Luciana G.de Godoi, Nátaly A. Jiménez Monroy, Alessandro J. Q. Sarnaglia, Bartolomeu Zamprogno, Fabio Fajardo Molinares & Mauro Campos

Introdução à probabilidade e estatística

Exemplos e aplicações no* `R`*.

Versão preliminar. Última atualização: 20 de dezembro de 2018.

1a Edição – 2018. Departamento de Estatística - Universidade Federal do Espírito Santo. Vitória - ES. Brasil. Material desenvolvido no projeto intitulado *Elaboração de material didático para o ensino da Estatística na UFES*, coordenado pela equipe do Núcleo de Ensino e Aperfeiçoamento da Estatística.

Apoio: Programa de aprimoramento e desenvolvimento do ensino (PRÓ-ENSINO).

Sumário

Introdução

Estatística transforma dado em informação

Questões básicas de contagens, enumerações e recenseamento sempre foram preocupações das mais antigas civilizações, tanto do ponto de vista econômico, quanto do ponto de vista social. Os imperadores, ou governantes, ordenavam recenseamentos com o intuito de conhecer sua população para assim realizar cobranças de impostos e para o recrutamento militar. Os registros históricos mais antigos indicam que o primeiro censo, para fins agrícolas e comerciais, foi realizado em 2238 a.C. pelo primeiro Imperador da China. Nessa época, a Estatística não era uma ciência e nem tinha um corpo para assim ser caracterizada.

A importância dos métodos de coleta de informações veio seguida da melhoria dos meios de análise dos dados obtidos, o que pode se chamar atualmente de **Estatística Descritiva**. No século XVI, o italiano Francesco Sansovini propôs uma orientação descritiva àqueles que lidavam com as coletas dos dados estatísticos e, no século XVIII, alemães aprimoraram o método italiano apresentando uma melhor sistematização e definição. Isso porque, cada vez mais, os estados necessitavam de mais informações para tomadas de decisões. Essa necessidade da administração pública leva à associação da palavra estatística à palavra latina *Status* (Estado). No entanto, a palavra estatística foi cunhada pelo acadêmico alemão Gottfried Achenwall (1719-1772) com o termo *Staatenkunde*, em 1746, que deu origem à designação atual. Na Enciclopédia Britânica, o verbete *Statistics* apareceu em 1797.

Embora a palavra estatística tenha tido associação com a palavra estado em sua origem, atualmente a sua definição abrange muito mais do que contagens, tabelas, gráficos e cálculo de medidas, como veremos no decorrer do texto.

Um pouco antes do aparecimento da palavra que origina o termo estatística, o estudo da sua base teórica, Probabilidade, ganhou status de ciência quando Jacques Bernoulli provou uma das versões da Lei dos Grandes Números. Isso evidência que, para uma análise precisa de Estatística, em toda sua dimensão, deve-se conhecer o desenvolvimento da **Teoria de Probabilidades**. Com uma boa base de Teoria de Probabilidades, fica mais fácil de entender a extrapolação de resultados relativos a um subconjunto da população, ou amostra, para toda população, o que é denominado de **Inferência Estatística**. Posteriormente, os modelos estatísticos, que tem como base a Teoria de Probabilidade, tornam-se mais maleáveis, pois certamente são melhores empregados dada a melhor compreensão de sob quais circunstâncias se aplicam, o que não implica em resolução imediata do problema em questão, visto que o comportamento probabilístico pode ser de difícil modelagem.

Atualmente, a Estatística é considerada uma ciência multidisciplinar, que tem um elo com a Matemática Aplicada e com a Informática, que fornece métodos para coleta, organização, descrição, análise e interpretação de dados e para a utilização dos mesmos na tomada de decisões. Seu objetivo é o estudo da variabilidade para a tomada de decisões frente a esse tipo de incerteza.

Como a variabilidade e a incerteza estão presentes em todas as áreas do conhecimento, a Estatística é um ciência crucial para resolver uma série de problemas com uso de metodologias apropriadas a partir da análise dos dados coletados.

A utilização de técnicas destinadas à análise de diversas situações, complexas ou não, tem aumentado e faz parte do nosso cotidiano. No passado, tratar uma grande massa de dados era uma tarefa custosa e cansativa, que exigia horas de trabalho tedioso. Recentemente, no entanto, grandes quantidades de informações podem ser analisadas rapidamente com um computador pessoal e softwares adequados. Desta forma, o computador contribui de maneira crucial na difusão e uso de métodos estatísticos. Por outro lado, o computador possibilita uma automação que pode levar um indivíduo sem preparo específico a utilizar técnicas inadequadas para resolver um dado problema. Assim, é necessária a compreensão dos conceitos básicos da Estatística em toda a sua abrangência, incluindo Estatística Descritiva, Probabilidade e Inferência Estatística, para posterior compreensão e interpretação de uma modelagem mais rebuscada.

História da estatística

A estatística de massa, ou contagens, iniciou-se nos grandes Impérios da Antiguidade, como Grécia, Roma, Egito, Índia e China, entre outros, tendo como principal objetivo a administração dos bens, homens, armas e obras públicas do Estado. Os registros históricos mais antigos indicam que o primeiro censo foi realizado em 2238 a.C. pelo primeiro Imperador da China. Os romanos registravam os nascimentos e as mortes para fazerem os censo do número de homens aptos a guerrear e também com objetivo de taxação e cobrança de impostos.

A partir do século XVIII, a Estatística começou a caminhar para a ciência que conhecemos hoje. Foi quando deu início à ligação entre Probabilidade e os conhecimentos estatísticos, surgindo a Inferência Estatística. É nesta época também que se originou-se a palavra "estatística" e o desenvolvimento da Demografia.

Na segunda década do século XIX e principalmente no início do século XX, acelera-se o desenvolvimento da Estatística, tendo como principal responsável, Sir Ronald A. Fisher, conhecido como o "Pai" da Estatística moderna. É na primeira metade do século XX que desenvolveu-se e sedimentou-se a grande parte da metodologia estatística, desde as bases axiomáticas das probabilidades, passando pela Inferência Estatística Clássica e Bayesiana, Análise de Regressão, Delineamento e Análise de Experimentos, Análise Multivariada, Análise de Sobrevivência, Análise Não-Paramétrica e Análise de Séries Temporais, consolidando-se aplicações importantes nas áreas biológicas, agrárias, industriais, econômicas, além dos levantamentos populacionais.

A maior revolução nessa ciência, ocorreu por volta de 1970, mudando o foco da Estatística para sempre: o rápido desenvolvimento e disponibilidade dos computadores e softwares mudou completamente o significado da realização de uma análise estatística. Influenciou também na facilidade com que cientistas e profissionais podem coletar e armazenar dados, dando início a uma nova era, com inúmeras possibilidades para a implementação de novas e antigas idéias, a partir de abordagens em maiores escalas e soluções computacionalmente intensivas. Como consequência, no final do século XX, foi possível uma explosão de novas possibilidades em termos de metodologias e aplicações, aumentando ainda mais o grau de interdisciplinaridade. De forma resumida, os avanços da estatística estão relacionado com a computação da seguinte maneira:

- 1960: máquinas de calcular manuais, elétricas e eletrônicas;
- 1960 $\rightarrow$ 1980: "grandes computadores": IBM 1620; CDC 360; VAX etc; cartões e discos magnéticos; FORTRAN;
- 1980 até recentemente: computadores pessoais; supercomputadores; computação paralela; "clouds"; C; C+; S; Pacotes estatísticos: S-Plus, SPSS, Minitab etc. Repositório R; Data Mining: redes neurais, support vector machines.
- Era do "Big Data": Análise de *microarrays* em Bioinformática; Dados de alta frequência em finanças; taxas de câmbio e taxa de juros; Dados meteorológicos, oceanográficos e astronômicos; Simulações.

Definições de estatística

Como mencionado anteriormente, a palavra "Estatística" foi originalmente derivada da mesma raiz da palavra "Estado", já que foi a função tradicional de governos centrais, no sentido de armazenar registros da população, nascimentos e mortes, produção das lavouras, taxas e muitas outras espécies de informação e atividades. A contagem e mensuração dessas quantidades gera todos os tipos de dados numéricos ou não que são úteis para o desenvolvimento de muitos tipos de funções governamentais e formulação de políticas públicas.

Conjunto de dados são de fato uma parte da Estatística, mas são apenas a matéria-prima, que precisa ser transformada pelos "métodos estatísticos" para posterior análise. A Estatística, como um método científico, refere-se ao planejamento de experimentos e a descrição e interpretação de observações que são feitas. De um ponto de vista moderno, a Estatística é frequentemente definida como um método de tomada de decisão em face da aleatoriedade dos fenômenos. Em uma perspectiva mais ampla, o escopo da Estatística pode ser pensado em termos de três áreas diferentes: a Estatística Descritiva; a Estatística Indutiva (Inferência); e a Teoria da Decisão Estatística.

De forma mais sucinta as seguintes definições podem ser consideradas:

1. A Estatística é um conjunto de técnicas que permite, de forma sistemática, organizar, descrever, analisar e interpretar dados oriundos de estudos ou experimentos, realizados em qualquer área do conhecimento.

2. A Estatística é uma coleção de métodos para planejar experimentos, obter dados e organizá-los, resumi-los, analisá- los, interpretá-los e deles extrair conclusões (Triola, 1998).

3. "Estatística é a Ciência que permite obter conclusões a partir de dados" (Paul Velleman).

4. Estatística é a ciência que transforma dados (registros) em informação.

A interdisciplinaridade da estatística

O emprego da estatística se dá em todas as áreas do conhecimento e, dessa forma, todas as instituições tem informações para serem tratadas, desde seu uso mais básico ao seu uso mais sofisticado. Instituições públicas e privadas como IBGE, EMBRAPA, Ministérios, Bancos e Seguradoras, Operadoras de plano de saúde, Indústrias, Hospitais e Instituições de Pesquisa na área médica e biológica, Universidades, Centros de Pesquisa, entre outros, são locais que fazem uso de ferramentas estatísticas.

A direção de uma empresa de qualquer tipo, incluindo as estatais e governamentais, exige de seu administrador a importante tarefa de tomar decisões, e o conhecimento e o uso da Estatística facilitarão seu tríplice trabalho de organizar, dirigir e controlar a empresa. Por meio de sondagem, de coleta de dados e de recenseamento de opiniões, os gestores podem conhecer a realidade social, os recursos naturais, humanos e financeiros disponíveis, as expectativas da comunidade sobre a empresa, estabelecer suas metas, e objetivos com maior possibilidade de serem alcançados a curto, médio ou longo prazo. A Estatística ajudará nas demandas de trabalho que surgem de forma mais simples no dia a dia, como também na seleção e organização da estratégia a ser adotada no empreendimento e ainda na escolha das técnicas de verificação e avaliação da qualidade e da quantidade do produto e mesmo dos eventuais lucros e/ou perdas.

A Estatística apontará caminhos mais viáveis e práticos. Na leitura de um banco de dados, é capaz de minerar os dados para uma tomada de decisão mais assertiva. Na área de seguros/planos de saúde, indica modelos de custo/preços dos seguros que garantem o equilíbrio financeiro das seguradoras, elabora pesquisas dos serviços ofertados, revisa contratos firmados, evitando prejuízos, entre outras atividades. Em órgãos públicos, a Estatística está inserida em todas etapas de levantamento de dados para estratégias de investimentos, melhorias, destino de verbas a locais mais necessitados, verifica regiões com maior incidência de crimes, identifica causas e propõe ações para eliminar ou sanar os problemas, além de atuar em diversos e amplos trabalhos de grande responsabilidade. Na área de saúde, pode participar tanto de testes de novos medicamentos como indicar caminhos para prevenir o aumento de custos das patologias no mais diversos aspectos. No mercado financeiro, pode atuar no suporte de modelos econômicos, caracterizar os clientes bancários em diferentes formas, fornecendo ao gestor o melhor meio de se relacionar com o clientes. Em lojas de varejo, a Estatística é capaz de sugerir promoções para atrair o cliente, pois consegue determinar o perfil de consumo do mesmo, aumentando o faturamento da empresa. Na área de marketing, é capaz de determinar a melhor maneira de elaborar uma propaganda (peça publicitária) e até indicar os melhores pontos de mercados para atuação da empresa e melhores nichos de mercado comprador do produto. Em uma industria, essa ciência aperfeiçoará os processos produtivos, irá medir, entre outras variáveis, a produtividade de uma máquina por hora, irá avaliar a qualidade dos produtos, checará as informações sobre itens fabricados com defeitos, entre outros.

Diante de tudo isso, a Estatística proporciona aumentar os lucros, otimizar processos e aumentar o ganho de produtividade das empresas. Na vida particular, em vários momentos de um dia, você de alguma forma utiliza ao menos um conhecimento estatístico que você tem sobre várias ocasiões.

Você toma decisões com estatística?

Na vida, qualquer pessoa mais prudente avalia uma série de informações e muitas vezes nem sabe que está lidando com uma decisão estatística.

"Amanhã vou sair de casa 15 minutos mais cedo para não chegar atrasado(a) na escola". Como é que eu decidi isso? Adivinhei que o trânsito estará pior? Dessa forma, em um único dia, você é capaz de utilizar uma forma simplificada de método estatístico. Nos casos apresentados, leva-se em consideração informações de um passado recente ou remoto para decidir a respeito de uma determinada situação e muitas vezes nem se repara que isso está ligado a um pensamento estatístico. O fato de sair de casa mais cedo pode estar atrelado a saber que amanhã é véspera de um feriado, que estará chovendo, que há a necessidade de passar em algum lugar antes, que passará por outro caminho? Todas essas informações são motivações para a decisão tomada e isso é pensar de forma estatística.

Por mais básica que possa ser a informação, ela é capaz de solucionar determinada atividade sem maiores problemas. Ter uma informação com alta chance de se tornar real lhe deixa um passo a frente de outras pessoas. Um dado não processado (analisado) não vale nada.

A Estatística Dedutiva e a Indutiva

Após emprego de técnica e coleta dos dados, geralmente obtidos de toda a população, via censos, ou de parte da população, através de amostras (com uso de métodos de amostragens apropriados), passamos a parte da estatística denominada de dedutiva, também conhecida como Estatística Descritiva, que se encarrega, como o próprio nome diz, de descrever o conjunto de dados, desde a elaboração da pesquisa, via tabelas, gráficos, até o cálculo de determinada medida.

Para análise dos dados provenientes de amostras, que de forma intrínseca contém a incerteza, a Estatística Descritiva não é suficiente para concluir, ou melhor, inferir ou induzir, o que há na amostra para sua totalidade (a população em estudo). Para isso, faz uso da Estatística Indutiva ou Inferencial, onde tem por base toda uma fundamentação teórica, a teoria de probabilidade, que se expande por toda área da Inferência.

Este material aborda a Estatística Descritiva e a base teórica de probabilidade para que se desenvolva de forma apropriada a Estatística Inferencial.

1

Análise descritivo de dados

A Estatística é a ciência dos dados. Fornece os princípios e os métodos para planejar, coletar, classificar, resumir, organizar, analisar e interpretar informações numéricas. O objetivo principal da Estatística é auxiliar a tomada de decisões em situações de incerteza com base em um conjunto de informações quantitativas.

1.1 Conceitos Fundamentais

- **Dado:** Fato e número coletado, analisado e sintetizado para apresentação e interpretação.

- **Unidade experimental:** Elemento (pessoa, objeto, transação, evento, etc.) a respeito do qual se coletam os dados.

- **População:** Conjunto de unidades experimentais que são objeto de estudo.

- **Amostra:** Subconjunto de unidades experimentais da população.

- **Variável:** Característica ou propriedade da unidade experimental. Chama-se assim devido ao fato de que qualquer característica pode variar entre unidades experimentais de uma população.

- **Medição:** Processo usado para atribuir números às variáveis de unidades experimentais distintas. A medição pode ser realizada usando equipamentos específicos, questionários ou contagens.

- **Observação:** Conjunto de medidas obtidas, correspondentes a determinada unidade experimental.

- **Censo:** Processo no qual as variáveis são medidas para cada unidade experimental da população. É vantajoso em situações onde a população é pequena, quando se exige precisão completa e quando se dispõe de toda a informação necessária para atingir os objetivos do estudo.

- **Pesquisa Amostral ou Amostragem:** Processo de realização de uma pesquisa para coletar dados de uma amostra. É vantajosa quando a população é infinita (isto é, muito grande), quando há necessidade de destruir a unidade de observação ou por motivos operacionais, entre outros.

1.2 Técnicas de seleção de amostras aleatórias

Definição 1.1. A ***amostragem aleatória simples*** seleciona uma amostra de tamanho n de forma que ***TODAS*** as possíveis amostras de tamanho n tenham a mesma chance de serem selecionadas.

Definição 1.2. A ***amostragem aleatória estratificada*** classifica a população em pelo menos dois grupos ou estratos para selecionar uma amostra de cada um.

Definição 1.3. A ***amostragem aleatória por conglomerados*** divide a população em grupos ou setores, selecionam-se aleatoriamente alguns desses grupos e todos os elementos dos mesmos são observados.

1.3 Áreas da Estatística

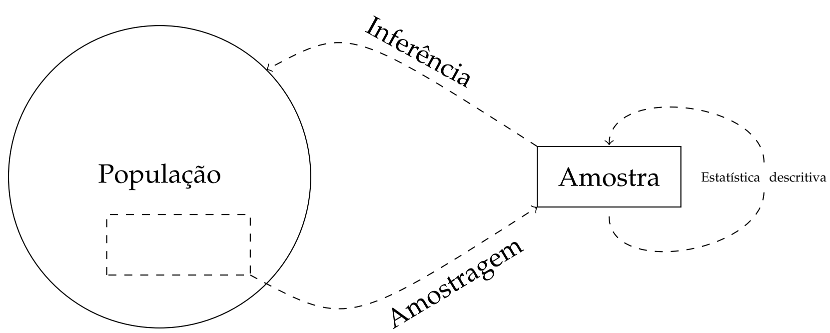
A Estatística divide-se basicamente em:

- Estatística Descritiva: É a primeira etapa da análise dos dados.

Compreende técnicas para a organização, o resumo e a apresentação das observações. Se utilizada em uma amostra essas técnicas apenas descrevem **os dados em questão**;

- Inferência Estatística: É um conjunto de técnicas que se baseia na teoria de probabilidade para extrapolar as informações de uma amostra para toda a população.

Um estudo estatístico pode ser representado por meio do seguinte esquema:



1.4 Estatística Descritiva

Frequentemente as pesquisas geram grandes quantidades de dados. A Estatística descritiva pode ser usada como uma ferramenta importante para o correto tratamento desses dados, através da qual podem ser obtidas conclusões válidas. É conduzida por meio de duas abordagens:

- Representações tabulares e gráficas.

- Medidas numéricas.

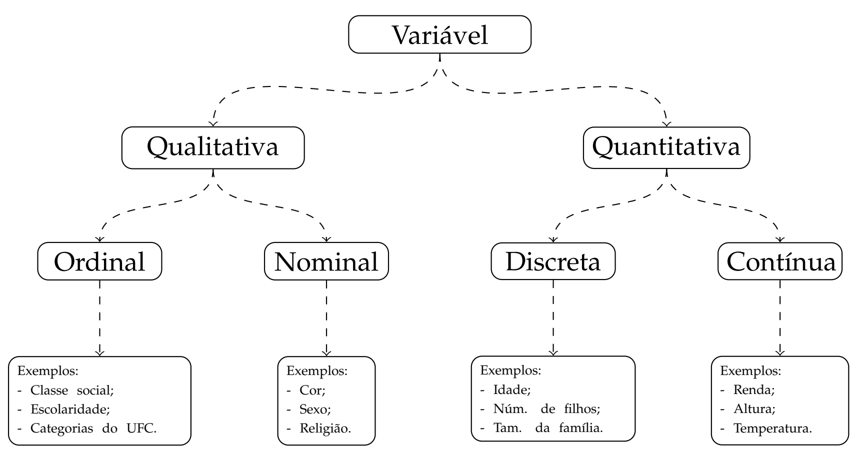
O tipo de variável sob estudo determina a abordagem a ser usada na análise.

