Teoria da Probabilidade

Augusto César Morgado e Ralph Costa Teixeira

30/6/2006

Contents

[1] Em alguns dos problemas desta seção, você vai precisar de uma calculadora ou computador que calcule a Distribuição Binomial Acumulada. Use a planilha EXCEL *Discrete.xls* para fazer os cálculos.

Chapter 1

Conceitos Básicos

1.1 O que é probabilidade?

"As questões mais importantes da vida são, em grande parte, nada mais do que problemas de probabilidade."

"A Teoria da Probabilidade nada mais é do que o cálculo do bom senso." – Pierre-Simon Laplace (1749-1827)

O objetivo da Teoria da Probabilidade é modelar matematicamente conceitos como *incerteza*, *risco*, *chance*, *possibilidade, verossimilhança, perspectivas* e, até mesmo, *sorte*. Considere as seguintes frases do nosso dia-a-dia:

- A probabilidade de uma moeda lançada "dar" coroa é de 50%;
- A previsão do tempo é de 40% de probabilidade de chuva amanhã;
- A radiografia indica uma moderada probabilidade de Tromboembolia Pulmonar;
- O Copom afirma que aumentou a probabilidade da convergência da inflação para a trajetória de metas;
- Depois da rodada de ontem, a probabilidade do Flamengo ser rebaixado aumentou muito.

Quase todos nós temos ao menos uma intuição do que estas frasem significam. No entanto, encontre a sua resposta para a seguinte pergunta: o que exatamente significa a palavra **probabilidade**? O que exatamente significam as frases acima? Pense nesta pergunta antes de ler os próximos parágrafos...

Seguem aqui duas interpretações comuns do conceito de probabilidade (ambas levam à mesma formulação matemática – apenas as maneiras de expressar e interpretar os resultados mudam com o ponto-de-vista escolhido):

A interpretação **freqüentista** imagina um *grande* número de situações semelhantes à apresentada e tenta descobrir em quantas delas o evento em questão realmente acontece; esta proporção seria a probabilidade do evento. Assim, "dividindo o número de coroas obtidas pelo número de lançamentos, a proporção se aproximará de 50% à medida que o número de lançamentos cresce". Esta interpretação pode precisar de um pouco de imaginação: "chove em 40% dos dias com características climáticas *semelhantes* às de amanhã".

A interpretação **subjetiva** (ou **Bayesiana**, ou **epistemológica**) diz que a probabilidade de um evento é apenas uma medida da fé que temos sobre a sua ocorrência. Assim, a probabilidade de um evento varia de indivíduo para indivíduo, dependendo das informações e crenças que ele tenha. Esta interpretação "maleável" nos permite discutir conceitos como a probabilidade de um evento passado ter ocorrido (como a probabilidade de uma pessoa *ter cometido* um crime)[1].

A Teoria da Probabilidade é apenas um **modelo**. Modelos não são "A REALIDADE" ou "A VERDADE". *Modelos são úteis exatamente porque simplificam a realidade para que possamos entendê-los*[2]. Se soubéssemos exatamente as características físicas da moeda, sua posição e velocidade iniciais, e as forças nela aplicadas (pelo seu dedão, pela gravidade da Terra, pela resistência do ar, etc.) seríamos capazes de predizer com exatidão se a

[1] Do matemático e mágico Persi Diaconis: "*probabilidades não fazem parte das moedas; probabilidades fazem parte das pessoas*".

[2] Num mapa de metrô, as estações aparecem alinhadas; o mapa não mostra todas as ruas, nem os jardins, nem a topografia da cidade. O mapa está *errado*? Não, o mapa é um *modelo*; ele é perfeito para a sua função (saber se a próxima estação é onde eu tenho que descer ou não), mas, se usado além de suas limitações (para planejar uma caminhada, por exemplo), ele falha miseravelmente.

moeda daria cara ou coroa[3]. Mas trabalhar com todas estas variáveis é impraticável[4] – é preferível inventar este "misterioso 50% de incerteza", jogando fora os outros detalhes da realidade. Como lidamos com nossa própria incerteza e ignorância desde que nascemos, o conceito de probabilidade até que não é tão misterioso assim[5].

1.1.1 Interpretação Freqüentista: as moedas se compensam?

Uma moeda *justa* deu 10 caras seguidas. Qual resultado é mais provável no próximo lançamento: cara ou coroa?

Dizer que "esta moeda provavelmente é viciada" não é válido no problema proposto – afinal, partimos da hipótese de que a moeda *é justa*. Então, "cara" não é a resposta.

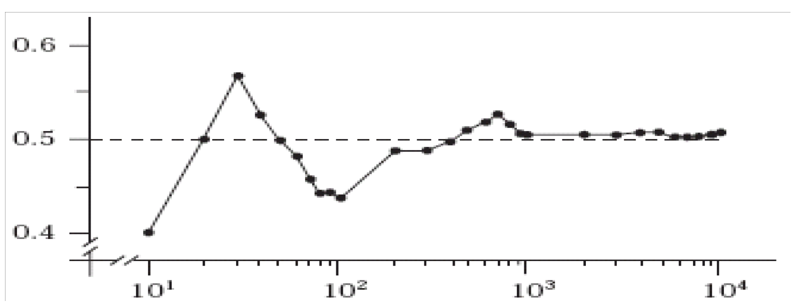
Por outro lado, como a moeda é justa, a proporção $\frac{\#(\text{caras})}{\#(\text{lançamentos})}$ deve se aproximar de 50% *a longo prazo* (esta é a interpretação freqüentista, a ser justificada mais adiante pela Lei dos Grandes Números). Note: *a longo prazo*! Assim, não há necessidade alguma da moeda "compensar as 10 caras lançadas" logo no próximo lançamento. Então coroa também não é a resposta!

Mas, se a proporção tem de se aproximar de 50%, mesmo a longo prazo, então no futuro as coroas vão ter que recuperar o terreno perdido para as caras, certo? Errado! Mesmo que nos próximos $2n$ lançamentos tivéssemos n caras e n coroas, a proporção nos $2n + 10$ lançamentos se aproximaria de 50% para n grande. Afinal:

$$\lim_{n\to\infty} \frac{n}{2n+10} = \frac{1}{2}$$

O problema é que "a proporção se aproxima de 0.5" não é o mesmo que "o número de caras se aproxima da metade do número de lançamentos"[6]! Considere o experimento de John Kerrich – um matemático sul-africano que, prisioneiro de guerra na Dinamarca durante a Segunda Guerra Mundial, lançou uma moeda 10000 vezes:

Lançamentos	10	40	100	200	400	800	2000	8000	10000
Caras	4	21	44	98	199	413	1013	4034	5067
Acima do Esp.	-1	1	-6	-2	-1	13	13	34	67
Proporção	0.4	0.525	0.44	0.49	0.4975	0.5163	0.5065	0.5043	0.5067



Note como o número de caras acima do "esperado" **parece oscilar e aumentar com o número de lançamentos.** Isto não contradiz a Interpretação Freqüentista: as **proporções** estão se aproximando de 0.5.

[3]Ou pelo menos é isso que a Física Clássica diria. Já a Mecânica Quântica (um dos pilares da Física Moderna) diria que as partículas que compõem o universo não *estão* em lugar algum – elas têm *probabilidades* de estar em lugares distintos ao mesmo tempo. Esta incerteza não seria devida à nossa incapacidade de criar instrumentos para medi-las, mas seria uma característica intrínseca da natureza do universo. Assim, é impossível ter conhecimento completo sobre o estado atual do universo – ou seja, há uma parcela de chance em todos os fenômenos físicos. Difícil de engolir? Você não está sozinho: até Einstein tinha dificuldades de aceitar este modelo, dizendo estar "convencido de que Deus não joga dados". Apesar disto, a Mecânica Quântica explica fenômenos observáveis que contradizem frontalmente a Física Clássica de Newton!

[4]*Quase* impraticável: no artigo "Dynamical Bias in the Coin Toss" (2004), Diaconis, Holmes e Montgomery analisam mais cuidadosamente o processo de lançar uma moeda e pegá-la com a mão. Conclusão do artigo: se a moeda mostrava cara no início do lançamento, a probabilidade de mostrar cara ao final é cerca de 50.8%!

[5]A pergunta realmente misteriosa é a seguinte: por que os conceitos básicos formais da Teoria da Probabilidade só aparecem no século XVII, quando Pascal e Fermat começaram sua célebre correspondência a respeito de jogos de azar? Afinal, ignorância, incerteza e jogos de azar existem há mais de 5000 anos...

[6]Matematicamente, se $f(n)$ é o número de caras em n lançamentos:

$$\lim_{n\to\infty} \frac{f(n)}{n} = \frac{1}{2} \text{ não significa que } \lim_{n\to\infty} \left(f(n) - \frac{n}{2}\right) = 0$$

1.2 Modelos de Probabilidade

Considere um experimento qualquer cujo resultado não seja conhecido (ou seja, um experimento **aleatório**). Chamaremos de **espaço amostral** o conjunto de todos os resultados possíveis deste experimento[7], comumente denotado por S. Um **evento** é representado por um subconjunto qualquer de S; diz-se que um evento **ocorre** se algum de seus elementos foi o resultado.

Exemplo 1 *Lança-se um dado e observa-se a face que cai voltada para cima. O espaço amostral é* $S = \{1, 2, 3, 4, 5, 6\}$. *Alguns eventos (que serão utilizados no resto desta seção) são:*
$A =$ *"o número observado é par"*$= \{2, 4, 6\}$
$B =$ *"o número observado é maior do que 3"*$=\{4, 5, 6\}$
$C =$ *"o número observado é maior do que 4"*$=\{5, 6\}$
Note que, se o resultado for 6, todos estes três eventos ocorrem.

A partir de eventos quaisquer, podemos construir novos eventos usando as operações de complemento, união e interseção:

- $\bar{A}$ é o evento "A **NÃO** ocorre";
- $A \cup B$ é o evento "A ocorre **OU** B ocorre";
- $A \cap B$ é o evento "A ocorre **E** B ocorre".

Exemplo 2 *Usando a notação do exemplo anterior:*
$\bar{A} = \{1, 3, 5\} =$ *"o número observado* não *é par"*
$A \cup B = \{2, 4, 5, 6\} =$ *"o número é par* ou *maior do que 3"*
$A \cap B = \{4, 6\} =$ *"o número é par* e *é maior do que 3".*

Definição 3 *Dois eventos A e B são chamados de* ***mutuamente excludentes*** *se não podem ocorrer simultaneamente, isto é, se $A \cap B = \emptyset$.*

Exemplo 4 *X e $\bar{X}$ são sempre mutuamente excludentes; no exemplo anterior, C e $\{1, 2\}$ são mutuamente excludentes – o número não pode ser maior do que 4 e menor do que 3 simultaneamente.*

Associaremos a cada evento um número, que chamaremos de **probabilidade do evento** e que traduzirá nossa confiança na capacidade do evento ocorrer.

Definição 5 *Uma* ***probabilidade*** *é uma função que associa a cada evento A um número $\Pr(A)$ de forma que:*
i) Para todo evento A, $0 \leq \Pr(A) \leq 1$;
ii) $\Pr(S) = 1$;
iii) Se A e B são eventos mutuamente excludentes então[8]

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

Não é difícil ver que, para atribuir probabilidades a um espaço amostral finito, basta atribuir probabilidades a cada um de seus eventos elementares (representados por conjuntos com um único elemento).

Exemplo 6 *Se acreditarmos que o dado é justo (todas as faces têm a mesma chance), então usaríamos*

$$\Pr(\{1\}) = \Pr(\{2\}) = \Pr(\{3\}) = \Pr(\{4\}) = \Pr(\{5\}) = \Pr(\{6\}) = \frac{1}{6}.$$

[7] Ao listar um espaço amostral, é importante que *todos* os resultados sejam listados, e que *nenhum deles apareça mais de uma vez.*

[8] Para espaços amostrais infinitos, deveríamos incluir uma condição semelhante com infinitos eventos mutuamente excludentes dois a dois:

$$\Pr(A_1 \cup A_2 \cup ...) = \Pr(A_1) + \Pr(A_2) + ...$$

Neste caso, teríamos

$$\begin{aligned}
\Pr(A) &= \Pr(\{2\})+\Pr(\{4\})+\Pr(\{6\}) = \frac{1}{6}+\frac{1}{6}+\frac{1}{6}=\frac{3}{6}=\frac{1}{2}=50\% \\
\Pr(B) &= \Pr(\{4\})+\Pr(\{5\})+\Pr(\{6\}) = \frac{1}{6}+\frac{1}{6}+\frac{1}{6}=\frac{3}{6}=\frac{1}{2}=50\% \\
\Pr(C) &= \Pr(\{5\})+\Pr(\{6\}) = \frac{1}{6}+\frac{1}{6}=\frac{2}{6}=\frac{1}{3}=33.333...\%
\end{aligned}$$

Mas, se você acredita que o dado é viciado, nada impede que você use outros modelos. Por exemplo, talvez eu acredite em

$$\begin{aligned}
\Pr(\{2\}) &= \Pr(\{3\}) = \Pr(\{4\}) = \Pr(\{5\}) = 10\% \\
\Pr(\{1\}) &= 20\%;\ \Pr(\{6\}) = 40\%
\end{aligned}$$

(caso em que o dado não é justo*). Fica a cargo do leitor ver como as probabilidades de A, B e C se alteram neste caso para* 60%, 60% *e* 50%.

As demonstrações das seguintes propriedades do cálculo de probabilidades são simples e deixadas como exercício para o leitor:

Proposição 7 (Lei do Complemento) $\Pr(\bar{A}) = 1-\Pr(A)$. *Em outras palavras, a probabilidade de um evento ocorrer mais a probabilidade de ele não ocorrer dá* 100%.

Proposição 8 $\Pr(\emptyset) = 0$, *isto é, se um evento é impossível, sua probabilidade deve ser* 0.

Proposição 9 (Lei da Adição)

$$\boxed{\Pr(A\cup B) = \Pr(A)+\Pr(B)-\Pr(A\cap B)}$$

isto é, a probabilidade de A ou B ocorrer é a probabilidade de A ocorrer, mais a probabilidade de B ocorrer, menos a probabilidade de A e B ocorrerem (pois esta "havia sido contada duas vezes"!).

Exemplo 10 *Nos exemplos anteriores, tínhamos* $\bar{A} = \{1,3,5\}$, $A\cup B = \{2,4,5,6\}$ *e* $A\cap B = \{4,6\}$. *Se o dado for justo, teremos:*

$$\begin{aligned}
\Pr(\bar{A}) &= \frac{3}{6} = 50\% = 1-\Pr(A) \\
\Pr(A\cup B) &= \frac{4}{6} = \frac{3}{6}+\frac{3}{6}-\frac{2}{6} = \Pr(A)+\Pr(B)-\Pr(A\cap B)
\end{aligned}$$

Se o dado for viciado como descrito no exemplo anterior, então teríamos

$$\begin{aligned}
\Pr(\bar{A}) &= 20\%+10\%+10\% = 40\% = 1-\Pr(A) \\
\Pr(A\cup B) &= 10\%+10\%+10\%+40\% = 70\% = \\
&= \Pr(A)+\Pr(B)-\Pr(A\cap B) = (10\%+10\%+40\%)+(10\%+10\%+40\%)-(10\%+40\%) = 70\%
\end{aligned}$$

e as leis continuam valendo.

