2)$ = (N^o resultados para 2 bolas vermelhas) / (N^o total de resultados)

Os denominadores serão os mesmos para os dois quocientes: há um total de 5 bolas ("objetos") disponíveis em número ilimitado (porque há reposição) para extrair em 2 retiradas ("espaços"), resultando **n** = 5 e **k** = 2, *então*:

$$N^o \text{ total de resultados} = \mathbf{n}^{\mathbf{k}} = 5^2 = 25$$

N^o de resultados para 2 bolas brancas: há um total de 2 bolas brancas ("objetos") disponíveis em número ilimitado (porque há reposição) para extrair em 2 retiradas ("espaços"), resultando **n** = 2 e **k** = 2, *então*:

$$N^o \text{ total de resultados} = \mathbf{n}^{\mathbf{k}} = 2^2 = 4$$

N^o de resultados para 2 bolas vermelhas: há um total de 3 bolas vermelhas ("objetos") disponíveis em número ilimitado (porque há reposição) para extrair em 2 retiradas ("espaços"), resultando **n** = 3 e **k** = 2, *então*:

$$N^o \text{ total de resultados} = \mathbf{n}^{\mathbf{k}} = 3^2 = 9$$

Então:

$P(B1 \cap B2)$ = (N^o resultados para 2 bolas brancas) / (N^o total de resultados) = 4 /25

$P(V1 \cap V2)$ = (N^o resultados para 2 bolas vermelhas) / (N^o total de resultados) = 9/25

Substituindo na fórmula:

$P(F) = P[(B1 \cap B2) \cup (V1 \cap V2)] = P(B1 \cap B2) + P(V1 \cap V2) = 4/25 + 9/25 = 13/25 = 0{,}52$ (52%)

Então se as retiradas forem feitas com reposição a probabilidade de que as 2 bolas sejam da mesma cor será igual a 0,52 (52%). Observe que é exatamente o mesmo resultado obtido no Exemplo 8.5.

Claro que para este caso extremamente simples (apenas 2 retiradas com 2 resultados possíveis em cada uma) o uso de Análise Combinatória não é necessário, mas permite chegar aos mesmos resultados que seriam obtidos com as técnicas anteriores. Se, porém, houver muitas retiradas e/ou muitas opções tornar-se-á indispensável.

Marcelo Menezes Reis

8.4.2 - Análise Combinatória sem Repetição

Continuam havendo **n** objetos para colocar em **k** espaços, mas os objetos não estão mais disponíveis em número ilimitado: não há repetição, ou não há reposição. A seleção de um dos objetos modifica a probabilidade de seleção dos outros: há <u>dependência</u>. Para calcular o número de maneiras possíveis de preencher os espaços é preciso relembrar os conceitos de Arranjos e Combinações.

Os Arranjos são utilizados para calcular o número de maneiras de dispor os **n** objetos nos **k** espaços quando a ORDEM E A NATUREZA dos objetos são importantes para o problema. O número de Arranjos de **n** objetos distintos tomados **k** a **k** será:

$$A_{n,k} = \frac{n!}{(n-k)!}$$ [9]

Exemplo 8.8 - Cinco carros, disputando os 3 primeiros lugares em uma corrida. Há quantas maneiras diferentes de classificá-los?
*Observe que há 5 objetos a dispor em 3 espaços, então **n** = 5 e **k** = 3. Os objetos não estão disponíveis em número ilimitado: uma vez definido o primeiro colocado ele não pode simultaneamente ocupar a terceira posição. Outro aspecto importante é que importam tanto a <u>ordem</u> quanto a <u>natureza</u> dos objetos: há diferença se o corredor A não chegar entre os 3 primeiros, mas também há diferença se o corredor chegar em primeiro ou segundo. Sendo assim serão usados Arranjos.*

$$A_{n,k} = \frac{n!}{(n-k)!} = \frac{5!}{(5-3)!} = \frac{5 \times 4 \times 3 \times 2!}{2!} = 60 \text{ maneiras.}$$

Então há 60 maneiras de classificar os 5 carros nos 3 primeiros lugares.

As Combinações são utilizadas para calcular o número de maneiras de dispor os **n** objetos nos **k** espaços quando apenas a NATUREZA dos objetos é importante para o problema. O número de Combinações de **n** objetos distintos tomados **k** a **k** será:

$$C_{n,k} = \frac{n!}{k! \times (n-k)!}$$

Exemplo 8.9 - De quantas maneiras diferentes podemos selecionar 3 dentre 5 pessoas para uma tarefa?
*Observe que novamente há 5 objetos a dispor em 3 espaços, então **n** = 5 e **k** = 3. Os objetos não estão disponíveis em número ilimitado: uma vez que uma pessoa seja selecionada não poderá novamente ser escolhida. Neste caso importa apenas a <u>natureza</u> dos objetos, apenas definir as pessoas que serão selecionados. Sendo assim serão usadas Combinações.*

$$C_{n,k} = \frac{n!}{k! \times (n-k)!} = \frac{5!}{3! \times (5-3)!} = \frac{5 \times 4 \times 3!}{3! \times 2!} = 10 \text{ maneiras.}$$

Então há 10 maneiras de selecionar 3 dentre 5 pessoas.

[9] n! significa fatorial de n: n x (n-1) x (n-2) x....x 1; lembrando que 0! = 1.

Exemplo 8.10 - Uma urna contém 18 bolas brancas, 15 vermelhas e 10 azuis. Serão retiradas X bolas, sem reposição, e observadas suas cores.
a) Seja X = 8 (oito bolas).Qual a probabilidade de que as bolas sejam da mesma cor?
b) Seja X = 6 Qual a probabilidade de que 2 sejam brancas, 2 sejam vermelhas e 2 sejam azuis?
Este problema seria extremamente trabalhoso para resolver usando uma Árvore de Probabilidades, por possuir várias retiradas com 3 resultados cada. Observe que não há reposição, portanto deve-se usar Análise Combinatória ***sem repetição****: repare que não há interesse na ordem das bolas retiradas (tanto no item a quanto no item b), mas apenas na cor das bolas (na sua "natureza"), sendo assim deve-se usar* *<u>combinações</u>* *para calcular o número de resultados necessários para calcular as probabilidades.*
a) Há uma grande quantidade de resultados possíveis para este problema, deve-se identificar o evento de interesse: 8 bolas da mesma cor. Neste caso 8 bolas brancas OU 8 bolas vermelhas OU 8 bolas azuis, evento UNIÃO 8 brancas com 8 vermelhas com 8 azuis. Chamando o evento 8 bolas da mesma cor de F: F = (8 brancas ∪ 8 vermelhas ∪ 8 azuis)
Observe que os 3 eventos acima são ***mutuamente exclusivos****: as 8 bolas retiradas não podem ser brancas e azuis SIMULTANEAMENTE. Então:*
P(F) = P (8 brancas ∪ 8 vermelhas ∪ 8 azuis) = P(8 brancas) + P(8 vermelhas) + P(8 azuis)

Para calcular as probabilidades dos eventos pode-se usar a definição clássica de probabilidade:
P(8 brancas) = (N^o. resultados para 8 brancas) / (N^o. total de resultados)
P(8 vermelhas) = (N^o. resultados para 8 vermelhas) / (N^o. total de resultados)
P(8 azuis) = (N^o. resultados para 8 azuis) / (N^o. total de resultados)

O denominador será o mesmo para todas as expressões. Há um total de 43 bolas (43 objetos, ***n*** *= 43) para colocar em 8 espaços (8 retiradas,* ***k*** *= 8), usando combinações:*

$$\text{N}^o\text{. total de resultados} = C_{n,k} = \frac{n!}{k! \times (n-k)!} = \frac{43!}{8! \times (43-8)!} = 145008513$$

Para as bolas brancas. Há 18 bolas brancas (18 objetos, ***n*** *= 18) para colocar em 8 espaços (8 retiradas,* ***k*** *= 8), usando combinações:*

$$\text{N}^o\text{. de resultados para 8 brancas} = C_{n,k} = \frac{n!}{k! \times (n-k)!} = \frac{18!}{8! \times (18-8)!} = 43758$$

Para as bolas vermelhas. Há 15 bolas vermelhas (15 objetos, ***n*** *= 15) para colocar em 8 espaços (8 retiradas,* ***k*** *= 8), usando combinações:*

$$\text{N}^o\text{. de resultados para 8 vermelhas} = C_{n,k} = \frac{n!}{k! \times (n-k)!} = \frac{15!}{8! \times (15-8)!} = 6435$$

Para as bolas azuis. Há 10 bolas azuis (10 objetos, ***n*** *= 10) para colocar em 8 espaços (8 retiradas,* ***k*** *= 8), usando combinações:*

$$\text{N}^o\text{. de resultados para 8 azuis} = C_{n,k} = \frac{n!}{k! \times (n-k)!} = \frac{10!}{8! \times (10-8)!} = 45$$

Substituindo os valores diretamente na fórmula geral:

P(F) = P(8 brancas) + P(8 vermelhas) + P(8 azuis)

$$P(F) = \frac{43758}{145008513} + \frac{6435}{145008513} + \frac{45}{145008513} = 0{,}000346$$

Marcelo Menezes Reis

Arredondando, a probabilidade de que as 8 bolas retiradas sejam da mesma cor é igual a 0,0003 (0,03%)[10].

b) *Neste caso há interesse em calcular a probabilidade de que 2 bolas sejam brancas E 2 sejam vermelhas E 2 sejam azuis, evento INTERSECÇÃO 2 brancas com 2 vermelhas com 2 azuis. Chamando este evento de* G: G = (2 brancas $\cap$ 2 vermelhas $\cap$ 2 azuis).
Para os casos de intersecção o cálculo do número de resultados associados precisa ser feito da seguinte forma: os números de resultados possíveis associados a cada "sub-evento" componente devem ser multiplicados para obter o número de resultados da intersecção. ISSO, PORÉM, NÃO SIGNIFICA QUE OS EVENTOS SEJAM INDEPENDENTES!
P(G) = (N^o. res. 2 brancas x N^o. res. 2 vermelhas x N^o. res. 2 azuis)/ (N^o. total de resultados)

*Há um total de 43 bolas (43 objetos, **n** = 43) para colocar em 6 espaços (6 retiradas, **k** = 6), usando combinações:*

$$N^o\text{. total de resultados} = C_{n,k} = \frac{n!}{k!\times(n-k)!} = \frac{43!}{6!\times(43-6)!} = 6096454$$

*N^o. res. 2 brancas: há 18 bolas brancas (18 objetos, **n** = 18) para colocar em 2 espaços (2 retiradas, **k** =2), usando combinações:*

$$N^o\text{. de resultados para 2 brancas} = C_{n,k} = \frac{n!}{k!\times(n-k)!} = \frac{18!}{2!\times(18-2)!} = 153$$

*N^o. res. 2 vermelhas: há 15 bolas vermelhas (15 objetos, **n** = 15) para colocar em 2 espaços (2 retiradas, **k** =2), usando combinações:*

$$N^o\text{. de resultados para 2 vermelhas} = C_{n,k} = \frac{n!}{k!\times(n-k)!} = \frac{15!}{2!\times(15-2)!} = 105$$

*N^o. res. 2 azuis: há 10 bolas azuis (10 objetos, **n** = 10) para colocar em 2 espaços (2 retiradas, **k** =2), usando combinações:*

$$N^o\text{. de resultados para 2 azuis} = C_{n,k} = \frac{n!}{k!\times(n-k)!} = \frac{10!}{2!\times(10-2)!} = 45$$

Substituindo na fórmula de P(G):

P(G) = (153 x 105 x 45)/ (6096454) = 0,11858

Arredondando, a probabilidade de que 2 bolas sejam brancas e 2 vermelhas e 2 azuis é igual a 0,12 (12%).

[10] Este valor tão baixo era esperado devido à quantidade de bolas e ao número total de combinações possíveis.

Marcelo Menezes Reis

8.5 - Variáveis Aleatórias

Uma pergunta que é normalmente feita a todos que trabalham com ciências exatas: "por que a obsessão em reduzir tudo a números"? Vimos em Análise Exploratória de Dados que uma variável QUANTITATIVA (intervalar) geralmente[11] apresenta mais informação que uma variável qualitativa (nominal ou ordinal), pode ser resumida não somente através de tabelas e gráficos mas também através de medidas de síntese, e possibilita a realização de previsões por meio de uma equação de regressão.

Nos exemplos anteriores sobre probabilidade os eventos foram geralmente definidos de forma **verbal**: bolas da mesma cor, 2 bolas vermelhas, soma das faces menor ou igual a 5, etc. Não haveria problema em definir os eventos através de **números**.

Variáveis Aleatórias são **funções** matemáticas que associam números reais aos resultados de um Espaço Amostral associado a um Experimento Aleatório. Se o Espaço Amostral for finito ou infinito numerável[12] a variável aleatória é dita **discreta**. Se o Espaço Amostral for infinito a variável aleatória é dita **contínua**.

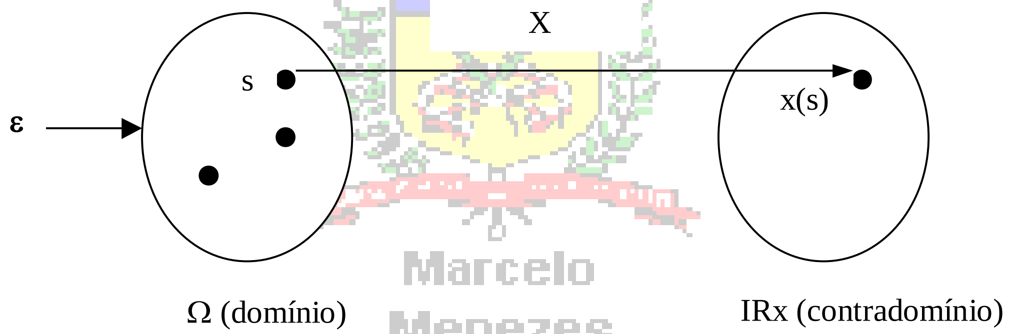


Figura 6 - Variável aleatória

Por exemplo, imaginemos o Experimento Aleatório jogar uma moeda honesta duas vezes e observar a face voltada para cima. O Espaço Amostral seria finito:

$$\Omega = \{CaraCara; CaraCoroa; CoroaCara; CoroaCoroa\}$$

Se houvesse interesse no número de caras obtidas, poderia ser definida uma variável aleatória discreta X, onde X = Número de caras em dois lançamentos. Os valores possíveis de X seriam:

$$X = \{0, 1, 2\}$$

O valor 0 é associado ao evento CoroaCoroa, o valor 1 é associado aos eventos CaraCoroa e CoroaCara, e o valor 2 é associado ao evento CaraCara.