1.4.0.1 Classificação das variáveis

As variáveis são classificadas de acordo com os valores que assumem. Elas podem ser:

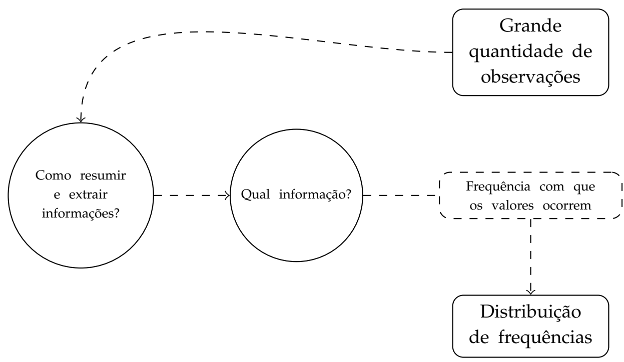
- Qualitativas (ou categóricas), se seus valores são categorias, qualidades ou atributos dos indivíduos. Ainda se subdividem em:
  - Ordinais - quando essas categorias admitem alguma **ordenação** lógica;
  - Nominais - quando esses atributos não podem ser ordenados logicamente;
- Quantitativas, se seus valores são números, geralmente resultados de contagens ou medições. Ainda se subdividem em:
  - Discretas - seus possíveis valores formam um conjunto finito, ou infinito **enumerável**. Um conjunto é ***enumerável*** se é possível estabelecer uma bijeção entre seus elementos e um subconjunto de $\mathbb{Z}$;
  - Contínuas - seus possíveis valores formam um conjunto não enumerável.

Esquematicamente:



1.4.1 Representação Tabular

1. **Distribuição de frequências:** Resumo tabular de dados mostrando o número (frequência absoluta) de elementos em cada uma das diversas classes ou categorias. No caso dos dados quantitativos, é necessário definir intervalos de valores não sobrepostos. Esquematicamente:



1.4.1.1 Casos qualitativo e quantitativo discreto (poucas observações diferentes)

Seja X uma variável. Suponha que $x_1, x_2, \ldots, x_n$ são observações de X. As observações $x_1, x_2, \ldots, x_n$ podem ser os valores de X para a população completa ou apenas uma amostra de X. Agora, considere que:

i. dos valores $x_1, x_2, \ldots, x_n$, apenas $k \leq n$ são diferentes;
ii. o restante são apenas repetições desses valores;
iii. sejam $x_1^*, x_2^*, \ldots, x_k^*$ as observações diferentes;
iv. suponha que o valor x_i^* se repetiu n_i vezes, $i = 1, 2, \ldots, k$.

Definição 1.4. A distribuição de frequências de X em $x_1, x_2, \ldots, x_n$ é dada por

X	x_1^*	x_2^*	$\ldots$	x_k^*	Total
Freq.	n_1	n_2	$\ldots$	n_k	$\sum_{i=1}^{k} n_i = n$

Observações:

- Se X for ordinal ou discreta, os x_i^* devem ser dispostos de forma que
$$x_1^* < x_2^* < \cdots < x_k^*;$$
- A distribuição de frequências fornece um resumo considerável dos dados;
- Neste caso, não há perda de informação ao utilizar essa técnica ;
- A frequência com que os valores de X ocorrem fica evidente;
- Utiliza-se esse tipo de distribuição de frequências no caso discreto quando $k \ll n$.

1.4.1.2 Exemplo 1

Seja X = "Tipo de música preferida". Neste caso $X \in \{p, r, s\}$, onde p = pagode, r = rock e s = sertanejo. Suponha que $n = 40$ pessoas foram entrevistadas e o valor de X para cada uma delas foi verificado. Os dados são

$$\begin{array}{cccccccccc}
s & p & p & p & s & p & s & s & r & p \\
s & r & p & s & r & p & s & s & p & p \\
p & s & r & s & s & p & p & p & s & s \\
p & r & s & s & r & s & p & p & p & s
\end{array}$$

Assim, a distribuição de frequências de X é apresentada em uma tabela da forma:

X	p	r	s	Total
Freq.				40

As frequências são obtidas contando o número de unidades que apresentam cada categoria (p, r e s), assim

$$\begin{array}{cccccccccc}
s & p & p & p & s & p & s & s & r & p \\
s & r & p & s & r & p & s & s & p & p \\
p & s & r & s & s & p & p & p & s & s \\
p & r & s & s & r & s & p & p & p & s
\end{array}$$

Assim, $n_p = 17$. Da mesma forma obtem-se $n_r = 6$ e $n_s = 17$. Portanto, a distribuição de frequências de X é

X	p	r	s	Total
Freq.	17	6	17	40

1.4.1.3 Exemplo 2

O RH de uma empresa com 600 funcionários deseja fazer um levantamento com respeito Ã“ escolaridade e número de filhos (variáveis E e N, respectivamente) dos mesmos. Uma amostra de $n = 30$ funcionários forneceu as seguintes observações:

i	E_i	N_i	i	E_i	N_i	i	E_i	N_i
1	f	3	11	f	2	21	f	2
2	s	2	12	s	0	22	s	2
3	f	2	13	f	2	23	f	1
4	m	1	14	s	1	24	f	2
5	s	1	15	f	3	25	f	2
6	m	1	16	s	0	26	s	1
7	f	2	17	m	1	27	s	2
8	s	3	18	f	1	28	s	1
9	m	1	19	m	1	29	f	3
10	m	2	20	m	2	30	m	3

f = "fundamental", m = "médio" e s = "superior"

– Tabela de frequências para a variável E: Realiza-se a contagem das vezes que se repete cada uma das categorias (f, m e s).

i	E_i	N_i	i	E_i	N_i	i	E_i	N_i
1	f	3	11	f	2	21	f	2
2	s	2	12	s	0	22	s	2
3	f	2	13	f	2	23	f	1
4	m	1	14	s	1	24	f	2
5	s	1	15	f	3	25	f	2
6	m	1	16	s	0	26	s	1
7	f	2	17	m	1	27	s	2
8	s	3	18	f	1	28	s	1
9	m	1	19	m	1	29	f	3
10	m	2	20	m	2	30	m	3

$n_f = 12$, $n_m = 8$ e $n_s = 10$

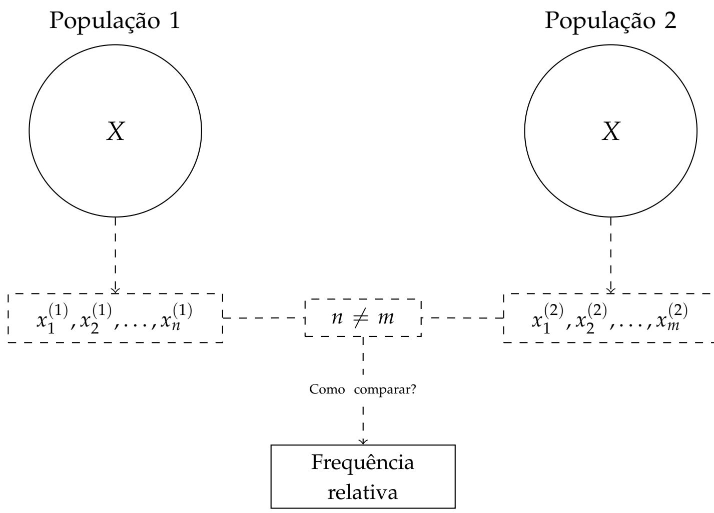
Assim, a distribuição de frequências de E é dada por

E	f	m	s	Total
Freq.	12	8	10	30

Análogamente, a distribuição de frequências da variável N obtem-se a partir da contagem das unidades observadas em cada categoria da variável (0, 1, 2, 3). A distribuição é dada por

N	0	1	2	3	Total
Freq.	2	11	12	5	30

2. **Distribuição de frequências relativas:** Constantemente interessa determinar a proporção ou porcentagem de items em cada classe da variável. A frequência relativa de uma classe é igual à fração ou proporção de unidades experimentais que pertencem a uma classe. Particularmente útil quando interessa comparar duas ou mais populações com diferentes tamanhos. Considere o seguinte esquema:



Definição 1.5. Suponha X uma variável com a seguinte distribuição de frequências:

X	x_1^*	x_2^*	$\ldots$	x_k^*	Total
Freq.	n_1	n_2	$\ldots$	n_k	$\sum_{i=1}^{k} n_i = n$

A frequência relativa do i-ésimo valor (x_i^*) é definida por $f_i = \frac{n_i}{n}$. A distribuição de frequências relativas é dada por

X	x_1^*	x_2^*	$\ldots$	x_k^*	Total
Freq.	f_1	f_2	$\ldots$	f_k	$\sum_{i=1}^{k} f_i = 1$

1.4.1.4 Exemplo 2 - Continuação

Suponha que todos 600 funcionários foram questionados e que as distribuições de frequências de E e N neste caso são:

E	f	m	s	Total
Freq.	230	180	190	600

e

N	0	1	2	3	Total
Freq.	30	230	235	105	600

As distribuições de frequências relativas de E e N para a amostra e para a população são dadas, respectivamente, por

E	f	m	s	Total
Freq.	0.4	0.27	0.33	1

E	f	m	s	Total
Freq.	0.38	0.3	0.32	1

e

N	0	1	2	3	Total
Freq.	0.07	0.37	0.4	0.16	1

N	0	1	2	3	Total
Freq.	0.05	0.38	0.39	0.18	1

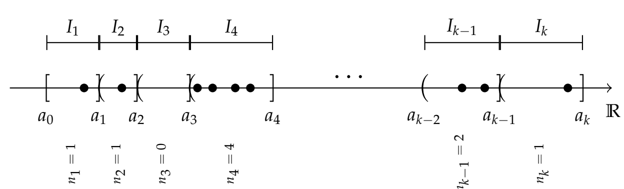
1.4.1.5 Casos quantitativos contínuo e discreto (muitas observações diferentes)

Neste caso: $x_1, \ldots, x_n$ tem muitas (ou todas) observações diferentes; o método anterior nem resume e nem extrai informação dos dados.

Alternativa - Agrupar valores próximos em intervalos, assim

1. Escolha $a_0 < a_k$ tais que $x_1, \ldots, x_n \in [a_0, a_k]$;
2. Fixe uma partição $I_1 = [a_0, a_1]$, $I_i = (a_{i-1}, a_i]$, $i = 2, \ldots, k$, isto é,
$$\bigcup_{i=1}^{k} I_i = [a_0, a_k] \text{ e } I_i \cap I_j = \emptyset, i \neq j;$$
3. A frequência n_i é o número de observações do intervalo I_i.

Esquematicamente:



Definição 1.6. A distribuição de frequências dos dados $x_1, \ldots, x_n$ é definida por

X	$[a_0, a_1]$	$(a_1, a_2]$	$\ldots$	$(a_{k-1}, a_k]$	Total
Freq.	n_1	n_2	$\ldots$	n_k	$\sum_{i=1}^{k} n_i = n$

Amostras de tamanhos diferentes são comparadas por meio de suas frequências relativas.

Definição 1.7. A distribuição de frequências relativas de $x_1, \ldots, x_n$ é definida por

X	$[a_0, a_1]$	$(a_1, a_2]$	$\ldots$	$(a_{k-1}, a_k]$	Total
Freq.	f_1	f_2	$\ldots$	f_k	$\sum_{i=1}^{k} f_i = 1$

onde $f_i = n_i/n$, $i = 1, 2, \ldots, k$.

Problema: Se os intervalos forem de grande amplitude, possivelmente haverá frequência relativa alta. Pode-se contornar esta situação por meio de duas alternativas:

A: Estabelecer intervalos de mesma amplitude da seguinte forma

- $*$ fixa-se a quantidade k de intervalos;
- $*$ toma-se $a_0 = \min(x_1, x_2, \ldots, x_n)$ e $a_k = \max(x_1, x_2, \ldots, x_n)$;
- $*$ obtem-se a amplitude dos intervalos por $h = \frac{a_k - a_0}{k}$;
- $*$ calcula-se o limite superior do i-ésimo intervalo por

$$a_i = a_{i-1} + h = a_0 + ih, i = 1, 2, \ldots, k.$$

Observações:

1. Quando k aumenta, a perda de informação e capacidade de resumo dos dados diminuem;
2. Quando k diminui, aumentam a perda de informação e a capacidade de resumo dos dados;
3. Sturges (1926) sugere a utilização de $k = 1 + 3.322 \log_{10} n$.

B: Utilizar a frequência relativa por unidade de medida da variável em estudo. Isto é, a **densidade**.

Definição 1.8. Um intervalo I_i com frequência relativa f_i tem densidade $d_i = \frac{f_i}{h_i}$, onde h_i é a amplitude deste intervalo

Assim:

- $*$ A densidade não é influenciada pela amplitude do intervalo;
- $*$ Quanto maior (menor) a densidade maior (menor) a frequência por unidade de medida da variável.

1.4.1.6 Exemplo 3

Um determinado fabricante de baterias de carro deseja determinar a vida útil (em anos) de seus produtos. Para isso, a vida útil de $n = 40$ baterias foi observada e os dados são apresentados abaixo.

2.0	4.1	3.5	4.5	3.2	3.7	3.0	2.6
3.4	1.7	3.1	3.3	3.8	3.1	4.7	3.7
1.9	4.3	3.4	3.6	2.9	3.3	3.9	3.1
3.3	3.1	3.7	4.4	3.2	4.1	1.9	3.4
4.7	3.8	3.2	2.6	3.9	3.0	4.2	3.5

Primeiramente, calcula-se a distribuição de frequências para os intervalos

$$I_1 = [1.5, 2.0], I_2 = (2.0, 2.5], \ldots, I_7 = (4.5, 5.0].$$

Tem-se que a distribuição de frequências é dada por

X	$[1.5, 2.0]$	$(2.0, 2.5]$	$(2.5, 3.0]$	$(3.0, 3.5]$	$(3.5, 4.0]$	$(4.0, 4.5]$	$(4.5, 5.0]$	Total
Freq.	4	0	5	15	9	6	1	40

A distribuição de frequências relativas é dada por

X	$[1.5, 2.0]$	$(2.0, 2.5]$	$(2.5, 3.0]$	$(3.0, 3.5]$	$(3.5, 4.0]$	$(4.0, 4.5]$	$(4.5, 5.0]$	Total
Freq.	0.100	0.000	0.125	0.375	0.225	0.150	0.025	1.000

Observações:

1. O intervalo $(2.0, 2.5]$ tem frequência relativa $f_2 = 0$;
2. Pode-se unir $[1.5, 2.0]$ e $(2.0, 2.5]$ para que não haja nenhum intervalo com frequência 0.

Utilizando as densidades, tem-se:

X	$[1.5, 2.5]$	$(2.5, 3.0]$	$(3.0, 3.5]$	$(3.5, 4.0]$	$(4.0, 4.5]$	$(4.5, 5.0]$
Freq.	0.100	0.250	0.750	0.450	0.300	0.050

Pode-se também determinar a distribuição de frequências com intervalos iguais.

1. Seja $k = 1 + 3.322 \log_{10} 40 = 6.322 \approx 6$;
2. Tem-se

$$a_0 = \min(x_1, \ldots, x_n) = 1.7 \text{ ''e ``} a_k = a_6 = \max(x_1, \ldots, x_n) = 4.7;$$

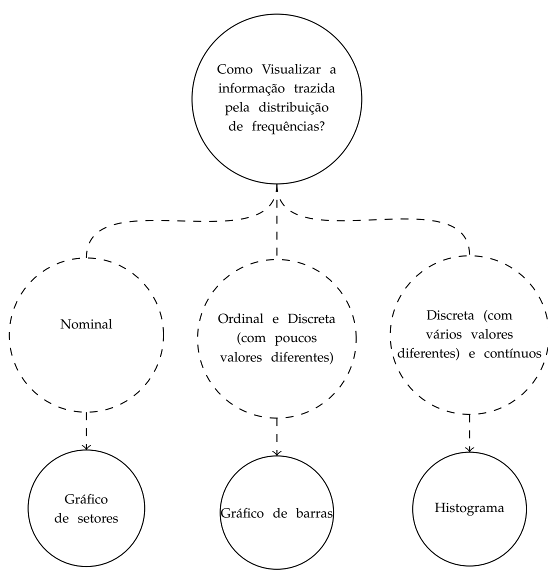
3. A amplitude dos intervalos será $h = \frac{a_k - a_0}{k} = \frac{4.7-1.7}{6} = \frac{3.0}{6} = 0.5$.

A distribuição de frequências neste caso é dada por

X	$[1.7, 2.2]$	$(2.2, 2.7]$	$(2.7, 3.2]$	$(3.2, 3.7]$	$(3.7, 4.2]$	$(4.2, 4.7]$	Total
Freq.	4	2	10	13	7	4	40

1.4.2 Representação Gráfica

A representação gráfica dos dados permite uma melhor visualização e uma análise detalhada dos dados analisados. Existem diversos tipos de gráficos que podem ser usados de acordo com o tipo de variável de interesse. Os gráficos mais usados são de setores, barras e histograma. Esquematicamente:

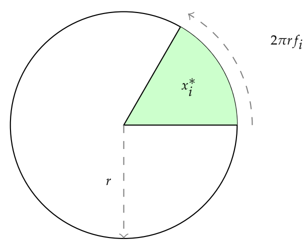


1. **Gráfico de setores:** É construido da seguinte maneira:

    1. suponha que os valores diferentes da variável são $x_1^*, \ldots, x_k^*$;

2. sejam $f_1, \ldots, f_k$ as frequências relativas desses valores;
3. desenhe um círculo de raio arbitrário;
4. represente no círculo desenhado o valor x_i^* através de um arco de circunferência de ângulo proporcional a sua frequência relativa. Isto é, um arco de comprimento $2\pi r f_i$, onde r é o raio arbitrário.

Esquematicamente:

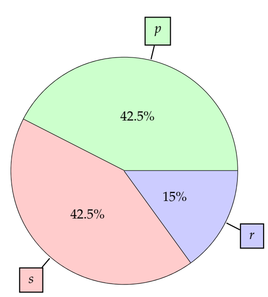


1.4.2.1 Exemplo 1 - Continuação

Relembrando, a distribuição de frequências relativas da variável X = "Tipo de música preferida" é dada por

X	p	r	s	Total
Freq.	0.425	0.150	0.425	1.000

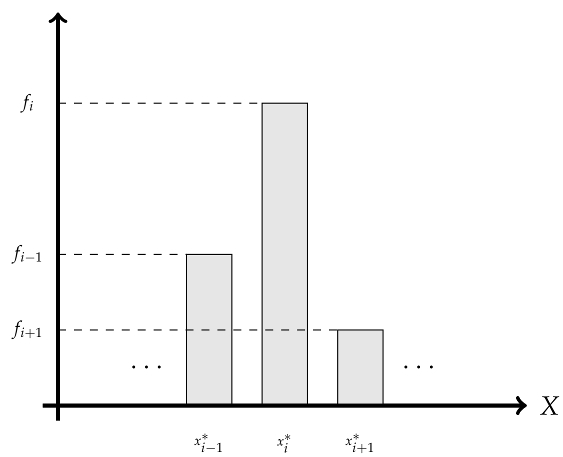
Assim, o gráfico de setores de X para a amostra observada no Exemplo 1 é dado abaixo.



2. **Gráfico de barras:** Construido da seguinte maneira:

1. suponha que os valores diferentes da variável são $x_1^* < \cdots < x_k^*$;
2. sejam $f_1, \ldots, f_k$ as frequências relativas desses valores;
3. desenhe o eixo cartesiano;
4. no eixo das abscissas, emcima dos valores $x_1^*, \ldots, x_k^*$, posicione retângulos (barras);
5. as bases de todos retângulos devem ter o mesmo comprimento escolhido arbitrariamente;
6. a altura de um retângulo é dada pela frequência relativa do valor de X que é representado por ele. Isto é, a altura da barra que representa o valor x_i^* é dada por f_i.