Da Lei da Adição, note que

$$\boxed{\Pr(A\cup B) = \Pr(A)+\Pr(B) \Leftrightarrow \Pr(A\cap B) = 0}$$

ou seja, você pode somar probabilidades **apenas** no caso em que os eventos sejam mutuamente excludentes[9] (bom, e quando você quiser calcular $\Pr(A\cup B)$, a probabilidade de pelo menos um deles ocorrer).

[9] Tecnicamente, isto não é bem verdade – mais tarde veremos que há eventos de probabilidade 0 que *podem* acontecer, e assim $\Pr(A\cap B) = 0$ não significa necessariamente "mutuamente excludentes"... Voltaremos a esta discussão no momento apropriado.

Exemplo 11 *Numa rotina clássica dos trapalhões, Didi argumenta que, sendo sua jornada apenas de 8 horas diárias, ele não precisa trabalhar nos outros $\frac{2}{3}$ do tempo do ano. Mas ele também não precisa trabalhar durante $\frac{2}{7}$ do ano (finais de semana), e a lei lhe garante um mês de férias – outros $\frac{1}{12}$ do ano em que não se trabalha. Somando tudo, a probabilidade do Didi não trabalhar num dia escolhido a esmo seria $\frac{2}{3}+\frac{2}{7}+\frac{1}{12}=\frac{29}{28}$, o que já deu mais de 100% (sem contar feriados, hora do almoço, Copa do Mundo, etc.)! Assim, o patrão do Didi tem que deixá-lo em casa o ano todo e ainda lhe pagar hora extra... Onde está o erro? Ora, não se podem simplesmente somar estas proporções pois os eventos não são mutuamente excludentes! Por exemplo, Didi contou horas de dormir, em finais de semana, durante as férias, três vezes!*

Um **modelo equiprobabilístico** num espaço amostral S com n elementos associa a cada evento elementar a probabilidade $\frac{1}{n}$. Se o modelo é equiprobabilístico, então a probabilidade de um evento é simplesmente[10]

$$\boxed{\Pr(A)=\frac{\#(A)}{\#(S)}=\frac{\text{“número de casos favoráveis”}}{\text{“número de casos totais”}}}$$

Nota 12 *Cuidado! Um erro muito muito muito comum é usar esta fórmula (ou este tipo de raciocínio) para modelos que não são equiprobabilísticos! Só porque o seu espaço amostral é $S=\{$ganho na loteria, não ganho na loteria$\}$ não significa que você tem 50% de chance de ganhar na loteria! Mais na frente veremos problemas (como o de Monty Hall) onde nossa intuição tem uma vontade terrível de fazer este tipo de raciocínio – e nossa intuição erra redondamente.*

1.2.1 Exercícios

Ex. 1 *Defina espaços amostrais razoáveis para os seguintes experimentos:*
a) Jogue uma moeda três vezes e anote a seqüência de caras (K) e coroas (C).
b) Jogue dois dados e anote a soma de seus pontos.
c) Jogue dois dados e anote a diferença de seus pontos.
d) Jogue um dado até que o número 6 apareça e anote quantas vezes ele foi jogado.
e) Jogue uma moeda 100 vezes e anote quantas caras foram obtidas.
f) Tire 6 bolas de uma urna com 100 bolas azuis e 200 bolas brancas e anote quantas bolas brancas foram retiradas.
g) Anote o lanterna do próximo campeonato brasileiro.
h) Anote o instante em que você recebe a primeira ligação telefônica do dia.
i) Anote a temperatura máxima do dia no seu quarto.
Em quais dos exemplos a-g acima é razoável usar um modelo eqüiprovável?

Ex. 2 *A partir dos três axiomas básicos da Probabilidade:*

$$\begin{aligned}
\textit{Para todo evento } A &: \quad 0\leq \Pr(A)\leq 1;\\
\textit{Para o espaço amostral } S &: \quad \Pr(S)=1;\\
\textit{Para quaisquer eventos mutuamente excludentes } A \textit{ e } B &: \quad \Pr(A\cup B)=\Pr(A)+\Pr(B)
\end{aligned}$$

Demonstre as seguintes propriedades:
a) A Lei do Complemento: $\Pr(\bar{A})=1-\Pr(A)$ *[Dica: A e $\bar{A}$ são mutuamente excludentes.]*
b) $\Pr(\emptyset)=0$ *[Dica: use o item anterior.]*
c) A Lei da Adição: $\Pr(A\cup B)=\Pr(A)+\Pr(B)-\Pr(A\cap B)$ *[Dica: $B-A$ e $A\cap B$ são mutuamente excludentes, assim como $B-A$ e A.]*
d) Se $A\subseteq B$ então $\Pr(A)\leq\Pr(B)$ *[Dica: $B-A$ e A são...]*

Ex. 3 *Mostre que*

$$\Pr(A\cup B\cup C)=\Pr(A)+\Pr(B)+\Pr(C)-\Pr(A\cap B)-\Pr(A\cap C)-\Pr(B\cap C)+\Pr(A\cap B\cap C).$$

Ex. 4 *Dados* $\Pr(A)=0.4$, $\Pr(B)=0.5$, $\Pr(C)=0.3$, $\Pr(A\cap B)=0.3$, $\Pr(A\cap C)=0$ *e* $\Pr(B\cap C)=0.1$, *determine:*
a) $\Pr(A\cap B\cap C)$

[10] Dado um conjunto X, a notação $\#(X)$ representa o número de elementos de X.

b) $\Pr(A \cup (B \cap C))$
c) $\Pr(A \cap (B - C))$
d) $\Pr(A - (B \cup C))$
e) $\Pr(A \cup B \cup C)$

Ex. 5 *Em certa escola a probabilidade de um aluno ser torcedor do Flamengo é* $0,6$*, de assistir novela é* $0,7$ *e de gostar de praia é* $0,8$*. Entre que valores está compreendida a probabilidade de um aluno dessa escola, simultaneamente, torcer pelo Flamengo, assistir novela e gostar de praia?*

Ex. 6 *Lança-se uma moeda* justa *três vezes e anota-se a seqüência de Caras (K) e Coroas (C) obtidas.*
a) Que modelo de probabilidade lhe parece razoável em S*?*
Sejam A *o evento "dois primeiros resultados são iguais",* B *o evento "o primeiro lançamento é uma cara" e* C *o evento "pelo menos um lançamento é uma cara".*
b) Escreva A, B *e* C *como subconjuntos de* S *e calcule as probabilidades de cada um.*
c) Interprete os seguintes eventos em linguagem comum e calcule as suas probabilidades:
i) $\bar{A}$ *ii)* $\bar{C}$ *iii)* $A \cap B$ *iv)* $B \cap C$ *v)* $B \cup C$ *vi)* $A \cup B$.

Ex. 7 *Lança-se uma moeda* justa *até obter-se duas caras ou duas coroas, não necessariamente consecutivas (ou seja, Kuerten e Coria disputam uma partida de tênis em três sets e têm chances iguais de vencer cada set). Anota-se a seqüência obtida (os vencedores de cada set). Repita os itens a-c do exercício anterior. Que respostas mudaram?*

Ex. 8 *Dois dados são lançados – um vermelho e um verde. Escreva um espaço amostral para este experimento, e calcule a probabilidade de a soma dos dois dados ser* 9*. O problema se altera se os dados forem da mesma cor?*

Ex. 9 *Os* 12 *times do campeonato do Rio são sorteados de forma completamente aletaória em dois grupos de* 6 *times cada. Qual a probabilidade de o Flamengo e o Fluminense acabarem no mesmo grupo?*

Ex. 10 *Em uma roda são colocadas* n *pessoas. Qual é a probabilidade de duas dessas pessoas ficarem juntas?*

Ex. 11 *Em uma fila são colocadas* n *pessoas. Qual é a probabilidade de duas dessas pessoas ficarem juntas?*

Ex. 12 *Laura e Telma retiram cada uma um bilhete numerado de uma urna que contém bilhetes numerados de* 1 *a* 100*. Determine a probabilidade do número de Laura ser maior que o de Telma, supondo a extração:*
a) sem reposição.
b) com reposição.

Ex. 13 *Três jogadores, A, B e C, disputam um torneio. Os três têm probabilidades iguais de ganhar o torneio; têm também probabilidades iguais de tirarem o segundo lugar e têm probabilidades iguais de tirarem o último lugar. É necessariamente verdadeiro que cada uma das seis ordens possíveis de classificação dos três jogadores tem probabilidade* $\frac{1}{6}$ *de ocorrer?*

Ex. 14 *Dois dados são lançados. Os eventos* $A =$ *"número do primeiro dado foi* a*" e* $B =$ *" a soma dos dados é* b*" são mutuamente excludentes (onde* $1 \leq a \leq 6$ *e* $2 \leq b \leq 12$*). Que outras conclusões você pode tirar sobre* a *e* b*?*

Ex. 15 ([EXCEL]) *Há* n *alunos em uma sala de aula. Qual a probabilidade de haver pelo menos um par que faça aniversário no mesmo dia (e mês)? Monte uma planilha mostrando os valores de* n *e as probabilidades correspondentes para* $1 \leq n \leq 366$*. Qual valor de* n *nos dá uma probabilidade de aproximadamente* 50%*?*

Ex. 16 (*) *a) Uma loteria semanal tem* 100 *bilhetes. Quem tem a maior chance de ganhar algum prêmio: quem compra* 10 *bilhetes numa semana ou quem compra* 1 *bilhete por semana durante* 10 *semanas?*
b) Seja $n > 1$*. Encontre o mínimo da função* $f(x) = (1+x)^n - (1+nx)$ *para* $x \in (-1, \infty)$*. Conclua que* $f(x) \geq 0$ *para* $x > -1$.
c) Generalize o item (a): suponha $1 < n < N$*. Numa loteria com* N *bilhetes, é melhor comprar* n *numa semana ou* 1 *por dia durante* n *semanas?*

1.3 Probabilidade Condicional

Se tivermos informação adicional sobre um experimento, podemos ser forçados a reavaliar as probabilidades dos eventos a ele associados.

Exemplo 13 *Como na seção anterior, jogue um dado e anote o valor de sua face superior. Então* $S = \{1, 2, 3, 4, 5, 6\}$. *Sejam* $A = \{2, 4, 6\}$, $B = \{4, 5, 6\}$ *e* $C = \{5, 6\}$. *Se o dado é justo, teremos:*

$$\Pr(A) = \frac{3}{6};\ \Pr(B) = \frac{3}{6};\ \Pr(C) = \frac{2}{6}$$

Agora, suponha que você sabe de alguma forma que o número rolado é par. Então seu novo universo é $A = \{2, 4, 6\}$. Sabendo-se que o número é par, *qual a probabilidade de ele ser maior do que* 3*? Ou seja, qual a chance de* B *ocorrer* na ceretza de que A ocorreu*? Esta é a chamada* ***probabilidade condicional de*** B ***dado*** A*; neste caso*

$$\Pr(B|A) = \frac{2}{3}$$

pois há apenas 2 *casos "favoráveis a* B*" dentre os* 3 *casos "possíveis em* A*". Analogamente, convença-se de que:*

$$\Pr(A|B) = \frac{2}{3};\ \Pr(A|C) = \frac{1}{2};\ \Pr(C|A) = \frac{1}{3};\ \Pr(B|C) = 1;\ \Pr(C|B) = \frac{2}{3}$$

Escreva estas probabilidades em linguagem comum: $\Pr(A|B) = \frac{2}{3}$ *significa que "sabendo-se que o número é maior que três, há* $\frac{2}{3}$ *de chance de ele ser par". Numa interpretação freqüentista, diríamos "se rolarmos o dado várias vezes, deu um número par cerca de* $\frac{2}{3}$ *das vezes em que o número foi maior do que* 3*".*
Note que $\Pr(B|C) = 100\%$, *isto é, "na certeza de que deu mais do que quatro, é óbvio que deu mais do que três", ou seja, "*B *acontece sempre que* C *acontece".*
Por outro lado, $\Pr(C|B) = \frac{2}{3}$ *apenas. Assim, "de cada* 3 *vezes em que* B *ocorre,* C *ocorre em apenas* 2*".*

O exemplo acima inspira a seguinte fórmula:

Definição 14 *Sejam* A *e* B *dois eventos com* $\Pr(A) \neq 0$. *A probabilidade condicional de* B *dado* A *é*

$$\boxed{\Pr(B|A) = \frac{\Pr(A \cap B)}{\Pr(A)}}$$

Exemplo 15 *Usando esta fórmula no exemplo anterior, temos*

$$\Pr(B|A) = \frac{\Pr(A \cap B)}{\Pr(A)} = \frac{2/6}{3/6} = \frac{2}{3}$$

Note como o número de elementos do espaço amostral S *(no caso,* 6*) desaparece e ficamos ao final apenas com a proporção dos elementos de* B *(que estão também em* A*) com relação aos elementos de* A*.*

Exemplo 16 *A tabela abaixo dá a distribuição dos alunos de uma turma, por sexo e por carreira pretendida:*

	M	*F*	*total*
ADM	*15*	*45*	*60*
ECO	*21*	*9*	*30*
total	*36*	*54*	*90*

Escolhe-se ao acaso um aluno. Sejam M, F, A *e* E *os eventos o aluno selecionado é do sexo masculino, é do sexo feminino, cursa ADM e cursa ECO, respectivamente. Temos:*
i) $\Pr(A) = \frac{60}{90}$, *isto é,* 66.67% *dos alunos cursam ADM; os outros* $\frac{30}{90} = 33.33\%$ *cursam ECO. Se você escolher um aluno ao acaso, há* 66.67% *de chance de ele ser de ADM.*
ii) $\Pr(A|M) = \frac{15}{36}$, *isto é,* 41.67% dos alunos homens *cursam ADM; os outros* 58.33% dos homens *estão em ECO. Se você escolher um aluno* homem *ao acaso, há* 41.67% *de chance de ele estudar ADM.*
iii) $\Pr(M|A) = \frac{15}{60}$, *isto é,* 25% dos alunos de ADM *são homens; se você escolher um aluno* de ADM *ao acaso, há* 25% *de chance deste aluno ser homem.*

A fórmula da probabilidade condicional é freqüentemente utilizada para **descobrir** $\Pr(A \cap B)$:

Proposição 17 (Lei da Multiplicação)

$$\boxed{\Pr(A \cap B) = \Pr(B|A) . \Pr(A) = \Pr(A|B) . \Pr(B)}$$

Exemplo 18 *Uma urna contém 4 bolas brancas e 6 bolas pretas. Sacam-se, sucessivamente e sem reposição, duas bolas dessa urna. Determine a probabilidade de ambas serem brancas.*
Solução: *Sejam* $B_1 = \{\text{primeira bola é branca}\}$ *e* $B_2 = \{\text{a segunda bola é branca}\}$. *Então*

$$\Pr(B_1 \cap B_2) = \Pr(B_1) . \Pr(B_2|B_1) = \frac{4}{10}\frac{3}{9} = \frac{12}{90} = \frac{2}{15}$$

Note que foi bastante simples o cálculo de $\Pr(B_2|B_1)$. *Realmente, na certeza de que a primeira bola foi branca, é fácil calcular a probabilidade da segunda bola ser branca, pois, para a segunda extração, a urna está com 3 bolas brancas e 6 pretas. De modo mais geral, é fácil calcular probabilidades condicionais quando as coisas estão na ordem certa, isto é, é fácil calcular probabilidades de coisas futuras na certeza de coisas passadas.*