Quando o Espaço Amostral é infinito muitas vezes já está definido de forma numérica, facilitando a definição da variável aleatória.

Os Modelos Probabilísticos são construídos para as variáveis aleatórias: assim haverá Modelos Probabilísticos Discretos e Modelos Probabilísticos Contínuos. Para construir um modelo probabilístico para uma variável aleatória é necessário definir os seus possíveis valores (contradomínio), e como a probabilidade total (do Espaço Amostral, que vale 1) *distribui-se* entre

[11] Geralmente porque NEM TUDO pode ser reduzido a números. Exemplos contundentes: inteligência e criatividade.

[12] Não há limite "superior", mas é possível identificar os incrementos entre os valores e o "início" do Espaço Amostral. Por exemplo, o número de acidentes ocorridos em uma rodovia em um mês: sabe-se que o menor valor é 0, mas não se sabe, antes do mês terminar, qual será o número máximo; sabe-se também que o incremento será igual a 1, haverá 0, 1, 2 ou mais acidentes (não há meio acidente).

eles: é preciso então definir a **distribuição de probabilidades**. Dependendo do tipo de variável aleatória haverá diferenças na construção da distribuição.

8.5.1 – Distribuição de probabilidades para Variáveis Aleatórias Discretas

Quando uma variável aleatória X é discreta (contradomínio finito ou infinito numerável), a construção da distribuição de probabilidades consiste em definir o conjunto de pares [xi, p(xi)], onde xi é o i-ésimo valor da variável X, e p(xi) é a probabilidade de ocorrência de xi, como na tabela abaixo:

X = xi	p(X = xi)
x1	p(x1)
x2	p(x2)
...	...
xn	p(xn)

Onde $p(xi) \geq 0$ e $\sum_{i=1}^{n} p(x_i) = 1{,}0$

EX.8.11 Construir a distribuição de probabilidades do número de acertos em 3 tentativas de cesta, sabendo que: P (encestar / encestado) = 0,6 P (encestar / não encestado) = 0,3

A variável aleatória X, número de acertos em três tentativas, é uma variável aleatória discreta: o seu contradomínio é finito, o jogador pode acertar 0, 1, 2 ou 3 vezes. Mas, para calcular as probabilidades associadas a esses valores é preciso estabelecer todos os eventos possíveis, pois mais de um evento contribui para as probabilidades de 1 e 2 acertos. Observando a árvore de eventos abaixo (onde A é acertar a cesta e E significa errar).

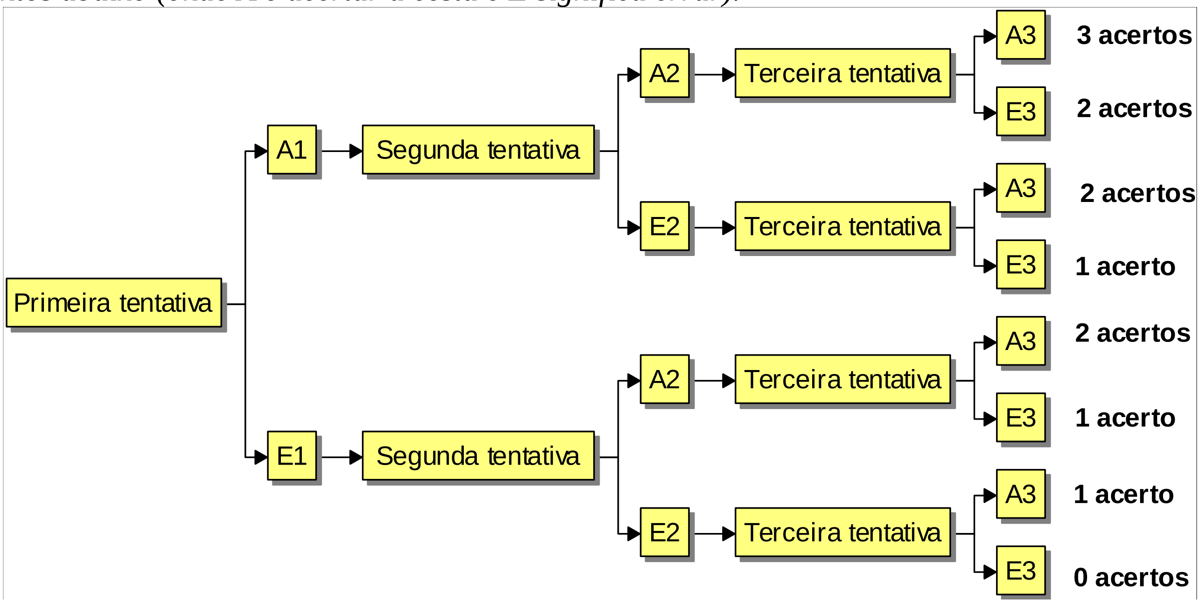


Figura 7 - Árvore de eventos

Observe que todos os eventos são mutuamente exclusivos, o jogador não pode, na mesma seqüência de 3 cestas, errar E acertar a primeira. É preciso explicitar os valores da variável, e os eventos em termos de teoria dos conjuntos.

Contradomínio: IRx = {0, 1, 2, 3} acertos. A equivalência entre os valores da variável e os eventos é estabelecida abaixo:

$X = 0 \Leftrightarrow [E1 \cap E2 \cap E3]$ $\quad X = 1 \Leftrightarrow [(A1 \cap E2 \cap E3) \cup (E1 \cap A2 \cap E3) \cup (E1 \cap E2 \cap A3)]$

$X = 2 \Leftrightarrow [(A1 \cap A2 \cap E3) \cup (E1 \cap A2 \cap A3) \cup (A1 \cap E2 \cap A3)]$ $\quad X = 3 \Leftrightarrow [A1 \cap A2 \cap A3]$

Então:

$P(X=0) = P[E1 \cap E2 \cap E3]$ $P(X=1) = P[(A1 \cap E2 \cap E3) \cup (E1 \cap A2 \cap E3) \cup (E1 \cap E2 \cap A3)]$

$P(X=2) = P[(A1 \cap A2 \cap E3) \cup (E1 \cap A2 \cap A3) \cup (A1 \cap E2 \cap A3)]$ $P(X=3) = P[A1 \cap A2 \cap A3]$

Assume-se que na primeira tentativa o jogador tem 50% de chance de acertar[13]*, então*

$P(A1) = 0{,}5$ e $P(E1) = 0{,}5$

Além disso estabeleceu-se que quando o jogador acertou a cesta na tentativa anterior a probabilidade de acertar a próxima é de 0,6, e caso tenha errado na anterior a probabilidade de acertar na próxima é de apenas 0,3. Tratam-se de duas probabilidades condicionais, estabelecidas em função de eventos já ocorridos.

Se o jogador acertou na tentativa i (qualquer uma), as probabilidades de acertar e errar na próxima tentativa serão:

$P(A_{i+1}|A_i) = 0{,}6$ *Pelo complementar obtém-se* $P(E_{i+1}|A_i) = 0{,}4$

Se o jogador errou na tentativa i, as probabilidades de acertar e errar na próxima tentativa serão:

$P(A_{i+1}|E_i) = 0{,}3$ *Pelo complementar obtém-se* $P(E_{i+1}|E_i) = 0{,}7$

Com estas probabilidades estabelecidas, lembrando da regra do produto, do teorema da probabilidade total, e considerando o fato de que os eventos são mutuamente exclusivos é possível calcular as probabilidades de ocorrência de cada valor da variável aleatória X.

$$P(X=0) = P[E1 \cap E2 \cap E3] = P(E1) \times P(E2|\, E1) \times P(E3|\, E1 \cap E2)$$

Como os resultados em uma tentativa só dependem daqueles obtidos na imediatamente anterior, o terceiro termo da expressão acima pode ser simplificado para P(E3| E2), e a probabilidade será:

$$P(X=0) = P(E1) \times P(E2|\, E1) \times P(E3|\, E2) = 0{,}5 \times 0{,}7 \times 0{,}7 = 0{,}245\ (24{,}5\%)$$

Estendendo o procedimento acima para os outros valores:

$$P(X=1) = P[(A1 \cap E2 \cap E3) \cup (E1 \cap A2 \cap E3) \cup (E1 \cap E2 \cap A3)]$$

$$P(X=2) = P[(A1 \cap A2 \cap E3) \cup (E1 \cap A2 \cap A3) \cup (A1 \cap E2 \cap A3)]$$

$$P(X=3) = P[A1 \cap A2 \cap A3]$$

Como os eventos são mutuamente exclusivos:

$$P(X=1) = P(A1 \cap E2 \cap E3) + P(E1 \cap A2 \cap E3) + P(E1 \cap E2 \cap A3)$$
$$P(X=1) = P(A1) x P(E2|A1) x\ P(E3|E2) + P(E1) x\ P(A2|E1) \times P(E3|A2) + P(E1) \times P(E2|E1) \times P(A3|E2)$$
$$P(X=1) = 0{,}5 x\ 0{,}4 \times 0{,}7 + 0{,}5 x\ 0{,}3 \times 0{,}4 + 0{,}5 \times 0{,}7 \times 0{,}3 = 0{,}305$$

$$P(X=2) = P(A1 \cap A2 \cap E3) + P(E1 \cap A2 \cap A3) + P(A1 \cap E2 \cap A3)$$
$$P(X=2) = P(A1) x P(A2|A1) x\ P(E3|A2) + P(E1) x\ P(A2|E1) \times P(A3|A2) + P(A1) x\ P(E2|A1) x\ P(A3|E2)$$
$$P(X = 2) = 0{,}5 \times 0{,}6 \times 0{,}4 + 0{,}5 \times 0{,}3 \times 0{,}6 + 0{,}5 \times 0{,}4 \times 0{,}3 = 0{,}27\ (27\%)$$

$$P(X=3) = P[A1 \cap A2 \cap A3] = P(A1) \times P(A2|A1) \times P(A3|A2) = 0{,}5 \times 0{,}6 \times 0{,}6 = 0{,}18\ (18\%)$$

[13] E1, errar a primeira cesta, é o evento complementar de A1, acertar a primeira cesta.

Com os valores calculados acima é possível construir a tabela com os pares valores-probabilidades.

X	p(X = xi)
0	0,245
1	0,305
2	0,270
3	0,180
Total	1,0

8.5.2 – Distribuição de probabilidades para Variáveis Aleatórias Contínuas

Uma variável aleatória contínua tem contradomínio infinito. Assim, a probabilidade de que a variável assuma exatamente um valor xi é zero, não havendo mais sentido em representar a distribuição pelos pares xi – p(xi). Utiliza-se então uma função, a função densidade de probabilidades, definida para todos os valores possíveis da variável aleatória: para calcular a probabilidade de uma variável aleatória contínua assumir valores entre **a** e **b** (dois valores quaisquer), basta calcular a integral[14] da função no intervalo de interesse. Em muitas situações de nosso interesse tais probabilidades podem ser calculadas através de fórmulas matemáticas relativamente simples, ou foram dispostas em tabelas, que são encontradas em praticamente todos os livros de estatística, e que serão vistas em seções posteriores deste texto.

Uma função densidade de probabilidade poderia ser apresentada graficamente da seguinte forma:

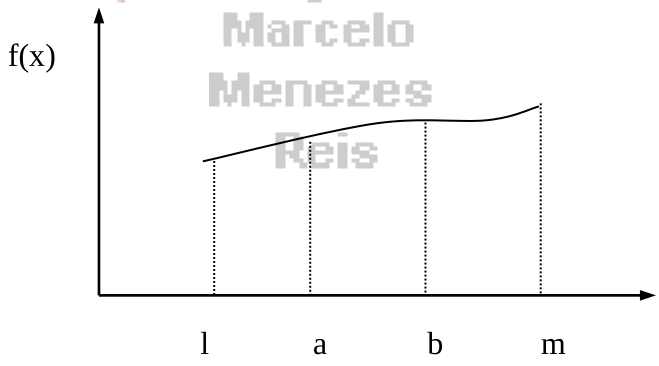


Figura 8 - Função densidade de probabilidades

Onde l e m são os limites da função de densidade de probabilidade (para valores menores do que l e maiores do que m a função vale zero).

8.5.3 – Valor Esperado e Variância

Todos os modelos probabilísticos apresentam duas medidas (dois momentos) que permitem caracterizar a variável aleatória para a qual eles foram construídos: o Valor Esperado e a Variância da variável aleatória. O Valor Esperado (simbolizado por E(X)) nada mais é do que a **média** vista em Análise Exploratória de Dados, utilizando *probabilidades* ao invés de *freqüências* no cálculo. Analogamente, a Variância (simbolizada por V(X)) é a **variância** vista anteriormente, utilizando *probabilidades*. Da mesma forma que em Análise Exploratória de Dados é também comum trabalhar com o Desvio Padrão, raiz quadrada positiva da Variância (que aqui será simbolizado por

[14] Serão necessários conhecimentos de cálculo integral.

Marcelo Menezes Reis

σ(X), "sigma de X"). A interpretação dos resultados obtidos pode ser feita de forma semelhante à Análise Exploratória de Dados, apenas recordando que se trata de uma *variável aleatória*, e estão sendo usadas probabilidades e não freqüências.

Para uma variável aleatória discreta o valor esperado e a variância podem ser calculados da seguinte forma:

$$E(X) = \sum_{i=1}^{n} x_i \times p(x_i) \qquad V(X) = E\left(X^2\right) - [E(X)]^2 \quad \text{onde} \quad E\left(X^2\right) = \sum_{i=1}^{n} x_i^2 \times p(x_i)$$

Para uma variável aleatória contínua a obtenção do valor esperado e da variância exigem o cálculo de integrais das funções de densidade de probabilidades. Para as distribuições mais importantes as equações encontram-se disponíveis nos livros de estatística, em função dos parâmetros da distribuição, e algumas serão vistas nas seções posteriores deste texto.

O valor esperado (média) e a variância apresentam algumas propriedades.