Esquematicamente:

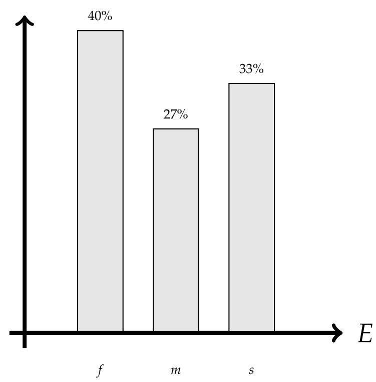


1.4.2.2 Exemplo 2 - Continuação

Relembrando a distribuição de frequências relativas de E:

E	f	m	s	Total
Freq.	0.4	0.27	0.33	1

O gráfico de barras para a variável E nesta amostra é dado por

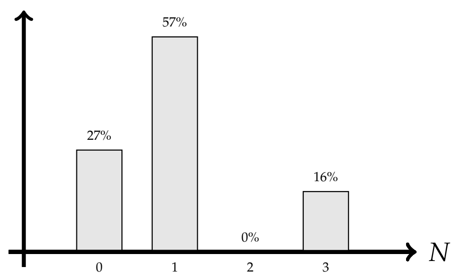


Observação: Em um gráfico de barras, um valor com frequência 0 deve ser representado.

Suponha que a variável número de filhos (N) do Exemplo 2 tivesse a seguinte distribuição de frequências relativas:

N	0	1	2	3	Total
Freq.	0.27	0.57	0.00	0.16	1

O gráfico de barras neste caso é dado abaixo.



3. **Histograma:** É construido da seguinte maneira:

   1. suponha que a variável X forneceu a seguinte distribuição de densidades:

X	$[a_0, a_1]$	$(a_1, a_2]$	$\ldots$	$(a_{k-1}, a_k]$
Freq.	d_1	d_2	$\ldots$	d_k

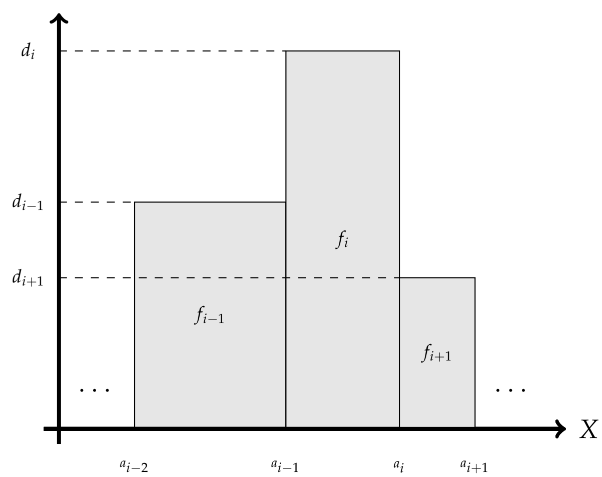
   2. desenhe o eixo cartesiano;

3. no eixo das abscissas, encima do intervalo $I_i = (a_{i-1}, a_i]$ posicione um retângulos;

4. a base desse retângulo deve corresponder ao intervalo I_i. Portanto, **não deve haver espaço entre as barras**;

5. a altura desse retângulo é dada pela densidade do i-ésimo intervalo, d_i.

Esquematicamente:



1.4.2.3 Exemplo 3 - Continuação

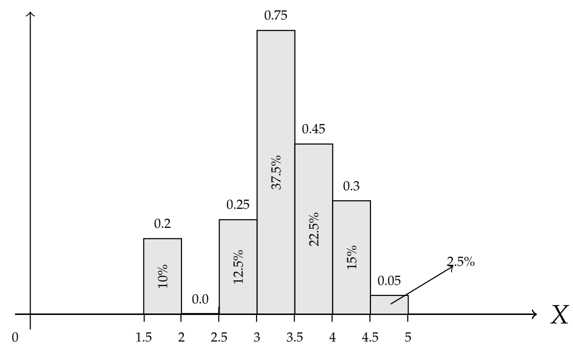
Relembrando a distribuição de frequências relativas do Exemplo 3.

X	$[1.5, 2.0]$	$(2.0, 2.5]$	$(2.5, 3.0]$	$(3.0, 3.5]$	$(3.5, 4.0]$	$(4.0, 4.5]$	$(4.5, 5.0]$	Total
Freq.	0.100	0.000	0.125	0.375	0.225	0.150	0.025	1.000

Assim, a distribuição de densidades é dada abaixo.

X	$[1.5, 2.0]$	$(2.0, 2.5]$	$(2.5, 3.0]$	$(3.0, 3.5]$	$(3.5, 4.0]$	$(4.0, 4.5]$	$(4.5, 5.0]$
Freq.	0.200	0.000	0.250	0.750	0.450	0.300	0.050

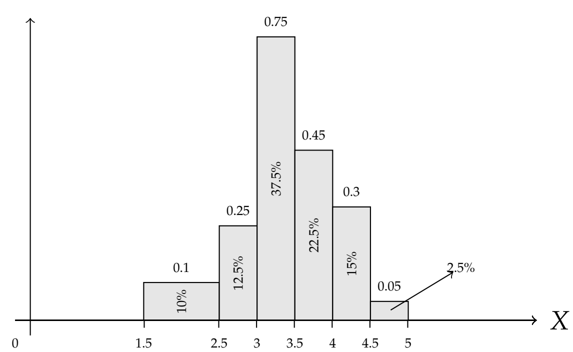
O histograma por sua vez é dado por



Relembrando, que os dois primeiros intervalos foram unidos a distribuição de densidades é dada abaixo.

X	$[1.5, 2.5]$	$(2.5, 3.0]$	$(3.0, 3.5]$	$(3.5, 4.0]$	$(4.0, 4.5]$	$(4.5, 5.0]$
Dens.	0.100	0.250	0.750	0.450	0.300	0.050

Neste caso, o histograma é dado abaixo. Note-se que não há mais intervalos de frequência 0.



1.4.3 Medidas Numéricas

O objetivo principal é expressar alguma característica dos elementos da amostra. As principais características de interesse dos dados são:

- centro (localização); e
- dispersão.

A vantagem das medidas numéricas é que os dados podem ser resumidos por meio de um único número que representa características importantes. Entretanto, como desvantagem tem-se perda de informação.

1. **Medidas de tendência central:** Rempresentam o valor típico dos dados observados. Esse valor pode ser o ponto de equilíbrio, o ponto central ou o ponto de maior frequência dos dados. As principais medidas de tendência central são:

   - média;
   - mediana;
   - moda.

1.4.3.1 Média

Definição 1.9. Dadas as observações $x_1, \ldots, x_n$, a média é definida por

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{x_1 + \cdots + x_n}{n}.$$

Suponha que os valores são apresentados na forma da seguinte distribuição de frequências:

X	x_1^*	x_2^*	$\ldots$	x_k^*	Total
Freq.	n_1	n_2	$\ldots$	n_k	$\sum_{i=1}^{k} n_i = n$

A média pode ser calculada como

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{k} n_j x_j^* = \sum_{i=1}^{k} f_j x_j^*,$$

onde $f_j = \frac{n_j}{n}$ é a frequência relativa do j-ésimo valor.

Suponha que as observações são dadas na forma da seguinte distribuição de frequências intervalar:

X	$[a_0, a_1]$	$(a_1, a_2]$	$\ldots$	$(a_{k-1}, a_k]$	Total
Freq.	n_1	n_2	$\ldots$	n_k	$\sum_{i=1}^{k} n_i = n$

Problema: Os valores observados são desconhecidos.

Para $i = 1, \ldots, k$, deve-se **aproximar** os valores do intervalo $(a_{i-1}, a_i]$ pelo ponto médio $x_i^m = \frac{a_{i-i}+a_i}{2}$ desse intervalo. Agora, a média é aproximada por

$$\bar{x} \approx \frac{1}{n}\sum_{i=1}^{k} n_j x_j^m = \sum_{i=1}^{k} f_j x_j^m,$$

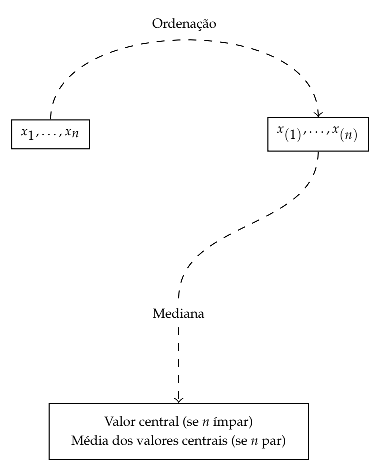
onde $f_j = \frac{n_j}{n}$ é a frequência relativa do j-ésimo intervalo.

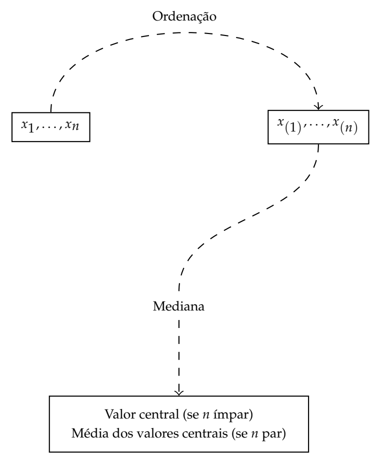
Características:

- é influenciada por valores atípicos;
- não recomendada em dados assimétricos;
- só é calculada em variáveis quantitativas;
- no caso intervalar, quanto menor for o número de intervalos, menos precisa será a aproximação para a média.

1.4.3.2 Mediana

A Mediana divide os dados de forma que 50% deles são menores ou iguais e 50% deles são maiores ou iguais que a mediana. Esquematicamente





Definição 1.10. Dadas as observações $x_1, \dots, x_n$, sejam $x_{(1)}, \dots, x_{(n)}$ esses valores ordenados, isto é, $x_{(1)} \leq \cdots \leq x_{(n)}$. A mediana é definida por

$$Med = \begin{cases} x_{\left(\frac{n+1}{2}\right)}, & \text{se } n \text{ é ímpar;} \\ \frac{x_{\left(\frac{n}{2}\right)} + x_{\left(\frac{n}{2}+1\right)}}{2}, & \text{se } n \text{ é par.} \end{cases}$$

Problema: No caso intervalar, o valor da mediana deve ser aproximado.

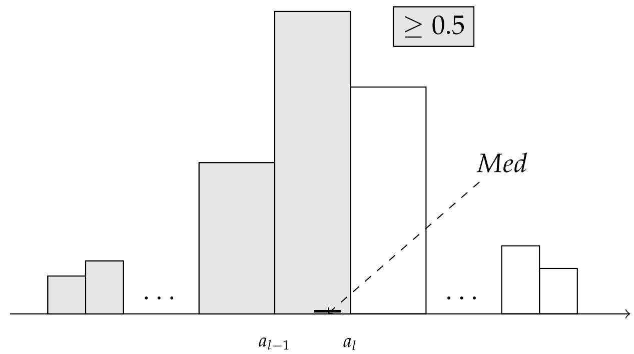
Solução: Suponha a distribuição de frequências relativas abaixo:

X	$[a_0, a_1]$	$(a_1, a_2]$	$\cdots$	$(a_{k-1}, a_k]$	Total
Freq.	f_1	f_2	$\cdots$	f_k	$\sum_{i=1}^{k} f_i = 1$
Freq. acum.	$F_1 = f_1$	$F_2 = F_1 + f_2$	$\cdots$	$F_k = F_{k-1} + f_k = 1$	$\times$

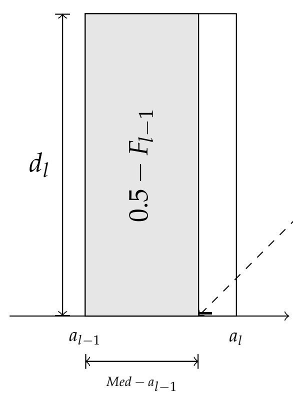
Denominamos $F_1 = f_1$ e $F_i = F_{i-1} + f_i$, $i = 2, \dots, k$, de frequências acumuladas. Seja I_l o primeiro intervalo tal que $F_l \geq 0.5$. Isto é, l é o menor valor em $\{1, \dots, k\}$ tal que $F_l \geq 0.5$. Deve-se aproximar a mediana por

$$Med \approx a_{l-1} + \frac{0.5 - F_{l-1}}{d_l}.$$

Esquematicamente:



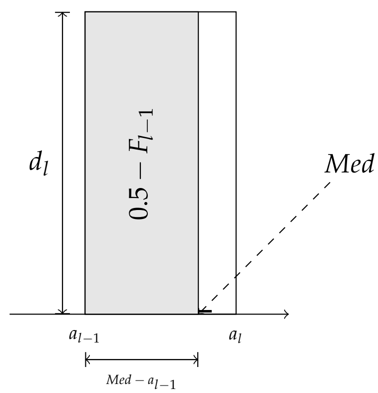
Olhando o retângulo de base $(a_{l-1}, a_l]$, tem-se o seguinte gráfico.



Assim,

$$0.5 - F_{l-1} = d_l \times (Med - a_{l-1}) \Rightarrow Med \approx a_{l-1} + \frac{0.5 - F_{l-1}}{d_l}.$$

Olhando o retângulo de base $(a_{l-1}, a_l]$, tem-se o seguinte gráfico.



Daí,

$$0.5 - F_{l-1} = d_l \times (Med - a_{l-1}) \Rightarrow Med \approx a_{l-1} + \frac{0.5 - F_{l-1}}{d_l}.$$

Observações:

1. sem interpretação física;
2. depende apenas da posição e não do valor;
3. tem menos influência de dados atípicos;
4. pode ser usada em variáveis qualitativas ordinais.

1.4.3.3 Moda

Definição 1.11. Dadas as observações $x_1, \ldots, x_n$, se $x_1^*, \ldots, x_k^*$ denotarem os k valores diferentes, a moda é dada pelo valor com maior frequência.

Observação: Um conjunto de dados pode ser amodal, unimodal, bimodal, etc.

Problema: em variáveis contínuas, frequentemente, observam-se poucos valores repetidos. Assim, na maioria dos casos, esse tipo de dado é amodal.

Alternativa: utilizar a classe com maior densidade, ou **classe modal**.

Aproximação: assim, a moda pode ser aproximada pelo valor médio da classe modal.

Observações: A moda de um conjunto de dados

1. representa o(s) valor(es) mais provável(eis);
2. é muito indicada em dados multimodais;
3. não é afetada por dados atípicos;
4. pode ser usada em variáveis qualitativas.

2. **Medidas de dispersão**

   São medidas para representar quão disperso os dados estão. As medidas de dispersão mais usadas são:

   – amplitude amostral;
   – variância;
   – desvio-padrão;
   – coeficiente de variação.

1.4.3.4 Amplitude

Sejam $x_1, \ldots, x_n$. que representam os elementos da amostra e sejam $x_{(1)}, \ldots, x_{(n)}$. os elementos ordenados:

Definição 1.12. A amplitude da amostra é definida por

$$Amp = x_{(n)} - x_{(1)}.$$

Vantagens:

– cálculo rápido;
– fácil interpretação;
– alto impacto de dados atípicos.

1.4.3.5 Variância

Definição 1.13. Variância populacional:

$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2.$$

Variância amostral:

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2.$$

Observações:

1. mais robusta a dados atípicos;
2. se os dados, por exemplo, são expressos em cm a variância é em cm^2;
3. é possível mostrar que $\sum_{i=1}^{n}(x_i - \bar{x})^2 = \sum_{i=1}^{n} x_i^2 - n\bar{x}^2$;
4. portanto $\sigma^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \bar{x}^2$;
5. e $s^2 = \frac{1}{n-1}\sum_{i=1}^{n} x_i^2 - \frac{n}{n-1}\bar{x}^2$;
6. para dados apresentados em frequências intervalares, temos que
$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{k} n_i (x_i^m)^2 - \bar{x}^2$$
e que
$$s^2 = \frac{1}{n-1}\sum_{i=1}^{k} n_i (x_i^m)^2 - \frac{n}{n-1}\bar{x}^2,$$
onde x_i^m é o ponto médio do i-ésimo intervalo.

1.4.3.6 Desvio-padrão

Para remediar o fato de a variância ser expressa na unidade de medida da variável ao quadrado, podemos calcular o **desvio-padrão**.

Definição 1.14. O desvio-padrão populacional é dado por

$$\sigma = \sqrt{\sigma^2} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2}.$$

O desvio-padrão amostral é dado por

$$s = \sqrt{s^2} = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2}.$$

A grande vantagem do desvio-padrão é que ele é expresso na mesma unidade de medida dos dados.

1.4.3.7 Coeficiente de variação

As medidas de dispersão apresentadas são influenciadas pela "grandeza" da variável estudada. É comum então representar o desvio-padrão como percentual da média. Denomina-se essa medida de coeficiente de variação.

Definição 1.15. O coeficiente de variação populacional é dado por

$$CV = \frac{\sigma}{\mu},$$

onde μ é a média populacional. O coeficiente de variação amostral é dado por

$$cv = \frac{s}{\bar{x}},$$

onde $\bar{x}$ é a média amostral.

Vantagem: Duas populações com médias muito diferentes podem ter suas disperções comparadas através do coeficiente de variação.

1.4.3.8 Exemplo 2 - Continuação

Retornando ao Exemplo 2. Tem-se que a distribuição de frequências é dada por

N	0	1	2	3	Total
Freq.	2	11	12	5	30

Assim,

$$\bar{x} = \frac{2 \cdot 0 + 11 \cdot 1 + 12 \cdot 2 + 5 \cdot 3}{30} \approx 1.67,$$

$$Med = \frac{x_{(15)} + x_{(16)}}{2} = \frac{2+2}{2} = 2,$$

e

$$Mod = 2.$$

$$Amp = x_{(30)} - x_{(1)} = 3 - 0 = 3,$$

$$s^2 \approx \frac{11 \cdot 1^2 + 12 \cdot 2^2 + 5 \cdot 3^2}{29} - \frac{30 \cdot 1.67^2}{29} = \frac{104}{29} - \frac{83.667}{29} \approx 0.701,$$

$$s \approx \sqrt{0.701} \approx 0.837$$

e

$$cv = \frac{s}{\bar{x}} = \frac{0.837}{1.67} \approx 0.501 = 50.1\%.$$

1.4.3.9 Exemplo 3 - Continuação

Retornando ao Exemplo 3. Tem-se

X	$[1.5, 2.5]$	$(2.5, 3.0]$	$(3.0, 3.5]$	$(3.5, 4.0]$	$(4.0, 4.5]$	$(4.5, 5.0]$
Pt. Méd.	2.00	2.75	3.25	3.75	4.25	4.75
Freq.	4	5	15	9	6	1
Freq.	0.100	0.125	0.375	0.225	0.150	0.025
F_i	0.100	0.225	0.600	0.825	0.975	1.000
Dens.	0.100	0.250	0.750	0.450	0.300	0.050

Assim, temos que

$$\bar{x} \approx 2 \cdot 0.1 + 2.75 \cdot 0.125 + \cdots + 4.75 \cdot 0.025 \approx 3.363$$

$$Med = a_{l-1} + \frac{0.5 - F_{l-1}}{d_l} = 3 + \frac{0.5 - 0.225}{0.750} \approx 3.367,$$

e

$$Classe``Modal = (3.0, 3.5] \Rightarrow Mod \approx \frac{3 + 3.5}{2} = 3.25.$$

$$Amp = x_{(40)} - x_{(1)} = 4.7 - 1.7 = 3$$

$$s^2 \approx \frac{4 \cdot 2^2 + 5 \cdot 2.75^2 + \cdots + 1 \cdot 4.75^2}{39} - \frac{40 \cdot 3.363^2}{39} = \frac{469.75}{39} - \frac{452.391}{39} \approx 0.445,$$

$$s \approx \sqrt{0.445} \approx 0.667$$

e

$$cv = \frac{s}{\bar{x}} = \frac{0.667}{3.363} \approx 0.198 = 19.8\%.$$

2

Probabilidade

Os primeiros estudos voltados à teoria das probabilidades tiveram como origem questões relacionadas aos jogos de azar. Cardano, Galileu, Pascal, Fermat e Huygens dedicaram estudos que contribuíram para o que hoje chamamos de enfoque clássico da probabilidade. Neste capítulo, apresentaremos as definições clássica, frequentista, subjetiva e axiomática de probabilidade e mostraremos que, para além das chances de ganho em jogos de azar, a teoria de probabilidades constitui-se na atualidade em um dos ramos da matemática mais explorados pelas ciências em geral. Para isso, iniciaremos este capítulo contrapondo os conceitos de experimento determinístico e aleatório.

2.1 Experimentos aleatórios e determinísticos

Suponha a situação em que um engenheiro elétrico pretende medir a corrente em um fio fino de cobre com o objetivo de estudar sua condutividade elétrica. A esta atividade ou processo denominaremos **experimento**.