Exemplo 19 *Uma urna contém* 4 *bolas brancas e* 6 *bolas pretas. Sacam-se, sucessivamente e sem reposição, duas bolas dessa urna. Determine a probabilidade da primeira bola ser branca sabendo que a segunda bola é branca.*
Solução: *Sejam* B_1 *e* B_2 *como no problema anterior. Queremos* $\Pr(B_1|B_2)$. *Note que essa é uma probabilidade do passado na certeza do futuro. Aqui usamos a fórmula da definição de probabilidade condicional:*

$$\Pr(B_1|B_2) = \frac{\Pr(B_1 \cap B_2)}{\Pr(B_2)}$$

Calculamos $\Pr(B_1 \cap B_2)$ *no exemplo anterior. Para calcular* $\Pr(B_2)$, *basta notar a simetria do problema – não há motivo para imaginar que a segunda bola seja branca mais ou menos freqüentemente do que a primeira! Se este argumento não lhe parece convincente, faça o seguinte: considere separadamente os casos em que a primeira bola é branca e os casos onde a primeira bola não é branca:*

$$\Pr(B_2) = \Pr(B_1 \cap B_2) + \Pr\left(\overline{B_1} \cap B_2\right)$$

A primeira parcela já foi calculada. Quanto à segunda:

$$\Pr\left(\overline{B_1} \cap B_2\right) = \Pr\left(\overline{B_1}\right) . \Pr\left(B_2|\overline{B_1}\right) = \frac{6}{10}.\frac{4}{9} = \frac{4}{15}$$

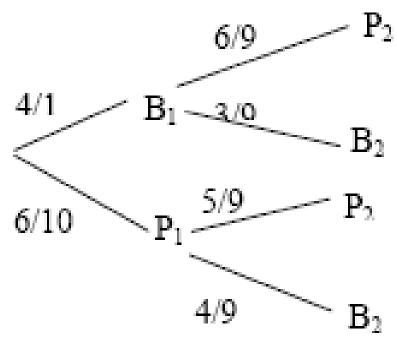
já que, após retirar uma bola preta, ficam 4 *brancas dentre* 9 *bolas. Juntando tudo,*

$$\Pr(B_2) = \frac{2}{15} + \frac{4}{15} = \frac{6}{15}$$

como havíamos afirmado anteriormente. Enfim:

$$\Pr(B_1|B_2) = \frac{\Pr(B_1 \cap B_2)}{\Pr(B_2)} = \frac{2/15}{6/15} = \frac{1}{3}$$

Uma maneira boa ilustrar esta solução é usar uma árvore de probabilidades:



Menos intuitiva (mas mais fácil de digitar) seria uma tabela como esta (de uma "população fictícia de 900 *retiradas*[11] *"):*

	B_1	$\overline{B_1}$	*Totais*
B_2	120	240	360
$\overline{B_2}$	240	300	540
Totais	360	540	900

Note que as posições das bolas (primeira e segunda) são intercambiáveis, isto é, $\Pr(B_1) = \Pr(B_2)$*, e mais,* $\Pr(B_1|B_2) = \Pr(B_2|B_1)$*, e assim por diante.*

Exemplo 20 *Você tem duas moedas, uma com duas caras e a outra justa. Escolha uma delas e a lance. O resultado é cara. Qual a chance de ela ser a moeda "viciada"?*
Solução: *seja* V *o evento "escolhemos a moeda justa" e* K *o evento "deu cara". Então:*

$$\Pr(V|K) = \frac{\Pr(V \cap K)}{\Pr(K)}$$

Mas

$$\Pr(V \cap K) = \Pr(V) . \Pr(K|V) = \frac{1}{2}\frac{1}{1} = \frac{1}{2}$$

e

$$\Pr(K) = \Pr(V \cap K) + \Pr\left(\bar{V} \cap K\right)$$

onde

$$\Pr\left(\bar{V} \cap K\right) = \Pr\left(\bar{V}\right) . \Pr\left(K|\bar{V}\right) = \frac{1}{2}\frac{1}{2} = \frac{1}{4}$$

Juntando tudo

$$\begin{aligned} \Pr(K) &= \frac{1}{2} + \frac{1}{4} \Rightarrow \\ &\Rightarrow \Pr(V|K) = \frac{\frac{1}{2}}{\frac{1}{2} + \frac{1}{4}} = \frac{2}{3} \end{aligned}$$

Exemplo 21 *Algumas pesquisas estatísticas podem causar constrangimentos aos entrevistados com perguntas do tipo "você usa drogas?" e correm o risco de não obter respostas sinceras ou não obter respostas de espécie alguma. Para estimar a proporção* p *de usuários de drogas em certa comunidade, pede-se ao entrevistado que, longe das vistas do entrevistador, jogue uma moeda: se o resultado for coroa, responda a "você usa drogas?" e, se o resultado for cara, responda "sim". Assim, caso o entrevistado diga sim, o entrevistador não saberá se ele é um usuário de drogas ou se a moeda deu cara.*
Se s *é a probabilidade de um entrevistado responder sim,* s *é facilmente estimado pela proporção de respostas sim obtidas nas entrevistas. Estime* p *a partir de* s*.*
Solução: *seja* D *o evento "usuário disse que usa drogas" e* K *o evento "moeda deu cara". Colocando tudo numa tabela, temos.*

	D	$\overline{D}$	
K			0.5
$\overline{K}$			0.5
	p	$1-p$	1

Como D *e* K *são independentes, podemos completar a tabela simplesmente multiplicando as probabilidades correspondentes:*

	D	$\overline{D}$	
K	$0.5p$	$0.5(1-p)$	0.5
$\overline{K}$	$0.5p$	$0.5(1-p)$	0.5
	p	$1-p$	1

[11] Não estamos dizendo que de cada 900 experimentos, extamente 360 terão a primeira bola branca – estamos dizendo que as proporções representadas pelos números da tabela são exatamente as probabilidades do problema.

Note que os entrevistados que dizem "sim" estão em 3 lugares da tabela acima – ambas da linha K e os usuários de drogas da célula $D\bar{K}$. Assim

$$s = 0.5p + 0.5\,(1-p) + 0.5p = 0.5\,(1+p) \Rightarrow p = 2s - 1$$

Por exemplo, se 60% dos entrevistados respondem sim, você pode estimar em 20% a proporção de usuários de drogas.

1.3.1 Probabilidade Total e Teorema de Bayes

Os problemas anteriores ilustram duas técnicas comuns para obtenção de probabilidades:

Proposição 22 (Lei da Probabilidade Total) *Suponha que B_1, B_2, ..., B_n formam uma partição*[12] *de S. Então*

$$\begin{aligned}\Pr(A) &= \Pr(A\cap B_1) + \Pr(A\cap B_2) + ... + \Pr(A\cap B_n)\\ &= \Pr(A|B_1)\,.\Pr(B_1) + \Pr(A|B_2)\,.\Pr(B_2) + ... + \Pr(A|B_n)\,.\Pr(B_n)\end{aligned}$$

Em particular, a partição $S = B\cup\bar{B}$ nos dá

$$\boxed{\Pr(A) = \Pr(A|B)\,.\Pr(B) + \Pr\left(A|\bar{B}\right).\Pr\left(\bar{B}\right)}$$

De fato, A é a união dos conjuntos (sem interseção dois a dois!) da forma $A\cap B_i$, justificando a primeira igualdade. A segunda igualdade vem simplesmente de aplicar a Lei da Multiplicação várias vezes. Compare esta "lei" com os exemplos da subseção anterior.

Proposição 23 (Teorema de Bayes) *Suponha que B_1, B_2, ..., B_n formam uma partição de S. Então:*

$$\Pr(B_1|A) = \frac{\Pr(A|B_1)\,.\Pr(B_1)}{\Pr(A|B_1)\,.\Pr(B_1) + \Pr(A|B_2)\,.\Pr(B_2) + ... + \Pr(A|B_n)\,.\Pr(B_n)}$$

Em particular

$$\boxed{\Pr(B|A) = \frac{\Pr(A|B)\,.\Pr(B)}{\Pr(A|B)\,.\Pr(B) + \Pr\left(A|\bar{B}\right).\Pr\left(\bar{B}\right)}}$$

O Teorema de Bayes nos dá a fórmula exata para calcular uma condicional quando temos as condicionais "na outra ordem". Apesar de muito útil, em geral ele é mais fácil de ser entendido com o auxílio de tabelas ou árvores – novamente, perceba como ele foi utilizado nos exemplos anteriores.

1.3.2 Independência

Em algumas ocasiões, o conhecimento sobre a ocorrência de um evento não muda a probabilidade de um outro – este é o conceito de independência estatística:

Definição 24 *Dois eventos (não impossíveis) A e B são ditos* ***independentes*** *se o conhecimento de um deles não afeta a probabilidade do outro ocorrer, isto é, se*

$$\Pr(B|A) = \Pr(B)$$

Intuitivamente, se tudo o que você quer saber é se B acontece ou não, informações sobre o evento A não vão lhe ajudar em nada (e, portanto, você não pagaria dinheiro algum pela informação de A ter acontecido ou não – mesmo que você tenha certeza de que A acontece, a probabilidade de B não muda).

Exemplo 25 *No caso dos dados do início desta seção, note que A e C são independentes, pois*

$$\Pr(A|C) = \frac{1}{2} = \Pr(A)\,;\quad \Pr(C|A) = \frac{1}{3} = \Pr(C)$$

No entanto, A e B não são independentes, muito menos B e C.

[12] Isto é, eles são mutuamente excludentes dois a dois e $B_1\cup B_2\cup ...\cup B_n = S$.

Note que, da definição de probabilidade condicional, concluímos que dois eventos são independentes se, e somente se, $\Pr(B|A) = \frac{\Pr(A\cap B)}{\Pr(A)} = \Pr(B)$, isto é

$$\boxed{A \text{ e } B \text{ são independentes} \Leftrightarrow \Pr(A\cap B) = \Pr(A).\Pr(B)}$$

De quebra, passando $\Pr(B)$ para o lado esquerdo, acabamos de mostrar que

$$\Pr(B|A) = \Pr(B) \Leftrightarrow \Pr(A\cap B) = \Pr(A) . \Pr(B) \Leftrightarrow \Pr(A|B) = \Pr(A)$$

Pode-se mostrar também que, se A e B são independentes, então

$$\Pr(B|A) = \Pr(B) = \Pr\left(B|\bar{A}\right)$$

Exemplo 26 *Suponha que, numa família com duas crianças, a probabilidade do filho estar gripado é 40% ($\Pr(H) = 0.4$) e a probabilidade da filha estar gripada é 60% ($\Pr(M) = 0.6$). É possível calcular a probabilidade de ambos estarem gripados?*
Se supusermos que estes dois eventos são independents, então é simples: $\Pr(H\cap M) = (0.4)(0.6) = 24\%$. *Mas será que esta suposição é razoável? Afinal, se um deles estiver gripado, imagina-se que a probabilidade do outro estar gripado aumenta. Matematicamente falando, acreditamos que* $\Pr(H|M) \geq \Pr(H) = 40\%$, *e a probabilidade condicional é que teria de ser usada:*

$$\Pr(H\cap M) = \Pr(H|M) . \Pr(M)$$

Sem mais dados, não é possível resolver o problema.

Exemplo 27 *Por outro lado, se no problema anterior forem dados* $\Pr(H) = 0.4$, $\Pr(M) = 0.6$ *e* $\Pr(H\cap M) = 0.3$, *é possível* verificar *se os eventos H e M são independentes! De fato, como* $\Pr(H\cap M) \neq \Pr(H) . \Pr(M)$, *os eventos não seriam independentes. Outras maneiras de chegar à mesma conclusão:*

$$\begin{aligned} \Pr(H|M) &= \frac{0.3}{0.6} = 0.5 > 0.4 = \Pr(H) \\ \Pr(M|H) &= \frac{0.3}{0.4} = 0.75 > 0.6 = \Pr(M) \end{aligned}$$

Neste caso, diz-se que o evento H ***atrai*** *o evento M ou que os eventos são* ***positivamente associados****.*

1.3.3 Estudo de Caso

Problema: Uma pessoa deseja saber se tem o vírus da AIDS ou não. Ao fazer o teste, o teste pode indicar POSITIVO (+) ou NEGATIVO (-). No entanto, nenhum teste é 100% correto − em algumas ocasiões, o teste pode ser + mesmo que esta pessoa não tenha a doença (o chamado *falso positivo*); em outras, apesar do paciente estar doente, o teste apresenta resultado - (um *falso negativo*). Digamos que você tem em mãos os seguintes dados a respeito do "Teste Elisa" para AIDS:

- Apenas 0.5% das pessoas no seu país têm AIDS (diz-se que a *prevalência* da doença é 0.5%)
- "Elisa" identifica corretamente (como +) 98% das pessoas que têm o vírus (diz-se que a *sensitividade* do teste é de 98%);
- "Elisa" identifica corretamente (-) 93% das pessoas que não têm o vírus (diz-se que a *especificidade* do teste é de 93%).

Um paciente escolhido aleatoriamente neste país é testado e testa +. Qual a probabilidade de ele ter o vírus?

Resposta: Este tipo de problema pode ser facilmente resolvido usando uma tabela com uma população "fictícia". Comece a tabela supondo que haja 10000 pessoas neste país, destas, 50 teriam AIDS e as outras 9950 não teriam:

	$AIDS$	$\overline{AIDS}$	Totais
$+$			
$-$			
Totais	50	9950	10000

Mas, daquelas 50, 98% (49 pessoas) testarão positivo; daquelas 9950, 7% (696.5 pessoas) serão falsos positivos. Use estes números para completar a tabela (não se preocupe com as populações fictícias e fracionárias – afinal, o que interessam são as proporções):

	$AIDS$	$\overline{AIDS}$	Totais
$+$	49	696.5	745.5
$-$	1	9253.5	9254.5
Totais	50	9950	10000

Agora podemos proceder a quaisquer respostas como no exemplo anterior. Por exemplo, se sabemos que o paciente testou $+$, qual a chance de ele ter $AIDS$? Seria:

$$\Pr(AIDS|+) = \frac{49}{745.5} = 6.573\%$$

Dizemos que o *poder preditivo positivo do teste* é de 6.573%.

Note como o Teorema de Bayes resolve o problema com uma fórmula só, mas escondendo um bocado a intuição do exemplo. Afinal, os dados são

$$\Pr(+|AIDS) = 98\%; \ \ \Pr\left(-|\overline{AIDS}\right) = 93\%; \ \ \Pr(AIDS) = 0.5\%$$

dos quais tiramos via Lei do Complemento:

$$\Pr\left(+|\overline{AIDS}\right) = 7\% \text{ e } \Pr\left(\overline{AIDS}\right) = 99.5\%$$

Enfiando tudo na fórmula de Bayes:

$$\Pr(AIDS|+) = \frac{\Pr(+|AIDS).\Pr(AIDS)}{\Pr(+|AIDS).\Pr(AIDS) + \Pr\left(+|\overline{AIDS}\right).\Pr\left(\overline{AIDS}\right)} = \frac{(0.98)(0.005)}{(0.98)(0.005) + (0.07)(0.995)} = 6.573\%$$

Análise: Os dados apresentados acima são compatíveis com os valores de sensitividade e especificidade do Teste Elisa para AIDS. A prevalência da AIDS varia muito de país para país e de ano para ano – em 1990, era de cerca de 1% nos Estados Unidos, mas apenas 0.2% na Austrália. Como um teste que parecia tão preciso pode errar tanto? O problema é que esta doença é muito pouco comum (apenas 0.5% da população a tem). É mais provável que esta pessoa seja um dos "proporcionalmente poucos falsos positivos" dentre a grande massa de pessoas sadias do que um dos "proporcionalmente muitos corretos positivos" dentre as poucas pessoas doentes!