Para o valor esperado E(X), sendo k uma constante:
a) $E(k) = k$ b) $E(k \pm X) = k \pm E(X)$ c) $E(k \times X) = k \times E(X)$
d) $E(X \pm Y) = E(X) \pm E(Y)$
e) Sejam X e Y duas variáveis aleatórias **independentes** $E(X \times Y) = E(X) \times E(Y)$

Para a variância V(X), sendo k uma constante:
a) $V(k) = 0$ b) $V(k \pm X) = V(X)$ c) $V(k \times X) = k^2 \times V(X)$
d) Sejam X e Y duas variáveis aleatórias **independentes** $V(X \pm Y) = V(X) \pm V(Y)$

EX.8.12 Calcular o valor esperado e a variância da distribuição do EX.8.11.
Para uma variável aleatória discreta é aconselhável acrescentar mais uma coluna à tabela com os valores e probabilidades, para poder calcular o valor de E(X^2):

X	p(X = xi)	xi × p(X = xi)	xi^2 × p(X = xi)
0	0,245	0	0
1	0,305	0,305	0,305
2	0,270	0,540	1,08
3	0,180	0,540	1,62
Total	1,0	1,385	3,005

Substituindo nas expressões de valor esperado e variância:

$$E(X) = \sum_{i=1}^{n} x_i \times p(x_i) = 1{,}385 \text{ cestas}$$

$$V(X) = \sum_{i=1}^{n} x_i^2 \times p(x_i) - \left[\sum_{i=1}^{n} x_i \times p(x_i)\right]^2 = 3{,}005 - (1{,}385)^2 = 1{,}087 \text{ cestas}^2$$

$$\sigma(X) = \sqrt{V(X)} = \sqrt{1{,}087} = 1{,}042 \text{ cestas}$$

*Observe que o valor esperado (1,385 cestas) é um valor que a variável aleatória **não pode assumir**! Não é o "valor mais provável", é o ponto de equilíbrio do conjunto. Repare que a*

Marcelo Menezes Reis

unidade da variância dificulta sua comparação com o valor esperado, mas ao se utilizar o desvio padrão é possível verificar que a dispersão dos resultados é quase do valor da média (valor esperado).

Nas próximas seções estudaremos várias distribuições de probabilidade (modelos probabilísticos) que são extremamente úteis para modelar muitas situações práticas, auxiliando na tomada de decisões. Na figura a seguir podemos ver algumas[15] das distribuições de probabilidade para variáveis aleatórias discretas e contínuas (em verde as que serão vistas neste Capítulo, e em vermelho as que serão vistas no Capítulo 9).

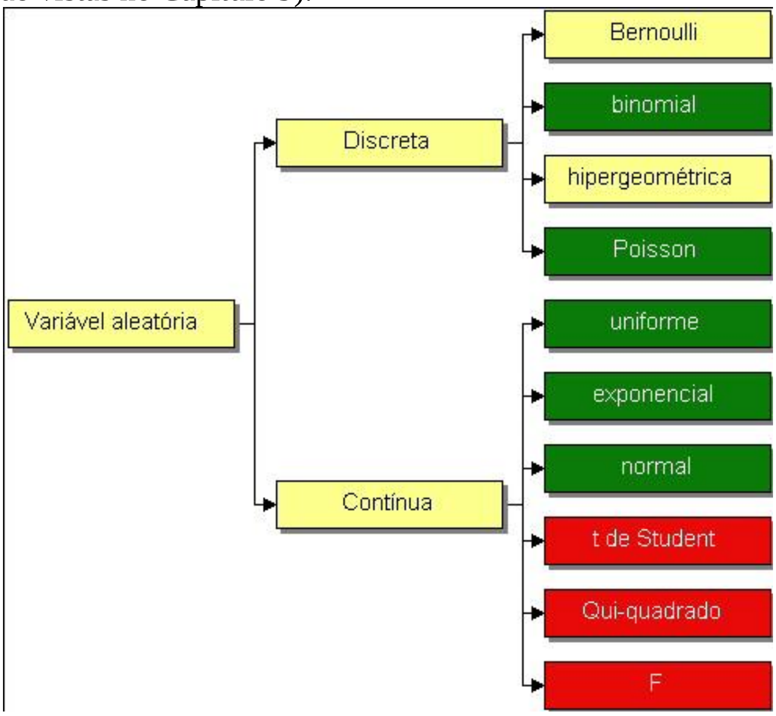


Figura 9 - Modelos probabilísticos

8.6 - Modelo Binomial (Distribuição Binomial)

Seja um Experimento Aleatório qualquer que apresenta as seguintes características:
- consiste na realização de um número ***finito*** e ***conhecido*** **n** de ensaios (ou repetições);
- cada um dos ensaios tem ***apenas dois*** resultados possíveis: "sucesso" ou "fracasso" (estão entre aspas porque a definição de sucesso não quer necessariamente algo "positivo", e também porque poderá incluir significar um grupo de resultados).
- os ensaios são ***independentes*** entre si, apresentando probabilidades de "sucesso" (**p**) e de "fracasso" (**1-p**) ***constantes***.

Neste caso estamos interessados no número de "sucessos" obtidos nos **n** ensaios: como o Espaço Amostral é finito (vai de 0 a **n**) uma variável aleatória associada seria <u>discreta</u>. Este tipo de experimento é chamado de BINOMIAL.

Então, a variável aleatória discreta X, número de "sucessos" nos **n** ensaios, apresenta uma distribuição binomial com os seguintes parâmetros:

n = número de ensaios **p = probabilidade de "sucesso"**

Com esses dois parâmetros é possível calcular as probabilidades de um determinado número de sucessos, bem como obter o Valor Esperado e a Variância da variável X[16]:

[15] Há muitas outras.
[16] O "x" nas expressões é o sinal de MULTIPLICAÇÃO.

$$E(X) = n \times p \qquad V(X) = n \times p \times (1 - p)$$

Exemplo 8.13 - Experimentos binomiais:
a) Observar o número de caras em 3 lançamentos imparciais de uma moeda honesta: n=3; p=0,5
b) Observar o número de meninos nascidos em 3 partos de uma família: n=3; p = x
c) Observar o número de componentes defeituosos em uma amostra de 10 componentes de um grande número de peças que apresentaram anteriormente 10% de defeituosos: n = 10; p= 0,1

Vamos ver com maiores detalhes o caso do número de meninos (e meninas) nascidos em uma família. Chamando menino de evento H, será o "sucesso", e menina de evento M, e sabendo pela história da família que P(H) = 0,52 e P(M) = 0,48 (então **p** = 0,52 e **1- p** = 0,48), quais serão as probabilidades obtidas para a variável aleatória número de meninos em 3 nascimentos?
Resolvendo usando os conceitos gerais de probabilidade é preciso primeiramente determinar o Espaço Amostral, como poderão ser os sexos das 3 crianças:

$$\mathbf{\Omega = \{H\cap H\cap H,\ H\cap H\cap M,\ H\cap M\cap H,\ M\cap H\cap H,\ H\cap M\cap M,\ M\cap H\cap M,\ M\cap M\cap H,\ M\cap M\cap M\}}$$

Supondo que os nascimentos sejam independentes, podemos calcular as probabilidades de cada intersecção simplesmente multiplicando as probabilidades individuais de seus componentes:
$P\{H\cap H\cap H\} = P(H) \times P(H) \times P(H) = p \times p \times p = p^3$
$P\{H\cap H\cap M\} = P(H) \times P(H) \times P(M) = p \times p \times (1- p) = p^2 \times (1 - p)$
$P\{H\cap M\cap H\} = P(H) \times P(M) \times P(H) = p \times (1 - p) \times p = p^2 \times (1 - p)$
$P\{M\cap H\cap H\} = P(M) \times P(H) \times P(H) = (1 - p) \times p \times p = p^2 \times (1 - p)$
$P\{H\cap M\cap M\} = P(H) \times P(M) \times P(M) = p \times (1 - p) \times (1 - p) = p \times (1 - p)^2$
$P\{M\cap H\cap M\} = P(M) \times P(H) \times P(M) = (1 - p) \times p \times (1 - p) = p \times (1 - p)^2$
$P\{M\cap M\cap H\} = P(M) \times P(M) \times P(H) = (1 - p) \times (1 - p) \times p = p \times (1 - p)^2$
$P\{M\cap M\cap M\} = P(M) \times P(M) \times P(M) = (1- p) \times (1 - p) \times (1 - p) = (1 - p)^3$

Observe que:
$P\{H\cap H\cap M\} = P\{H\cap M\cap H\} = P\{M\cap H\cap H\} = p^2 \times (1 - p)$ = Prob. de 2 "sucessos"
$P\{H\cap M\cap M\} = P\{M\cap H\cap M\} = P\{M\cap M\cap H\} = p \times (1 - p)^2$ = Prob. de 1 "sucesso"

Importa apenas a "natureza" dos sucessos, não a ordem em que ocorrem: com a utilização de combinações é possível obter o número de resultados iguais para cada número de sucessos. Supondo que o número de ensaios **n** é o número de "objetos" disponíveis, e que o número de "sucessos" em que estamos interessados (doravante chamado **k**) é o número de "espaços" onde colocar os objetos (um objeto por espaço), o número de resultados iguais será:

$$C_{n,k} = \frac{n!}{k! \times (n-k)!}$$

Para o caso acima, em que há 3 ensaios (**n** =3):
- para 2 sucessos (**k** =2) $C_{3,2} = \dfrac{3!}{2! \times (3-2)!} = 3$ (o mesmo resultado obtido por enumeração)
- para 1 sucesso (**k** =1) $C_{3,1} = \dfrac{3!}{1! \times (3-1)!} = 3$ (o mesmo resultado obtido por enumeração)

O procedimento acima poderia ser feito para quaisquer valores de **n** e **k** (desde que **n** > **k**), permitindo obter uma expressão geral para calcular a probabilidade associada a um resultado qualquer.

A probabilidade de uma variável aleatória discreta X, número de sucessos em **n** ensaios, com distribuição binomial de parâmetros **n** e **p**, assumir um certo valor **k** (0 ≤ **k** ≤ **n**) será:

$$P(X=k)=C_{n,k}\times p^{k}\times(1-p)^{n-k} \qquad \text{onde} \qquad C_{n,k}=\frac{n!}{k!\times(n-k)!}$$

É importante lembrar que a probabilidade de ocorrer **k** sucessos é igual à probabilidade de ocorrer **n - k** fracassos, e que todos os axiomas de probabilidade continuam válidos.

Exemplo 8.14 - Admitamos que a probabilidade de que companhia não entregue seus produtos no prazo é igual a 18%. Quais são as probabilidades de que em 3 entregas 1, 2 ou todas as 3 entregas sejam feitas no prazo. Calcular também valor esperado, variância e desvio padrão do número de entregas no prazo.

Para cada entrega ("ensaio") há apenas dois resultados: no prazo ou não. Há um número limitado de realizações, **n** = 3. *Definindo "sucesso" como no prazo, e supondo as operações independentes, a variável aleatória X, número de entregas no prazo em 3 terá distribuição binomial com parâmetros*

n = 3 e **p** = 0,82 (*e* **1- p** = 0,18).

Então:

$$P(X=0)=C_{3,0}\times 0{,}82^{0}\times(0{,}18)^{3}=\frac{3!}{0!\times(3-0)!}\times 0{,}82^{0}\times(0{,}18)^{3}=0{,}006$$

$$P(X=1)=C_{3,1}\times 0{,}82^{1}\times(0{,}18)^{2}=\frac{3!}{1!\times(3-1)!}\times 0{,}82^{1}\times(0{,}18)^{2}=0{,}080$$

$$P(X=2)=C_{3,2}\times 0{,}82^{2}\times(0{,}18)^{1}=\frac{3!}{2!\times(3-2)!}\times 0{,}82^{2}\times(0{,}18)^{1}=0{,}363$$

$$P(X=3)=C_{3,3}\times 0{,}82^{3}\times(0{,}18)^{0}=\frac{3!}{3!\times(3-3)!}\times 0{,}82^{3}\times(0{,}18)^{0}=0{,}551$$

Somando todas as probabilidades o resultado é igual a 1, como teria que ser[17]*. O Valor Esperado, Variância e o Desvio Padrão serão:*

$$E(X)=n\times p=3\times 0{,}82=2{,}46\ \text{entregas}$$

$$V(X)=n\times p\times(1-p)=3\times 0{,}82\times 0{,}18=0{,}4428\ \text{entregas}^{2}.$$

$$\sigma(X)=\sqrt{V(X)}=\sqrt{0{,}4428}=0{,}665\ \text{entregas}$$

A média é quase igual ao número de operações devido à alta probabilidade de sucesso.

Exemplo 8.15- Estudos anteriores mostraram que há 73% de chance de consumidores do sexo feminino apresentar uma reação positiva a anúncios publicitários com crianças. Uma agência está conduzindo um estudo, apresentando um novo anúncio para 5 consumidoras. Qual é a probabilidade

[17] Lembre-se que a soma das probabilidades de TODOS os eventos que compõem o Espaço Amostral é igual a 1. E que 0! = 1, e que um número diferente de 0 elevado a zero é igual a 1.

Marcelo Menezes Reis

de que pelo menos 3 das 5 consumidores apresentem reação positiva? Calcular também o valor esperado, variância e desvio padrão do número de consumidoras que apresentam reação positiva.
Para cada consumidora ("ensaio") há apenas dois resultados: reação positiva ou não. Há um número limitado de realizações, **n** = 5. *Definindo "sucesso" como reação positiva, e supondo as consumidoras "independentes", a variável aleatória X, número de consumidoras com reação positiva em 5 que assistiram o novo anúncio terá distribuição binomial com parâmetros*
n = 5 e **p** = 0,73 (*e* **1- p** = 0,27).
O evento de interesse é a recuperação de pelo menos 3 ratos (3 ou mais): **P(X ≥ 3).**

P(X ≥ 3) = P(X = 3) + P(X = 4) + P(X = 5)

É preciso calcular as três probabilidades acima e somá-las, então:

$$P(X=3) = C_{5,3} \times 0{,}73^3 \times (0{,}27)^2 = \frac{5!}{3! \times (5-3)!} \times 0{,}73^3 \times (0{,}27)^2 = 0{,}284$$

$$P(X=4) = C_{5,4} \times 0{,}73^4 \times (0{,}27)^1 = \frac{5!}{4! \times (5-4)!} \times 0{,}73^4 \times (0{,}27)^1 = 0{,}383$$

$$P(X=5) = C_{5,5} \times 0{,}73^5 \times (0{,}27)^0 = \frac{5!}{5! \times (5-5)!} \times 0{,}73^5 \times (0{,}27)^0 = 0{,}207$$

P(X ≥ 3) = P(X = 3) + P(X = 4) + P(X = 5) = 0,284 + 0,383 + 0,207 = 0,874

A probabilidade de que pelo menos 3 das 5 consumidoras apresentem reação positiva é igual a 0,874 (87,4%).