Se repetíssemos o experimento acima diversas vezes, observaremos que os resultados diferem levemente de uma repetição para outra, por mais controlada que fosse a situação. Exemplos de variáveis que podem influenciar no experimento acima:

- variações na temperatura ambiente no momento da realização;
- variações nos equipamentos utilizados para realizar a medição;
- impurezas na composição química do fio, se a medição é realizada em diversas localidades;

- impulsos na fonte da corrente;
- entre uma infinidade de outros fatores.

Não importa quão cuidadosamente tenha sido conduzido o experimento, sempre existem variáveis de perturbação (ou ruído) que não são controladas. Isso provoca aleatoriedade dos resultados obtidos em diferentes realizações do experimento.

Definição 2.1. Dizemos que um experimento é **aleatório** se, mesmo quando repetido sob condições idênticas, não é possível predizer com absoluta certeza o seu resultado. Frequentemente, experimentos aleatórios são denotados pela letra E. Em contrapartida, um experimento é denominado **determinístico** se, quando repetido sob as mesmas condições, conduz a resultados idênticos, sempre.

Alguns outros exemplos de experimentos aleatórios:

1. $E_1 =$ "uma peça é fabricada em uma linha de produção e, depois de inspecionada, é classificada como 'defeituosa' (D) ou 'não defeituosa' (N)";
2. $E_2 =$ "o número de ligações que chega em determinado dia a um *call center* é observado";
3. $E_3 =$ "o tempo em minutos necessário para realizar uma reação química é observado".

2.2 Espaço amostral de um experimento e eventos

Uma vez identificado que um experimento realizado é aleatório, voltamos a nossa atenção aos possíveis resultados que esse experimento pode fornecer. Note que, embora não saibamos o resultado que um experimento fornecerá, devemos poder listar todos os seus possíveis resultados.

Definição 2.2. O CONJUNTO de TODOS os possíveis resultados de um experimento aleatório é denominado **espaço amostral** do experimento. Frequentemente, o espaço amostral é denotado pela letra Ω.

Nos exemplos anteriores os espaços amostrais seriam:

1. $\Omega_{E_1} = \{D, N\}$;
2. $\Omega_{E_2} = \{0, 1, 2 \ldots\}$;
3. $\Omega_{E_3} = \{\omega : \omega > 0\} = (0, \infty)$.

Suponha que os exemplos anteriores sejam repetidos 2 vezes. Os espaços amostrais agora passam a ser:

1. $\Omega_1^* = \Omega_{E_1} \times \Omega_{E_1} = \{DD, DN, ND, NN\}$;
2. $\Omega_2^* = \Omega_{E_2} \times \Omega_{E_2} = \{(0,0), (0,1), (0,2), \ldots, (1,0), (1,1), (1,2), \ldots\}$;
3. $\Omega_3^* = \Omega_{E_3} \times \Omega_{E_3} = \{(\omega_1, \omega_2) : \omega_1, \omega_2 > 0\}$.

Em geral, quando conduzimos um experimento, não estamos interessados apenas em um resultado em particular, mas sim em uma coleção destes. Isto é, em um **subconjunto** do espaço amostral.

Definição 2.3. Um **evento** é qualquer subconjunto do espaço amostral de um experimento aleatório. Frequentemente, eventos são denotados por letras iniciais do alfabeto maiúsculas: $A, B, C, \ldots$.

Exemplos de eventos dos espaços amostrais Ω_1^*, Ω_2^* e Ω_3^*:

1. $A_1 =$ "Pelo menos uma peça defeituosa" $= \{DD, DN, ND\}$;
2. $A_2 =$ "Receber 4 ligações no total" $= \{(0,4), (1,3), (2,2), (3,1), (4,0)\}$ $= \{(\omega_1, \omega_2) \in \Omega_2^* : \omega_1 + \omega_2 = 4\}$;
3. $A_3 =$ "a soma dos tempos de duração das reações é de 5 minutos ou mais" $= \{(\omega_1, \omega_2) \in \Omega_3^* : \omega_1 + \omega_2 \geq 5\}$

2.2.1 Operações com eventos

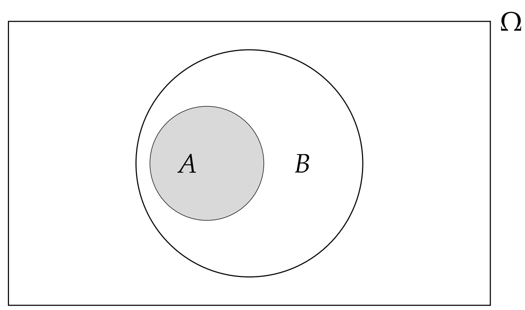
Um evento é essencialmente um conjunto, de forma que as relações e resultados da teoria elementar dos conjuntos podem ser usados para o estudo dos eventos.

Definição 2.4. Sejam A, B eventos de Ω, dizemos que A está contido em B se todos elementos de A pertencem a B, ou seja,

$$\omega \in A \Rightarrow \omega \in B.$$

Neste caso, escrevemos $A \subset B$. Alternativamente, podemos dizer que o evento B contém A ($B \supset A$).

Um sistema de organização e visualização de eventos e manipulações de eventos bastante útil é o diagrama de Venn. Nele, o espaço amostral Ω é representado por um retângulos e quaisquer outros eventos em Ω são representados por curvas fechadas, geralmente círculos. Por exemplo, no figura abaixo, apresentamos o Diagrama de Venn representando $A \subset B$ ou, similarmente, $B \supset A$.



Dois eventos A e B, ambos contidos em Ω, são considerados iguais se, e somente se, $A \subset B$ e $B \subset A$. Denotamos essa situação por $A = B$. O conjunto vazio é representado simbologicamente por $\varnothing$ e está contido em qualquer outro evento do espaço amostral.

Sejam $A, B \subset \Omega$ dois eventos. Podemos criar novos eventos a partir de A e B através de operações de conjuntos. Operações básicas:

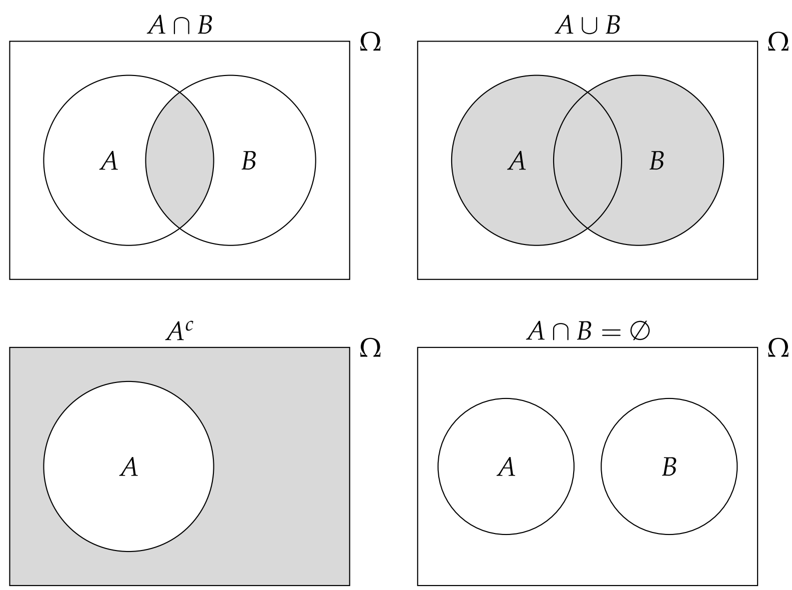
- A **intersecção** de A e B é denotada por $A \cap B$ e representa a ocorrência **simultânea** dos eventos A **e** B;
- A **união** de A e B é denotada por $A \cup B$ e representa a ocorrência dos eventos A **ou** B, **ou** de ambos;
- O **complementar** de A é denotado por A^c e representa a **não** ocorrência do evento A.

Formalmente, temos que:

- $A \cap B = \{\omega \in \Omega : \omega \in A \text{ e } \omega \in B, \text{simultaneamente}\}$;
- $A \cup B = \{\omega \in \Omega : \omega \in A, \text{ ou } \omega \in B, \text{ ou } \omega \in A \cap B\}$;
- $A^c = \{\omega \in \Omega : \omega \notin A\}$.

Definição 2.5. Dois eventos $A, B \subset \Omega$ são disjuntos ou mutuamente exclusivos se, e somente se, eles não puderem ocorrer simultaneamente, ou seja, $A \cap B = \emptyset$.

Na figura a seguir, a área sombreada no Diagrama de Venn representa o evento resultante da operação entre dois eventos A e B, ambos contidos em Ω.

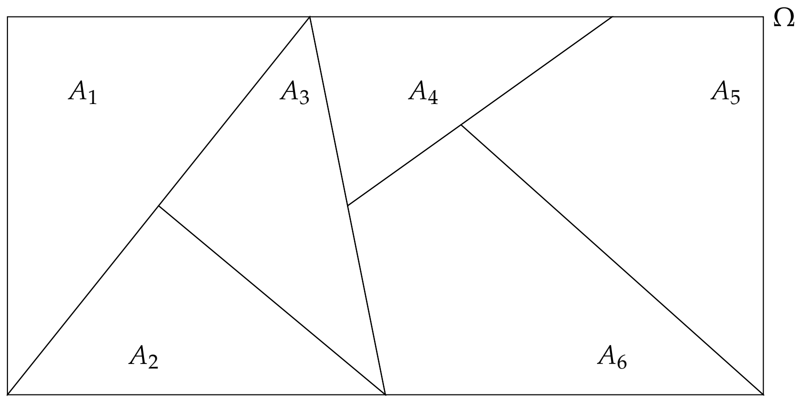


Definiremos agora o conceito de partição de um evento bastante útil no cálculo de probabilidades.

Definição 2.6. (Partição de um evento) Seja A um evento não-vazio. Uma família de eventos A_1, A_2, ..., A_k, todos não-vazios, é uma partição de A se, e somente se, satisfaz as seguintes condições:

1. $A_1 \cup A_2 \cup ... \cup A_k = A$ (eventos exaustivos) e

2. $A_1 \cap A_j = \emptyset, \forall\, i \neq j$ (eventos mutuamente exclusivos).



Figura"2.2.1
Partição do evento A em eventos mutuamente exclusivos.

Outros importantes resultados com respeito a operação entre dois eventos são dadas a seguir:

- $(A^c)^c = A$;
- $(A \cap B)^c = A^c \cup B^c$;
- $(A \cup B)^c = A^c \cap B^c$;
- $A \cup A^c = \Omega$ e $A \cap A^c = \emptyset$;
- $A \cup \emptyset = A$ e $A \cup \Omega = \Omega$;
- $A \cap \emptyset = \emptyset$ e $A \cap \Omega = A$;
- $\emptyset^c = \Omega$ e $\Omega^c = \emptyset$;
- $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$;
- $(\bigcap_{i=1}^{n} A_i)^c = (A_1 \cap \cdots \cap A_n)^c = A_1^c \cup \cdots \cup A_n^c = \bigcup_{i=1}^{n} A_i^c$;
- $(\bigcup_{i=1}^{n} A_i)^c = (A_1 \cup \cdots \cup A_n)^c = A_1^c \cap \cdots \cap A_n^c = \bigcap_{i=1}^{n} A_i^c$.

2.3 Probabilidade

Como visto na seção anterior, a todo experimento aleatório temos associado uma espaço amostral Ω. O objetivo da probabilidade é atribuir a cada evento A em Ω um número que nos forneça a informação de quão verossímil é a ocorrência desse evento A. Tal número, denotado por $P(A)$, nos dará uma medida da chance do evento A ocorrer. De maneira geral, podemos atribuir probabilidades aos diferentes resultados de experimento aleatório e estas devem satisfazer aos seguintes axiomas (propriedades básicas) de probabilidade.

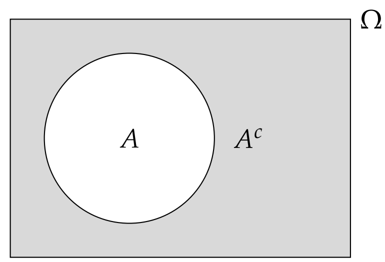
Definição 2.7. Uma medida de probabilidade é qualquer função de eventos $P(\cdot)$ que satisfaça as seguintes propriedades:

1. $P(\Omega) = 1$;
2. $P(A) \geq 0$, se $A \subset \Omega$ é evento;
3. Se $A_1, A_2, ..., A_n$ são eventos mutuamente exclusivos (disjuntos) em Ω então
$$P(A_1 \cup A_2 \cup ... \cup A_n) = \sum_{i=1}^{n} P(A_i)$$
.

Sejam Ω espaço amostral e $A, B \subset \Omega$ eventos. Uma função de probabilidade definida de acordo com os axiomas de probabilidade satisfaz as seguintes propriedades:

1. $P(A^c) = 1 - P(A)$;
2. $P(\emptyset) = 0$;
3. $P(A) \leq P(B)$, se $A \subset B$;
4. $P(B - A) = P(B) - P(A \cap B)$, onde $B - A = B \cap A^c$;
5. $P(B - A) = P(B) - P(A)$, se $A \subset B$;
6. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

Todas 6 propriedades anteriormente apresentadas podem ser demonstradas a partir do conhecimento da operação entre eventos e da Definição 2.7. Por exemplo, mostremos que se $A \subset B$ então $P(A^c) = 1 - P(A)$. Para isso, considere o seguinte seguinte diagrama de Venn:



Note que $A \cup A^c = \Omega \Rightarrow P(A \cup A^c) = P(\Omega)$. Da Definição 2.7, temos que $P(\Omega) = 1$. Considerando $n = 2$, $A_1 = A$, $A_2 = A^c$ e lembrando que os eventos A e A^c são disjuntos, fazemos uso da Propriedade 3 também da Definição 2.7 e obtemos que

$$\begin{aligned} P(A \cup A^c) &= P(\Omega) \Leftrightarrow \quad (2.1)\\ P(A) + P(A^c) &= 1 \Leftrightarrow \quad (2.2)\\ P(A) &= 1 - P(A^c). \end{aligned}$$

Os axiomas de probabilidade não determinam probabilidades, mas apenas estabelecem regras que garantem que as medidas de probabilidade sejam consistentes com noções intuitivas da medida. Como atribuir então probabilidades aos eventos do espaço amostral? Para isso, vejamos as interpretações clássica, frequentista e subjetiva de probabilidade.

2.3.1 Interpretação clássica

Baseada no conceito de que todos os resultados do experimento aleatório tem a mesma chance de ocorrência e, portanto, seriam equiprováveis. O cálculo de probabilidades por esta interpretação passa por identificar todos os resultados do experimento aleatório em um número finito de casos mutuamente exclusivos, simétricos e igualmente possíveis de ocorrer tendo sua origem associada aos jogos de azar. É esta interpretação que nos leva a afirmar que a probabilidade de obter uma "cara" no lançamento de uma moeda é $1/2$. De maneira geral, se $\Omega = \{\omega_1, \ldots, \omega_n\}$ então

$$P(\omega_i) = \frac{1}{n}, \quad i = 1, \ldots, n.$$

Interessante notar também que as técnicas de análise combinatória são amplamente utilizadas aqui já que muitas vezes se requer a contagem de número de casos favoráveis aos evento dentre o número de casos possíveis do experimento aleatório para se poder determinar a medida de probabilidade via esta interpretação.

2.3.2 Interpretação frequentista

Quando o espaço amostral Ω for infinito enumerável, ou seja, $\Omega = \{\omega_1, \omega_2, \ldots\}$, a interpretação clássica de probabilidade não pode ser utilizada. A interpretação frequentista sustenta que a probabilidade de um certo evento pode ser obtida avaliando a frequência relativa com que este evento ocorre se o experimento aleatório pudesse ser replicado infinitas vezes, de maneira idêntica e independente. No séc. XVIII, Buffon lançou 4040 vezes uma moeda, das quais 2048 desses resultou em "cara", obtendo uma frequencia relativa de 0,5046. Karl Pearson repetiu o mesmo experimento 24000 vezes e obteve uma frequência relativa de caras de 0,5005.

Observe que para a interpretação frequentista não são necessárias as hipóteses de equiprobabilidade dos eventos elementares, tão pouco da finitude do espaço dos resultados, superando-se portanto as duas restrições impostas pela definição clássica. No entanto, vale ressaltar que, na prática, nem todos os experimentos aleatórios podem ser replicados por diversas vezes sob as mesmas condições.

2.3.3 Interpretação subjetiva

Baseada no conceito de que a probabilidade de um evento A, $P(A)$, representa o "grau de crença" que o observador tem a respeito da ocorrência do evento A. Nesse ponto de vista, indivíduos diferentes podem atribuir probabilidades diferentes para o mesmo evento A. Por exemplo, um paciente doente submetido a um novo tipo de cirurgia quer saber a sua probabilidade de cura. Não podemos obter essa probabilidade baseada na repetição do experimento (cirurgia), uma vez que a técnica é pioneira. Esta probabilidade será atribuída com base no conhecimento do cirurgião e sua experiência pessoal.

Independente de qual interpretação fundamente a obtenção da medida de probabilidade, TODAS devem satisfazer a definição axiomática de probabilidade apresentada na Definição 2.7.

2.3.4 Probabilidade Condicional

Imagine a seguinte situação:

- um canal digital de comunicação tem uma taxa de erro de um *bit* por cada mil transferidos;
- embora raros, erros dessa natureza tendem a ocorrer em sequência;
- portanto, se um único *bit* for transmitido, podemos atribuir a probabilidade de erro de comunicação de 1/1000;
- entretanto, se soubermos que o *bit* anterior apresentou erro, o bom senso diz que o próximo *bit* terá probabilidade de erro maior do que 1/1000.

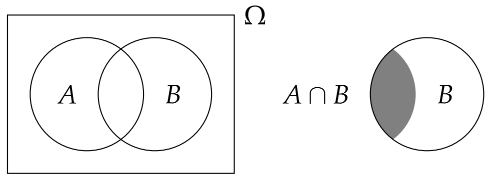
No exemplo acima, o conhecimento da ocorrência de erro no *bit* anterior, altera a probabilidade de erro no próximo *bit*.

Em outras palavras, é possível "atualizar" a probabilidade de um evento quando temos informação da ocorrência de outro.

Definição 2.8. Sejam $A, B \subset \Omega$ eventos. A probabilidade do evento A ocorrer, dada a ocorrência do evento B, é dada por

$$P(A|B) = \begin{cases} \frac{P(A \cap B)}{P(B)}, & \text{se } P(B) > 0, \\ P(A), & \text{se } P(B) = 0. \end{cases}$$

Via diagrama de Venn, podemos identificar o eventos associados a probabilidade condicional,no caso em que $P(B) > 0$, como segue.



Exemplo 1: Suponha que um escritório possui 100 máquinas de calcular. Algumas dessas máquinas são elétricas (E), manuais (M), novas (M) e usadas (U). Um funcionário pega uma máquina e vê que é nova. Qual a probabilidade de que ela seja elétrica?

Tipo	Elétrica	Manual	Total
Nova	40	30	70
Usada	20	10	30
Total	60	40	100

Resolução:

Notação: $P(E \mid N)$

$$P(E|N) = \frac{P(E \cap N)}{P(N)} = \frac{40/100}{70/100} \approx 0.57.$$

Assim, concluímos que, saber que o disco tem ao saber que a máquina é nova, a probabilidade dela ser elétrica é 0.57. Note que se não soubéssemos que a máquina é nova, teríamos como probabilidade da máquina ser apenas elétrica dada por 0.6.

■

Exemplo 2: Discos de policarbonato plástico são analisados com relação a resistência a arranhões e choque. Os resultados de 100 discos são resumidos a seguir:

Res. a arranhão	Res. a choque		Total
	Alta	Baixa	
Alta	80	9	89
Baixa	6	5	11
Total	86	14	100

Sejam os eventos A = "alta resistência a arranhões" e B = "alta resistência a choques". Qual a probabilidade de que um disco selecionado ao acaso tenha:

(a) alta resistência para choques e arranhões?