Este tipo de probabilidade tem de ser divulgada às pessoas que são testadas! É por este motivo que, ao testar + para uma doença, você deve realizar um segundo teste!

Por outro lado, não é que o primeiro teste foi "inútil" não. Um resultado positivo no primeiro teste aumenta a probabilidade de doença de 0.5% (**a priori**) para uns 7% (**a posteriori**). O paciente deve sim se preocupar muito mais do que antes do teste. O poder do teste não está em *determinar* a probabilidade de se estar doente, mas em *aumentá-la* a partir duma probabilidade a priori. A propósito, note como estes cálculos podem variar terrivelmente dependendo do que se sabe sobre este indivíduo *antes* de ele se testar. Por exemplo, se o indivíduo pertence a um grupo de risco (digamos, no caso da AIDS, se é hemofílico) não é razoável usar o 0.5% como probablidade a priori – usar-se-ia um número maior que refletisse a percentagem de hemofílicos que contraiu a doença. Aliás, é por este motivo que é impossível divulgar os tais 7% exatamente – este número depende da probabilidade a priori de cada indivíduo.

Para administradores, este tipo de raciocínio é algo que deve ser levado em conta antes de se decidir por testar ou não membros de uma organização (associado às reações psicológicas dos testados que não sabem a diferença entre $\Pr(+|AIDS)$ e $\Pr(AIDS|+)$; preconceitos que possam estar associados aos resultados dos testes; etc.).

Postcript: O seguinte texto foi retirado do site do "Superior Tribunal de Justiça", no momento sob o link HTTP://WWW.STJ.GOV.BR/WEBSTJ/NOTICIAS/DETALHES_NOTICIAS.ASP?SEQ_NOTICIA=6425. Textos semelhantes foram publicados em vários jornais do país em Setembro de 2002.

Quinta-feira, 26 de setembro de 2002
09:33 - Fundação Pró-Sangue terá de pagar indenização por erro em exame de HIV

Por causa de diagnóstico errado para HIV positivo, a Fundação Pró-Sangue Hemocentro de São Paulo terá de pagar uma indenização no valor de R$ 40 mil ao torneiro P.G.S.. No entendimento unânime da Terceira Turma do Superior Tribunal de Justiça (STJ), instituição que emite laudo sobre o vírus da Aids sem ressalva quanto à falibilidade do diagnóstico, tem de se responsabilizar se houver uma falha no resultado.

Com isso, os ministros do STJ mantiveram a decisão do Tribunal de Justiça de São Paulo (TJ-SP) que condenou a Fundação Pró-Sangue a pagar a indenização. De acordo com o acórdão do TJ-SP, o laudo feito pelo maior hemocentro da América Latina não trouxe nenhuma ressalva, observação ou advertência de que o resultado deveria ser confirmado para que houvesse certeza do diagnóstico. Para o Tribunal, a falibilidade do teste é de conhecimento notório de pessoas bem informadas. Não seria o caso, entretanto, de P.G.S., "um modesto operário".

Ele propôs ação de indenização por ato ilícito contra a Fundação Pró-Sangue para reparação dos danos causados pela notícia equivocada. Ao doar sangue ao hemocentro em junho de 1996, o torneiro teve de submeter-se ao teste para detectar se era soropositivo. O resultado apontou que ele era portador do vírus da Aids. Incoformado, Paulo Gomes fez outro exame que constou um diagnóstico "indeterminado".

Durante dois meses, o operário viveu um inferno, segundo relatou no processo. Em conseqüência do choque, passou a faltar várias vezes ao trabalho, sobreviveu à base de calmantes e adquiriu gastrite nervosa. O erro do laboratório também teria lhe causado insônia, depressão e ansiedade. Com as reiteradas faltas ao trabalho, o torneiro foi advertido pelo chefe. Ao saber do que se passava, o chefe o aconselhou a procurar um laboratório particular para fazer um novo exame. Foi, então, que ele descobriu a verdade: não era portador do vírus.

A Fundação Pró-Sangue diz que não agiu com imperícia ou negligência. A instituição somente teria efetuado os testes sorológicos, mas não transmitido os resultados. O argumento é de que a coleta do sangue, triagem e os demais contatos teriam sido feitos com o doador no Núcleo de Hematologia de São Caetano do Sul, onde os funcionários deveriam ter orientado o operário sobre a falibilidade do laudo.

Segundo a Fundação, no exame em que constou o resultado positivo foi realizado o "Teste Elisa". Depois, quando o diagnóstico foi "indeterminado" foi aplicado o "Teste Western Blot". A Pró-Sangue explica que todos os métodos sorológicos possuem uma faixa de resultados falsos-positivos. Por isso, são realizados testes para confirmação ou não do resultado inicial.

O pedido da ação de indenização proposta pelo operário foi julgado improcedente na primeira instância. O TJ-SP, entretanto, reverteu a decisão. Tampouco foi acolhido recurso apresentado pela Fundação àquele Tribunal. Foi então que a instituição propôs agravo regimental que foi julgado improcedente pela ministra Nancy Andrighi, em despacho monocrático.

A Fundação Pró-Sangue decidiu apresentar novo recurso (Agravo Regimental em Agravo de Instrumento), que foi apreciado pela Terceira Turma do STJ. Por unanimidade, os ministros negaram provimento sob o argumento de que para mudar a conclusão do TJ-SP seria necessário rever as provas existentes nos autos, o que é vedado ao Superior Tribunal de Justiça, de acordo com o que estabelece a súmula n^o 7/STJ.

Da mesma forma, os ministros negaram pedido de redução do valor da indenização arbitrado pelo TJ-SP. Segundo a relatora, ministra Nancy Andrighi, a quantia estabelecida é razoável. Na inicial, o autor da ação pediu uma indenização de mil salários mínimos, ou seja, R$ 200 mil. Este valor foi considerado excessivo pelos desembargadores de São Paulo.

1.3.4 Exercícios

Ex. 17 *Joga-se um dado não-viciado duas vezes. Determine a probabilidade condicional de obter* 3 *na primeira jogada sabendo que a soma dos resultados foi* 7.

Ex. 18 *Um estudante resolve um teste de múltipla escolha de* 10 *questões, com* 5 *alternativas por questão. Ele sabe* 60% *da matéria do teste. Quando ele sabe uma questão, ele acerta, e, quando não sabe, escolhe a resposta ao acaso. Se ele acerta uma questão, qual é a probabilidade de que tenha sido por acaso?*

Ex. 19 *Na sua capa de 21/3/2001, a revisa VEJA afirma que "47% dos brasileiros não sentem vontade de fazer sexo". Dentro da revista, encontramos que "35% das mulheres não sentem nenhuma vontade de ter relações" e "entre os homens, apenas 12% se queixam de falta de desejo". Explique porque a informação da capa é incompatível com a do texto da reportagem; que dados você precisaria para estimar corretamente o número da capa?*

Ex. 20 *Na seção anterior, Kuerten e Coria jogavam uma partida de tênis em 3 sets. Cada um deles tem 50% de chance de vencer cada set (e supõe-se os sets independentes entre si). Considere os eventos $A =$"Kuerten vence a partida" e $B =$"o jogo termina em 2 sets". Calcule as probabilidades de cada um deles. Eles são independentes? Mutuamente excludentes?*

Ex. 21 *Repita o problema anterior onde Ralph joga contra Kuerten – agora, a probabilidade de Kuerten vencer um set é 70%, mas os sets ainda são independentes entre si.*

Ex. 22 *Lança-se um dado 3 vezes. Cada vez você tirar 5 ou 6, você ganha \$1, caso contrário, você paga \$1. Seja $A =$"você teve algum lucro ao final do jogo" e $B =$"você perdeu \$1 no primeiro lançamento". Calcule $\Pr(A)$, $\Pr(B)$, $\Pr(A \text{ e } B)$ e $\Pr(A|B)$. Os eventos A e B são independentes? Mutuamente excludentes?*

Ex. 23 *Ralph está na FGV 70% do horário comercial, enquanto Morgado está na FGV 20% do horário comercial. Sabe-se também que, em 20% do horário comercial, nenhum dos dois está presente à FGV. Os eventos "Ralph está na FGV" e "Morgado está na FGV" são independentes?*

Ex. 24 *Um dos problemas analisados no século XVII por Pascal e Fermat e que deu origem à Teoria da Probabilidade é o chamado "Problema do Cavalheiro de Méré". Num dos jogos em questão, jogavam-se 4 dados e apostava-se que ao menos um 1 ocorreria. O cavalheiro (Antoine Gombaud, que propôs o problema a Pascal) argumentava que a probabilidade disto ocorrer seria $\frac{1}{6}$ para cada dado, somando um total de $\frac{4}{6} = \frac{2}{3}$ nos 4 dados. O que está errado com este argumento? Qual é a probabilidade correta?*

Ex. 25 *A outra parte do "Problema do Cavalheiro de Méré" era calcular a probabilidade de conseguir um duplo 6 em 24 lançamentos de um par de dados. Novamente, o "cavalheiro" propunha que a probabilidade era de $\frac{1}{36}$ para um lançamento, então deveria ser $\frac{24}{36} = \frac{2}{3}$ para 24 lançamentos. Corrija este argumento.*

Ex. 26 *Quantas vezes, no mínimo, se deve lançar um dado para que a probabilidade de obter algum seis seja superior a 90%?*

Ex. 27 (Bertrand´s Box) *Você tem à sua frente três caixas; uma delas tem duas bolas brancas, uma outra tem duas bolas pretas e a terceira tem uma bola de cada cor. Você escolhe uma caixa ao acaso e, dela, retira uma bola ao acaso, verificando que ela é branca. Qual a chance de ela ter vindo da caixa com duas bolas brancas?*

Ex. 28 (Monty Hall) *a) Em um programa da televisão, o candidato devem escolher uma dentre três portas. Atrás de uma dessas portas há um prêmio e atrás de cada uma das outras duas portas há um bode. Escolhida uma porta pelo candidato, o apresentador abre uma das outras portas (nota: o apresentador nunca abre a porta do candidato e nunca abre a porta com o prêmio), e pergunta ao candidato se ele quer ficar com a porta que escolheu ou se prefere trocá-la pela outra porta que ainda está fechada. Você acha que o candidato deve trocar, não deve trocar ou que tanto faz?*
b)Agora suponha que os prêmios são um carro, um bode ou uma caixa de sabão em pó ESPUMOSO. O candidato escolhe uma porta ao acaso. O apresentador nunca abre a porta do carro nem a do candidato; no entanto, se as regras acima ainda permitirem, ele abre a do sabão em pó ESPUMOSO, que limpa mais branco, faz mais bolhinhas, e você lava lava lava esfrega esfrega esfrega e hmmmm! que cheirinho de limão!
i) O candidato escolhe uma porta e o apresentador abre a porta do bode. Qual a chance do carro estar na outra porta?
ii) E se a porta aberta pelo apresentador tiver o sabão ESPUMOSO?

Ex. 29 *Um juiz de futebol meio trapalhão tem no bolso um cartão amarelo, um cartão vermelho e um cartão com uma face amarela e uma face vermelha. Depois de uma jogada violenta, o juiz mostra um cartão, retirado do bolso ao acaso, para um atleta. Se a face que o jogador vê é amarela, qual é a probabilidade da face voltada para o juiz ser vermelha?*

Ex. 30 *O Departamento de Justiça dos Estados Unidos reportou que o número de adultos no Estados Unidos sob algum tipo de supervisão judiciária (prisões, casas de detenção ou prisão condicional) chegava a* 6.5 *milhões em* 2000*, dos quais* 3.4 *milhões eram brancos e* 2.15 *milhões eram negros. Outro relatório um pouco anterior dizia que* 9% *da população negra adulta dos Estados Unidos estavam sob algum tipo de supervisão judiciária, comparados com* 2% *da população adulta branca e* 1.3% *das outras raças. Usando estas informações, calcule:*
a) x *tal que "1 em cada* x *adultos nos Estados Unidos esteja sob supervisão judiciária".*
b) A probabilidade de um negro adulto estar sob supervisão. E um branco?
c) A probabilidade de um adulto sob supervisão ser negro. E ser branco?
d) A probabilidade de um adulto que não esteja sob supervisão *ser negro. E branco?*
e) Os eventos "ser branco" e "estar sob supervisão" são independentes?

Ex. 31 *No estudo de caso acima, suponha que um paciente que já testou + para AIDS faz um segundo teste independente do primeiro mas com os mesmos valores de sensitividade e especificidade. Se ele testa + de novo, qual a chance de ter AIDS? [Dica: é como se este indivíduo pertencesse a um grupo onde a prevalência da doença fosse* 6.573%*]*

Ex. 32 *Três eventos* A, B *e* C *são ditos independentes quando são independentes dois a dois e, além disto,* $\Pr(A \cap B \cap C) = \Pr(A)\Pr(B)\Pr(C)$*. Jogue um dado duas vezes. Sejam* $A = \{\text{primeiro número é par}\}$, $B = \{\text{segundo número é par}\}$ *e* $C = \{\text{a soma dos números é par}\}$*. Pergunta-se:*
a) A *e* B *são independentes?*
b) A *e* C *são independentes?*
c) B *e* C *são independentes?*
d) A, B *e* C *são independentes?*

Ex. 33 *Mostre que*

$$\Pr(A|B) = \Pr(A) \Rightarrow \Pr\left(A|\bar{B}\right) = \Pr(A)$$

isto é, se A *e* B *são independentes, então* A *e* $\bar{B}$ *também são independentes.*

Ex. 34 (*) *Se* A *joga uma moeda honesta* $n+1$ *vezes e* B *joga* n *vezes, determine a probabilidade de* A *obter mais caras do que* B*.*

Ex. 35 *Sejam* A *e* B *eventos não-impossíveis. Vimos que* A *é* ***independente*** *de* B *quando* $\Pr(B|A) = \Pr(B)$*. Dizemos que* A ***atrai*** B *(denotado* $A \uparrow B$*) quando* $\Pr(B|A) > \Pr(B)$ *e que* A ***repele*** B *(denotado* $A \downarrow B$*) quando* $\Pr(B|A) < \Pr(B)$*.*
a) Mostre que, se $0 < \Pr(A) < 1$*, então* $A \uparrow A$*.*
b) Mostre que

$$\begin{aligned} (A \uparrow B) &\quad\Leftrightarrow\quad (B \uparrow A) \\ (A \downarrow B) &\quad\Leftrightarrow\quad (B \downarrow A) \end{aligned}$$

ou seja, podemos dizer que A *e* B ***se atraem*** *(ou* ***repelem****)* ***mutuamente.***
c) Intuitivamente, quais pares de eventos abaixo se atraem e quais se repelem?
i) Time A ser campeão e time B (diferente de A) ser campeão.
ii) Time A ser rebaixado e time B (diferente de A) ser rebaixado.
iii) Irmão ter gripe e irmã, na mesma casa, ter gripe.
iv) Irmão ter olhos azuis e irmã ter olhos azuis.
v) Muitos sorvetes serem vendidos num dia e haver muitos afogamentos no mesmo dia.
d) Mostre que as seguintes afirmações ***não são necessariamente verdadeiras****:*

$$\begin{aligned} (A \uparrow B) \text{ e } (B \uparrow C) &\quad\Rightarrow\quad (A \uparrow C) \\ (A \uparrow B) \text{ e } (B \uparrow C) &\quad\Rightarrow\quad ((A \cap C) \uparrow B) \\ (A \downarrow B) \text{ e } (B \downarrow C) &\quad\Rightarrow\quad (A \downarrow C) \\ (A \downarrow B) \text{ e } (B \downarrow C) &\quad\Rightarrow\quad ((A \cap C) \downarrow B) \end{aligned}$$

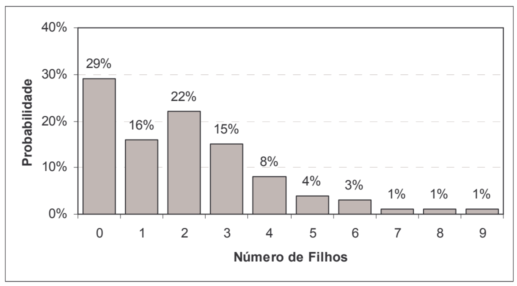
(Intuição do último item: eu não gosto da minha irmã nem no namorado dela. Eu raramente saio com ela, e quase nunca com o namorado dela – é mais provável eu sair sozinho! Mas, nas raríssimas ocasiões em que os dois saem, meus pais me forçam a ir com eles, então ***certamente*** *eu vou. Você consegue formalizar esta idéia com probabilidades?)*

1.4 Exercícios de Provas

Ex. 36 (A1 2004.2) *Um dado honesto tem duas de suas faces pintadas de vermelho e as demais de azul. O dado é lançado três vezes, anotando-se a cor da face obtida.*
a) Qual é a probabilidade de que a cor obtida no 1o. lançamento seja igual à obtida no 3o?
b) Dado que a mesma cor foi obtida no 1o e 2o lançamentos, qual é a probabilidade de que no 3o lançamento saia esta mesma cor?