Há duas outras formas de chegar ao mesmo resultado:

- *através do complementar:* **P(X ≥ 3) = 1 - P(X<3) = 1 - [P(X=0) + P(X=1) + P(X=2)]**
- *mudando a definição de sucesso, de reação positiva para* <u>*reação negativa*</u> (**p** = 0,27), *se pelo menos 3 consumidoras apresentam reação positiva então no máximo 2 apresentam reação negativa.*

Marcelo Menezes Reis

8.7 - Modelo de Poisson (Distribuição de Poisson)

Vamos supor um experimento "binomial", com apenas dois resultados possíveis, mas com uma das seguintes características:
1) O valor da probabilidade de sucesso **p** é muito pequeno, significando que o sucesso é um evento raro (implicando geralmente em um alto valor de **n**).
2) A probabilidade de sucesso **p** não é constante, sendo relacionada ao número de ensaios **n**: quanto maior **n**, menor **p**.
3) Situação em que apesar da probabilidade **p** ser constante o valor de **n** teoricamente é infinito.

Nas três situações acima o modelo binomial não proporcionará bons resultados (caso 1) ou mesmo não poderá ser utilizado (casos 2 e 3). Nestes casos deve ser utilizado o modelo de **Poisson**.

Como seria a solução para os casos acima?

Casos 1 e 2 - se os valores de **n** e **p** variam (ou são muito discrepantes) talvez fosse melhor usar uma quantidade constante[18] para analisar o problema, como o Valor Esperado E(X), que será chamado de **m**.

$$E(X) = n \times p = m$$

Caso 3 - como **n** é "infinito" deve-se fazer a análise das ocorrências em um período contínuo (de tempo, de espaço, etc.) subdividido em um certo número de subintervalos (número tal que a probabilidade de existir mais de uma ocorrência em uma subdivisão é desprezível, e supondo ainda que as ocorrências em subdivisões diferentes são independentes); novamente é preciso trabalhar com uma quantidade constante que será chamada de **m** também:

$$m = \lambda \times t$$

onde $\boldsymbol{\lambda}$ é uma taxa de ocorrência do evento em um período contínuo (igual ou diferente do período sob análise), e **t** é justamente o período contínuo sob análise[19].

Se uma variável aleatória discreta X, número de ocorrências de um evento, segue a distribuição de Poisson, a probabilidade de X assumir um valor **k** será:

$$P(X = k) = \frac{e^{-m} \times m^k}{k!}$$

Onde **e** é uma constante: $\mathbf{e} \cong 2{,}71$. E $m = n \times p$ ou $m = \lambda \times t$.

Uma particularidade interessante da distribuição de Poisson é que o Valor Esperado e a Variância de uma variável aleatória que siga tal distribuição serão iguais:

$$E(X) = m = n \times p \quad \text{ou} \quad E(X) = m = \lambda \times t$$

$$V(X) = m = n \times p \quad \text{ou} \quad V(X) = m = \lambda \times t$$

Exemplo 8.16 - Experimentos e fenômenos que seguem a distribuição de Poisson:

a) Número mensal de acidentes de tráfego em um cruzamento.

Observe que é uma variável aleatória discreta, pode assumir apenas valores inteiros (0, 1, 2, 3,...). Cada realização do "experimento" (acidente) pode ter apenas 2 resultados: ocorre o acidente ou não ocorre o acidente. Mas, o número máximo de realizações é desconhecido! Assim, a distribuição binomial não pode ser usada, e a análise do número de acidentes precisa ser feita em um período contínuo (no caso, período de tempo, 1 mês), exigindo o uso da distribuição de POISSON.

b) Número de itens defeituosos produzidos por hora em uma indústria.

Novamente, uma variável aleatória discreta (valores inteiros: 0,1, 2, 3, ...), cada realização só pode ter dois resultados possíveis (peça sem defeito ou peça defeituosa). Se o número máximo de realizações for conhecido, provavelmente a probabilidade de uma peça ser defeituosa será reduzida e apesar de ser possível a utilização da distribuição binomial o uso da distribuição de POISSON obterá resultados muitos próximos. Se o número máximo de realizações for desconhecido a distribuição binomial não pode ser usada, e a análise do número de acidentes precisa ser feita em um período contínuo (no caso, período de tempo, 1 hora), exigindo o uso da distribuição de POISSON.

c) Desintegração dos núcleos de substâncias radioativas: contagem do número de pulsações radioativas a intervalos de tempo fixos.

Situação semelhante a dos acidentes em um cruzamento, só que o "grau de aleatoriedade" deste experimento é muito maior. O número máximo de pulsações também é desconhecido, obrigando a realizar a análise em um período contínuo, utilizando a distribuição de POISSON.

[18] Se **n** e **p** estão relacionados, ao se aumentar **n**, **p** diminui, mas o produto **n** x **p** permanece constante.

[19] Apesar do símbolo t, o período contínuo NÃO É NECESSARIAMENTE um intervalo de tempo.

Marcelo Menezes Reis

Exemplo 8.17 - As estatísticas mostram que dentre os clientes de mais de 35 anos e menos de 45 anos há 0,12% de probabilidade de ocorrência de mal de Alzheimer. Qual é a probabilidade de que dentre 3000 clientes exatamente 3 apresentem a doença?
Cada cliente pode apresentar ou não a doença (apenas 2 resultados possíveis para cada realização). Definindo "sucesso" como apresentar a doença, podemos definir a variável aleatória ***X*** *como o número de sucessos em 3000 realizações (clientes). Observe que o número máximo de realizações é* <u>*conhecido*</u> *(3000) e que a probabilidade de sucesso é bastante pequena. Como não há nada que nos indique o contrário os clientes são supostos independentes.*
Com as condições acima podemos usar a distribuição binomial para calcular a probabilidade de ocorrência de 3 sucessos:

$P(X = k) = C_{n,k} \times p^k \times (1-p)^{n-k}$ **n = 3000** **p = 0,0012** **k = 3**

$$P(X = 3) = \frac{3000!}{3! \times (3000-3)!} \times (0{,}0012)^3 \times (1-0{,}0012)^{3000-3} = 0{,}2126$$

*Este é o resultado "**exato**".*

Observe que este problema apresenta uma das situações em que seria possível a utilização da distribuição de POISSON: ***p*** *muito pequena e* ***n*** *grande. Neste caso o valor esperado da distribuição de POISSON (igualado ao da binomial) seria:*

$$\mathbf{m} = \mathbf{n}\mathrm{x}\mathbf{p} = 3000\mathrm{x}0{,}0012 = 3{,}6$$

E a probabilidade de ocorrência de 3 sucessos, usando a distribuição de POISSON:

$$P(X = k) = \frac{e^{-m} \times (m)^k}{k!} = \frac{e^{-3{,}6} \times (3{,}6)^3}{3!} = 0{,}2125$$

Observe como o resultado é próximo do valor "exato", comprovando a eficácia da aproximação (em alguns casos o valor de ***n*** *é tal que o cálculo da combinação extrapola a capacidade dos meios disponíveis, nestes casos a solução é fazer o cálculo através da distribuição de POISSON).*

Exemplo 8.18 - Uma telefonista recebe cerca de 0,20 chamadas por minuto (valor obtido de medições anteriores).
a) Qual é a probabilidade de receber exatamente 5 chamadas nos primeiros 10 minutos?
b) Qual é a probabilidade de receber até 2 chamadas nos primeiros 12 minutos?
c) Qual é o desvio padrão do número de chamadas em meia hora?
Há interesse no número de chamadas ocorridas em um período contínuo (de tempo no caso). Para cada "ensaio" há apenas dois resultados possíveis: a chamada ocorre ou não. Observe que não há um limite para o número de chamadas no período (sabe-se apenas que o número mínimo pode ser 0): por esse motivo a utilização da binomial é inviável... Contudo há uma taxa de ocorrência ($\boldsymbol{\lambda}$ *= 0,20 chamadas/minuto) e isso permite utilizar a distribuição de Poisson.*

a) Neste caso o período ***t*** *será igual a 10 minutos (**t** = 10 min.), e há interesse em P(X = 5).*

$$m = \lambda \times t = 0{,}20 \times 10 = 2 \text{ chamadas}$$

$$P(X = k) = \frac{e^{-m} \times m^k}{k!} = P(X = 5) = \frac{e^{-2} \times 2^5}{5!} = 0{,}0361$$

Então a probabilidade de que a telefonista receba exatamente 5 chamadas em 10 minutos é igual a 0,0361 (3,61%).

b) Neste caso o período ***t*** *será igual a 12 minutos (**t** = 12 minutos). O evento de interesse é até 2 chamadas em 12 minutos (*$X \le 2$*).*

$$m = \lambda \times t = 0{,}20 \times 12 = 2{,}4 \text{ chamadas}$$

P(X ≤ 2) = P(X = 0) + P(X = 1) + P(X = 2)

$$P(X=k)=\frac{e^{-m}\times m^{k}}{k!}=P(X=0)=\frac{e^{-2,4}\times 2,4^{0}}{0!}=0,0907$$

$$P(X=k)=\frac{e^{-m}\times m^{k}}{k!}=P(X=1)=\frac{e^{-2,4}\times 2,4^{1}}{1!}=0,2177$$

$$P(X=k)=\frac{e^{-m}\times m^{k}}{k!}=P(X=2)=\frac{e^{-2,4}\times 2,4^{2}}{2!}=0,2613$$

P(X ≤ 2) = P(X = 0) + P(X = 1) + P(X = 2) = 0,0907 + 0,2177 + 0,2613 = 0,5697

Então a probabilidade de que a telefonista receba até 2 chamadas em 12 minutos é igual a 0,5697 (56,97%).

*c) Neste caso o período **t** será igual a 30 minutos (**t** = 30 minutos). Primeiro calcula-se a variância:*

$$V(X)=m=\lambda\times t=0,2\times 30=6\ \text{chamadas}^{2}$$

O Desvio Padrão é a raiz quadrada positiva da variância:

$$\sigma(X)=\sqrt{V(X)}=\sqrt{6}\cong 2,45\,\text{chamadas}$$

8.8 – Modelo Uniforme (Distribuição Uniforme)

Quando o Espaço Amostral associado a um Experimento Aleatório é infinito torna-se necessário o uso de uma Variável Aleatória Contínua para associar números reais aos resultados. Os modelos probabilísticos vistos anteriormente não podem ser empregados: a probabilidade de que uma variável aleatória contínua assuma EXATAMENTE um determinado valor é **zero**.

Para entender melhor a declaração acima vamos relembrar a definição clássica de probabilidade: a probabilidade de ocorrência de um evento será igual ao quociente entre o número de resultados associados ao evento pelo número total de resultados possíveis. Ora, se o número total de resultados é infinito, ou tende ao infinito para ser mais exato, a probabilidade de ocorrência de um valor específico é igual a zero. Por esse motivo, quando se lida com Variáveis Aleatórias Contínuas calcula-se a probabilidade de ocorrência de eventos formados por **intervalos** de valores. Uma outra conseqüência disso é que os símbolos > e ≥ (< e ≤ também) são equivalentes para variáveis aleatórias contínuas.

Vamos ver uma definição do prof. Pedro Barbetta (Barbetta, 2002):

"A distribuição de probabilidades de uma variável aleatória contínua pode ser representada por uma função não negativa, com a área formada entre o eixo das abcissas e a curva desta função igual a 1 (probabilidade total do Espaço Amostral): a função densidade de probabilidades, vista na seção 8.5.2. Os eventos podem ser representados por intervalos nos eixos das abcissas (eixo X), enquanto as correspondentes probabilidades por áreas sob a curva".

Seja uma variável aleatória contínua qualquer X que possa assumir valores entre A e B. Todos os valores entre A e B têm a mesma probabilidade de ocorrer, resultando no gráfico abaixo:

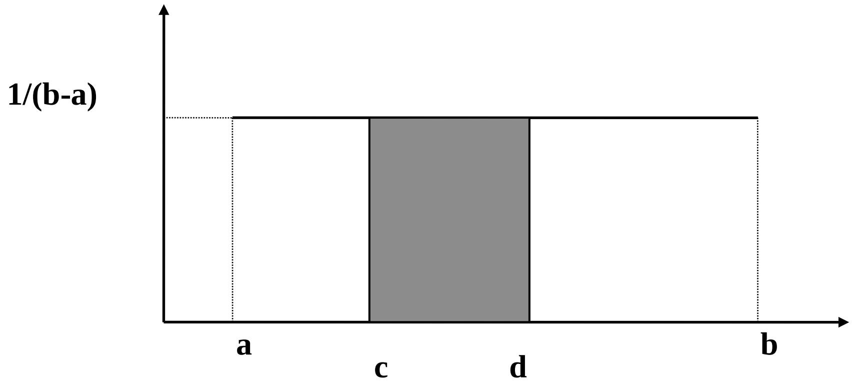


Figura 10 - Distribuição de probabilidades para uma variável aleatória contínua

Para que a área entre a e b seja igual a 1 o valor da ordenada precisa ser igual a **1/(b - a)**. A área escura representa a probabilidade da variável X assumir valores no intervalo **c** - **d**. Trata-se do modelo uniforme.

Intuitivamente podemos supor que muitas variáveis aleatórias contínuas terão um comportamento diferente do caso acima: em algumas delas haverá maior probabilidade de ocorrências de valores próximos ao limite inferior ou superior, etc.: para cada caso deverá ser ajustado um modelo probabilístico contínuo adequado.

O modelo uniforme é provavelmente o mais simples modelo probabilístico para variáveis aleatórias contínuas, mas que encontra várias aplicações práticas. Dois intervalos de valores da variável aleatória contínua, que tenham o mesmo tamanho, tem a mesma probabilidade de ocorrer (desde que dentro da faixa de valores para os quais a função de densidade de probabilidades não é nula). Formalmente, uma variável aleatória contínua X tem distribuição uniforme, com parâmetros **a** e **b** reais (sendo **a** menor do que **b**), se sua função densidade de probabilidades for tal como a da figura 10.