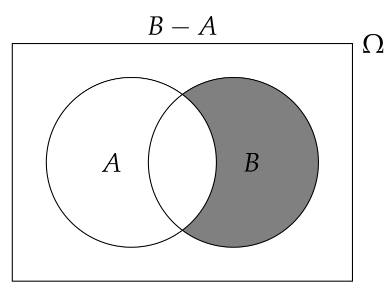
(b) alta resistência para choques ou arranhões?

Calcule também a $P(B-A)$ e $P(A|B^c)$. Interprete.

Resolução:

- $P(A \cap B) = 0.8$;
- $P(A \cup B) = P(A) + P(B) - P(A \cap B) = 0.89 + 0.86 - 0.80 = 0.95$.

O evento B-A está representado na área sombreada do diagrama abaixo:



Logo, $P(B-A) = P(B) - P(A \cap B) = 0.86 - 0.8 = 0.06$. Note que 0.06 corresponde a probabilidade de ukm disco selecionado ter alta resistência a choques e baixa resistência a arranhões. Por outro lado,

$$P(A|B^c) = \frac{P(A \cap B^c)}{P(B^c)} = \frac{0.09}{0.14} \approx 0.64.$$

Assim, concluímos que, saber que o disco tem baixa resistência a choque diminui a probabilidade dele ter alta resistência a arranhão de 0.89 para 0.64.

■

2.3.5 Regra do produto de probabilidades

Uma das consequências da probabilidade condicional é a regra do produto de probabilidades que corresponde ao cálculo da probabilidade de intersecção de dois eventos A e B em Ω, a partir da probabilidade condicional entre esses dois eventos. Ou seja,

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} \Longrightarrow P(A \cap B) = P(A \mid B)P(B)$$

e

$$P(B \mid A) = \frac{P(B \cap A)}{P(A)} \Longrightarrow P(A \cap B) = P(B \mid A)P(A).$$

Como seria a regra do produto entre 3 eventos?

$$\begin{aligned} P(A \cap B \cap C) &= P(C \mid A \cap B)P(A \cap B) && (2.3) \\ &= P(C \mid A \cap B)P(B \mid A)P(A). && (2.4) \end{aligned}$$

Generalização: Sejam $A_1, A_2, \ldots, A_n$ eventos em Ω, tem-se que

$$P(A_1 \cap A_2 \cap \ldots \cap A_n) = P(A_n \mid A_1 \cap A_2 \cap \ldots \cap A_{n-1}) \ldots P(A_3 \mid A_1 \cap A_2)P(A_2 \mid A_1)P(A_1).$$

Para demostrar essa generalização basta considerar o princípio da indução fraca.

Exemplo: Duas lâmpadas queimadas (Q) foram acidentalmente misturadas em uma gaveta com 4 lâmpadas boas (B). Suponha que duas lâmpadas são retiradas ao acaso dessa gaveta, sem reposição.

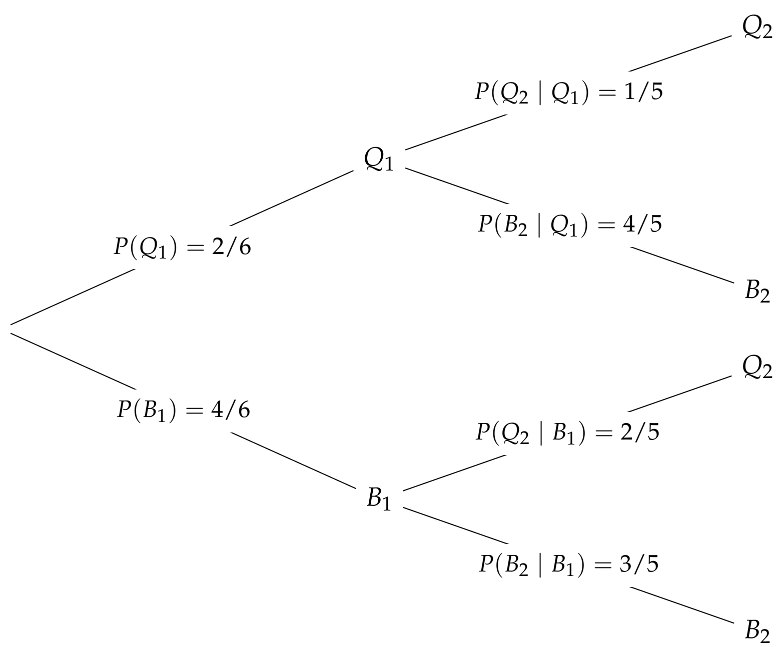
(a) Determine o espaço amostral associado a esse experimento.

(b) Qual a probabilidade de que ambas as lâmpadas sorteadas estejam queimadas?

(c) Qual a probabilidade de aparecer exatamente uma lâmpada queimada?

Resolução:

Antes de responder aos itens, vamos apresentar o diagrama de árvores que ilustra as possibilidades de retirada. Sejam Q_i e B_i, respectivamente, a obtenção de lâmpada queimada ou boa na i-ésima retirada, com $i = 1, 2$.



(a) $\Omega = \{(Q_1, Q_2), (Q_1, B_2), (B_1, Q_2), (B_1, B_2)\}$.

(b) $P(Q_1, Q_2) = P(Q_1 \cap Q_2) = P(Q_2 \mid Q_1)P(Q_1) = \frac{1}{5} \times \frac{2}{6} = \frac{1}{15}$.

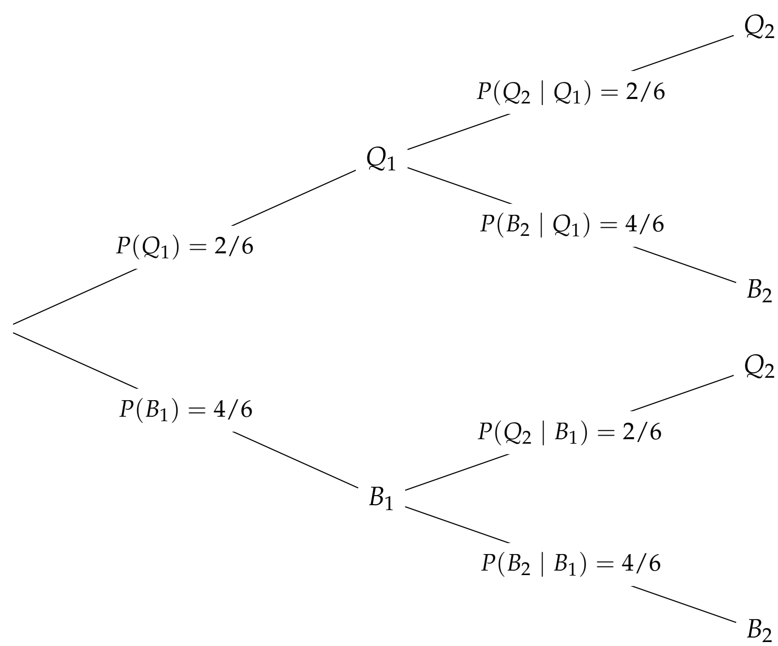
(c) Seja C o evento "aparecer exatamente uma lâmpada queimada na retirada de duas lâmpadas da gaveta, sem reposição", então

$$\begin{aligned} P(C) &= P((Q_1, B_2) \cup (B_1, Q_2)) \\ &= P((Q_1 \cap B_2) \cup (B_1 \cap Q_2)) \\ &= P(Q_1 \cap B_2) + P(B_1 \cap Q_2) \\ &= P(B_2 \mid Q_1)P(Q_1) + P(Q_2 \mid B_1)P(B_1) \\ &= \frac{4}{5} \times \frac{2}{6} + \frac{2}{5} \times \frac{4}{6} \\ &= \frac{8}{15}. \end{aligned} \tag{2.5}$$

$\blacksquare$

2.3.6 Independência

Considere a mesma gaveta do exercício anterior, com 2 lâmpadas queimadas (Q) e 4 boas (B). Novamente retiraremos duas lâmpadas dessa gaveta, mas agora sem reposição. O novo diagrama de árvores fica dado por:



Note que

- $P(Q_2 \mid Q_1) = 2/6$
- $P(Q_2 \mid B_1) = 2/6$
- $P(Q_2) = P(Q_1 \cap Q_2) + P(B_1 \cap Q_2) = P(Q_2 \mid Q_1)P(Q_1) + P(Q_2 \mid B_1)P(B_1) = \frac{4}{6} \times \frac{2}{6} + \frac{2}{6} \times \frac{4}{6} = \frac{2}{6}$.

Então $P(Q_2 \mid Q_1) = P(Q_2 \mid B_1) = 2/6 = P(Q_2)$. Portanto, não importa se saiu uma lâmpada boa ou queimada na primeira retirada da gaveta, a probabilidade de sair uma lâmpada queimada na segunda retirada é 2/6, indicando independência entre os eventos.

Nos casos em que a probabilidade condicional de um evento A dado ocorrência de B é idêntica a probabilidade incondicional de A, dizemos que os eventos A e B são independentes.

Definição 2.9. Sejam $A, B \subset \Omega$ eventos. Dizemos que A e B são independentes se qualquer uma das seguintes afirmações for verdadeira:

- $P(A|B) = P(A)$;
- $P(B|A) = P(B)$.

Como consequência, a regra do produto para eventos independentes pode ser simplificada da seguinte forma:

Proposição 2.3.1. *Sejam $A, B \subset \Omega$ eventos. Dizemos que A e B são independentes se, e somente se,*

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

Não é difícil demonstrar essa proposição, por isso a deixamos a cargo do leitor.

Observação: É importante ressaltar que eventos independentes não necessariamente são disjuntos. De fato, se dois eventos $A, B \in \Omega$ são disjuntos, eles são fortemente DEPENDENTES. Afinal, a ocorrência de um exclui completamente a possibilidade da ocorrência do outro. Em outras palavras, como A e B são disjuntos, temos que

$$P(A \cap B) = 0 \Rightarrow P(A|B) = \frac{P(A \cap B)}{P(B)} = 0 \text{ e } P(B|A) = \frac{P(A \cap B)}{P(A)} = 0.$$

A única maneira de A e B serem disjuntos e independentes é quando pelo menos um dos dois eventos tem probabilidade zero. Para ver isso, note que

$$0 = P(A \cap B) = P(A)P(B) \Rightarrow P(A) = 0 \text{ ou } P(B) = 0.$$

O conceito de independência pode ser estendido para mais de dois eventos, no entanto, alguns cuidados devem ser tomados. Por exemplo, considere como experimento o lançamento de dois dados. Defina os eventos:

$$A = \{\text{o primeiro dado mostra um n}^{\text{o}} \text{ par}\},$$

$$B = \{\text{o segundo dado mostra um n}^{\text{o}} \text{ ímpar}\},$$

$$C = \{\text{os dados mostram simultaneamente números pares ou ímpares}\}$$

. Não é difícil ver que $P(A) = P(B) = P(C) = 1/2$ e que que os pares de eventos A e B, A e C, e B e C são dois a dois independentes. Por outro lado, note que

$$P(C \mid A \cap B) = P(\emptyset) = 0 \neq P(C).$$

Assim, o conhecimento de que o evento $A \cap B$ ocorreu, nos diz que não há possibilidade do evento C ocorrer. Portanto, os três eventos (A, B, C) não são mutuamente independentes, ainda que os eventos dois a dois o sejam.

Definição 2.10. Dizemos que os eventos $A_1, A_2, ..., A_n \subset \Omega$ são mutuamente independentes se, e somente, se:

$$P(A_1, A_2, ..., A_n) = P(A_1)P(A_2)\dots P(A_n)$$

e para qualquer sub-coleção contendo pelo menos dois eventos estes sejam mutuamente independentes.

Exemplo: Em uma empresa de informática, 2% dos equipamentos produzidos tem problemas elétricos, 0.8% tem defeitos na carcaça e 0.4% tem defeitos na carcaça e elétricos, simultaneamente. Os eventos problemas elétricos e defeitos na carcaça são independentes?

Resolução:

Sejam os eventos E = "equipamento com problemas elétricos" e C = "equipamento com defeitos na carcaça". Do enunciado temos: $P(E) = 0.02$, $P(C) = 0,008$ e $P(E \cap C) = 0.004$. Note que

$$P(E)P(C) = 0.02 \times 0.008 = 0.0016 \neq P(E \cap C) = 0.004.$$

Portanto, os eventos E e C não são independentes.

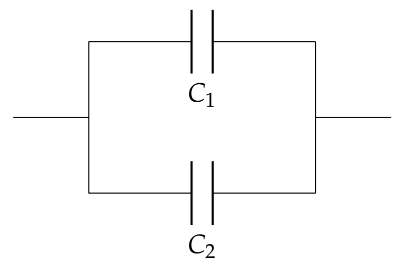
2.3.7 Aplicação: confiabilidade

Confiabilidade é a probabilidade de um sistema cumprir sem falhas uma missão com determinada duração. Sistemas mecânicos, elétricos e biológicos podem ser estudados através desta teoria.

Por exemplo, considere os seguintes mecanismos:

Figura"2.3.1
*

(a) Série



Figura"2.3.2
*

(b) Paralelo

Sejam os eventos:

- S: o sistema funciona
- C_i: o componente i funciona com probabilidade p_i, i = 1, 2.

Admita que os componentes funcionem um independentemente dos outros.

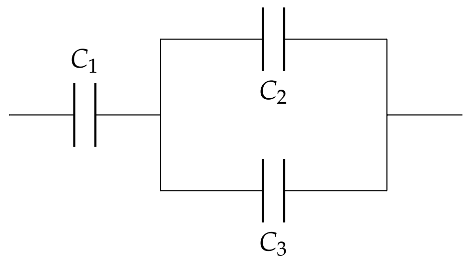
Em (a), o sistema funciona quando ambos os componentes 1 e 2 funcionam simultaneamente. Assim,

$$P(S) = P(C_1 \cap C_2) \overset{indep.}{=} P(C_1)P(C_2) = p_1 p_2.$$

Em (b), o sistema funciona se pelo menos um dos componentes funcionar. Assim,

$$P(S) = P(C_1 \cup C_2) = P(C_1) + P(C_2) - P(C_1 \cap C_2) = p_1 + p_2 - p_1 p_2.$$

Sistemas mais complexos podem ser analisados como uma composição de vários sub-sistemas em série e/ou em paralelo. Por exemplo, assuma que o sistema a seguir foi construído a partir de 3 componentes independentes.



Assim,

$$\begin{aligned} P(S) &= P((C_1) \cap (C_2 \cup C_3)) \\ &= P(C_1)P(C_2 \cup C_3) \\ &= p_1[P(C_2) + P(C_3) - P(C_2 \cap C_3)] \\ &= p_1[p_2 + p_3 - p_2 p_3]. \end{aligned}$$

2.3.8 Teorema da Probabilidade Total

Antes de apresentarmos formalmente o teorema da probabilidades total, vejamos um problema cuja a solução nos ajudará a entender intuitivamente o teorema.

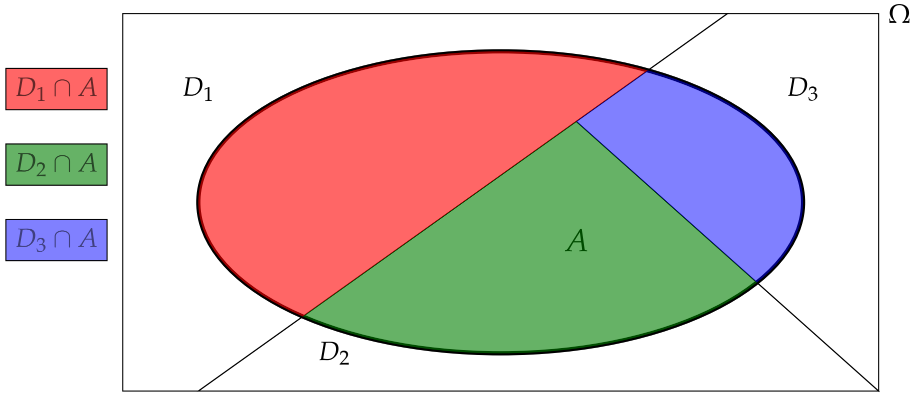
Problema: Uma empresa de transportes recebe 20% de todo o combustível que utiliza da distribuidora D_1, 30% da distribuidora D_2 e 50% da distribuidora D_3. Admita que as 3 distribuidoras são distintas e não tem relação uma com as outras. Em um fiscalização da prefeitura, observou-se que 20% dos galões de combustível da distribuidora D_1 estavam adulterados, enquanto que para D_2 e D_3 essa proporção era de 5% e 2%, respectivamente. Se selecionar ao acaso um galão de combustível, qual a probabilidade de ele estar adulterado?

Seja A o evento "o galão está adulterado". Sabemos que

$$\begin{array}{lll} P(A \mid D_1) = 0.2 & P(A \mid D_2) = 0.05 & P(A \mid D_3) = 0.02 \\ P(D_1) = 0.2 & P(D_2) = 0.3 & P(D_3) = 0.5 \end{array}$$

Queremos obter $P(A)$.

Para isso, considere a seguinte esquematização:



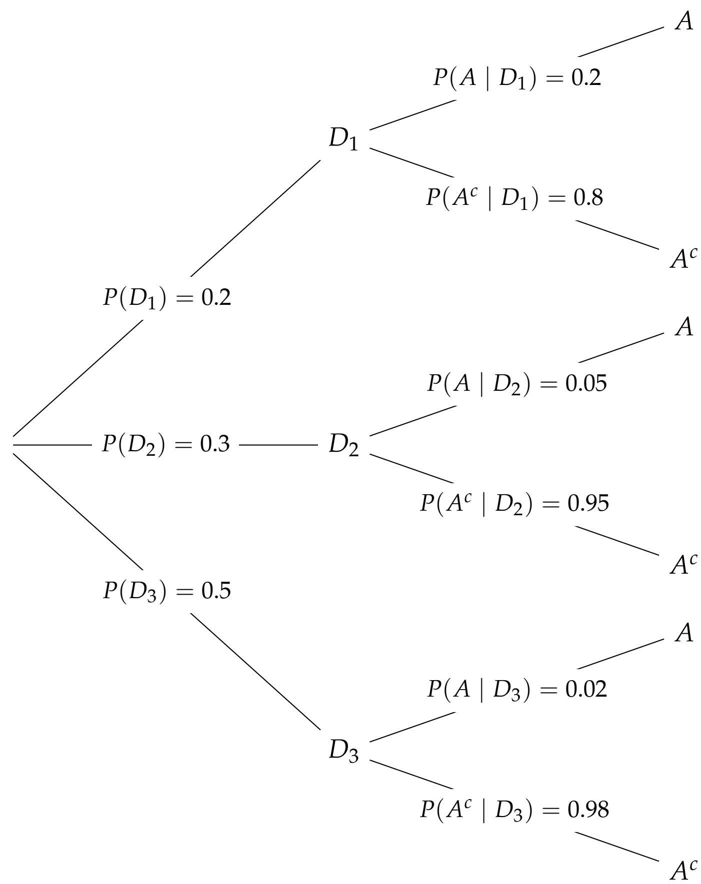
Através do diagrama de Venn, notamos que o evento A pode ser representado como a união dos eventos disjuntos $A \cap D_1$, $A \cap D_2$ e $A \cap D_3$. Tais eventos são disjuntos pois as distribuidoras D_1, D_2 e D_3 são distintas e não tem relação uma com as outras. Sendo assim,

$$A = (A \cap D_1) \cup (A \cap D_2) \cup (A \cap D_3).$$

Dessa forma,

$$\begin{aligned}
P(A) &= P[(A \cap D_1) \cup (A \cap D_2) \cup (A \cap D_3)] \\
&\overset{\text{disjuntos}}{=} P(A \cap D_1) + P(A \cap D_2) + P(A \cap D_3) \\
&\overset{\text{regra do produto}}{=} P(A \mid D_1)P(D_1) + P(A \mid D_2)P(D_2) + P(A \mid D_3)P(D_3) \\
&= 0.2 \times 0.2 + 0.05 \times 0.3 + 0.02 \times 0.5 \\
&= 0.065
\end{aligned}$$

O diagrama de árvores para o exemplo em questão ficaria dado por:



Definição 2.11. (Teorema da probabilidade total) Considere o espaço amostral Ω particionado em K eventos $D_1, D_2, \ldots, D_k$ conforme a Definição 2.6 e que $P(D_i) > 0$, $\forall i = 1, \ldots, k$. Então, para qualquer evento A referente ao espaço amostral Ω, temos que:

$$P(A) = \sum_{i=1}^{k} P(A \mid D_i) P(D_i).$$

Idéia da demonstração: Considere $A = (A \cap D_1) \cup (A \cap D_2) \cup \ldots \cup (A \cap D_k)$, observe que os eventos $(A \cup D_i)$, $i = 1, \ldots, k$ são disjuntos e aplique a regra do produto, de maneira similar ao procedimento realizado para resolver o exemplo anterior.