Ex. 37 (A1 2004.2) *A figura abaixo mostra a probabilidade de que uma pessoa, escolhida ao acaso em um grupo de mulheres com idades de 25 a 35 anos, tenha um certo número de filhos, ou seja,*

$$Probabilidade = \frac{Número\ de\ mulheres\ com\ um\ certo\ número\ de\ filhos}{Número\ total\ de\ mulheres\ na\ amostra}$$



a) Se uma mulher é escolhida ao acaso neste grupo, é mais provável que ela tenha quantos filhos? Qual é a probabilidade correspondente?
b) Se uma mãe é escolhida ao acaso neste grupo, é mais provável que ela tenha quantos filhos? Qual é a probabilidade correspondente?
c) Suponhamos que, dentre todos os filhos das mulheres da amostra, um seja escolhido ao acaso. Qual é a probabilidade de que ele seja filho único?

Ex. 38 (AS 2004.2) *Os alunos de um certo período de uma faculdade fazem 5 matérias. As provas finais serão marcadas para uma única semana (de segunda a sábado). Admitindo que cada professor escolha ao acaso e independentemente dos demais a data de sua prova, qual é a probabilidade:*
a) de que não haja provas no sábado?
b) de que os alunos não façam mais de uma prova por dia?
c) de que haja algum dia com 4 ou mais provas?

Ex. 39 (AS 2005.2) *A probabilidade do tenista Berrando Gemigemi vencer um set contra Maria Xaropova é de* 70%. *Eles disputam uma partida de* 3 *sets (suponha que os sets são independentes uns dos outros).*
a) Qual a probabilidade do Berrando vencer a partida?
b) Qual a probabilidade de a partida terminar em 2 *sets?*
c) Qual a probabilidade de Berrando vencer em 2 *sets?*
d) Qual a probabilidade de uma partida vencida pelo Berrando ter terminado em 2 *sets?*

Ex. 40 (T1 2006.1) *Segundo uma pesquisa de opinião do IBOPE*[13] *de 2003, 2% dos brasileiros torciam para o Botafogo, 15% torciam para o Flamengo, 2% para o Fluminense e 5% para o Vasco. Restringindo a população a apenas brasileiros que tivessem grau superior (que eram apenas 8% do total), as porcentagens mudavam para 4%, 10%, 2% e 8%, respectivamente. Escolha um brasileiro da amostra total do IBOPE ao acaso.*
a) Qual a chance de ele não torcer para nenhum dos quatro grandes clubes cariocas?
b) Qual a chance de ele ser um flamenguista com grau superior?
c) Que porcentagem dos torcedores do Fluminense tem grau superior? E do Botafogo?
d) Se um brasileiro não tem grau superior, qual a chance de ele ser flamenguista?
e) De acordo com estes dados, os eventos "torcer para o Fluminense" e "ter grau superior" são independentes? Mutuamente excludentes?

[13] http://www.ibope.com.br/opp/pesquisa/opiniaopublica/download/imprensa_torcidas_1_mencao.pdf

Chapter 2

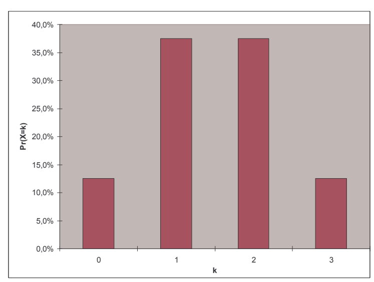
Variáveis Aleatórias Discretas

2.1 Função de Probabilidade e Função de Probabilidade Conjunta

Muitas vezes estamos interessados não no resultado de uma experiência aleatória e sim em certa função numérica do resultado. Por exemplo, jogamos um par de dados e estamos interessados na soma dos resultados; ou jogamos uma moeda até obtermos uma "cara" e estamos interessados no número de lançamentos que tivemos que fazer. Essas funções são chamadas de **variáveis aleatórias**. Mais precisamente, variáveis aleatórias são funções reais definidas no espaço amostral.

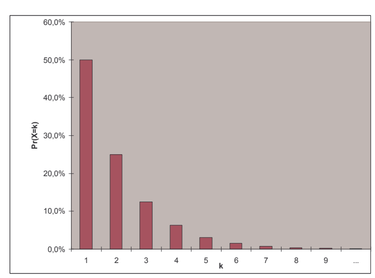
Exemplo 1 *Jogamos uma moeda honesta três vezes. Seja K o número de caras obtidas nesses três lançamentos*[1]*.*
K é uma variável aleatória cujo conjunto de valores é $\{0,1,2,3\}$. Usando K para cara e C para coroa:
K assume o valor 0 quando o resultado é CCC; portanto, $\Pr(K=0)=\frac{1}{8}$*;*
K assume o valor 1 quando o resultado é KCC ou CKC ou CCK; portanto, $\Pr(K=1)=\frac{3}{8}$*;*
K assume o valor 2 quando o resultado é KKC ou KCK ou CKK; portanto, $\Pr(K=2)=\frac{3}{8}$*;*
K assume o valor 3 quando o resultado é KKK; portanto, $\Pr(K=3)=\frac{1}{8}$*.*
As probabilidades associadas a cada valor de K estão na tabela e no gráfico abaixo:

k	0	1	2	3
$\Pr(K=k)$	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$



Exemplo 2 *Jogamos uma moeda honesta até a obtenção da primeira cara. Seja G o número de lançamentos efetuados. G é uma variável aleatória cujo conjunto de valores é o conjunto dos inteiros positivos $\{1,2,3,...\}$. Para $k=1,2,3,...$, vê-se que* $\Pr(G=k)$ *é a probabilidade de os $k-1$ primeiros lançamentos resultarem em coroa e o k-ésimo, em cara. Logo,* $\Pr(G=k)=\left(\frac{1}{2}\right)^{k-1}\left(\frac{1}{2}\right)=\frac{1}{2^k}$*:*

k	1	2	3	4	...	k	...
$\Pr(G=k)$	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{16}$	...	$\frac{1}{2^k}$	...



[1] Usaremos sempre letras maiúsculas para variáveis aleatórias.

Exemplo 3 *Jogamos dois dados honestos. Seja D_1 o resultado do primeiro dado e D_2 o resultado do segundo dado. É de se esperar que todos os* 36 *resultados deste experimento sejam igualmente prováveis, levando à seguinte tabela de probabilidades*

$D_1 \backslash D_2$	1	2	3	4	5	6	Total
1	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{6}$
2	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{6}$
3	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{6}$
4	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{6}$
5	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{6}$
6	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{6}$
Total	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	1

Definição 4 *Se X é uma variável aleatória discreta, definimos a* ***função de probabilidade de*** *X por*

$$\boxed{p_X(x) = \Pr(X = x)}$$

É claro que $p_X(x) \geq 0$ para todo x e $\sum p_X(x) = 1$.

Definição 5 *Se X e Y são variáveis aleatória discretas, definimos a* ***função de probabilidade conjunta de*** *X e Y por*

$$\boxed{p_{X,Y}(x, y) = \Pr(X = x \text{ e } Y = y)}$$

É claro que $p_{X,Y}(x, y) \geq 0$ para todo x, y e $\sum p_{X,Y}(x, y) = 1$.

Exemplo 6 *No exemplo da moeda acima, a* função de probabilidade *é a da tabela, isto é,*

$$\begin{aligned} p_K(0) &= \frac{1}{8}; \, p_K(1) = \frac{3}{8}; \, p_K(2) = \frac{3}{8}; \, p_K(3) = \frac{1}{8}; \\ p_K(k) &= 0 \text{ para } k \notin \{0, 1, 2, 3\} \end{aligned}$$

No exemplo dos dois dados, a função de probabilidade conjunta *é a que está na própria tabela. Por exemplo*

$$p_{D_1, D_2}(3, 4) = \Pr(D_1 = 3; D_2 = 4) = \frac{1}{36}$$

2.1.1 Distribuição Marginal e Condicional

Se tivermos uma função de probabilidade conjunta para X e Y mas quisermos a função de probabilidade de apenas uma das variáveis, basta somarmos os valores correspondentes nas linhas ou colunas, obtendo a **distribuição marginal** daquela variável.

Caso o valor de X seja conhecido (digamos $X = x$), a distribuição de Y dado aquele particular valor de X é chamada de **distribuição condicional** de Y dado $X = x$.

Exemplo 7 *Suponha que as variáveis X e Y têm a seguinte função de probabilidade conjunta:*

$Y \backslash X$	1	2	3
0	0.4	0.1	0.1
1	0	0.2	0.2

por exemplo, $\Pr(X = 1; Y = 1) = 0$ *(X e Y nunca são* 1 *ao mesmo tempo), enquanto* $\Pr(X = 1; Y = 0) = 40\%$. *Fazendo o total dentro de cada coluna encontramos a* distribuição marginal *de X:*

x	1	2	3
$\Pr(X = x)$	0.4	0.3	0.3

Por outro lado, dado um valor específico de Y, podemos encontrar uma distribuição condicional *de X para aquele valor de Y. No exemplo acima, a distribuição condicional de X dado que $Y = 0$ é simplesmente*

x	1	2	3
$\Pr(X = x \mid Y = 0)$	$\frac{4}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

obtidas pela definição de probabilidade condicional

$$\Pr(X = x \mid Y = 0) = \frac{\Pr(X = x \text{ e } Y = 0)}{\Pr(Y = 0)}$$

isto é, a distribuição condicional de X dado $Y = 0$ é obtida dividindo termos da distribuição conjunta (na linha $Y = 0$) pelo termo correspondente da distribuição marginal de Y ($\Pr(Y = 0) = 0.6$).

Exemplo 8 *Em uma urna há 4 bolas, numeradas 1, 2, 3 e 3. Sacam-se, sem reposição, duas bolas dessa urna. Sejam B_1 e B_2 os números da primeira e segunda bolas sacadas, respectivamente. Os valores da função de probabilidade conjunta de B_1 e B_2 encontram-se na tabela a seguir:*

$B_2 \backslash B_1$	1	2	3	*Total*
1	0	$\frac{1}{12}$	$\frac{2}{12}$	$\frac{1}{4}$
2	$\frac{1}{12}$	0	$\frac{2}{12}$	$\frac{1}{4}$
3	$\frac{2}{12}$	$\frac{2}{12}$	$\frac{2}{12}$	$\frac{2}{4}$
Total	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{2}{4}$	1

Por exemplo, $\Pr(B_1 = 1; B_2 = 2) = \Pr(B_1 = 1) . \Pr(B_2 = 2 \mid B_1 = 1) = \frac{1}{4}\frac{1}{3} = \frac{1}{12}$. Sugerimos que o leitor confira os demais valores e observe que a soma de todos os valores (excluídos os totais) é 1. As probabilidades às margens da tabela, que são probabilidades de uma só das variáveis, são as probabilidades marginais.

2.1.2 Independência de Variáveis Aleatórias Discretas

Definição 9 *Dizemos que X e Y são variáveis* ***independentes*** *exatamente quando*

$$\boxed{\Pr(X = i \text{ e } Y = j) = \Pr(X = i) . \Pr(Y = j)}$$

para quaisquer i e j.

Proposição 10 *X e Y são independentes se, e somente se, a distribuição condicional de X dado $Y = j$ é idêntica à distribuição marginal de X (qualquer que seja y possível). De fato, temos:*

$$\Pr(X = i \mid Y = j) = \frac{\Pr(X = i \text{ e } Y = j)}{\Pr(Y = j)} = \Pr(X = i) .$$

para todo i e j.

Em suma, X e Y são independentes quando o conhecimento sobre o valor de X não afeta a distribuição de probabilidades de Y. Neste capítulo, até aqui, há apenas um exemplo de distribuição conjunta com duas variáveis independentes – o exemplo dos dois dados justos cujos resultados são D_1 e D_2. Note como, naquele exemplo, cada probabilidade da distribuição conjunta $\left(\frac{1}{36}\right)$ é o produto das correspondentes probabilidades marginais $\left(\frac{1}{6} \text{ para cada variável}\right)$.

Exemplo 11 *Suponha que as variáveis X e Y têm a seguinte distribuição de probabilidade conjunta*

$Y \backslash X$	0	1	2	*Marginal de Y*
7	0.32	0.24	0.24	0.8
15	0.08	0.06	0.06	0.2
Marginal de X	0.4	0.3	0.3	1

Note que X e Y são independentes. Note também que todas as condicionais de X dado $Y = y$ são iguais à marginal de X (e todas as condicionais de Y são iguais à marginal de Y).

2.1.3 Funções de Variáveis Aleatórias Discretas

Às vezes temos a distribuição de uma variável aleatória discreta X e gostaríamos de encontrar a distribuição de uma nova variável $Z = f(X)$. Também é comum ter a distribuição conjunta das variáveis discretas X e Y e procurar a distribuição de uma terceira variável $Z = f(X, Y)$ (ou até a distribuição conjunta de $Z = f(X, Y)$ e $W = g(X, Y)$). Os seguintes exemplos mostram como "coletar" as probabilidades de X (ou X e Y) para obter a distribuição de Z (ou Z e W).

Exemplo 12 *Como no exemplo acima, seja K o número de caras obtidas em três lançamentos independentes de uma moeda justa. Suponha que você ganha \$10 para cada cara mas perde \$6 para cada coroa lançada. Então sua receita é*

$$R_1 = 10.K - 6.(3 - K) = 16K - 18$$

Qual a função de probabilidade de R_1? Ora, basta fazer a conversão para cada valor de K:

k	0	1	2	3
r	-18	-2	14	30
$\Pr(K = k) = \Pr(R_1 = r)$	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$

Isto é, $\Pr(R_1 = -18) = \Pr(R_1 = 30) = \frac{1}{8}$ *e* $\Pr(R_1 = -2) = \Pr(R_1 = 14) = \frac{3}{8}$. *Para qualquer valor de* $r \notin \{-18, -2, 14, 30\}$ *temos* $\Pr(R_1 = r) = 0$.