Para calcular a probabilidade de que a variável assuma valores entre c e d (sendo a < c < d < b), basta calcular a área compreendida entre c e d:

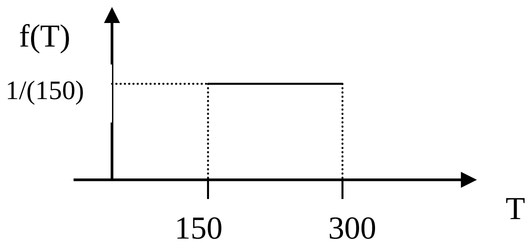
$$P(c < X < d) = (d - c) \times \frac{1}{(b - a)}$$

Seu valor esperado e variância são:

$$E(X) = \frac{a + b}{2} \qquad V(X) = \frac{(b - a)^2}{12}$$

EX.8.19 A temperatura T de destilação do petróleo é crucial para determinar a qualidade final do produto. Suponha que T seja considerada uma variável aleatória contínua com distribuição uniforme de 150 a 300°C. Suponha que o custo para produzir um galão de petróleo seja de 50 u.m.. Se o óleo é destilado a menos de 200°C, o galão é vendido a 75 u.m., se a temperatura for superior a 200°C, o produto é vendido a 100 u.m..
a) Fazer o gráfico da função densidade de probabilidade de T.
b) Qual é o lucro médio esperado por galão?

a) Os parâmetros **a** *e* **b** *definem completamente uma distribuição uniforme, para fazer o gráfico basta encontrá-los no enunciado acima. Identifica-se que o limite inferior,* **a**, *vale 150º C, e o superior,* **b**, *vale 300º C, resultando no gráfico a seguir:*



b) A variável aleatória de interesse, lucro, é discreta, somente pode assumir dois valores: 25 u.m. (caso o óleo seja destilado a menos de 200º C, posto que o galão custa 50 u.m. para ser produzido e será vendido a 75 u.m. nestas condições), ou 50 u.m. (caso o óleo seja destilado a mais de 200º C, posto que o galão custa 50 u.m. para ser produzido e será vendido a 100 u.m.). Sendo assim seu contradomínio será: $IR_{lucro} = \{25, 50\}$ *sendo os resultados mutuamente exclusivos.*

Lembrando das definições de distribuições de probabilidades, e de valor esperado e variância para variáveis aleatórias discretas (itens 8.5.2 e 8.5.3), para obter o lucro médio (valor esperado da variável lucro), é preciso obter as probabilidades de ocorrência dos seus dois valores (25 e 50). Relacionando com os valores de T:

$$\mathbf{P(Lucro = 25) = P(T \leq 200)} \qquad \mathbf{P(Lucro = 50) = P(T > 200)}$$

Marcelo Menezes Reis

Os valores das probabilidades acima correspondem às áreas abaixo da curva da função densidade de probabilidades para cada intervalo, calculando as áreas:

$$P(T \leq 200) = (200 - 150) \times \frac{1}{(300-150)} = \frac{50}{150} \qquad P(T > 200) = (300 - 200) \times \frac{1}{(300-150)} = \frac{100}{150}$$

Então a distribuição de probabilidades da variável lucro será:

Lucro	Probabilidade
25	50/150
50	100/150
Total	1,0

Calculando o valor esperado:

$$\mathbf{E(Lucro) = \sum Lucro_i \times P(Lucro_i)}$$

$$\mathbf{E(Lucro) = 25 \times \frac{50}{150} + 50 \times \frac{100}{150} = 41{,}67} \textit{ u.m.}$$

O lucro médio é de 41,67 u.m.. Repare que a variável lucro NÃO PODE assumir este valor, o que significa que o valor esperado (a média) NÃO É o valor mais provável. Neste problema o valor mais provável, a moda (ver Capítulo 2), vale 50 u.m., pois tem a maior probabilidade de ocorrência (66,67%).

8.9 – Modelo Exponencial (Distribuição Exponencial[20])

O modelo exponencial tem uma forte relação com o modelo de Poisson. A distribuição de Poisson modelava o número de ocorrências em um período contínuo (de tempo, de comprimento, de

[20] Exponencial negativa.

área, de volume). A ***distância*** entre estas ocorrências (seja medida em minutos, metros, metros quadrados) também é uma variável aleatória, mas agora contínua, que pode ser modelada pela distribuição exponencial.

Formalmente, "uma variável aleatória contínua X que é igual à distância entre contagens sucessivas de um processo que segue uma distribuição de Poisson, cuja média vale λ, segue uma distribuição exponencial com parâmetro λ". Sua função densidade de probabilidades será:

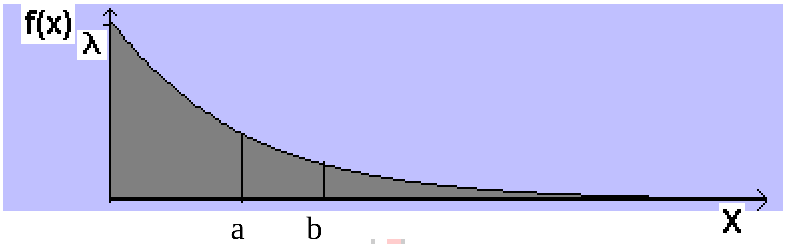


Figura 11 - Função densidade de probabilidade - Distribuição Exponencial

Para calcular a probabilidade de que a variável X assuma valores entre a e b é preciso a utilização de cálculo integral. Contudo, vamos apresentar apenas os resultados, bastando que o leitor substitua o parâmetro da distribuição exponencial, e os valores de interesse nas equações.

$$P(a < X < b) = 1 - e^{-\lambda \times b} - (1 - e^{-\lambda \times a}) = e^{-\lambda \times a} - e^{-\lambda \times b}$$

$$P(X \le a) = 1 - e^{-\lambda \times a} \quad P(X \le b) = 1 - e^{-\lambda \times b} \quad P(X \ge a) = e^{-\lambda \times a} \quad P(X \ge b) = e^{-\lambda \times b}$$

Onde e é uma constante, que vale aproximadamente 2,71

Lembrando que λ é uma constante positiva, que representa uma taxa de ocorrência (uma taxa de falha, número de falhas a cada 1000 horas, uma taxa de saída, número de saídas a cada 10 minutos). O valor esperado e a variância da distribuição exponencial são:

$$E(X) = \frac{1}{\lambda} \qquad V(X) = \frac{1}{\lambda^2}$$

EX.8.20 Certo componente eletrônico apresenta uma média de 500 horas de tempo T de vida útil, e pressupõe-se que T siga uma distribuição exponencial. Qual é a probabilidade de que T seja maior do que a média?

A variável aleatória contínua T (tempo de vida em horas do componente) segue uma distribuição exponencial, mas com qual parâmetro λ? Sabe-se também que a média (valor esperado) do tempo vale 500 horas, então como o valor esperado de uma distribuição exponencial vale $1/\lambda$:

$$\lambda = \frac{1}{E(T)} = \frac{1}{500}$$

Há interesse em obter a probabilidade de que o tempo de vida (T) seja maior do que a média (500 horas). Essa probabilidade poderia servir como base para a determinação de um prazo de garantia, por exemplo. O evento de interesse então é T > 500, usando as fórmulas vistas anteriormente:

$$P(T > 500) = e^{-\lambda \times 500} = e^{-\frac{1}{500} \times 500} = e^{-1} = 0{,}3679$$

Conclui-se então que a probabilidade de que o tempo de vida seja maior do que 500 horas é igual a 0,3679 (36,79%). Se fosse estabelecido um prazo de garantia de 500 horas para os transistores,

Marcelo Menezes Reis

isto é transistores que falhassem em até 500 horas seriam substituídos gratuitamente, o fabricante teria um grande prejuízo, pois apenas 36,79% duram mais do que 500 horas.

8.10 - Modelo Normal (Distribuição Normal, Distribuição de De Moivre - Laplace - Gauss ou Distribuição gaussiana).

Há casos em que há maior probabilidade de ocorrência de valores situados em intervalos “centrais” da função densidade de probabilidades da variável aleatória contínua, e esta probabilidade diminui à medida que os valores se afastam deste centro (para valores menores ou maiores) o modelo probabilístico contínuo mais adequado seja o modelo Normal ou gaussiano[21]. Isso é especialmente encontrado em variáveis biométricas, resultantes de medidas corpóreas em seres vivos.

8.10.1 - Características do Modelo Normal

O Modelo Normal é extremamente adequado para medidas numéricas em geral, descrevendo vários fenômenos, e permitindo fazer aproximações de modelos discretos. É extremamente importante também para a Estatística Indutiva (mais detalhes no próximo capítulo). O gráfico da distribuição de probabilidades de uma variável aleatória contínua que siga o modelo Normal (distribuição Normal) será como a figura abaixo:

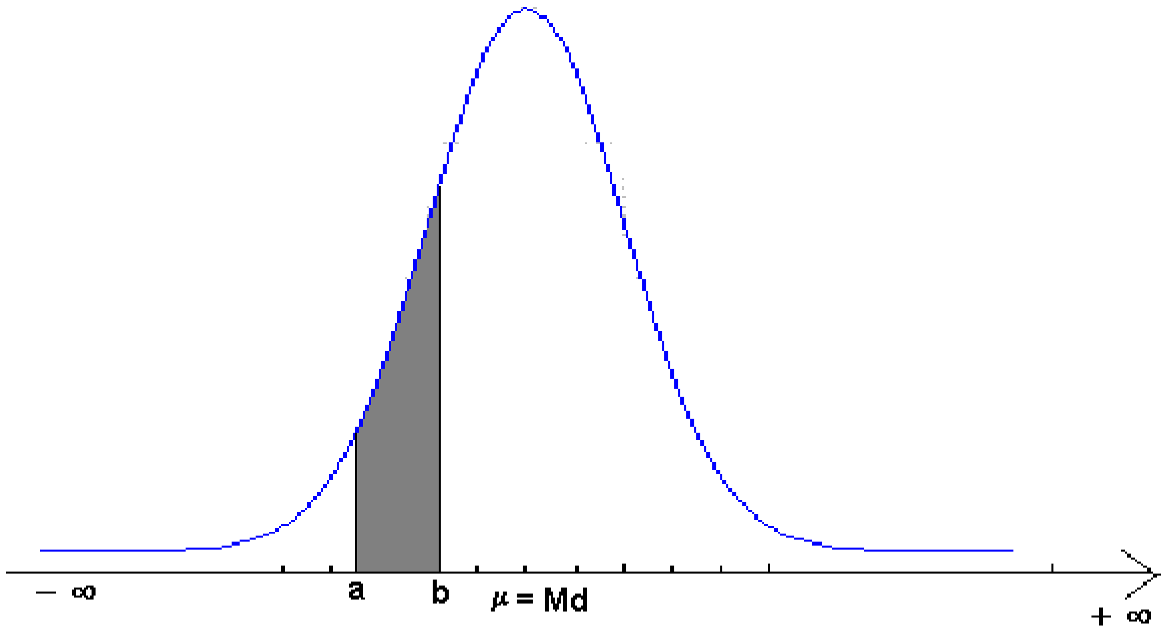


Figura 12 - Distribuição Normal

Características:
- a curva apresenta forma de sino, há maior probabilidade da variável assumir valores próximos do centro.
- os valores de média (μ) e de mediana (**Md**) são iguais, significando que a curva é SIMÉTRICA em relação à média.
- teoricamente a curva prolonga-se de - ∞ a + ∞ , então a área total sob a curva é igual a 1 (100%).

[21] O matemático alemão Gauss utilizou amplamente este modelo no tratamento de erros experimentais, embora não tenha sido o seu “descobridor”.

- qualquer distribuição normal é perfeitamente especificada por seus parâmetros média (μ) e variância (σ^2)[22] => **X**: **N** (μ , σ^2) significa que a variável X tem distribuição normal com média μ e variância σ^2.
- a área escura na figura 7 é a probabilidade de uma variável que siga a distribuição normal assumir valores entre **a** e **b**: esta área é calculada através da **integral** da função Normal de **a** a **b**.
- cada combinação (μ , σ^2) resulta em uma distribuição Normal diferente, portanto há uma família infinita de distribuições.
- a função Normal citada acima tem a seguinte (e aterradora...) fórmula para sua função densidade de probabilidade:

$$f(x) = \frac{1}{\sqrt{2 \times \pi \times \sigma^2}} \times e^{\left(\frac{-1}{2} \times \left[\frac{x-\mu}{\sigma}\right]^2\right)} \qquad -\infty < x < +\infty$$

NÃO EXISTE solução analítica para uma integral da expressão acima: qualquer integral precisa ser resolvida usando métodos numéricos de integração, que são extremamente trabalhosos quando implementados manualmente (somente são viáveis se usarem meios computacionais). Gauss desenvolveu seu trabalho entre o fim do século XVIII e início do século XIX, e os computadores começaram a se popularizar a partir da década de 60, do século XX...[23]

Porém todas as distribuições normais apresentam algumas características em comum, independentemente de seus valores de média e de variância:
- 68% dos dados estão situados entre a média menos um desvio padrão (**μ - σ**) e a média mais um desvio padrão (**μ + σ**);
- 95,5% dos dados estão situados entre a média menos dois desvios padrões (**μ - 2σ**) e a média mais dois desvios padrões (**μ + 2σ**);
- 99,7% dos dados estão situados entre a média menos três desvios padrões (**μ - 3σ**) e a média mais três desvios padrões (**μ + 3σ**).