Exemplo: Num total de 25 peças, 5 delas sofreram excessivo encolhimento. Se duas peças são selecionadas ao acaso, qual será a probabilidade de que a segunda tenha sofrido excessivo encolhimento?

Seja $E_i =$ "i-ésima peça ter sofrido excessivo encolhimento", $i = 1, 2$. Note que

$$P(E_1) = 5/25, P(E_1^c) = 20/25, P(E_2|E_1) = 4/24 \text{ e } P(E_2|E_1^c) = 5/24.$$

Além disso, E_1, E_1^c formam uma partição de Ω, portanto o teorema anterior pode ser utilizado para encontrar $P(E_2)$ e fornece

$$\begin{aligned} P(E_2) = P(E_2|E_1)P(E_1) + P(E_2|E_1^c)P(E_1^c) &= \frac{5}{25}\frac{4}{24} + \frac{20}{25}\frac{5}{24} \\ &= \frac{1}{30} + \frac{1}{6} = \frac{1}{5}. \end{aligned}$$

2.3.9 Teorema de Bayes

Em muitas situações, precisamos calcular a probabilidade condicional de um evento, por exemplo, $P(A \mid B)$, quando conhecemos a $P(B \mid A)$. Voltemos ao problema da empresa de transportes e o uso de galões de gasolina adulterados que nos motivou a intuir o teorema da probabilidade total. Pergunta-se: sabendo que o galão de gasolina escolhido é adulterado, qual a probabilidade de que tenha sido fornecido pela distribuidora D_1?

De fato, queremos obter a $P(D_1 \mid A)$. Do enunciado do problema, vimos que

$$\begin{array}{lll} P(A \mid D_1) = 0.2 & P(A \mid D_2) = 0.05 & P(A \mid D_3) = 0.02 \\ P(D_1) = 0.2 & P(D_2) = 0.3 & P(D_3) = 0.5 \end{array}$$

Da teoria de probabilidade condicional, sabemos que:

$$P(D_1 \mid A) = \frac{P(D_1 \cap A)}{A}.$$

Por outro lado, pela regra do produto de probabilidades, temos que

$$P(D_1 \cap A) = P(A \mid D_1)P(D_1) = 0.2 \times 0.2 = 0.04$$

e, pelo teorema da probabilidade total a $P(A) = 0.065$, pois, como visto,

$$P(A) = P(A \mid D_1)P(D_1) + P(A \mid D_2)P(D_2) + P(A \mid D_3)P(D_3).$$

Sendo assim,

$$P(D_1 \mid A) = \frac{P(D_1 \cap A)}{A} = \frac{0.04}{0.065} = 0.615.$$

Vimos com este exemplo que é possível inverter a ordem que condicionamos os eventos e calcularmos a sua probabilidade. A teoria que relaciona probabilidades condicionais da forma $P(A \mid B)$ com probabilidades condicionais da forma $P(B \mid A)$ é conhecida por Teorema de Bayes e será formalizada a seguir.

Definição 2.12. (Teorema de Bayes) Considere o espaço amostral Ω particionado em K eventos $D_1, D_2, \ldots, D_k$ conforme a Definição 2.6 e que $P(D_i) > 0, \forall i = 1, \ldots, k$, seja conhecidas. Suponha ainda que para um evento A referente ao espaço amostral Ω, se conheçam as probabilidades $P(A \mid D_i), \forall i = 1, \ldots, k$. Então,

$$P(D_i \mid A) = \frac{P(A \mid D_i)P(D_i)}{\sum_{i=1}^{k} P(A \mid D_i)P(D_i)}, i = 1, \ldots, k.$$

A demonstração desse teorema é similar aos paços executados para resolver $P(D_1 \mid A)$ no problema da empresa de transportes e o uso de galões de gasolina adulterados e, portanto, será deixado como exercício.

2.4 Variável aleatória

Em qualquer experimento, há diversas características que podem ser observadas ou medidas. Denominaremos como variável aleatória toda função que nos permite associar aos resultados do experimento um número real. Por exemplo, considere os experimentos:

$E_1 =$ "duas peças de uma linha de produção são inspecionadas e classificadas como 'defeituosa' (D) ou 'não defeituosa' (N)";

$E_2 =$ "uma linha de produção é observada até a produção da primeira peça defeituosa";

$E_3 =$ "um ponto em um círculo de raio unitário é escolhido ao acaso".

Fixado cada experimento, suponha que se tenha interesse em observar certas características específicas associadas a ele. Por exemplo,

- em E_1, X_1 = "número de peças defeituosas observadas";
- em E_2, X_2 = "número de peças produzidas até a interrupção";
- em E_3, X_3 = "distância do ponto sorteado ao centro do círculo".

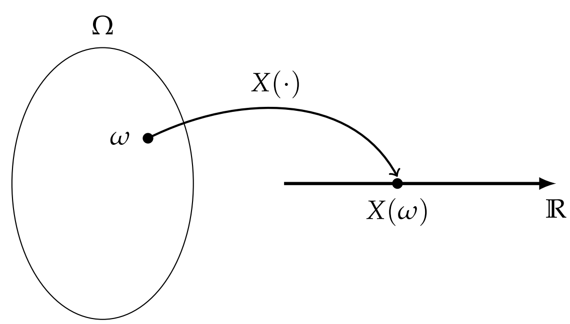
Observe que X_1, X_2 e X_3 são funções (numéricas) dos resultados (não necessariamente numéricos) do experimento aleatório ao qual estão associadas.

Neste caso, dizemos que X_1, X_2 e X_3 são **variáveis aleatórias**.

Definição 2.13. Sejam Ω espaço amostral de um experimento aleatório E e $\omega \in \Omega$ um de seus resultados. Se

$$X: \begin{array}{l} \Omega \to \mathbb{R} \\ \omega \mapsto X(\omega) \end{array}$$

dizemos que X é variável aleatória (v.a.).



Há dois tipos fundamentalmente diferentes de variáveis aleatórias: variáveis aleatórias discretas e variáveis aleatórias contínuas. Para se fazer essa classificação, será necessário conhecer o conceito imagem associado a uma variável aleatória.

Definição 2.14. O conjunto de todos os possíveis valores que uma variável aleatória X pode assumir é denominado **imagem** de X e denotado por $Im(X)$.

Dessa forma, para os experimentos E_1, E_2 e E_3 e as respectivas varáveis aleatórias associadas, X_1, X_2, X_3 definidas anteriormente, podemos identificar as seguintes imagens para cada uma das variáveis aleatórias:

1. $Im(X_1) = \{0, 1, 2\}$, pois X_1 representa o número de peças defeituosas observadas no experimento E_1;

2. $Im(X_2) = \{1, 2, 3, \ldots\}$, pois X_2 representa o número de peças produzidas até a interrupção e

3. $Im(X_3) = \{x : 0 \leq x \leq 1\} = [0, 1]$, pois $X_3 =$ representa a distância do ponto sorteado ao centro do círculo.

Há diferenças entre $Im(X_1)$, $Im(X_2)$ e $Im(X_3)$:

1. $Im(X_1)$ é um conjunto finito;

2. $Im(X_2)$ é um conjunto infinito contável (ou enumerável);

3. $Im(X_3)$ é um conjunto infinito não contável (ou não enumerável).

Definição 2.15. Dizemos que uma variável aleatória X é **discreta** se $Im(X)$ é um conjunto finito ou infinito enumerável. Caso a $Im(X)$ seja um conjunto não-enumerável, dizemos que uma variável aleatória X é **contínua**.

Portanto, X_1 e X_2 são variáveis aleatórias discretas, enquanto X_3 é uma variáveis aleatórias contínua.

Apresentamos agora alguns exemplos de variáveis aleatórias e sua classificação:

- O tempo que um projétil gasta para retornar à Terra $\rightarrow$ contínua;

- O número de vezes que um transistor em uma memória de computador muda de estado em uma operação $\rightarrow$ discreta;

- O volume de gasolina que é perdido por evaporação, durante o enchimento de um tanque de gasolina $\rightarrow$ contínua;

- O diâmetro de externo de um eixo usinado $\rightarrow$ contínua;

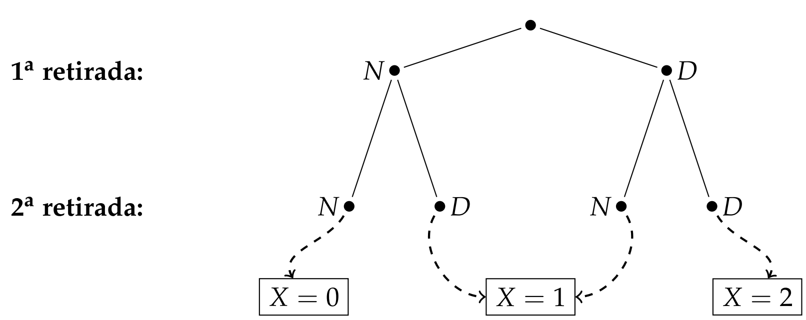
- O número de fraturas que excedem meia polegada em 10 milhas de uma auto-estrada interestadual $\rightarrow$ discreta;

- O peso de uma peça plástica moldada por injeção → contínua;
- O número de moléculas em uma amostra de gás → discreta;
- A concentração de saída de um reator → contínua;
- A corrente em um circuito elétrico → contínua.

Lembre-se que um experimento é considerado aleatório se, mesmo quando repetido sob as mesmas condições, não é possível predizer com absoluta certeza o seu resultado. Nessa situação, definimos o conceito de espaço amostral Ω e identificamos cada um dos possíveis resultados do experimento aleatório por ω. Note que o resultado ω é desconhecido *a priori* e, como uma variável aleatória é uma função de Ω, então o valor resultante da variável aleatória $X(\omega)$ também é desconhecido, dessa forma, devemos tratar a variável aleatória X probabilisticamente. Abordagens diferentes devem ser consideradas nos casos discreto e contínuo.

2.5 Variáveis aleatórias discretas

Relembremos o experimento E_1 em que duas peças de uma linha de produção são inspecionadas e classificadas como 'defeituosa' (D) ou 'não defeituosa' (N) e a v.a. X_1 representando o número de peças defeituosas observadas em E_1. Vimos que que $Im(X_1) = \{0, 1, 2\}$. No diagrama abaixo apresentamos os possíveis resultados do experimento aleatório e a variável aleatória associada.



Perceba que cada valor $x \in Im(X_1)$ induz a um evento em Ω e esses eventos formam uma partição do espaço amostral. Isto é, $A_x = \{\omega \in \Omega : X(\omega) = x\}$, em que $x = 0, 1, 2$, formam uma partição de Ω. Em outras palavras:

- $A_0 = \{(N, N)\}$;
- $A_1 = \{(N, D), (D, N)\}$;
- $A_2 = \{(D, D)\}$.

Portanto, é natural definir a probabilidade da v.a. assumir um valor x como a probabilidade do evento induzido por x. Isto é,

$$P(X = x) = P(A_x), \text{``}x = 0, 1, 2.$$

Adicionalmente, suponha agora que as peças são classificadas como defeituosas com probabilidade $p \geq 0$ e de que as classificações das peças são independentes. Assim, temos que

$$\begin{aligned} P(X = 0) = P(A_0) &= P(\{(N, N)\}) = P(\{N\})P(\{N\}) \\ &= (1 - p)(1 - p) = (1 - p)^2, \end{aligned}$$

$$\begin{aligned} P(X = 1) = P(A_1) &= P(\{(N, D), (D, N)\}) \\ &= P(\{(N, D)\}) + P(\{(D, N)\}) \\ &= P(\{N\})P(\{D\}) + P(\{D\})P(\{N\}) \\ &= p(1 - p) + (1 - p)p = 2p(1 - p) \end{aligned}$$

e

$$\begin{aligned} P(X = 2) = P(A_2) &= P(\{(D, D)\}) = P(\{D\})P(\{D\}) \\ &= p \cdot p = p^2. \end{aligned}$$

Ou seja, associado a cada possível resultado da variável aleatória, conseguimos especificar a probabilidade de observar aquele valor quando o experimento aleatório for realizado. de maneira geral, chamamos de função de probabilidade de uma variável aleatória X a função que associa a cada possível valor da v. a. a sua probabilidade de ocorrência.

Definição 2.16. Sejam X v.a. discreta com imagem $Im(X)$ e A_x, $x \in Im(X)$, a partição de Ω induzida pelos possíveis valores de X. A função de probabilidade (f.p.) de X é definida por

$$P_X(x) = P(X = x) = P(A_x), x \in Im(X).$$

2.5.0.1 Propriedades de uma f.p.

Seja X v.a. discreta com imagem $Im(X)$. A função de probabilidade $P_X(x)$, $x \in Im(X)$, satisfaz as seguintes propriedades:

1. $P_X(x) = P(A_x) \geq 0, \forall x \in Im(X)$, pois A_x é evento de Ω;
2. $\sum_{x \in Im(X)} P_X(x) = \sum_{x \in Im(X)} P(A_x) = 1$, afinal A_x, $x \in Im(X)$, é uma partição de Ω.

Retornando à v.a. X_1, é fácil ver que $P_X(x) \geq 0$, $x \in Im(X)$. Além disso, temos que

$$\begin{aligned}
\sum_{x \in Im(X)} P_X(x) &= \sum_{x=0}^{2} P(X = x) \\
&= P(X = 0) + P(X = 1) + P(X = 2) \\
&= (1-p)^2 + 2p(1-p) + p^2 = (1-p)(1-p+2p) + p^2 \\
&= (1-p)(1+p) + p^2 = 1 - p^2 + p^2 \\
&= 1.
\end{aligned}$$

Exercício: Obtenha a função de probabilidade da v.a. X_2 e mostre que as propriedades (1) e (2) acima também valem para a f.p. da v.a. X_2.

Podemos imaginar a função de probabilidades de uma variável aleatória como sendo um **modelo** para a distribuição de frequências de uma variável quantitativa discreta.

Por exemplo, suponha que um dado honesto é arremessado. Qual a função de probabilidade da v.a. discreta X = "face do dado voltada para cima"?

Temos que $Im(X) = \{1, 2, \ldots, 6\}$. Como supomos que o dado é honesto, podemos postular que todas as faces ocorrem com a mesma probabilidade. Assim, a f.p. de X é dada por

$$P_X(x) = \frac{1}{6}, \text{“} x = 1, 2, \ldots, 6.$$

Portanto, se repetíssemos o arremesso do dado uma quantidade grande de vezes, esperaríamos que a frequência relativa da face x se aproximasse do modelo postulado $P_X(x)$.

Na tabela a seguir mostramos um estudo em que simulamos n = 50, 200 e 10000 arremessos de um dado equilibrado.

		Face					
	n	1	2	3	4	5	6
Prob.	-	0.167	0.167	0.167	0.167	0.167	0.167
	50	0.220	0.200	0.120	0.100	0.160	0.200
Freq.	200	0.145	0.150	0.195	0.185	0.180	0.145
	10000	0.159	0.166	0.165	0.175	0.168	0.168

Note que a medida em que se aumenta o número de lançamentos do dado, a frequência relativa de se sair cada face tende a se aproximar da probabilidade teórica conhecida. Tal construção está associada a ideia frequentista de probabilidade, como já discutido anteriormente.

2.5.0.2 Função de distribuição cumulativa

Suponha que no exemplo anterior (arremesso do dado), estejamos interessados em $P(X \leq 2)$. Essa probabilidade é facilmente obtida. De fato

$$
\begin{aligned}
P(X \leq 2) = P\left(\bigcup_{k \leq 2} \{X = k\}\right) &= \sum_{k \leq 2} P(X = k) \\
&= P(X = 1) + P(X = 2) = \frac{2}{6} = \frac{1}{3}.
\end{aligned}
$$

Podemos generalizar a probabilidade **cumulativa** acima.

Definição 2.17. Seja X v.a. discreta com f.p. $P_X(x)$. A função de distribuição cumulativa (f.d.c.) de X é definida por

$$F_X(x) = P(X \leq x) = \sum_{k \leq x} P(X = k), \text{“} x \in \mathbb{R}.$$

Note que a f.d.c. $F_X(x)$ de uma v.a. discreta X é definida para todo $x \in \mathbb{R}$ e não somente para aqueles $x \in Im(X)$.

Vejamos no caso do exemplo do arremesso do dado como ficaria sua função de distribuição cumulativa:

$$F_X(x) = \sum_{k \leq x} P(X = k) = \begin{cases} 0, & x < 1, \\ \frac{\lfloor x \rfloor}{6}, & 1 \leq x \leq 6, \\ 1, & x > 6, \end{cases}$$

onde $\lfloor x \rfloor$ denota o maior inteiro menor ou igual a x. Por exemplo, $\lfloor 3.5 \rfloor = 3$.

É possível utilizar a f.d.c. de uma v.a. discreta para determinar a sua função de probabilidade.

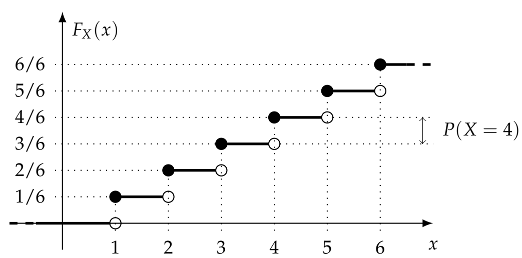
Observe no gráfico abaixo que

$$P(X = x) = F_X(x) - F_X(x^-), \text{“} x \in Im(X),$$

isto é, a f.p. de X no ponto x, $P_X(x)$, é o **salto** que ocorre na f.d.c. de X neste ponto. Na equação acima, a expressão

$$F_X(x^-) := \lim_{t \to x^-} F_X(t),$$

isto é, o limite de $F_X(t)$ quando t tende a x pela esquerda.



Outras propriedades da f.d.c. são:

- $F_X(x) \to 0$, quando $x \to -\infty$;
- $F_X(x) \to 1$, quando $x \to \infty$;
- $x \leq y \Rightarrow F_X(x) \leq F_X(y)$.

Exercício: Calcule e esboce o gráfico da função de distribuição cumulativa da variável X_2.

2.5.0.3 Esperança e variância

Apresentemos agora o conceito de esperança e média de uma variável aleatória.

Definição 2.18. Seja X v.a. discreta com f.p. $P_X(x)$. A **esperança** e a **variância** de X são definidas, respectivamente, por

$$\mu = E(X) = \sum_{x \in Im(X)} x P_X(x)$$

e

$$\sigma^2 = Var(X) = \sum_{x \in Im(X)} (x-\mu)^2 P_X(x) = \sum_{x \in Im(X)} x^2 P_X(x) - \mu^2.$$

Nas duas equações acima vimos que valores com maior probabilidade têm mais peso no cálculo da esperança e da variância. Vale ressaltar que na literatura estatística a esperança também é chamada por valor esperado ou média de uma variável aleatória.

Da mesma forma que a f.p. representa um modelo teórico para as frequências relativas, a esperança μ e a variância σ^2 representam valores teóricos para a média $\bar{x}$ e a variância s^2.

Dessa forma, esperamos que ao repetir o experimento uma quantidade muito grande de vezes, esperamos que os valores de $\bar{x}$ e de s^2 se aproximem de μ e σ^2, respectivamente.

Retornemos ao exemplo do arremesso do dado. Temos que

$$\mu = E(X) = \sum_{x=1}^{6} x P_X(x) = 1\frac{1}{6} + \cdots + 6\frac{1}{6} = \frac{1+\cdots+6}{6} = 3.5$$

e

$$\begin{aligned}\sigma^2 = Var(X) &= \sum_{x=1}^{6} x^2 P_X(x) - \mu^2 = 1^2\frac{1}{6} + \cdots + 6^2\frac{1}{6} - 3.5^2 \\ &= \frac{1^2+\cdots+6^2}{6} - 12.25 = \frac{91}{6} - 12.25 \approx 15.167 - 12.25 = 2.917.\end{aligned}$$

Na tabela a seguir simulamos o lançamento de um dado por 50, 200 e 10000 vezes. Uma vez lançado os dados e colhida a amostra, foi calculado a média amostral e a variância amostral.