Exemplo 13 *Por outro lado, se tivéssemos $R_2 = \left(K - \frac{3}{2}\right)^2$, então:*

k	0	1	2	3
r	2.25	0.25	0.25	2.25
$\Pr(K = k)$	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$

Mas esta tabela ainda não é a função de probabilidade de R_2 – devemos primeiro juntar as probabilidades para valores iguais de r. A função de probabilidade correta é

r	0.25	2.25
$\Pr(R_2 = r)$	$\frac{6}{8}$	$\frac{2}{8}$

Exemplo 14 *Como no exemplo acima, sejam D_1 e D_2 os números rolados por dois dados justos e independentes. Qual a distribuição de probabilidade da soma $S = D_1 + D_2$? Bom, a distribuição conjunta de D_1 e D_2 era*

$D_1 \backslash D_2$	*1*	*2*	*3*	*4*	*5*	*6*	*Total*
1	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{6}$
2	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{6}$
3	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{6}$
4	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{6}$
5	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{6}$
6	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{6}$
Total	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	1

Anote o valor de S em cada uma das 36 células acima. Note que

$$\begin{aligned}
S &= 2 \Leftrightarrow (D_1, D_2) = (1,1) \\
S &= 3 \Leftrightarrow (D_1, D_2) \in \{(1,2), (2,1)\} \\
S &= 4 \Leftrightarrow (D_1, D_2) \in \{(1,3), (2,2), (3,1)\} \\
&\ldots \\
S &= 12 \Leftrightarrow (D_1, D_2) = (6,6)
\end{aligned}$$

Para cada valor fixo s de S, localize as células onde $D_1 + D_2 = s$ e some as suas probabilidades

$$\Pr(S = s) = \sum_{d_1 + d_2 = s} \Pr(D_1 = d_1; D_2 = d_2)$$

Assim, encontramos a função de probabilidade de S:

s	2	3	4	5	6	7	8	9	10	11	12
$\Pr(S=s)$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

Exemplo 15 *Ainda no exemplo dos dados D_1 e D_2, seja D o módulo da diferença entre eles (isto é, $D = |D_1 - D_2|$). Qual é a distribuição* conjunta *de D e D_2? Para cada célula na tabela D_1D_2, considere o valor de D_2 e o valor de D. Juntando as células com pares de valores iguais, chegamos a*

$D \backslash D_2$	1	2	3	4	5	6	f_D
0	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{6}{36}$
1	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{2}{36}$	$\frac{2}{36}$	$\frac{2}{36}$	$\frac{1}{36}$	$\frac{10}{36}$
2	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{2}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{8}{36}$
3	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{6}{36}$
4	$\frac{1}{36}$	$\frac{1}{36}$	0	0	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{4}{36}$
5	$\frac{1}{36}$	0	0	0	0	$\frac{1}{36}$	$\frac{2}{36}$
f_{D_2}	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	1

Por exemplo, $\Pr(D=1; D_2=2) = \frac{2}{36}$ *pois esta célula corresponde a duas células da tabela anterior: $(D_1, D_2) = (1,2)$ ou $(D_1, D_2) = (3,2)$. Note que a marginal de D_2 continua sendo equiprovável entre os valores de 1 a 6, mas a distribuição marginal de D favorece a diferença 1 mais do que as outras, sendo 5 a diferença menos provável. Note como a* distribuição condicional *de D_2 varia bastante de acordo com o valor dado de D. Por exemplo, dado que $D = 1$ teríamos*

k	1	2	3	4	5	6
$\Pr(D_2 = k \mid D = 1)$	10%	20%	20%	20%	20%	10%

favorecendo os números "do meio". Porém, dado que $D = 5$, apenas os números extremos são válidos para D_2:

k	1	2	3	4	5	6
$\Pr(D_2 = k \mid D = 5)$	50%	0%	0%	0%	0%	50%

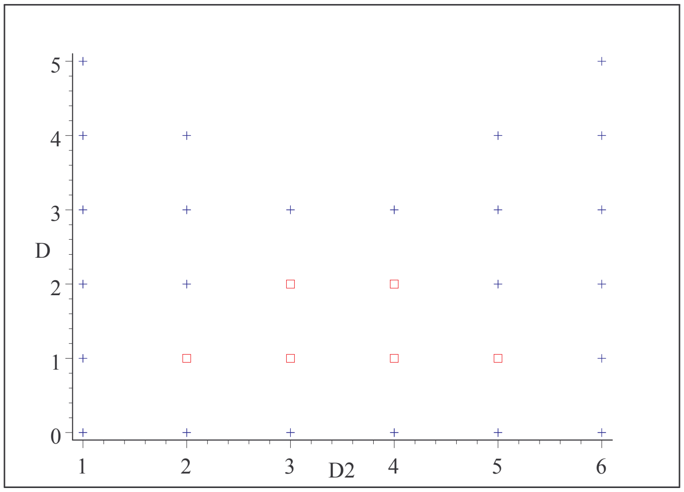
Exemplo 16 *Deixamos ao leitor verificar que a distribuição conjunta de D_1 e S é dada pela tabela*

$D_1 \backslash S$	2	3	4	5	6	7	8	9	10	11	12	f_{D_1}
1	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$						$\frac{1}{6}$
2		$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$					$\frac{1}{6}$
3			$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$				$\frac{1}{6}$
4				$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$			$\frac{1}{6}$
5					$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$		$\frac{1}{6}$
6						$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{6}$
f_S	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$	1

onde células vazias correspondem à probabilidade 0.

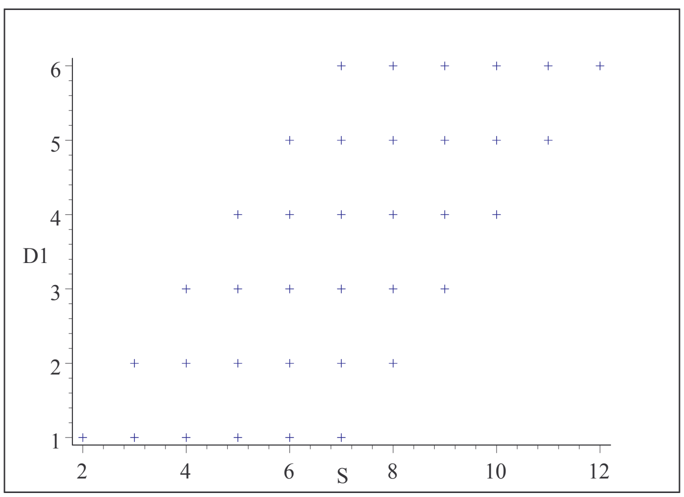
2.1.4 Diagrama de Dispersão

Às vezes, vale a pena marcar os possíves eventos de uma distribuição conjunta no plano para melhor visualizá-la. Por exemplo, a distribuição conjunta de D_2 e $D = |D_1 - D_2|$ seria representada por:



onde as cruzes azuis correspondem a pontos com probabilidade $\frac{1}{36}$ e os quadrados vermelhas a pontos com probabilidade $\frac{2}{36}$ (todos os outros pontos têm probabilidade 0). Este tipo de gráfico é chamado **diagrama de dispersão**.

Exemplo 17 *O diagrama de dispersão de D_1 e $S = D_1 + D_2$ é (cada cruz é uma probabilidade de $\frac{1}{36}$)*



2.1.5 Exercícios

Ex. 1 *Uma urna contém 4 bolas brancas e 6 bolas pretas. Retiram-se, sem reposição, duas bolas dessa urna. Seja X_1 o número de bolas brancas sacadas. Determine a função de probabilidade de X_1.*

Ex. 2 *Uma urna contém 4 bolas brancas e 6 bolas pretas. Retiram-se, sem reposição, três bolas dessa urna. Seja X_2 o número de bolas brancas sacadas. Determine a função de probabilidade de X_2.*

Ex. 3 *Sabe-se que uma moeda viciada mostra a face cara quatro vezes mais do que a face coroa. Lança-se esta moeda quatro vezes. Seja X_3 o número de caras que aparecem nestes quatro lançamentos. Determine a função de probabilidade de X_3 e calcule $\Pr(1 \leq X_3 < 3)$.*

Ex. 4 *Seja X_4 o número de sets jogados em uma partida de tênis (melhor de três sets) entre dois jogadores de igual habilidade. Encontre a função de probabilidade de X_4.*

Ex. 5 *Considere a distribuição conjunta dada pela tabela a seguir:*

$Y_5 \backslash X_5$	1	2	3
0	0.1	0.2	0.3
1	0.3	0	0.1

a) Encontre as distribuições marginais de X_5 e Y_5.
b) Encontre a distribuição condicional de X_5 dado que $Y_5 = 1$.
c) Calcule $\Pr(Y_5 = 0 \mid X_5 \geq 2)$.
d) As variáveis X_5 e Y_5 são independentes?
e) Encontre a distribuição de $Z_5 = 2X_5 - Y_5$.

Ex. 6 *Lançam-se simultaneamente uma moeda e um dado. Seja X_6 o número de caras obtidas e Y_6 o número obtido no dado.*
a) Encontre a distribuição conjunta de X_6 e Y_6.
b) Encontre a distribuição condicional de X_6 dado que $Y_6 = 1$.
c) Calcule $\Pr(X_6 = 0 \mid Y_6 \leq 4)$.
d) As variáveis X_6 e Y_6 são independentes?
e) Encontre a distribuição de $Z_6 = X_6 Y_6$.

Ex. 7 *Cada pessoa de um casal escolhe independentemente um número do conjunto* $\{-2, -1, 0, 1, 2\}$ *(todos os números têm a mesma chance de serem escolhidos). Seja X_7 o menor dos dois números e Y_7 o maior (se forem iguais, então $X_7 = Y_7$).*
a) Mostre que a distribuição conjunta de X_7 e Y_7 é dada pela tabela

$Y_7 \downarrow$ $X_7 \rightarrow$	-2	-1	0	1	2
-2	0.04	0	0	0	0
-1	0.08	0.04	0	0	0
0	0.08	0.08	0.04	0	0
1	0.08	0.08	0.08	0.04	0
2	0.08	0.08	0.08	0.08	0.04

b) Encontre a distribuição marginal de X_7.
c) Encontre a distribuição condicional de X_7 dado que $Y_7 = 0$.
d) Calcule $\Pr\left(X_7 \leq 0 \mid (Y_7)^2 = 1\right)$
e) Seja $Z_7 = Y_7 - X_7$. Encontre a distribuição de Z_7.

Ex. 8 *Um dado honesto é lançado duas vezes e D_1 e D_2 são os respectivos resultados.*
a) Encontre a função de probabilidade de $M = \max(D_1, D_2)$;
b) Encontre a função de probabilidade de $m = \min(D_1, D_2)$;
c) Encontre a distribuição conjunta de m e M e esboce seu diagrama de dispersão;

Ex. 9 *Seja X uma variável aleatória qualquer. Explique porque vale*

$$\Pr(a < X \leq b) = \Pr(X \leq b) - \Pr(X \leq a)$$

Ex. 10 *Sejam X e Y variáveis aleatórias quaisquer, e sejam* $M = \max(X, Y)$ *e* $m = \min(X, Y)$. *Mostre que*

$$\Pr(M = a) + \Pr(m = a) = \Pr(X = a) + \Pr(Y = a)$$

Verifique esta propriedade no problema anterior.

Ex. 11 *Seja X o número de caras obtidas em n lançamentos de uma moeda justa. Obtenha a função de probabilidade de X para $n = 1, 2, 3, 4, 5$. Você consegue determinar algum padrão nas funções obtidas?*
a) E se a moeda for viciada, com probabilidade p de dar cara a cada lançamento? [Esta distribuição será chamada de ***binomial com parâmetros*** *n e p.]*

Ex. 12 *Uma moeda viciada tem probabilidade p de dar cara a cada lançamento. Seja G o número de lançamentos a serem feitos até obtermos a primeira cara. Encontre a função de probabilidade de G. [Esta distribuição será chamada de* ***geométrica com parâmetro*** *p.]*

2.2 Função de Probabilidade Acumulada

2.2.1 Definição

Definição 18 *A **função de distribuição** (ou **função de probabilidade acumulada**) de X é definida por*

$$\boxed{F_X(x) = \Pr(X \leq x)}$$

para todo x real[2].

Exemplo 19 *No exemplo da moeda acima, a* função de distribuição *é obtida "acumulando" $p(x)$:*

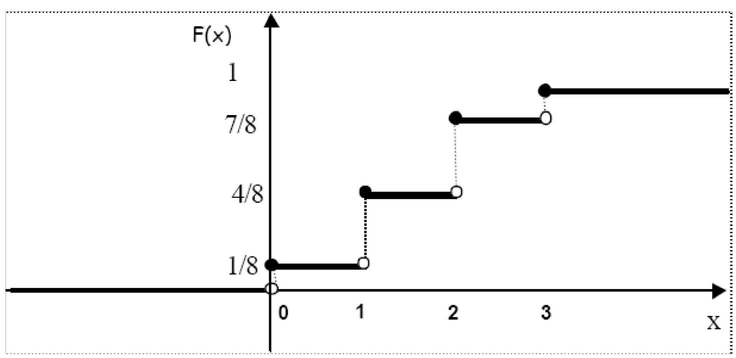
$$\begin{aligned} F(0) &= \Pr(X \leq 0) = p(0) = \frac{1}{8} \\ F(1) &= \Pr(X \leq 1) = p(0) + p(1) = \frac{4}{8} \\ F(2) &= \Pr(X \leq 2) = p(0) + p(1) + p(2) = \frac{7}{8} \\ F(3) &= \Pr(X \leq 3) = p(0) + p(1) + p(2) + p(3) \end{aligned}$$

Para ser exato, a função de distribuição pode ser calculada para x real! Mais exatamente:

$$\begin{aligned} \textit{Para } x &\in (-\infty, 0), \ F(x) = \Pr(X \leq x) = 0 \\ \textit{Para } x &\in [0, 1), \ F(x) = \Pr(X \leq x) = \Pr(X = 0) = \frac{1}{8} \\ \textit{Para } x &\in [1, 2), \ F(x) = \Pr(X \leq x) = \Pr(X \leq 1) = \frac{4}{8} \\ \textit{Para } x &\in [2, 3), \ F(x) = \Pr(X \leq x) = \Pr(X \leq 2) = \frac{7}{8} \\ \textit{Para } x &\in [3, \infty), \ F(x) = \Pr(X \leq x) = \Pr(X \leq 3) = 1 \end{aligned}$$

ou seja, resumindo tudo, a função e seu gráfico são

$$F(x) = \begin{cases} 0, \ \textit{para } x \in (-\infty, 0) \\ \frac{1}{8}, \ \textit{para } x \in [0, 1) \\ \frac{4}{8}, \ \textit{para } x \in [1, 2) \\ \frac{7}{8}, \ \textit{para } x \in [2, 3) \\ 1, \ \textit{para } x \in [3, \infty) \end{cases}$$



Proposição 20 *Se F é a função de distribuição acumulada de uma variável aleatória discreta:*
i) F é não-decrescente;
ii) $F(-\infty) = 0$ e $F(+\infty) = 1$ (ou seja, $\lim_{x \to -\infty} F(x) = 0$ e $\lim_{x \to +\infty} F(x) = 1$);
iii) F é constante por partes (isto é, uma função-escada).

Proposição 21 *Se F é a função de distribuição acumulada de uma variável aleatória, então*

$$\Pr(a < X \leq b) = F(b) - F(a)$$

[2] Para duas variáveis X e Y, é possível definir a função de probabilidade acumulada conjunta

$$F(x, y) = \Pr(X \leq x \text{ e } Y \leq y)$$

mas esta é raramente usada e por isto está relegada a esta nota de pé de página.