Marcelo Menezes

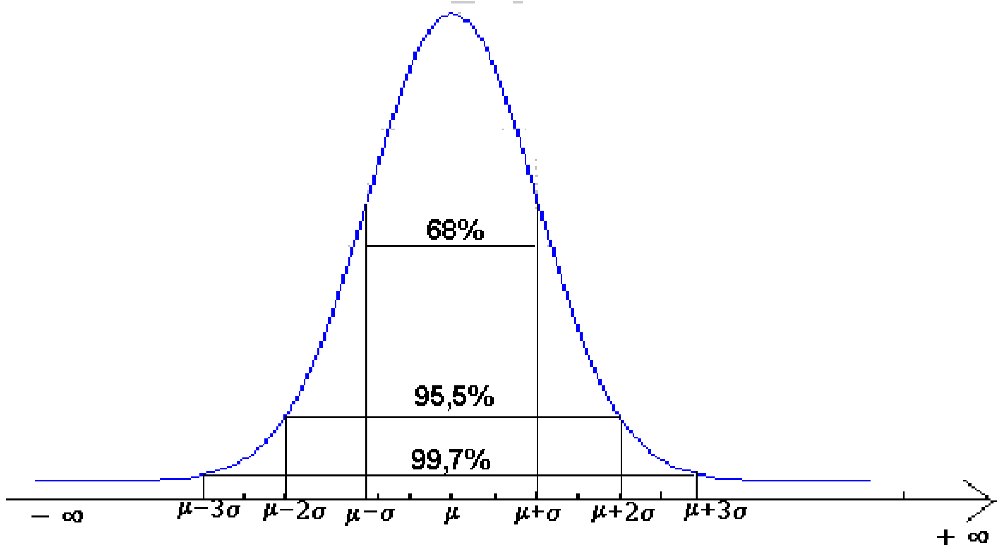


Figura 13 - Percentuais de dados e números de desvios padrões

Por causa dessas características alguém teve a idéia de criar uma distribuição Normal

[22] É comum a utilização de letras do alfabeto grego para representar algumas medidas. Não se esqueça que o desvio padrão (σ) é a raiz quadrada positiva da variância.
[23] Gauss, e todas as outras pessoas que usavam a distribuição Normal para calcular probabilidades até recentemente, resolviam as integrais usando métodos numéricos MANUALMENTE.

padrão: uma variável Z com distribuição normal de média igual a zero e desvio padrão igual a 1 [Z: **N** (**0** , **1**)]. As probabilidades foram calculadas para esta distribuição padrão e registradas em uma tabela. Através de uma transformação de variáveis é possível converter os valores de qualquer distribuição Normal em valores da distribuição Normal padrão e assim obter suas probabilidades - calcular o número de desvios padrões, a contar da média a que está um valor da variável, através da seguinte expressão:

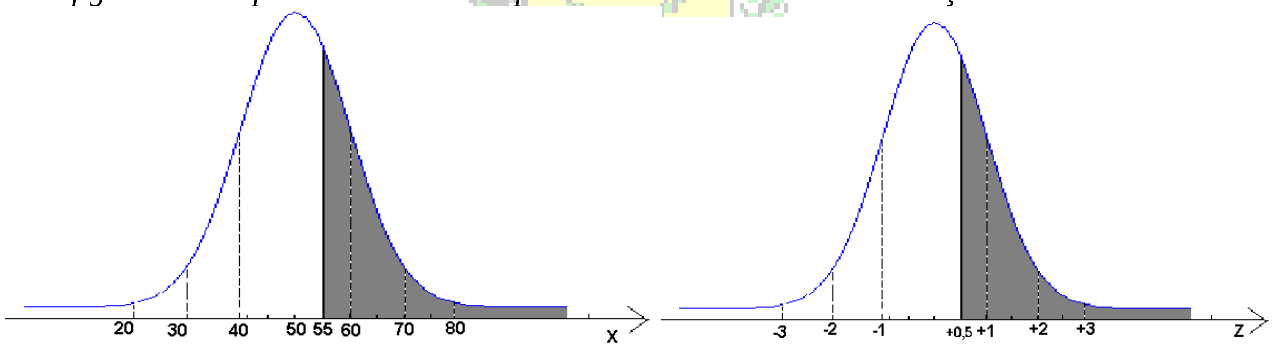
$$Z = \frac{x - \mu}{\sigma}$$

Z - número de desvios padrões a partir da média **x** - valor de interesse

μ - média da distribuição normal de interesse **σ** - desvio padrão da distribuição normal

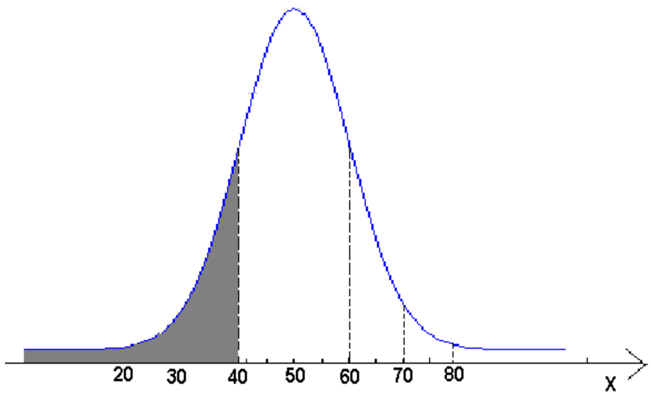
Z é um valor relativo: será negativo para valores de **x** menores do que a média, e será positivo para valores de **x** maiores do que a média. Pela transformação uma distribuição Normal qualquer **X**: **N** (μ , σ^2) passa a ser equivalente à distribuição Normal padrão Z: **N** (**0** , **1**), um valor de interesse x pode ser convertido em um valor z.

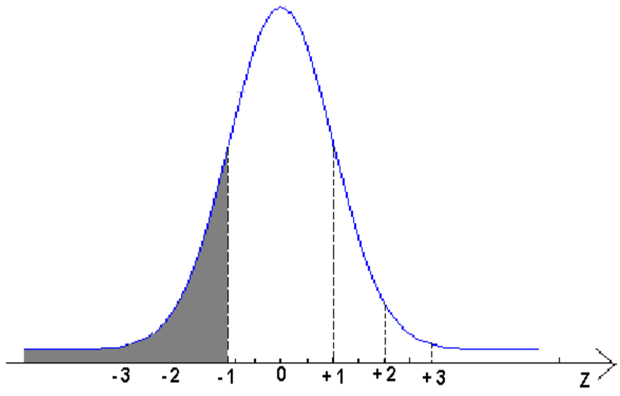
Exemplo 8.21 - Suponha uma variável aleatória X com média 50 e desvio padrão 10. Há interesse em calcular a probabilidade do evento X > 55.
Primeiro precisamos calcular o valor de Z correspondente a 55. Z = (55 - 50)/ 10 = + 0,5.
Pela figura abaixo pode-se ver a correspondência entre as duas distribuições:



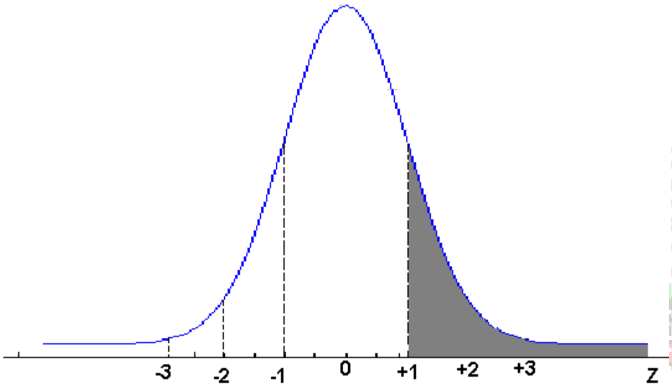
O evento P (X>55) *é equivalente ao evento* P (Z> 0,5). *Este valor pode ser obtido na tabela da distribuição Normal padrão (ver Apêndice). Os valores de Z são apresentados com dois decimais: o primeiro na coluna da extrema esquerda e o segundo na linha do topo da tabela. Observe pelas figuras que estão no alto da tabela que as probabilidades são para eventos do tipo do da figura acima* [P(Z> z1)]. *Assim, poderíamos procurar a probabilidade do evento que nos interessa,* P(Z > 0,5): *fazendo o cruzamento do valor 0,5 (na coluna) com o valor 0,00 (na linha do topo) encontramos o valor 0,3085 (30,85%). Portanto,* P(X>55) *é igual a 0,3085. Observe a coerência entre o valor encontrado e as áreas na figura: a área é menor do que a metade da figura (metade da figura significaria 50%), e a probabilidade encontrada vale 30,85%.*

Exemplo 8.22 - Supondo a mesma variável aleatória X com média 50 e desvio padrão 10. Agora há interesse em calcular a probabilidade de que X seja MENOR do que 40.
Primeiro precisamos calcular o valor de Z correspondente a 40. Z = (40 - 50)/ 10 = -1,00.
Pela figura abaixo pode-se ver a correspondência entre as duas distribuições:



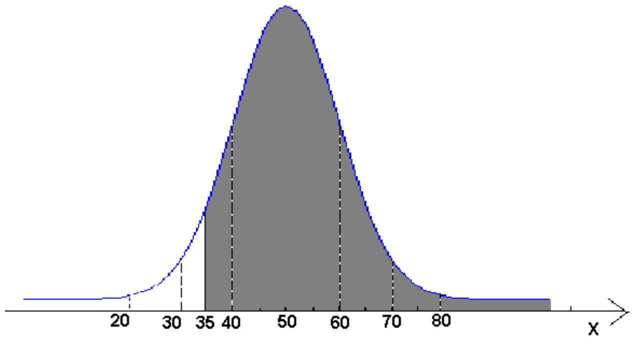


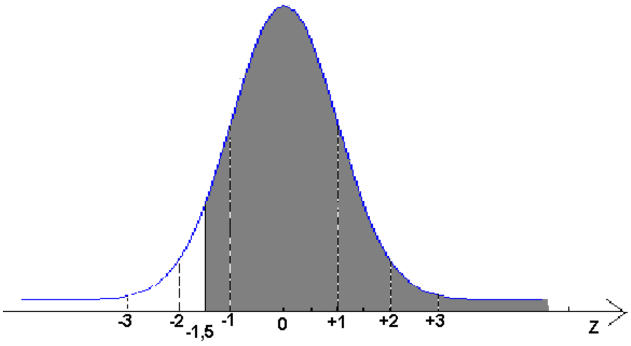
O evento P (X<40) *é equivalente ao evento* P (Z < -1,00). *Repare porém que queremos encontrar* P (Z < -1,00), *e a tabela nos apresenta valores apenas para* P (Z > 1,00). *Contudo, se rebatermos a figura da distribuição normal para a direita teremos o seguinte resultado:*



Ou seja a área P(Z < -1) = P(Z > 1). Esta probabilidade nós podemos encontrar diretamente pela tabela, fazendo o cruzamento do valor 1,0 (na coluna) com o valor 0,00 (na linha do topo) encontramos o valor 0,1587 (15,87%). Portanto, P(X<40) = P(Z<-1) = P(Z>1), *que é igual a 0,1587.*

Exemplo 8.23 - Supondo a mesma variável aleatória X com média 50 e desvio padrão 10. Agora há interesse em calcular a probabilidade de que X seja MAIOR do que 35.
Primeiro precisamos calcular o valor de Z correspondente a 35. Z = (35 - 50)/ 10 = -1,50.
Pela figura abaixo pode-se ver a correspondência entre as duas distribuições:

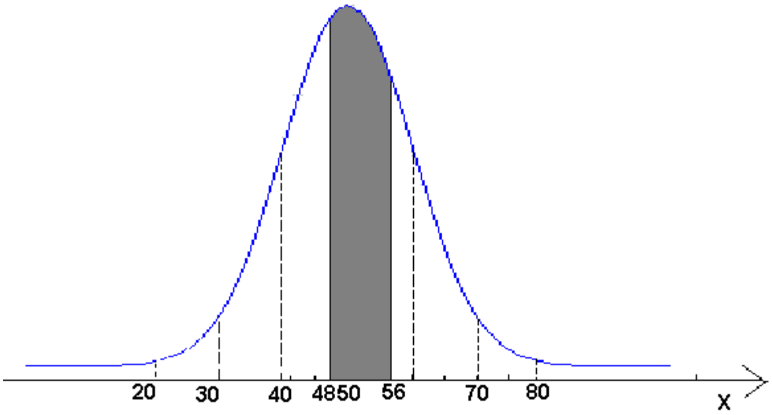




Não podemos obter a probabilidade P(Z>-1,50) diretamente, pois a tabela do Apêndice apenas apresenta resultados para valores positivos de Z. Sabemos que a probabilidade total vale 1,0, podemos então considerar que P(Z > -1,50) = 1 - P(Z < -1,50). *Usando o raciocínio descrito no Exemplo 8.22 (rebatendo as figuras para a direita), vamos obter:* P(Z<-1,50) = P(Z>1,50). *Esta última probabilidade pode ser facilmente encontrada na tabela da distribuição normal padrão:* P(Z>1,50) = P(Z<-1,50) = 0,0668. *Basta substituir na expressão:* P(Z > -1,50) = 1 - P(Z < -1,50) = 1 - 0,0668 = 0,9332 *(93,32%). Observe novamente a coerência entre as áreas da figura acima e o valor da probabilidade: a área na figura compreende mais do que 50% da probabilidade total, aproximando-se do extremo inferior da distribuição, perto de 100%, e a probabilidade encontrada realmente é próxima de 100%.*

Exemplo 8.24 - Supondo a mesma variável aleatória X com média 50 e desvio padrão 10. Agora há interesse em calcular a probabilidade de que X assuma valores entre 48 e 56.

Calcular P (48 < X < 56), *veja a figura abaixo*:



Novamente precisamos calcular os valores de Z correspondentes a 48 e a 56.

Z1 = (48 - 50)/ 10 = - 0,20

Z2 = (56 - 50)/ 10 = 0,60

Então:

P (48 < X < 56) = P (-0,20<Z<0,60)

Repare que a área entre 48 e 56 é igual à área de 48 até $+\infty$ *MENOS a área de 56 até* $+\infty$:

P(48 < X < 56) = P(X > 48) - P(X > 56)= P(-0,20< Z <0,60) = P(Z > - 0,20) - P(Z > 0,60)

E os valores acima podem ser obtidos na tabela da distribuição normal padrão:

P(Z > 0,60) = 0,2743 P(Z > -0,20) = 1- P(Z > 0,20) = 1- 0,4207 = 0,5793

P(48 < X < 56)= P(-0,20< Z <0,60) = P(Z > -0,20) - P(Z > 0,60) = 0,5793 - 0,2743 = 0,3050

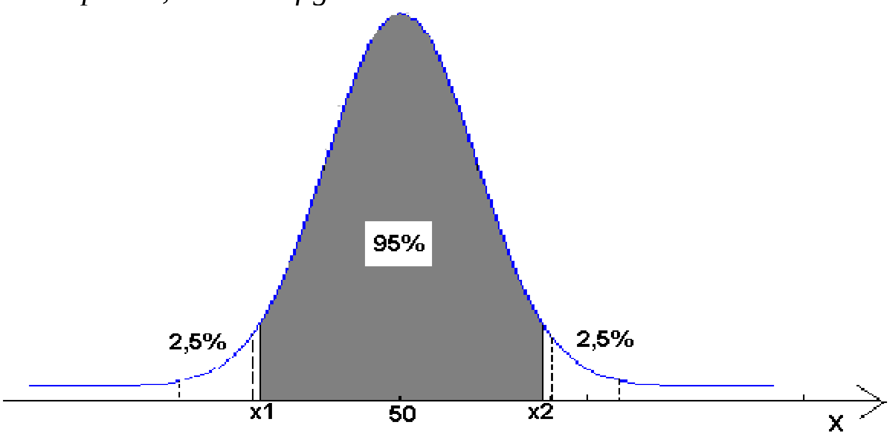
Então a probabilidade da variável X assumir valores entre 48 e 56 é igual a 0,305 (30,5%).