Valores	μ	σ^2
teóricos	3.5	2.917
n	$\bar{x}$	s^2
50	3.380	3.424
200	3.630	2.948
10000	3.499	2.904

Note que a medida que o número de lançamentos aumenta, a média e a variância amostrais se aproximação da média μ e da variância σ^2 da v.a.. Dessa forma, a esperança pode ser entendida com a média aritmética dos resultados da várias aleatória se o experimento pudesse ser repetido infinitas vezes sob as mesmas circunstâncias. A variância nos informaria sobre a dispersão desses valores.

Exercício: Encontre os valores de $E(X_2)$ e $Var(X_2)$.

2.5.0.4 Propriedades importantes de esperança e variância

Sejam X e Y v.a. discretas e a e b constantes. As seguintes propriedades são satisfeitas:

- $E(a) = a$;
- $Var(a) = 0$;
- $E(aX + b) = aE(X) + b$;
- $Var(aX + b) = a^2 Var(X)$;
- $E(X + Y) = E(X) + E(Y)$;
- Se X e Y são v.a. discretas independentes, $Var(X + Y) = Var(X) + Var(Y)$.

Exemplo: Suponha que o número de falhas X em uma liga tem a seguinte f.p.:

x	0	1	2	3
$P_X(x)$	0.4	0.3	0.2	0.1

O lucro L obtido na venda da liga depende do número de falhas X na mesma da seguinte forma, $L = 4 - 2X$. Na venda da liga, qual o lucro esperado e a sua variância?

Resolução: Temos que

$$\mu = E(X) = 0.3 + 2 \cdot 0.2 + 3 \cdot 0.1 = 1$$

e

$$\sigma^2 = Var(X) = \sum_{x=0}^{3} x^2 P_X(x) - \mu^2 = 0.3 + 4 \cdot 0.2 + 9 \cdot 0.1 - 1 = 1.$$

Logo $E(L) = 4 - 2E(X) = 2$ e $Var(L) = 2^2 Var(X) = 4$.

Ou seja, esperamos um lucro médio de 2 unidades monetárias (u.m.), com uma variância de 4 u.m^2.

■

2.6 Alguns modelos discretos

Vamos estudar alguns modelos probabilísticos padrões que podem ser usados em diversas situações práticas.

2.6.1 Modelo uniforme discreto

Definição 2.19. Seja X v.a. discreta com imagem $Im(X) = \{a, a+1, \ldots, b-1, b\}$, onde $a \leq b \in \mathbb{Z}$. Dizemos que X tem distribuição uniforme entre a e b se

$$P_X(x) = \frac{1}{b-a+1}, \text{``} x = a, a+1, \ldots, b-1, b.$$

Notação: $X \sim UD(a, b)$.

Temos que se $X \sim UD(a, b)$, então

$$\mu = E(X) = \frac{b+a}{2}$$

e

$$\sigma^2 = Var(X) = \frac{(b-a)(b-a+2)}{12}$$

Exemplo 1: Uma empresa tem 48 linhas telefônicas a para atendimento ao consumidor. Em determinado instante o sistema de atendimento é observado. Defina $X =$ "número de linhas em uso" e suponha que X tem distribuição uniforme discreta. Qual a esperança e a variância de X?

Resolução:

É óbvio que, neste caso, $a = 0$ e $b = 48$.

Portanto, temos que $\mu = E(X) = \frac{48}{2} = 24$ e que $\sigma^2 = Var(X) = \frac{48 \cdot 50}{12} = 200$.

■

Exercício: Considere ainda o exemplo acima e defina $Y =$ "percentual de linhas em uso". Qual a esperança e a variância de Y?

2.6.2 Modelo uniforme binomial

Muitos experimentos são tais que os resultados da variável aleatória está a ocorrência ou não de determinada característica (docotômica/indicadora). Por exemplo,

$$\begin{aligned}
&(1) \quad \text{lançamento de um moeda:} \begin{cases} \text{cara} \\ \text{ou} \\ \text{coroa;} \end{cases} \\
&(2) \quad \text{lançamento de um dado:} \begin{cases} \text{face 5 ocorre} \\ \text{ou} \\ \text{não“}(1,2,3,4,6); \end{cases} \\
&(3) \quad \text{peça selecionada em um lote:} \begin{cases} \text{tem defeito} \\ \text{ou} \\ \text{não.} \end{cases}
\end{aligned} \tag{2.6}$$

Definição 2.20. Considere um experimento E o qual pode ocasionar apenas dois resultados: s = "sucesso"; e f = "fracasso". O espaço amostral desse experimento é:

$$\Omega_E = \{s, f\}.$$

Denominamos E de **experimento de Bernoulli**.

Observe mais alguns experimentos de Bernoulli a seguir:

- E = "observar uma amostra de ar e verificar se ela possui alguma molécula rara";
- E = "observar um bit transmitido através de um canal digital e verificar se ele foi recebido com erro";
- E = "em uma questão de múltipla escolha um candidato tenta adivinhar a resposta correta".

Seja E o experimento que consiste em n repetições **independentes** de um experimento de Bernoulli nos quais a probabilidade de sucesso $P(\{s\}) := p$ permanece **inalterada**. Defina X = "o número de sucessos observados nas n réplicas".

É evidente que $Im(X) = \{0, 1, \ldots, n\}$. Vamos encontrar a f.p. de X, ou seja, desejamos obter $P_X(x) = P(X = x)$, $x = 0, 1, \ldots, n$.

- Fixe algum $x \in Im(X)$;
- Em n repetições, x sucessos (e $n - x$ fracassos) ocorrem com probabilidade $p^x(1-p)^{n-x}$;
- Não importa a ordem em que os x sucessos e os $n - x$ fracassos ocorrem, a probabilidade acima permanece inalterada. Por exemplo:

$$P(\{\overbrace{s \ldots ss}^{x}\overbrace{ff \ldots f}^{n-x}\}) = P(\{\overbrace{s \ldots s}^{x-1} fs \overbrace{f \ldots f}^{n-x-1}\}) = p^x(1-p)^{n-x};$$

A pergunta que surge é: de quantas maneiras diferentes podemos ter x sucessos em n repetições?

Posição escolhida	1ª	2ª	...	x-ésima
Posições disponíveis	n	$n-1$	...	$n-x+1$

A princípio poderiamos imaginar **erroneamente** que temos

$$n(n-1)(n-2)\dots(n-x+1) = \frac{n!}{(n-x)!}$$

maneiras de ter x sucessos em n repetições. Na equação acima $a! = a(a-1)\dots 2\cdot 1$. Define-se $0! = 1$. Entretanto, existem redundâncias nestas combinações.

Logo, devemos dividir $\frac{n!}{(n-x)!}$ pela quantidade de permutações que podemos fazer com as x posições escolhidas, a saber: $x!$. Portanto, observamos que obtém-se x sucessos em n repetições de

$$\frac{n!}{x!(n-x)!} := \binom{n}{x}$$

maneiras possíveis.

Dessa forma,

$$\begin{aligned} P_X(x) = P(X=x) &= \overbrace{p^x(1-p)^{n-x} + \cdots + p^x(1-p)^{n-x}}^{\binom{n}{x} \text{ vezes}} \\ &= \binom{n}{x} p^x (1-p)^{n-x}, \end{aligned}$$

onde $x = 0, 1, \dots, n$.

Definição 2.21. Seja X v.a. discreta com conjunto imagem $Im(X) = \{0, 1, \dots, n\}$ e f.p. dada por

$$P_X(x) = \binom{n}{x} p^x (1-p)^{n-x}, \text{“}x = 0, 1, \dots, n.$$

Dizemos que X tem distribuição binomial com n repetições e probabilidade de sucesso p. Notação: $X \sim B(n, p)$.

Se $X \sim B(n, p)$, então sua esperança e sua variância são dadas, respectivamente, por

$$\mu = E(X) = np$$

e

$$\sigma^2 = Var(X) = np(1-p).$$

Exemplo: Um estudante faz um teste de múltipla escolha com 25 questões,

cada uma com 4 alternativas, apenas chutando as respostas. Qual a probabilidade de o estudante acertar mais do que 20 questões? Quantas questões ele acerta em média e com qual variância?
Resolução:

Estamos interessados em $X =$ "número de questões certas das 25 do teste". Temos que $X \sim B(25, \frac{1}{4})$. Logo,

$$P(X > 20) = \sum_{x=21}^{25} P(X = x) = \sum_{x=21}^{25} \binom{25}{x} (0.25)^x (0.75)^{25-x} = 9 \cdot 10^{-10},$$

$$E(X) = np = 25 \cdot 0.25 = 6.25$$

e

$$Var(X) = np(1-p) = 25 \cdot 0.25 \cdot 0.75 = 4.6875.$$

■

2.6.3 Modelo geométrico e binomial negativo

Considere o experimento E que consiste em repetições independentes de um experimento de Bernoulli com probabilidade de sucesso $P(\{s\}) = p$ até a ocorrência do primeiro sucesso. Defina $X =$ "número de réplicas realizadas". Qual o conjunto imagem de X? $Im(X) = \{1, 2, \ldots\}$.
Exercício: Mostre que a f.p. de X é dada por:

$$P_X(x) = P(X = x) = (1-p)^{x-1} p, \text{``} x = 1, 2, \ldots.$$

Definição 2.22. Seja X v.a. discreta com conjunto imagem $Im(X) = \{1, 2, \ldots\}$ e f.p. dada por

$$P_X(x) = P(X = x) = (1-p)^{x-1} p, \text{``} x = 1, 2, \ldots.$$

Dizemos que X tem distribuição geométrica com probabilidade de sucesso p. Notação: $X \sim Geo(p)$.

Se $X \sim Geo(p)$, então

$$\mu = E(X) = \frac{1}{p} \text{ e } \sigma^2 = Var(X) = \frac{1-p}{p^2}$$

Exercício: Defina Y = "número de fracassos até a ocorrência do primeiro sucesso". Verifique que $P_Y(y) = (1-p)^y p$ e encontre $E(Y)$ e $Var(Y)$. Dica: $Y = X - 1$, logo

$$P_Y(y) = P(Y = y) = P(X - 1 = y) = P(X = y + 1) = P_X(y + 1).$$

A v.a. geométrica pode ser generalizada. Considere o experimento E que consiste em repetições independentes de um experimento de Bernoulli com probabilidade de sucesso $P(\{s\}) = p$ até a ocorrência do k-ésimo sucesso. Defina X = "Número de réplicas realizadas". Qual o conjunto imagem de X? $Im(X) = \{k, k+1, \ldots\}$.

É possível mostrar que

$$P_X(x) = P(X = x) = \binom{x-1}{k-1} p^k (1-p)^{x-k}, x = k, k+1, \ldots.$$

Definição 2.23. Seja X v.a. discreta com conjunto imagem $Im(X) = \{k, k+1, \ldots\}$ e f.p. dada por

$$P_X(x) = P(X = x) = \binom{x-1}{k-1} p^k (1-p)^{x-k}, x = k, k+1, \ldots.$$

Dizemos que X tem distribuição binomial negativa com k sucessos e probabilidade de sucesso p. Notação: $X \sim BN(k, p)$.

Se $X \sim BN(k, p)$, então

$$\mu = E(X) = \frac{k}{p} \text{ e } \sigma^2 = Var(X) = k\frac{(1-p)}{p^2}.$$

Observações:

- é possível mostrar que uma v.a. binomial negativa com k sucessos e probabilidade de sucesso p é a soma de k v.a. geométricas independentes com probabilidade de sucesso p;
- a diferença entre uma v.a. binomial e uma binomial negativa é que, na primeira o número de réplicas é constante e o número de sucessos é aleatório, enquanto na segunda o número de réplicas é aleatório e o número de sucessos é constante;
- se $X \sim BN(k, p)$, com $k = 1$, então $X \sim Geo(p)$.

Exemplo: Um sistema de tolerâncias de defeitos, que processa transações para uma financeira usa três computadores e age da seguinte forma: se o computador em operação falha, um dos dois de reserva o substituem; caso este venha a falhar, ele é substituído pelo último. Considere que uma falha durante qualquer transação ocorre com probabilidade 10^{-2}. Qual o número médio de transações até que todos os computadores falhem?

Resolução:

Defina X = "número de transações até que os três computadores falhem". Temos que $X \sim BN(3, 10^{-2})$. Logo,

$$E(X) = \frac{3}{10^{-2}} = 3 \cdot 10^2 = 300.$$

■

Exercício: Qual o número médio de transações sem falha?

2.6.4 Modelo hipergeométrico

Considere um lote de N peças com K defeituosas. Suponha que n peças serão sorteadas **sem** reposição desse lote e serão inspecionadas. Seja X = "número de peças defeituosas encontradas nas n sorteadas". Vale ressaltar alguns comentários:

1. podemos enxergar o processo de inspeção de cada peça como uma realização de um experimento de Bernoulli, onde ela será classificada como defeituosa (sucesso) ou não defeituosa (fracasso);

2. porém, as repetições de cada experimento não são mais independentes;

3. portanto, embora parecida com a v.a. binomial, pelo fato das repetições dos experimentos de Bernoulli não serem independentes, X tem outra função de probabilidade.

Qual o conjunto imagem de X?

$$Im(X) = \{\max(0, n - (N - K)), \ldots, \min(K, n)\}.$$

Afinal:

- $K \geq n \Rightarrow X \leq n$ e $K < n \Rightarrow X \leq K$, logo
$$X \leq \min(K, n);$$
- $N - k \geq n \Rightarrow X \geq 0$ e $N - k < n \Rightarrow X \geq n - (N - k)$, logo
$$\max(0, n - (N - K)).$$

É possível mostrar que

$$P_X(x) = \frac{\binom{K}{x}\binom{N-K}{n-x}}{\binom{N}{n}}, \text{“}x \in Im(X).$$

Definição 2.24. Seja X v.a. discreta com conjunto imagem $Im(X) = \{\max(0, n - (N - K)), \ldots, \min(K, n)\}$ e f.p. dada por

$$P_X(x) = \frac{\binom{K}{x}\binom{N-K}{n-x}}{\binom{N}{n}}, \text{“}x \in Im(X).$$

Dizemos que X tem distribuição hipergeométrica com N elementos, K do tipo 1 e n sorteados. Notação: $X \sim HG(N, K, n)$.

É possível mostrar que, se $X \sim HG(N, K, n)$, então

$$\mu = E(X) = np\text{“}'' \text{ e } \text{“}''\sigma^2 = Var(X) = np(1-p)\,\overbrace{\frac{(N-n)}{(N-1)}}^{(*)},$$

onde $p = K/N$. O termo $(*)$ é conhecido como fator de correção.

Exemplo: Em uma loteria, sorteiam-se 6 números de 40 sem reposição. Um jogador escolhe seis números. Qual a probabilidade de que o jogador acerte os seis números? Qual a probabilidade de quecinco dos seis números escolhidos pelo jogador sejam sorteados?

Resolução:

Temos $N = 40$ números no total, $n = 6$ são escolhidos pelo jogador e $K = 6$ são os que de fato são sorteados (tipo 1). Queremos encontrar, primeiramente, a probabilidade de que o jogador acerte os seis números

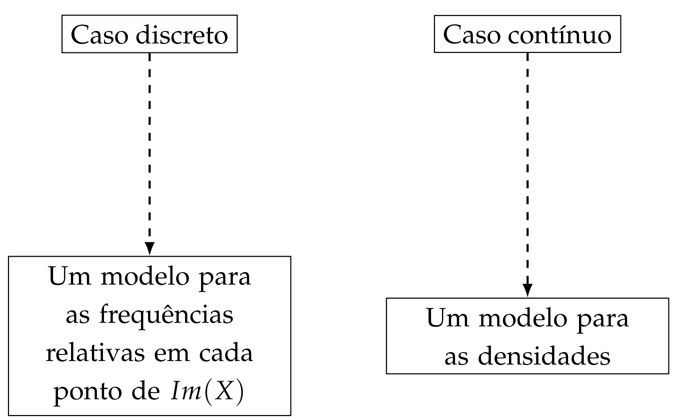
$$P_X(6) = P(X = 6) = \frac{\binom{6}{6}\binom{34}{6-6}}{\binom{40}{6}} = \frac{34!6!}{40!} = \frac{6 \cdot 5 \cdot 4 \cdot 3 \cdot 2}{40 \cdot 39 \cdot 38 \cdot 37 \cdot 36 \cdot 35}$$
$$= \frac{1}{10 \cdot 13 \cdot 37 \cdot 16 \cdot 6 \cdot 7} = \frac{1}{3232320} = 3.1 \cdot 10^{-7}.$$

■

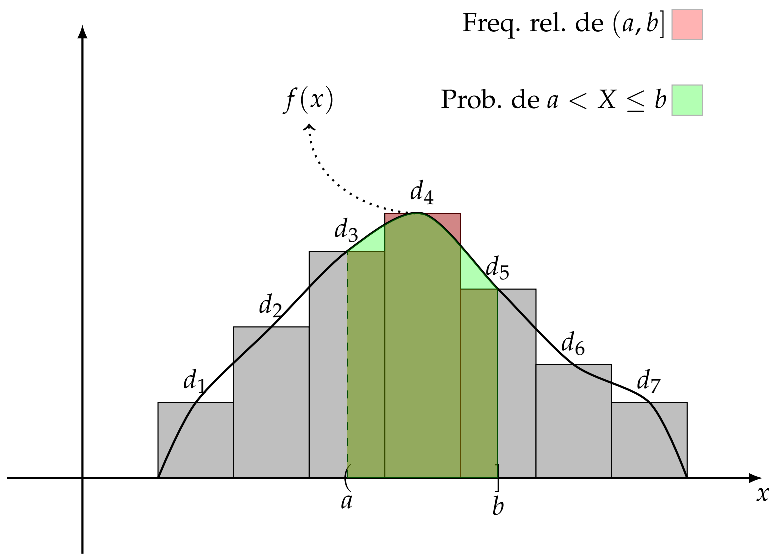
Exercício: Responder a segunda questão.

2.7 Variáveis aleatórias contínuas

Suponha que X é uma v.a. contínua. Devido a natureza não contável de $Im(X)$ neste caso, não devemos estabelecer probabilidades "ponto-a-ponto" conforme no caso discreto.



Devemos considerar então o histograma de densidades e aproximar a $P(a < X \leq b)$ através da área do hostrograma na região de (a, b].



2.7.1 Função densidade de probabilidade

Definição 2.25. Seja $f(x)$ uma função, tal que:

- $f(x) \geq 0$, para todo $x \in \mathbb{R}$;
- a área entre $f(x)$ e o eixo horizontal é igual a 1.

Dizemos que $f(x)$ é função densidade de probabilidade (f.d.p.) de alguma v.a. contínua.

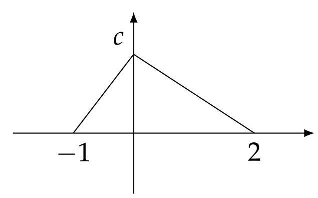
Exemplo: Seja

$$f(x) = \begin{cases} c(x+1), & -1 < x < 0; \\ c(1 - x/2), & 0 \leq x \leq 2; \\ 0, & \text{caso contrário.} \end{cases}$$

Para qual valor de c, $f(x)$ é densidade?

Resolução:

Note que o gráfico associado a $f(x)$ é dada por:



Note que, c deve ser maior que 0. A área abaixo de $f(x)$ por sua vez é $\frac{c}{2} + c = \frac{3c}{2}$. Logo, devemos ter $c = \frac{2}{3}$.