Exemplo 22 *Lance uma moeda* 100 *vezes. Como descobrir a probabilidade de termos entre* 40 *e* 60 *caras? Mais tarde, aprenderemos a calcular a distribuição acumulada usando a função "BinomialDist":*

$$\begin{aligned} F(60) &= \Pr(X \leq 60) = \text{BinomialDist}(60; 100, 0.5) = 0.9824 \\ F(39) &= \Pr(X \leq 39) = \text{BinomialDist}(39; 100, 0.5) = 0.0176 \end{aligned}$$

Portanto

$$\Pr(40 \leq X \leq 60) = \Pr(39 < X \leq 60) = 0.9824 - 0.0176 = 96.48\%$$

2.2.2 Quantis

Suponha que a variável aleatória X represente a altura de um brasileiro escolhido ao acaso, ou a nota de um aluno de uma turma escolhido ao acaso. Gostaríamos de dizer frases do tipo "25% dos brasileiros têm altura menor ou igual a x" ou "60% da turma tirou x ou mais". Os valores de x a colocar nas frases acima são chamados de *quantis* da variável aleatória X (mais exatamente, o 25%-quantil e o 60%-quantil). Mais formalmente:

Definição 23 *O q-quantil de uma variável aleatória X é qualquer valor x_q onde a função acumulada "acerta" q ou "passa" por q. Formalmente*[3]*:*

$$F(x_q-) \leq q \leq F(x_q)$$

Na prática, para encontrar o q-quantil, vá ao gráfico de $F(x)$ e procure onde o valor de F "igualou ou passou" por q. O valor de x naquele salto (ou os valores de x no patamar $F(x) = q$ caso ele exista, incluindo o extremo direito) é (são) o q-quantil.

Exemplo 24 *No exemplo dos três lançamentos de moedas da seção anterior (X era o número de caras), observe no gráfico de $F(x)$ que:*
O 0.2*-quantil de X é* 1 *(em $x = 1$, $F(x)$ pula de* 0.125 *para* 0.5*, saltando o valor* 0.2*).*
O 0.6*-quantil de X é* 2 *(em $x = 2$, $F(x)$ pula de* 0.5 *para* 0.875*, saltando o valor* 0.6*).*
O 0.5*-quantil de X é* 1.5 *(ou qualquer número em $[1, 2]$).*
O 0.875*-quantil de X é* 2.5 *(ou qualquer número em $[2, 3]$).*

Na prática, os quantis mais usados são o 0.25-quantil (chamado de **primeiro quartil**), o 0.5-quantil (chamado de **mediana** ou **segundo quartil**) e o 0.75-quantil (**terceiro quartil**).

Exemplo 25 *Seja X a nota de um aluno da sua turma escolhido ao acaso. Seu professor pode anunciar que os quartis de X são, respectivamente, $x = 3.5$, $x = 6.5$ (a mediana) e $x = 8.1$. O que isto significa? Se você tirasse menos que* 3.5*, você estaria nos* 25% *piores da turma, e se você tirasse mais que* 3.5*, você estaria nos* 75% *melhores (os outros números são análogos). Aproximadamente, isto significa que* 25% *da turma tirou nota abaixo de* 3.5*, outros* 25% *ficaram entre* 3.5 *e* 6.5*, outros* 25% *entre* 6.5 *e* 8.1 *e os melhores* 25% *estão de* 8.1 *para cima*[4]*.*

2.2.3 Exercícios

Ex. 13 *Faça o gráfico das funções de distribuição acumulada das variáveis $X_1 - X_7$ dos exercícios $1 - 7$ deste capítulo, e encontre os três quartis de cada uma delas.*

Ex. 14 *Determine a fórmula para a função de distribuição acumulada da variável G do exercício* 12 *e encontre uma fórmula aproximada para seu q-quantil (em função de q).*

Ex. 15 *Dê um exemplo de variável aleatória X cujos primeiro e terceiro quartis sejam iguais.*

[3] Usamos $F(x_q-)$ para representar o seguinte limite pela esquerda: $\lim_{x \to x_q-} f(x)$.

[4] Dissemos "aproximadamente" porque, de fato, pode haver um pouco mais do que 25% da turma abaixo ou igual a 3.5, caso 25% da turma não seja um número inteiro de alunos ou caso vários tenham empatado com 3.5 (por exemplo, talvez 20% da turma tenha 3.4 ou menos, e 20% da turma tenha empatado com exatamente 3.5 – o primeiro quartil ainda seria 3.5).

2.3 Valor Esperado

2.3.1 Intuição e Definição

Freqüentemente, estamos interessados na "média a longo prazo" de uma variável aleatória.

Exemplo 26 *Suponha que sua loja vende armários a \$100 cada. Digamos que a receita diaria é uma variável aleatória R com a seguinte distribuição:*

r	0	100	200	300	400
$\Pr(R=r)$	20%	10%	40%	20%	10%

Qual seria a sua receita média diária? Num longo período de N dias, haverá aproximadamente 20% dos dias em que você não vende nada, $0.1N$ dias com receitas de \$100 cada, e assim por diante. Nestes N dias, sua receita total seria (aproximadamente):

$$\textit{Receita Total} = (0)(0.2N) + (100)(0.1N) + (200)(0.4N) + (300)(0.2N) + (400)(0.1N) = 190N$$

ou seja, uma média de uns \$190 por dia (a longo prazo). Podemos dispensar o N e encontrar diretamente:

$$\textit{Receita "Média"} = 0(0.2) + (100)(0.1) + (200)(0.4) + (300)(0.2) + (400)(0.1) = 190$$

Definição 27 *Se X é uma variável aleatória discreta, definimos o* ***valor esperado*** *(ou* ***esperança matemática****, ou* ***expectância****, ou* ***média****, ou* ***valor médio****) de X por*

$$\boxed{E(X) = \sum_{x\in S} x.p(x)}$$

isto é, $E(X)$ é uma média ponderada dos valores de X, com pesos iguais às respectivas probabilidades deste valores. Ocasionalmente, escreveremos $\mu_X = E(X)$.

A esperança é uma medida de posição ou de tendência central (valores grandes de X acarretam $E(X)$ grande; valores pequenos de X acarretam $E(X)$ pequeno).

Exemplo 28 *No exemplo inicial deste capítulo, o valor esperado de K é de 1.5 caras (em 3 lançamentos), pois:*

$$E(K) = 0.p(0) + 1.p(1) + 2.p(2) + 3.p(3) = 0 + 3/8 + 6/8 + 3/8 = 1.5$$

Exemplo 29 *Se X é o número rolado por um dado justo,*

$$E(X) = \frac{1}{6}(1+2+3+4+5+6) = 3.5$$

2.3.2 Propriedades (Caso Unidimensional)

Proposição 30 *Se $Y = f(X)$, temos*

$$\boxed{E(Y) = E(f(X)) = \sum_{x\in S} f(x).p(x)}$$

Prova. A idéia intuitiva é que a distribuição de Y terá as mesmas probabilidades da de X só trocando cada valor x de X pelo correspondente $f(x)$ de Y. Formalmente, seja $y = f(x)$ e note que

$$\Pr(Y=y) = \sum_{f(x)=y} \Pr(X=x)$$

então

$$E(Y) = \sum_{y\in S} y.\Pr(Y=y) = \sum_{y} y.\left(\sum_{f(x)=y} \Pr(X=x)\right) = \sum_{y}\sum_{f(x)=y} f(x)\Pr(X=x) = \sum_{x} f(x)\,p(x)$$

■

Exemplo 31 *No exemplo 13, poderíamos calcular* $E(R_2) = E\left((X-1.5)^2\right)$ *sem calcular primeiro a distribuição de* R_2*, usando apenas a distribuição de* X*:*

$$E(R_2) = \sum_x (x-1.5)^2 p(x) = (2.25)\frac{1}{8} + (0.25)\frac{3}{8} + (0.25)\frac{3}{8} + (2.25)\frac{1}{8} = 0.75$$

Note que o somatório já faz o trabalho de juntar $\Pr(X=0) = \Pr(X=3) = \frac{1}{8}$ *num único termo correspondente a* $\Pr(R_2 = 2.25) = \frac{2}{8}$.

A fórmula acima fica bem mais simples quando a função f é afim. De fato:

Proposição 32 *Sejam* a *e* b *constantes quaisquer. Então:*

$$\boxed{E(aX+b) = aE(X) + b}$$

Em particular:

$$\begin{aligned} E(b) &= b \\ E(aX) &= aE(X) \\ E(X-\mu_X) &= 0 \end{aligned}$$

Prova. De fato, seja $Y = aX + b$ a outra variável aleatória. Então:

$$E(Y) = \sum_x (ax+b).p(x) = a\sum_x xp(x) + b\sum_x p(x) = aE(X) + b$$

■

Exemplo 33 *No exemplo inicial desta seção, seja* A *o número de armários vendidos num dia. Como* $R = 100A$*, temos* $E(R) = 100E(A)$*, isto é,* $E(A) = 1.9$ *armários. A grosso modo, isto significa que "esperamos vender* 1.9 *armários por dia". Note também que* $E(R-190) = E(R) - 190 = 0$.

Exemplo 34 *No exemplo 12, temos*

$$E(R_1) = E(16K - 18) = 16E(K) - 18 = 16(1.5) - 18 = 6$$

ou seja, uns \$6 de lucro por jogo (na média, a longo prazo).

2.3.3 Propriedades (Caso Bidimensional)

A idéia de "deixar o somatório do valor esperado coletar as probabilidades da nova variável sem calcular explicitamente a nova distribuição" funciona também com várias variáveis.

Exemplo 35 *Sejam* D_1 *e* D_2 *os resultados sucessivos do lançamento de dois dados justos. Seja* $Z = D_1D_2$*. Os valores de* Z *e suas respectivas probabilidades em cada célula são:*

$D_1\backslash D_2$	*1*	*2*	*3*	*4*	*5*	*6*
1	$1; \frac{1}{36}$	$2; \frac{1}{36}$	$3; \frac{1}{36}$	$4; \frac{1}{36}$	$5; \frac{1}{36}$	$6; \frac{1}{36}$
2	$2; \frac{1}{36}$	$4; \frac{1}{36}$	$6; \frac{1}{36}$	$8; \frac{1}{36}$	$10; \frac{1}{36}$	$12; \frac{1}{36}$
3	$3; \frac{1}{36}$	$6; \frac{1}{36}$	$9; \frac{1}{36}$	$12; \frac{1}{36}$	$15; \frac{1}{36}$	$18; \frac{1}{36}$
4	$4; \frac{1}{36}$	$8; \frac{1}{36}$	$12; \frac{1}{36}$	$16; \frac{1}{36}$	$20; \frac{1}{36}$	$24; \frac{1}{36}$
5	$5; \frac{1}{36}$	$10; \frac{1}{36}$	$15; \frac{1}{36}$	$20; \frac{1}{36}$	$25; \frac{1}{36}$	$30; \frac{1}{36}$
6	$6; \frac{1}{36}$	$12; \frac{1}{36}$	$18; \frac{1}{36}$	$24; \frac{1}{36}$	$30; \frac{1}{36}$	$36; \frac{1}{36}$

Para calcular $E(Z)$*, poderíamos primeiro montar a função de probabilidade de* Z *para então calcular* $\sum zp(z)$*. Ao invés disso, é mais fácil simplesmente usar direto a tabela acima com todos os valores de* Z*:*

$$E(Z) = \frac{1}{36}\begin{pmatrix} 1+2+3+4+5+6+ \\ +2+4+6+8+10+12+ \\ +3+6+9+12+15+18+ \\ ... \\ +6+12+18+24+30+36 \end{pmatrix} = \frac{49}{4}$$

Formalmente, temos:

Proposição 36 *Se* $Z = f(X, Y)$ *então*

$$\boxed{E(Z) = E(f(X,Y)) = \sum_{x,y} f(x,y)\, p_{X,Y}(x,y)}$$

Prova. A idéia intuitiva é a mesma do exemplo anterior; para encontrar cada $p(z)$ da função de probabilidade de Z, precisaríamos coletar as probabilidades de todas as "células" da distribuição conjunta que dessem aquele z. Mas o somatório acima já coleta estas probabilidades nos lugares certos e soma $zp(z)$ também! Formalmente:

$$p_Z(z) = \sum_{f(x,y)=z} p_{X,Y}(x,y)$$

Então

$$\begin{aligned} E(Z) &= \sum_z zp(z) = \sum_z z\left(\sum_{f(x,y)=z} p_{X,Y}(x,y)\right) = \sum_z \sum_{f(x,y)=z} zp_{X,Y}(x,y) = \\ &= \sum_z \sum_{f(x,y)=z} f(x,y)\, p_{X,Y}(x,y) = \sum_{x,y} f(x,y)\, p_{X,Y}(x,y) \end{aligned}$$

■

Analogamente, se Z é uma função afim de X e Y, a propriedade acima fica ainda mais simples:

Proposição 37 *Sejam* a, b *e* c *constantes quaisquer. Então*

$$\boxed{E(aX + bY + c) = aE(X) + bE(Y) + c}$$

Em particular,

$$\boxed{E(X+Y) = E(X) + E(Y)}$$

Prova. Seja $Z = aX + bY + c$. Então:

$$\begin{aligned} E(Z) &= \sum_{x,y} (ax + by + c)\, p(x,y) = a\sum xp(x,y) + b\sum yp(x,y) + c\sum p(x,y) = \\ &= aE(X) + bE(Y) + c \end{aligned}$$

■

Note que as duas últimas propriedades acima são válidas **mesmo que** X **e** Y **não sejam independentes**!

Exemplo 38 *Seja* $S = D_1 + D_2$ *a soma de dois dados justos. Para calcular* $E(S)$ *basta fazer*

$$E(D_1 + D_2) = E(D_1) + E(D_2) = 3.5 + 3.5 = 7$$

Para a diferença $D_1 - D_2$ *temos*

$$E(D_1 - D_2) = E(D_1) - E(D_2) = 0$$

Já para $D = |D_1 - D_2|$ *não podemos usar a fórmula mais simples – é mais fácil voltar à distribuição de* D *(veja exemplo na seção anterior) para encontrar*

$$E(D) = \frac{6}{36}0 + \frac{10}{36}1 + \frac{8}{36}2 + \frac{6}{36}3 + \frac{4}{36}4 + \frac{2}{36}5 = \frac{35}{18}$$

Proposição 39 $\boxed{\text{Se } X \text{ e } Y \text{ são independentes, então } E(XY) = E(X)\,E(Y)}$.

Prova. De fato, se X e Y são independentes

$$\begin{aligned} E(XY) &= \sum_x \sum_y xy \Pr(X = x; Y = y) = \sum_x \sum_y xy \Pr(X = x) \Pr(Y = y) = \\ &= \left(\sum_x x \Pr(X = x)\right) \cdot \left(\sum_y y \Pr(Y = y)\right) = E(X)\,E(Y) \end{aligned}$$

■

2.4 Variância e Outras Medidas de Dispersão

2.4.1 Definição

Queremos medir a dispersão da variável aleatória X, ou seja, queremos medir se os valores de X estão concentrados próximo à sua média (dispersão pequena) ou distantes da sua média (dispersão grande). Para isso, devemos considerar os desvios (resíduos ou afastamentos) de X em relação à sua média, isto é, os resíduos $X - E(X)$. Se estiverem próximos de zero, a dispersão será pequena. Como medir se os desvios estão ou não próximos a zero? Uma má idéia que ocorre a todos, ainda que por breves momentos, é usar a média dos desvios. Essa idéia é má porque, calculando a média, desvios positivos cancelam desvios negativos e a média dos desvios é igual a 0 (pois $E(X - E(X)) = 0$), mesmo quando os desvios não ficam próximos a zero. Como "sumir" com os sinais negativos dos resíduos?