A distribuição Normal também pode ser utilizada para encontrar valores da variável de interesse correspondentes a uma probabilidade fixada.

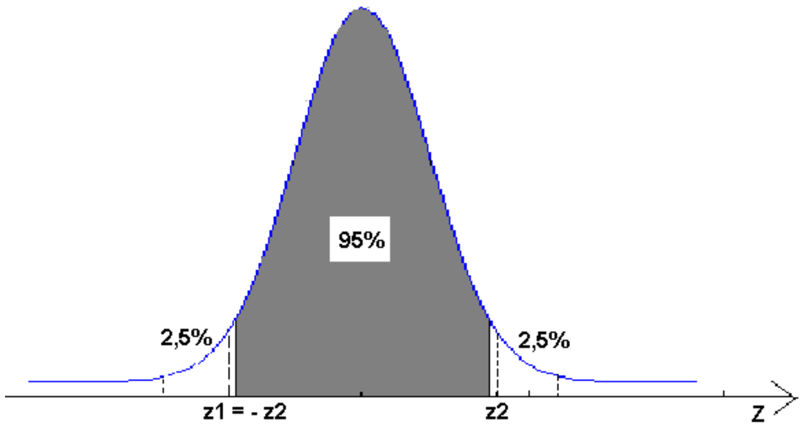
Marcelo Menezes Reis

Exemplo 8.25 - Supondo a mesma variável aleatória X com média 50 e desvio padrão 10. Encontre os valores de X, situados à mesma distância abaixo e acima da média que contém 95% dos valores da variável.

Como a distribuição Normal é simétrica em relação à média, e como neste problema os valores de interesse estão situados à mesma distância da média "sobram" 5% dos valores, 2,5% na cauda inferior e 2,5% na superior, como na figura abaixo:



É preciso encontrar os valores de Z (na tabela da distribuição Normal padrão) correspondentes às probabilidades da figura acima, e a partir daí obter os valores de ***x1*** *e* ***x2****. Passando para a distribuição Normal padrão* ***x1*** *corresponderá a um valor* ***z1****, e* ***x2*** *a um valor* ***z2****, como na figura a seguir:*



*Repare que a média da distribuição Normal padrão é igual a zero, fazendo com que **z1** e **z2** sejam iguais em módulo.*

*Podemos encontrar **z2**, já que* P(Z > **z2**) = 0,025
É necessário encontrar o valor da probabilidade na tabela da distribuição Normal padrão (ou o valor mais próximo) e obter o valor de Z associado.

*Para o caso de **z2**, ao procurar pela probabilidade 0,025 encontramos o valor exato 0,025, e por conseguinte o valor de **z2** que é igual a 1,96:* P (Z > 1,96) = 0,025.

*Como **z1** = -**z2**, encontramos facilmente o valor de z1: **z1** = -1,96.* P (Z < -1,96) = 0,025.
Observe que os valores são IGUAIS em módulo, mas corresponderão a valores diferentes da variável X. A expressão usada para obter o valor de Z, em função do valor da variável X, pode ser usada para o inverso:

$$Z = \frac{x - \mu}{\sigma} \quad => \quad x = \mu + Z \times \sigma$$

*E assim obteremos os valores de **x1** e **x2**[24], que correspondem a **z1** e **z2**, respectivamente:*

$$\mathbf{x1 = \mu + (z1 x\ \sigma) = 50 + [(-1{,}96)\ x\ 10] = 30{,}4}$$
$$\mathbf{x2 = \mu + (z2 x\ \sigma) = 50 + (1{,}96\ x\ 10) = 69{,}6}$$

Observe que os resultados obtidos são coerentes: 30,4 está abaixo da média (1,96 desvios padrões) e 69,6 acima (também 1,96 desvios padrões). O intervalo definido por estes dois valores compreende 95% dos resultados da variável X.

Todo este trabalho poderia ter sido poupado se houvesse um programa computacional que fizesse esses cálculos. Há vários softwares disponíveis no mercado, alguns deles de domínio público, que calculam as probabilidades associadas a determinados eventos, como também os valores associados a determinadas probabilidades.

8.10.2 - Modelo Normal como aproximação do modelo Binomial

Já se sabe que o modelo Binomial pode ser aproximado pelo modelo de Poisson (ambos são modelos discretos).

Contudo, o modelo Binomial (discreto) pode ser aproximado também pelo modelo Normal (contínuo) se certas condições forem satisfeitas:
- quando o valor de **n** (número de ensaios) for tal que os cálculos binomiais trabalhosos demais[25].
- quando o produto **n x p** (o valor esperado do modelo Binomial) e o produto **n x (1 - p)** forem AMBOS maiores ou iguais a 5.

[24] É muito importante que se preste atenção no SINAL do valor de **z** ao obter o valor de **x**. Observe se o resultado obtido faz sentido.
[25] Para os que pensam que o advento dos computadores eliminou este problema um alerta: em alguns casos os números envolvidos são tão grandes que sobrepujam suas capacidades.

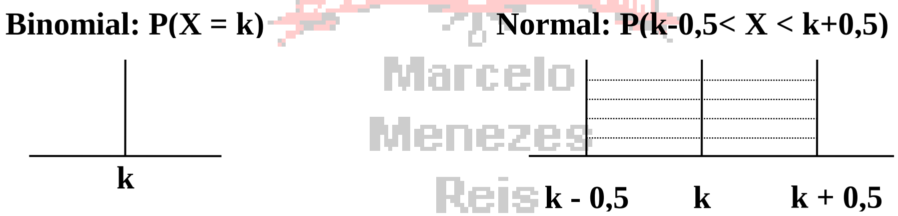
Então, uma binomial de parâmetros **n** e **p** pode ser aproximada por uma normal com:

média = μ = n x p (valor esperado do modelo Binomial)
variância = σ^2 = n x p x (1- p) (variância do modelo Binomial)

De uma maneira geral deve-se aproximar o modelo Binomial por Poisson quando a probabilidade de sucesso **p** for muito pequena, evento raro (e portanto **1 - p** for próxima de 1), ou quando p for próxima de 1 (e portanto **1 - p** for muito pequena). Se a probabilidade de sucesso tiver valores em torno de 0,5 deve-se fazer a aproximação pelo modelo Normal.

Usando o modelo Normal (contínuo) para aproximar o Binomial (discreto) é necessário fazer uma correção de continuidade: associar um **intervalo** ao valor discreto, para que o valor da probabilidade calculada pelo modelo contínuo seja mensurável. Este intervalo deve ser centrado no valor discreto, e deve ter uma amplitude igual à diferença entre dois valores consecutivos da variável discreta: se por exemplo a diferença for igual a 1 (a variável somente pode assumir valores inteiros) o intervalo deve ter amplitude igual a 1, 0,5 abaixo do valor e 0,5 acima. ESTA CORREÇÃO DE CONTINUIDADE PRECISA SER FEITA PARA GARANTIR A COERÊNCIA DA APROXIMAÇÃO.

Seja uma variável aleatória X com distribuição Binomial.
1) Há interesse em calcular a probabilidade de X assumir um valor **k** genérico, P(X = **k**), ao fazer a aproximação pela Normal será: P(**k - 0,5** < X < **k + 0,5**).



Marcelo Menezes Reis

Figura 14 - Correção de continuidade da aproximação do modelo Binomial pelo Normal - 1o caso.

2) Há interesse em calcular a probabilidade de X assumir valores menores ou iguais a um valor **k** genérico, P(X ≤ **k**), ao fazer a aproximação pela Normal será: P(X < **k + 0,5**), todo o intervalo referente a k será incluído.

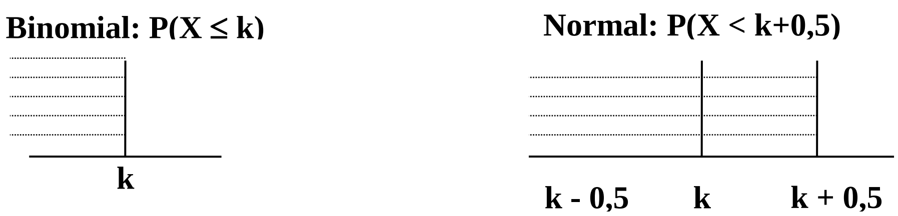


Figura 15 - Correção de continuidade da aproximação do modelo Binomial pelo Normal - 2o caso.

3) Há interesse em calcular a probabilidade de X assumir valores maiores ou iguais a um valor **k** genérico, P(X ≥ **k**), ao fazer a aproximação pela Normal será: P(X > **k - 0,5**), todo o intervalo referente a k será incluído.

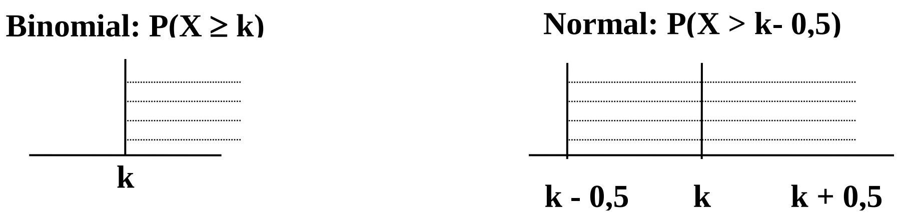


Figura 16 - Correção de continuidade da aproximação do modelo Binomial pelo Normal - 3o caso.

4) Há interesse em calcular a probabilidade de X assumir valores menores do que um valor **k** genérico, P(X < **k**), ao fazer a aproximação pela Normal será: P(X < **k - 0,5**), todo o intervalo referente a k será excluído.

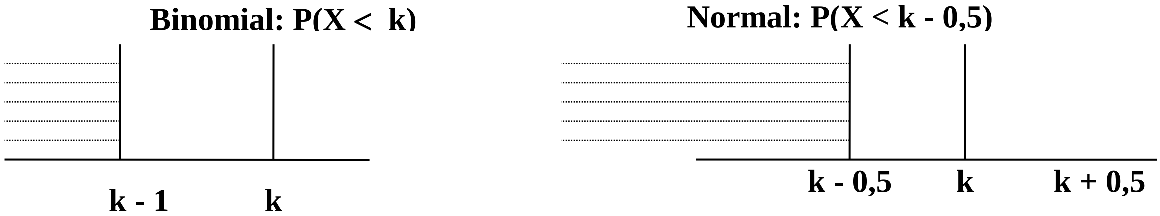


Figura 17 - Correção de continuidade da aproximação do modelo Binomial pelo Normal - 4o caso.

5) Há interesse em calcular a probabilidade de X assumir valores maiores do que um valor **k** genérico, P(X > **k**), ao fazer a aproximação pela Normal será: P(X > **k + 0,5**), todo o intervalo referente a k será excluído.

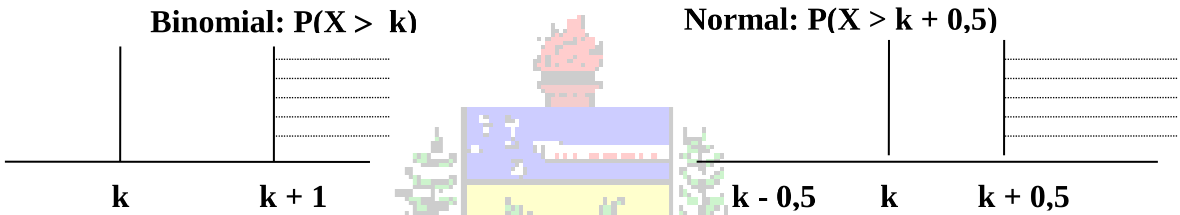


Figura 18 - Correção de continuidade da aproximação do modelo Binomial pelo Normal - 5o caso.

EX. 8.26 - Um município tem 40000 eleitores. Para uma pesquisa de opinião eleitoral uma amostra aleatória de 1500 pessoas foi selecionada. Qual é a probabilidade de que pelo menos 500 dos eleitores sejam menores de 25 anos se 35% dos 40000 são menores do que 25 anos?
Este problema poderia ser resolvido usando o modelo Binomial. Há apenas dois resultados possíveis para cada eleitor: menor de 25 anos ("sucesso") e maior ou igual a 25 anos ("fracasso"). Existe um limite superior de realizações, no caso os 1500 eleitores da amostra, e há independência entre as retiradas, pois a amostra foi retirada de forma aleatória (e a amostra representa menos de 5% dos 40000 eleitores).

Então: "sucesso" = menor de 25 anos **p** = 0,35 **1 - p** = 0,65 **n** = 1500
A variável aleatória discreta X, número de eleitores menores de 25 anos em 1500, terá distribuição binomial com parâmetros **n** = 1500 e **p** = 0,35.

O evento "pelo menos 500 menores de 25 anos" seria definido como 500 ou mais eleitores:

$$P(X \geq 500) = P(X = 500) + P(X = 501) + ...+ P(X = 1500)$$

Há cerca de 1000 expressões binomiais...

Vamos ver se é possível aproximar pelo modelo Normal.

O valor de **n** *é grande*: **n** x **p** = 1500 x 0,35 = 525 > 5 e **n** x (**1 - p**) = 1500 x 0,65 = 975 > 5.

Como as condições foram satisfeitas é possível aproximar por um modelo Normal:

$$\text{média} = \mu = \mathbf{n} \times \mathbf{p} = \mathbf{1500 \times 0{,}35 = 525}$$

$$\text{desvio padrão} = \boldsymbol{\sigma} = \sqrt{n x p x (1-p)} = \sqrt{1500 x 0{,}35 x 0{,}65} = 18{,}47$$

Pelo modelo Binomial: P (X ≥ 500). *Pelo modelo Normal será*: P (X ≥ 499,5).

$P(X \geq 499{,}5) = P(Z > z1)$ $z1 = (499{,}5 - 525)/18{,}47 = -1{,}38$ $P(Z > -1{,}38) = 1 - P(Z > 1{,}38)$

Procurando na tabela da distribuição Normal padrão: $P(Z > 1{,}38) = 0{,}0838$

Então: $P(X \geq 500) \cong P(X \geq 499{,}5) = P(Z > -1{,}38) = 1 - P(Z > 1{,}38) = 1 - 0{,}0838 = 0{,}9162$.

A probabilidade de que pelo menos 500 dos eleitores da amostra sejam menores de 25 anos é igual a 0,9162 (91,62%).

8.11 – Gráfico de Probabilidade.