■

Uma vez definida uma v.a. contínua X com função densidade de probabilidade $f(x)$, como obter probabilidades a partir dela? A probabilidade $P(a < X \leq b)$ é dada pela área entre a f.d.p. $f(x)$ e o eixo horizontal compreendida no intervalo $(a, b]$. Assim, a probabilidade é calculada fazendo-se

$$P(a < X \leq b) = \int_a^b f(x)dx.$$

Por este motivo

$$P(a < X \leq b) = P(a \leq X < b) = P(a < X < b) = P(a \leq X \leq b)$$

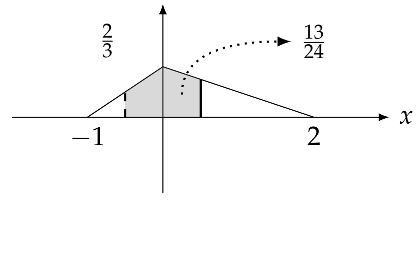
e, além disso,

$$P(X = x) = 0, \text{ para todo } x \in \mathbb{R}.$$

Exercício: Mostre que se X tem a f.d.p. do exemplo anterior, então

$$P(-0.5 < X < 0.5) = \frac{13}{24}.$$

Uma dica: sempre que possível faça o gráfico da função de densidade e localize a região para qual quer se calcular a probabilidade. Por exemplo, aqui,



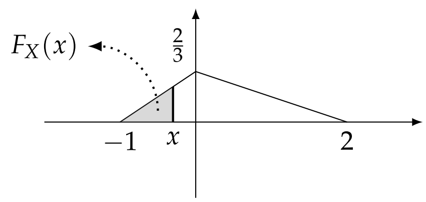
2.7.2 Função de distribuição cumulativa

Definição 2.26. Seja X v.a. contínua com f.d.p. $f(x)$. Definimos a função de distribuição cumulativa de X por $F_X(x) = P(X \leq x)$.

Exemplo: Retornando ao exemplo anterior, vamos obter a função de distribuição cumulativa (f.d.c.) da variável aleatória X, ou seja, $F_X(x) = P(X \leq x)$.

Resolução:

Suponha que $-1 < x < 0$. Considerando o gráfico abaixo da função de densidade e associando a f.d.c. a área sombreada, temos que

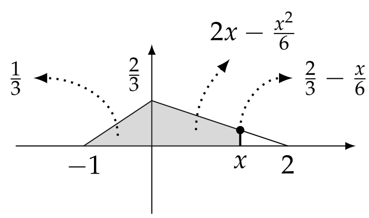


Assim,

$$F_X(x) = P(X \leq x) = \frac{(x+1)^2}{3},$$

se $-1 < x < 0$.

Suponha agora que queiramos obter a função de distribuição cumulativa de um x localizado entre 0 e 2. Analogamente ao que fizemos para $-1 < x < 0$, temos que a $F_X(x)$, para um $0 < x < 2$ é dad pela soma de duas áreas, a de um triângulo (2/3) e do trapézio ($2x - \frac{x^2}{6}$), como segue.



Assim,

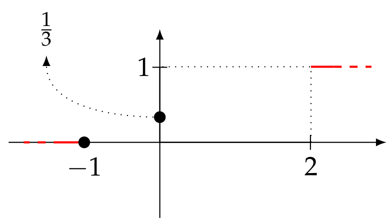
$$F_X(x) = P(X \leq x) = \frac{1}{3} + \frac{2x}{3} - \frac{x^2}{6},$$

se $0 \leq x < 2$.

Não se esqueçam, no entanto, que a f.d.c. é definida para todo $x \in \mathbb{R}$ e não somente para aqueles $x \in Im(X)$. Assim, a função de distribuição cumulativa é dada por

$$F_X(x) = \begin{cases} 0, & x \leq -1; \\ (x+1)^2/3, & -1 < x < 0; \\ \frac{1}{3} + \frac{2x}{3} - \frac{x^2}{6}, & 0 \leq x < 2; \\ 1, & x \geq 2 \end{cases}$$

e tem como respectivo gráfico



■

Alternativamente, a função de distribuição cumulativa de uma v.a. contínua X com f.d.p. $f(x)$ pode ser definida como segue:

$$F_X(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)dt.$$

Uma das propriedades da f.d.c. $F_X(x)$ é que é possível recuperar a f.d.p. da v.a. X através de

$$f(x) = \frac{d}{dx} F_X(x).$$

Dessa forma, no exemplo anterior, temos que

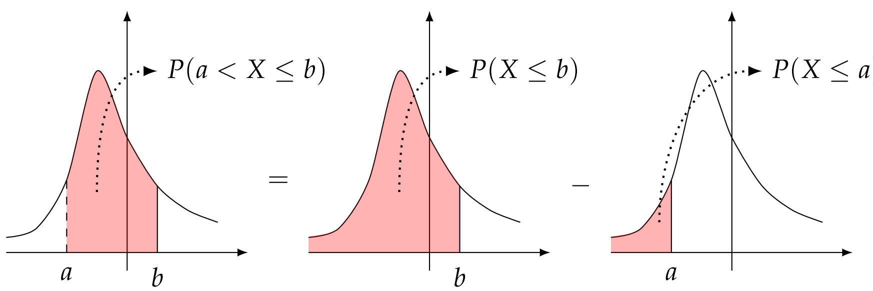
$$f(x) = \frac{d}{dx} F_X(x) = \begin{cases} \frac{2}{3}(x+1), & -1 < x < 0; \\ \frac{2}{3}(1 - \frac{x}{2}), & 0 \leq x < 2; \\ 0, & \text{c. c.} \end{cases}$$

No caso contínuo, a função de distribuição cumulativa satisfaz as mesmas propriedades que no caso discreto. A saber:

- $F_X(x) \to 0$, quando $x \to -\infty$;
- $F_X(x) \to 1$, quando $x \to \infty$;

- $x \leq y \Rightarrow F_X(x) \leq F_X(y)$;
- $P(a < X \leq b) = F_X(b) - F_X(a)$.

A última propriedade enunciada anteriormente pode ser vista através da seguinte figura:



Logo,

$$P(a < X \leq b) = F_X(b) - F_X(a).$$

Além disso, a f.d.c. de uma v.a. contínua é uma função contínua.

Assim como as medidas de probabilidade, a esperança e a variância são definidas no caso contínuo através do uso de integrais.

Definição 2.27. Seja X v.a. contínua com f.d.p. $f(x)$. Definimos a esperança e a variância de X por

$$\mu = E(X) = \int_{-\infty}^{\infty} x f(x) dx$$

e

$$\sigma^2 = Var(X) = \int_{-\infty}^{\infty} (x-\mu)^2 f(x) dx = \int_{-\infty}^{\infty} x^2 f(x) dx - \mu^2.$$

Exercício: Encontre a esperança e a variância da v.a. contínua com f.d.p.

$$f(x) = \begin{cases} \frac{2}{3}(x+1), & -1 < x < 0; \\ \frac{2}{3}(1-x/2), & 0 \leq x \leq 2; \\ 0, & \text{caso contrário.} \end{cases}$$

2.8 Alguns modelos contínuos

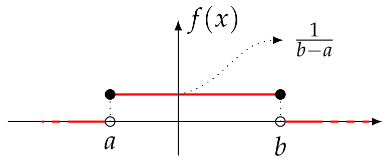
2.8.1 Modelo uniforme contínuo

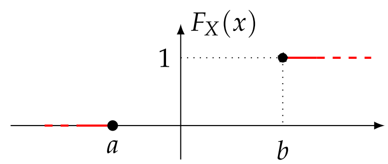
Definição 2.28. Seja X v.a. contínua com conjunto imagem $Im(X) = [a,b]$ e com f.d.p. dada por

$$f(x) = \begin{cases} \frac{1}{b-a}, & x \in [a,b] \\ 0, & \text{c. c.} \end{cases}$$

Então, dizemos que X tem distribuição uniforme contínua no intervalo $[a,b]$. Notação: $X \sim U[a,b]$.

A seguir apresentamos os gráficos da função de densidade e da função de distribuição cumulativa quando $X \sim U[a,b]$.





Exercício: Mostre que

$$F_X(x) = \begin{cases} 0, & \text{se“x¡a,} \\ \frac{x-a}{b-a}, & \text{se“}x \in [a,b], \\ 1, & \text{se“x¿b.} \end{cases}$$

Suponha que $X \sim U[a,b]$. É possível mostrar que a esperança e a variância de X são dadas, respectivamente, por

$$E(X) = \frac{a+b}{2}\text{“ e ″}Var(X) = \frac{(b-a)^2}{12}.$$

Exemplo: Suponha que o tempo em segundos requerido para completar uma operação de montagem seja $X \sim U[30,40]$. Determinemos:

1. a proporção de operações que duram mais do que 37.5 segundos;

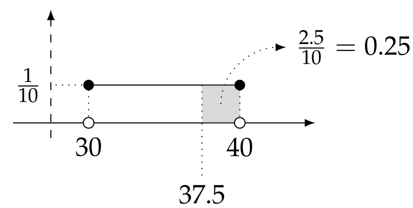
2. o tempo que é excedido por 90% das montagens;

3. a média e a variância da duração das montagens.

Resolução:

Em (1), temos que

$$P(X > 37.5) = \frac{40 - 37.5}{40 - 30} = 0.25.$$



Em (2) queremos encontrar x, tal que $P(X > x) = 0.9$. Note que

$$0.9 = P(X > x) = 1 - P(X \leq x) = 1 - F_X(x) = 1 - \frac{x - 30}{40 - 30}$$
$$\Rightarrow \frac{x - 30}{10} = 0.1 \Rightarrow x - 30 = 1 \Rightarrow x = 31.$$

Portanto, 90% das operações de montagem duram mais do que 31 segundos.

Em (3), basta utilizarmos as fórmulas de esperança e variância para esse modelo. Logo, temos que $\mu = E(X) = \frac{30+40}{2} = 35$ e $\sigma^2 = Var(X) = \frac{(40-30)^2}{12} = \frac{100}{12} = 8.33$.

■

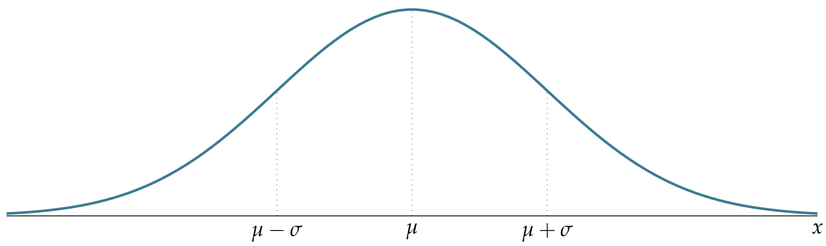
2.8.2 Modelo normal (ou gaussiano)

Definição 2.29. Dizemos que X tem distribuição normal com média μ e variância σ^2 se sua f.d.p. é dada por

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}.$$

Notação: $X \sim N(\mu, \sigma^2)$.

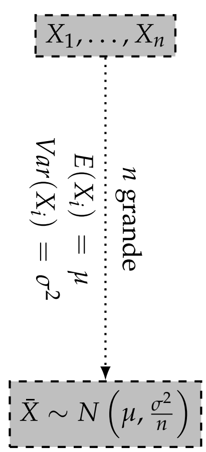
Abaixo apresentamos os gráfico das funções de densidade de uma variável aleatória com distribuição normal de média μ e desvio-padrão σ.



Algumas observações relevantes:

- $f(x)$ é simétrica com relação a μ, isto é, $f(x-\mu) = f(x+\mu)$,
- $f(x) \to 0$, quando $x \to \overset{+}{-} \infty$,
- o valor máximo de x ocorre para $x = \mu$, isto é, $f(x) = \frac{1}{\sigma\sqrt{2\pi}}$,
- $\mu - \sigma$ e $\mu + \sigma$ são os pontos de inflexão da curva.

Principal importância do modelo normal: grande aplicabilidade devido ao Teorema Central do Limite (TCL).



- Se fosse possível observar R amostras de tamanho n;
- Estas gerariam R médias amostrais;
- Se n for muito grande o histograma das R médias amostrais teria o formato aproximado de uma normal;
- Os parâmetros seriam a média teórica, μ, e variância teórica dividida pelo tamanho amostral, $\frac{\sigma^2}{n}$.

O TCL também nos fornece base teórica para afirmar que, em algumas situações, somas de **muitas** v.a.'s têm distribuição aproximadamente normal.

Exemplo: O desvio do comprimento de uma peça usinada do valor em sua especificação pode ser pensado como soma de um grande número de efeitos:

- pulsos na temperatura e na umidade;
- vibrações;
- variações no ângulo de corte;
- desgaste da ferramenta de corte;
- desgaste do mancal;
- variações na velocidade rotacional;
- variações de montagem e fixação;
- variações nas inúmeras características das matérias-primas.

Se os efeitos forem independentes, então se pode mostrar que o desvio total tem distribuição aproximadamente normal.

2.8.3 Padronização

É possível mostrar que, se $X \sim N(\mu, \sigma^2)$, então a esperança e a variância de X são, respectivamente,

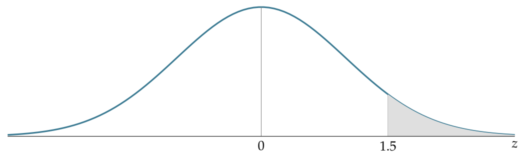
$$E(X) = \mu \text{ e } Var(X) = \sigma^2.$$

Definição 2.30. Seja $Z \sim N(0,1)$. Então dizemos que Z tem distribuição normal padrão.

Teorema 1. *Se $X \sim N(\mu, \sigma^2)$. Então $Z = \frac{X-\mu}{\sigma}$ tem distribuição normal padrão.*

Exemplo: Seja X a corrente em um pedaço de fio medida em miliampéres. Suponha que $X \sim N(10, 4)$. Qual a probabilidade de realizarmos uma medida nesse fio que supere 13 miliampéres?.

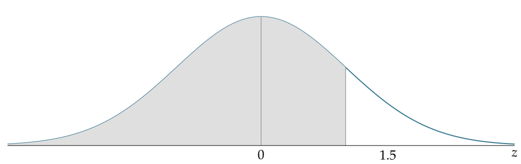
Resolução:
De fato, queremos encontrar $P(X > 13)$.
Observe que $P(X > 13) = P\left(\frac{X-\mu}{\sigma} > \frac{13-10}{2}\right) = P(Z > 1.5)$ e que a área sombreada no gráfico da função de densidade a seguir está associada a medida de probabilidade a se calcular.



Como a função de densidade da distribuição normal é não integrável, faremos uso da tabela da distribuição normal padrão gerada por meio de integração numérica. Existem várias versões desta tabela na literatura, neste trabalho faremos uso dessa:

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998

Na tabela são fornecidas as $P(Z \leq z)$, em que $z \leq 0$, como pode ser visto na figura a seguir.



Note que, como esperado, a $P(Z \leq 0) = 0.5$. Voltemos ao nosso problema, como calcular, fazendo uso da tabela da distribuição normal padrão, a $P(X > 13)$?

Note que, a probabilidade da corrente ultrapassar 13 miliampéres é

$$P(X > 13) = P(Z > 1.5) = 1 - P(Z \leq> 1.5) = 1 - 0.9332 = 0.0668.$$

Afinal, $P(Z > 1.5) + P(Z \leq> 1.5) = 1$ e a tabela fornece a $P(Z \leq> 1.5)$ como 0.9332.

■

2.8.4 Aproximação da binomial pela normal

Exemplo: Considere que uma linha de fabricação produz peças defeituosas com probabilidade de 0.05. Seja $Y =$ "número de peças defeituosas num total de 5000 observadas". Percebemos que $Y \sim B(n, p)$, onde $n = 5000$ e $p = 0.05$. Se fosse perguntado a probabilidade dessa amostra conter menos do que 230 peças defeituosas, como calcular essa probabilidade?

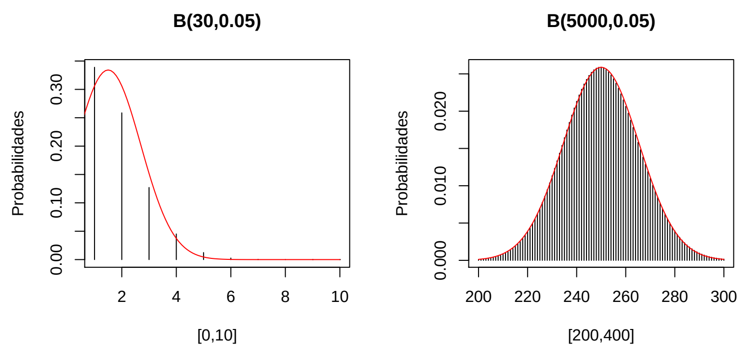
A rigor, deveríamos calcular

$$P(Y < 230) = \sum_{x=0}^{229} \binom{5000}{x} 0.05^x (0.95)^{5000-x}.$$

Essa tarefa pode ser facilitada percebendo que, para n suficientemente grande, o gráfico da função de probabilidades de uma v.a. $B(n, p)$ se assemelha muito com a f.d.p. de uma v.a. $N(\mu, \sigma^2)$!

Mas quais valores utilizar para μ e σ^2? Para termos uma ideia a respeito disso, apresentamos os gráficos das funções de probabilidade de duas v. a.´s, uma com distribuição B(30, 0.05)e outra B(5000, 0.5) e a aproximação pela distribuição normal em vermelho.

As curvas em vermelho foram geradas considerando $N(np, np(1-p))$. Notem que np corresponde a esperança da binomial e $np(1-p)$ a variância da binomial.



Portanto, a probabilidade do exemplo anterior pode ser aproximada da seguinte maneira:

$$P(Y < 230) \approx P(X < 230),$$

onde $X \sim N(\mu, \sigma^2)$, com

$\mu = np = 5000 \cdot 0.05 = 250$ e $\sigma^2 = np(1-p) = 5000 \cdot 0.05 \cdot 0.95 = 237.5.$

Logo,

$$P(Y < 230) \approx P(X < 230) = P(\frac{X-\mu}{\sigma} < \frac{230-250}{\sqrt{237.5}}) = P(Z < -1.29),$$

onde Z é a v.a. normal padrão. Olhando a tabela temos que

$$P(Y < 230) \approx 0.0985$$

O valor verdadeiro é 0.0904.

2.8.5 Modelo exponencial

Uma distribuição muito utilizada para representar o "tempo" (ou a distância) até a ocorrência de determinado evento geralmente é modelada pela variável aleatória exponencial.

Definição 2.31. Seja X v.a. contínua com conjunto imagem $Im(X) = [0, \infty)$. Suponha que a f.d.p de X seja

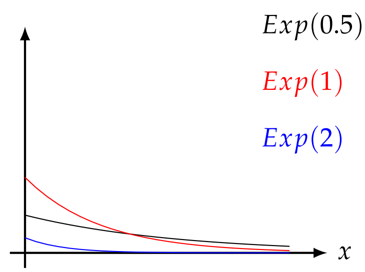
$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & x > 0; \\ 0, & \text{c.c.}, \end{cases}$$

onde $\lambda > 0$. Dizemos que X tem distribuição exponencial com parâmetro λ. Notação: $X \sim Exp(\lambda)$.

É possível mostrar que se $X \sim Exp(\lambda)$. Então,

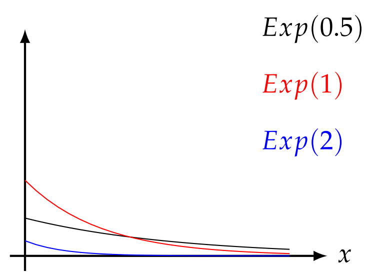
$$\mu = E(X) = \frac{1}{\lambda} \text{ e } \sigma^2 = Var(X) = \frac{1}{\lambda^2}.$$

O gráfico da função de densidade para diferentes valores de λ é dado abaixo:



Além disso, é possível mostrar que a função de probabilidade cumulativa de X é dada por

$$F_X(x) = \begin{cases} 1 - e^{-\lambda x}, & x > 0; \\ 0, & \text{c.c.} \end{cases}$$



Exemplo: Seja X = "tempo em anos decorrido até a falha de determinado equipamento mecânico". Suponha que $X \sim Exp(\lambda)$ e que em **média** o equipamento demora 2 anos até falhar. Qual a probabilidade de esse equipamento não falhe antes de 3 anos?

Temos que encontrar primeiramente o valor de λ. Como $\mu = E(X) = \frac{1}{\lambda} = 2$. Então $\lambda = 0.5$. Agora, temos que

$$\begin{aligned} P(X \geq 3) &= 1 - P(X < 3) = 1 - P(X \leq 3) = 1 - F_X(3) \\ &= 1 - (1 - e^{-0.5 \cdot 3}) = e^{-1.5} \approx 0.22. \end{aligned}$$