Em Análise Exploratória de Dados (ver Capítulo 2) foi apresentado o histograma, um gráfico de barras justapostas que permite descrever o comportamento de uma variável quantitativa discreta (histograma não agrupado) ou contínua (geralmente um histograma agrupado em classes). Às vezes há interesse em saber se uma variável segue um determinado modelo probabilístico: isso será especialmente importante no Capítulo 9 – Inferência Estatística, onde muitas técnicas exigem que os dados tenham distribuição normal, sendo comum examinar o formato do histograma para avaliar se há indícios da variável seguir algum modelo. O problema do histograma, especialmente o agrupado em classes é a sua inadequação para representar pequenos conjuntos de dados (menos de 100 observações), ou que mudanças, muitas vezes arbitrárias, no número de classes podem modificar substancialmente a forma da distribuição. Torna-se necessário um gráfico que possibilite uma análise mais objetiva, tanto para grandes quanto pequenos conjuntos de dados: o gráfico de probabilidade.

O gráfico de probabilidade é um diagrama de dispersão (ver Capítulo 3), como mostrado na Figura 19, em um dos eixos é colocada a frequência acumulada observada ("probabilidade" observada) a partir dos dados, e no outro é colocada a probabilidade acumulada *esperada* de acordo com algum modelo de probabilidade.[26] Em seguida ajusta-se uma equação de reta (ver Capítulo 3) aos dados do diagrama de dispersão, com coeficiente linear (intersecção) igual a zero (linha preta na Figura 19). Se os pontos ficarem aproximadamente sobre a reta (linha azul na Figura 19) é indicação que os dados são provenientes da distribuição de probabilidade esperada. Se, porém, os pontos afastarem-se substancialmente da reta esperada (linha vermelha na Figura 19) é indício que os dados não provêm do modelo esperado, e que talvez outra distribuição de probabilidades seja mais apropriada para os dados.

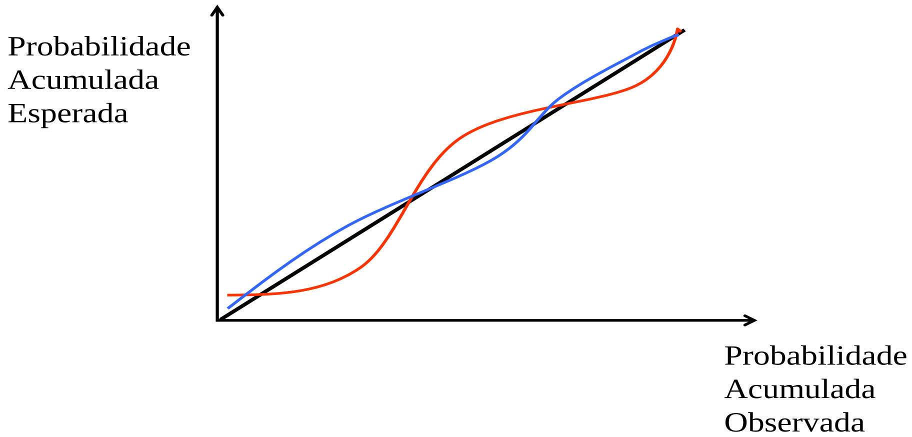


Figura 19 - Gráfico de probabilidade - caso genérico

[26] Esta é a ideia dos testes de aderência, mas ao invés de usar um gráfico são calculadas as diferenças entre as probabilidades observadas e esperadas. Mais detalhes no Capítulo 10 de BARBETTA, P.A., REIS, M.M., BORNIA, A.C. Estatística para Cursos de Engenharia e Informática, 3ª ed., São Paulo: Atlas, 2010.

O modelo probabilístico para o qual geralmente é feito um gráfico de probabilidade é o normal, especialmente para os casos que veremos no Capítulo 9, em que várias técnicas exigem que os dados sejam provenientes de uma distribuição normal. O procedimento para construir um gráfico de probabilidade normal na planilha eletrônica Microsoft Excel ® pode ser visto em:
https://www.youtube.com/watch?v=HRhKR30en70

EX. 8.27 – A figuras a seguir mostra os histogramas agrupados em 15 classes de 1000 observações das variáveis X_1 e X_2.

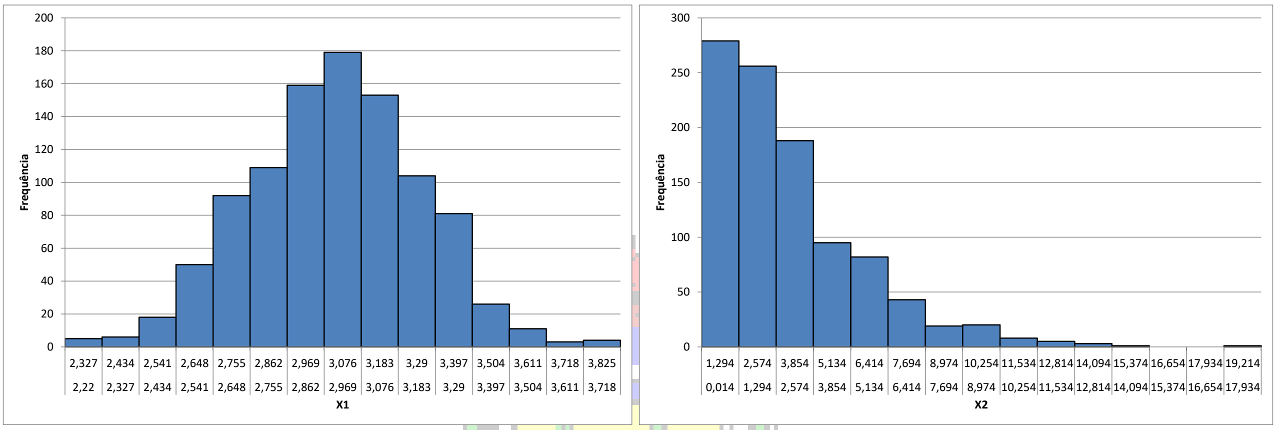


Figura 20 - Histogramas agrupados em 15 classes de 1000 observações das variáveis X_1 e X_2

Observando os histogramas é razoável afirmar que a distribuição das 1000 observações da variável X_1 é simétrica, e pode ser considerada normal, porque as maiores frequências estão nas classes centrais, reduzindo as frequências proporcional e simetricamente quanto mais distantes do centro, e o formato do histograma lembra o de um sino. Já a distribuição das 1000 observações da variável X_2 é claramente assimétrica, com as maiores frequências nas menores classes, reduzindo substancialmente as frequências à medida que aumentam os valores de X_2. Espera-se, então, que o gráfico de probabilidade normal para a variável X_1 tenha os pontos bem próximos da reta teórica, e no da variável X_2 os pontos apresentarão substancial desvio do comportamento esperado se houvesse aderência ao modelo normal. Os gráficos são mostrados nas Figuras 21 e 22.

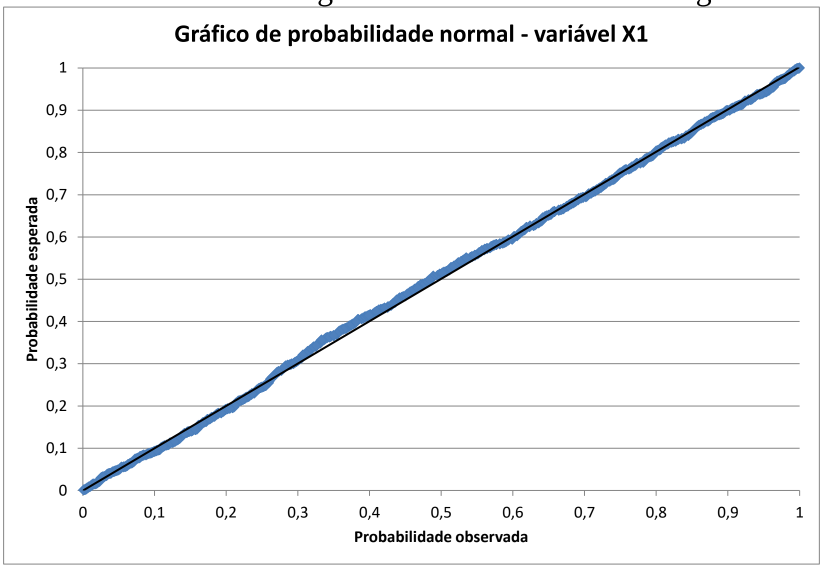


Figura 21 - Gráfico de probabilidade normal de 1000 observações da variável X_1

Observem que os pontos, mostrados em azul na Figura 21, estão praticamente todos sobre a linha teórica em preto, o que indica que as observações da variável X_1 devem realmente ser originárias de uma distribuição normal, corroborando a análise feita pelo histograma.

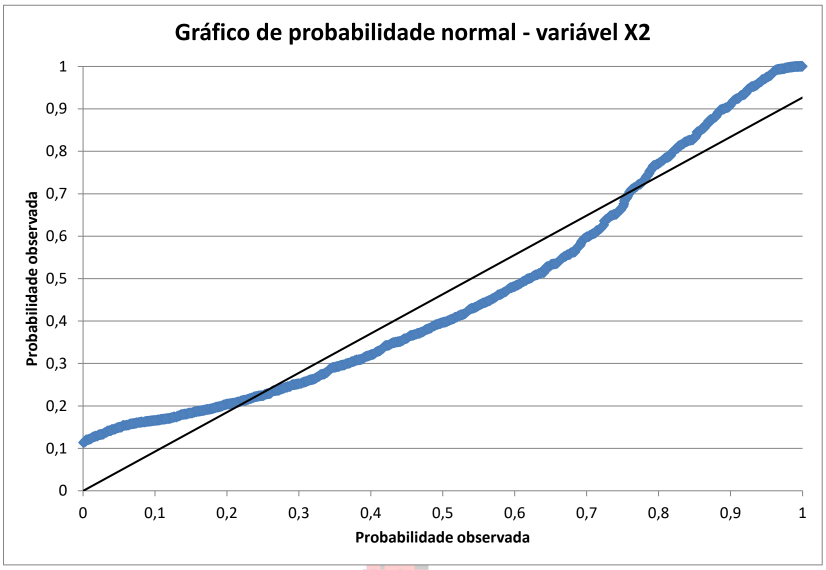


Figura 22 - Gráfico de probabilidade normal de 1000 observações da variável X_1

Observem que os pontos, mostrados em azul na Figura 22, afastam-se bastante da linha teórica em preto, o que indica que as observações da variável X_2 NÃO devem realmente ser originárias de uma distribuição normal, corroborando a análise feita pelo histograma.

Marcelo Menezes Reis

A seguir um diagrama com o resumo deste capítulo.

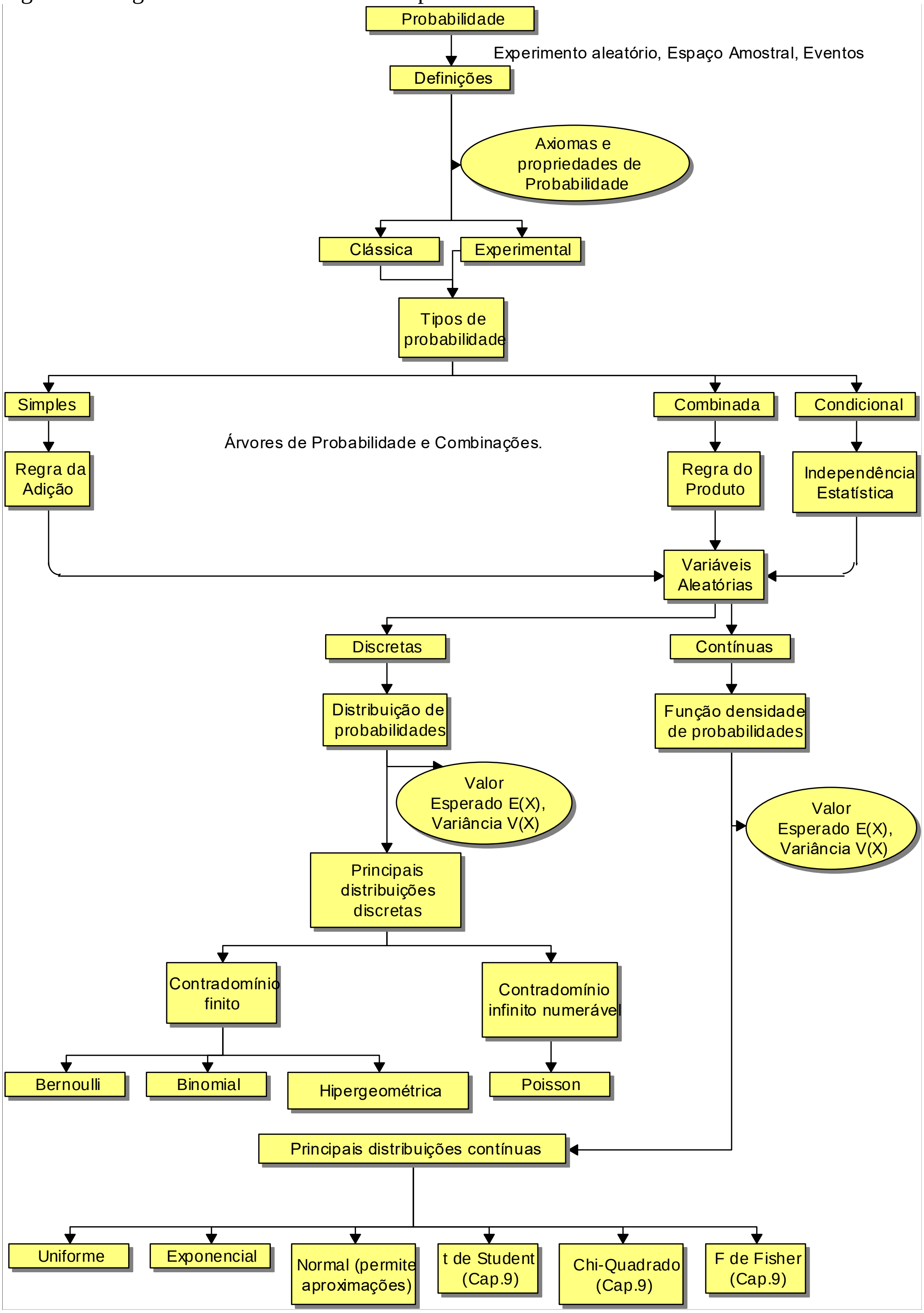


Figura 23 - Resumo do Capítulo (adaptado de LEVINE, D. M., BERENSON, M. L. e STEPHAN, D. – *Estatística: Teoria e Aplicações usando o Excel.* Rio de Janeiro: LTC, 2000)