Definição 40 *Duas medidas de dispersão comuns são o* ***desvio médio****, definido por*

$$DM(X) = E(|X - E(X)|)$$

e a ***variância****, definida por*

$$\boxed{Var(X) = E[(X - E(X))^2]}$$

Ao invés da variância, podemos medir a dispersão de X pelo seu ***desvio-padrão***

$$\boxed{\sigma(X) = \sqrt{Var(X)}}$$

Exemplo 41 *No exemplo 1 deste capítulo, $E(X) = 1.5$ e portanto*

$$\begin{aligned}
Var(X) &= \sum_{x=0}^{3} (x-1.5)^2 p(x) = (-1.5)^2 \frac{1}{8} + (-0.5)^2 \frac{3}{8} + (0.5)^2 \frac{3}{8} + (1.5)^2 \frac{1}{8} = 0.75 \text{ “caras quadradas”} \\
\sigma(X) &= \sqrt{Var(X)} = \sqrt{0.75} \approx 0.866 \text{ caras} \\
DM(X) &= \sum_{x=0}^{3} |x-1.5|\ p(x) = (1.5)\frac{1}{8} + (0.5)\frac{3}{8} + (0.5)\frac{3}{8} + (1.5)\frac{1}{8} = 0.75 \text{ caras}
\end{aligned}$$

Exemplo 42 *No exemplo dos armários, tem-se $E(R) = 1900$. Assim*

$$\begin{aligned}
Var(R) &= (-190)^2(0.2) + (-90)^2(0.1) + (10)^2(0.4) + (110)^2(0.2) + (210)^2(0.1) = 14900 \text{ “reais quadrados”} \\
\sigma(X) &= \sqrt{1490} = R\$38.60 \\
DM(X) &= (190)(0.2) + (90)(0.1) + (10)(0.4) + (110)(0.2) + (210)(0.1) = R\$94.00
\end{aligned}$$

Se você preferir, trabalhe com o número de armários:

$$\begin{aligned}
Var(A) &= (-1.9)^2(0.2) + (-0.9)^2(0.1) + (0.1)^2(0.4) + (1.1)^2(0.2) + (2.1)^2(0.1) = 1.49 \text{ “armários ao quadrado”} \\
\sigma(X) &= \sqrt{1.49} = 3.86 \text{ armários} \\
DM(X) &= (1.9)(0.2) + (0.9)(0.1) + (0.1)(0.4) + (1.1)(0.2) + (2.1)(0.1) = R\$0.94
\end{aligned}$$

Observe que a unidade que mede X também mede $E(X)$, $DM(X)$ e $\sigma(X)$, mas a unidade de $Var(X)$ é o quadrado da unidade que mede X. Observe também que $Var(X)$, $\sigma(X)$ e $DM(X)$ são sempre maiores que ou iguais a 0.

Proposição 43 *Sejam a e b constantes quaisquer. Então:*

$$\boxed{Var(aX + b) = a^2 Var(X)}$$

$$\boxed{\sigma(aX + b) = |a| .\sigma(X)}$$

$$\boxed{DM(aX + b) = |a| .DM(X)}$$

Em particular

$$\begin{aligned} Var(b) &= 0 \\ Var(X+b) &= Var(X) \\ Var(aX) &= a^2 Var(X) \end{aligned}$$

Prova. Seja $Y = aX + b$. Escreva, para facilitar, $\mu_X = E(X)$ e $\mu_Y = E(Y)$. Já sabemos que $\mu_Y = a\mu_X + b$. Então:

$$\begin{aligned} Var(Y) &= E\left((Y-\mu_Y)^2\right) = E\left(((aX+b)-(a\mu_X+b))^2\right) = \\ &= E\left(a^2(X-\mu_X)^2\right) = a^2 E\left((X-\mu_X)^2\right) = a^2 Var(X) \end{aligned}$$

Conseqüentemente,

$$\sigma(Y) = \sqrt{Var(Y)} = |a| . \sqrt{Var(X)} = |a| . \sigma(X)$$

Enfim

$$\begin{aligned} DM(Y) &= E(|Y-\mu_Y|) = E(|aX+b-(a\mu_X+b)|) = \\ &= E(|a(X-\mu_X)|) = E(|a||X-\mu_X|) = |a| E(|X-\mu_X|) = |a| .MD(X) \end{aligned}$$

■

A fórmula a seguir é a maneira comumente mais simples para encontrar a variância de X.

Proposição 44

$$\boxed{Var(X) = E\left(X^2\right) - (E(X))^2}$$

Prova. De fato,

$$\begin{aligned} Var(X) &= E\left((X-\mu_X)^2\right) = \sum_{x\in S} (x-\mu_X)^2 .p(x) = \\ &= \sum_{s\in S} \left(x^2 - 2x\mu_X + \mu_X^2\right).p(x) = \\ &= \sum_{s\in S} x^2 p(x) - 2\mu_X \sum_{x\in S} xp(x) + \mu_X^2 \sum_{x\in S} p(x) = \\ &= E\left(X^2\right) - 2\mu_X\mu_X + \mu_X^2 = E\left(X^2\right) - \mu_X^2 \end{aligned}$$

■

Proposição 45 $\boxed{\textit{Se } X \textit{ e } Y \textit{ são independentes, } Var(X+Y) = Var(X) + Var(Y)}$

Prova.

$$\begin{aligned} Var(X+Y) &= E\left((X+Y)^2\right) - (E(X+Y))^2 = \\ &= E\left(X^2+2XY+Y^2\right) - (E(X)+E(Y))^2 = \\ &= E\left(X^2\right) - (E(X))^2 + E\left(Y^2\right) - (E(Y))^2 + 2(E(XY) - E(X)E(Y)) = \\ &= Var(X) + Var(Y) \end{aligned}$$

pois, como X e Y são independentes, temos $E(XY) = E(X)E(Y)$. ■

2.4.2 Desigualdade de Chebyshev

Como interpretar a variância (ou o desvio-padrão) de uma variável aleatória? Intuitivamente, quanto menor a variância, mais perto de $E(X)$ estão os valores prováveis de X. A formalização desta intuição pode ser feita pelo teorema a seguir.

Teorema 46 (Desigualdade de Chebyshev) *Seja X uma variável aleatória com valor esperado $\mu = E(X)$ e desvio-padrão $\sigma = \sigma(X)$. Seja $P = \{x \in \mathbb{R} \mid |x-\mu| < k\sigma\}$ (isto é, P é o intervalo aberto $(x - k\sigma, x + k\sigma)$, um conjunto de valores de x que estão "perto da média" pelo menos k desvios-padrão). Então, para qualquer $k > 0$, tem-se*

$$\Pr(X \notin P) \leq \frac{1}{k^2}$$

ou seja

$$\boxed{\Pr(|X-\mu| \geq k\sigma) \leq \frac{1}{k^2}}$$
$$\boxed{\Pr(|X-\mu| < k\sigma) \geq 1 - \frac{1}{k^2}}$$

Prova. A demonstração é um belíssimo truque, que se inicia na fórmula da variância

$$Var(X) = E\left((X-\mu)^2\right) = \sum_{x\in S} (x-\mu)^2 p(x)$$

Separe o somatório acima em duas partes:

$$Var(X) = \sum_{x\in P} (x-\mu)^2 p(x) + \sum_{x\notin P} (x-\mu)^2 p(x)$$

O primeiro somatório é maior ou igual a 0; no segundo somatório, vale que $(x-\mu)^2 \geq k^2\sigma^2$. Então:

$$\sigma^2 = Var(X) \geq 0 + k^2\sigma^2 \sum_{x\notin P} p(x) = k^2\sigma^2 \Pr(X \notin P) \Rightarrow \Pr(X \notin P) \leq \frac{1}{k^2}$$

■

Nota 47 *A Desigualdade de Chebyshev significa o seguinte: em* ***qualquer*** *distribuição de probabilidade:*

- Há no máximo $\frac{1}{4} = 25\%$ de chance de X estar a 2 ou mais desvios-padrão da média; assim, há pelo menos 75% de chance de X estar a menos de 2 desvios-padrão da média.
- Há no máximo $\frac{1}{9}$ de chance de X estar a 3 ou mais desvios-padrão da média.
- Há no máximo $\frac{1}{16}$ de chance de X estar a 4 ou mais desvios-padrão da média.
- Há no máximo $\frac{1}{25}$ de chance de X estar a 5 ou mais desvios-padrão da média.
- Em geral, há no máximo $\frac{1}{k^2}$ de chance de X estar a mais de k desvios-padrão da média (k não precisa ser inteiro!).

2.4.3 Exercícios

Ex. 16 *Para cada uma das variáveis $X_1 - X_7$ dos exercicios $1 - 7$ deste capítulo, encontre seu valor esperado $E(X_i)$, o número $E\left(X_i^2\right)$, o desvio médio, a variância e o desvio-padrão.*

Ex. 17 *Sabendo que $E(X) = 5$ e $Var(X) = 7$, calcule $E\left(X^2 + 2X + 5\right)$ e $Var(2X + 5)$.*

Ex. 18 *Uma variável aleatória X tem $E(X) = \mu$ e $Var(X) = \sigma^2$. A variável $Z = \frac{X-\mu}{\sigma}$ é chamada de* ***variável aleatória padronizada*** *associada a X. Calcule $E(Z)$ e $Var(Z)$.*

Ex. 19 *Uma moeda justa é lançada e observa-se se ela deu cara ou coroa. Considere os seguintes jogos baseados nesta moeda:*
A) Se deu cara, você ganha $100 *mil; caso contrário, você não ganha nada;*
B) Se deu cara, você ganha $100 *milhões; caso contrário, você paga* $10 *milhões.*
a) Sendo X *o seu prêmio, qual o valor esperado e desvio-padrão de* X *em cada caso?*
b) Você tem de escolher um dos dois jogos, mas só pode jogar uma vez. Que jogo você escolheria?
c) E se você pudesse jogar um deles várias vezes (digamos, umas 100 *vezes), qual você escolheria?*

Ex. 20 *Uma roleta em Las Vegas tem todos os números de* 1 *a* 36 *mais um* 0 *e um* 00*. Destes números,* 18 *são vermelhos,* 18 *são pretos e o* 0 *e o* 00 *são verdes. A banca roda a roleta com uma bolinha dentro, que pára em um dos* 38 *números. Suponha que a roleta é justa e os* 38 *números são eqüiprováveis.*
a) Se você aposta $1 *no vermelho e o número é vermelho, você tem um lucro de* $1*; caso contrário você perde sua aposta. Calcule o lucro esperado (e sua variância) ao apostar* $1 *no vermelho.*
b) E se você apostar $1 *no vermelho e* $1 *no preto ao mesmo tempo, quais são a distribuição, valor esperado e variância do seu lucro?*
c) Se você aposta $1 *no número* 13 *e a bolinha cai no* 13*, você tem um lucro de* $35*; caso contrário, você perde sua aposta. Calcule o lucro esperado (e a variância) de apostar* $1 *no número* 13.
d) Como mudam as respostas acima em Monte Carlo, onde há apenas 37 *números na roleta (eles não têm o* 00*) mas os pagamentos se mantêm?*

Ex. 21 *Um vendedor de carros se comunica diariamente com* 0, 1 *ou* 2 *clientes, com probabilidades* 20%, 30% *e* 50%*, respectivamente. A cada cliente, a probabilidade de ele conseguir fazer uma venda é de* 40%*. Seja* X *o número diário de vendas deste vendedor.*
a) Encontre a função de probabilidade de X.
b) Calcule $E(X)$, $E(X^2)$, $Var(X)$ *e* $\sigma(X)$.

Ex. 22 *Quando jogamos uma moeda várias vezes, uma seqüência maximal de lançamentos iguais é chamada de* ***corrida****. Assim, se a lançássemos* 11 *vezes e o resultado fosse* $CCCKKCCCCKC$*, diríamos que há nesta seqüência* 5 *"corridas" (o* CCC*, o* KK*, o* $CCCC$*, o* K *e o* C*).*
Jogue uma moeda justa 3 *vezes. Seja* X *o número de corridas nesta seqüência de três lançamentos.*
a) Encontre um espaço amostral para este experimento.
b) Encontre $E(X)$, $E(X^2)$ *e* $Var(X)$.

Ex. 23 (Aposta de Pascal) *Pascal usava o seguinte argumento a favor de acreditar em Deus: seja* p *a probabilidade de Deus existir.*
Suponha que você acredita em Deus; se Ele existir, você iria ao paraíso, e seu ganho seria a *(praticamente infinito); se Ele não existir, você deixaria de lado alguns prazeres materiais, e teria um ganho de* $-b$ *(finito).*
Por outro lado, se você não acreditar em Deus, e Ele existir, você iria para o inferno – ganho de $-c$ *onde* c *é muito grande. Se Ele não existir, o ganho é* 0.
Calcule os valores esperados do seu ganho se você acreditar e se você não acreditar em Deus (em função de a, b, c *e* p*). Por que o valor esperado de acreditar seria maior? Qual o valor mínimo de* p *para que isto aconteça?*

Ex. 24 *Você tem uma loja que vende tortas caseiras. O custo de fazer uma torta é de* $20.00 *e você as vende no mesmo dia em que foram feitas por* $50.00*. Os clientes não costumam avisar com antecedência se querem uma torta ou não, então você tem de decidir quantas tortas levar para a sua loja* ***antes de saber a demanda de cada dia****. No entanto, se uma torta não é vendida no dia em que foi feita, você tem de descartá-la. Suponha que o número* X *de tortas demandadas por dia tem a seguinte distribuição:*

x	0	1	2	3
$\Pr(X = x)$	0.2	0.3	0.3	0.2

Quantas tortas você deve fazer diariamente para maximizar o valor esperado do seu lucro?

Ex. 25 *Seja* X *uma variável aleatória qualquer e seja* $f(t) = E\left((X-t)^2\right)$*. Mostre que* $f(t)$ *é uma função quadrática cujo mínimo ocorre para* $t = E(X)$.

Ex. 26 *Sejam X e Y duas variáveis aleatórias independentes com a mesma média μ mas diferentes varIâncias σ_1^2 e σ_2^2, respectivamente. Seja $Z = \alpha X + (1-\alpha) Y$. Mostre que $E(Z) = \mu$ e encontre $Var(Z)$ em função de α. Que valor de α minimiza $Var(Z)$?*
[Este problema se aplica à seguinte situação: X e Y podem ser duas medições independentes de uma mesma quantidade μ mas feitas com instrumentos de precisões distintas – daí as variâncias distintas; quanto maior a variância, pior o instrumento. A partir de X e Y queremos fazer uma média ponderada para encontrar uma medida ainda mais precisa Z, e para tanto queremos que $Var(Z)$ seja o mínimo possível – como fazer a ponderação?]

Ex. 27 *Seja X uma variável aleatória com distribuição uniforme discreta, isto é, o conjunto de valores de X é $\{1, 2, ..., n\}$ e X assume esses valores com probabilidades iguais. Determine $E(X)$, $E(X^2)$ e $Var(X)$. [Dica: você vai precisar da igualdade*

$$1^2 + 2^2 + ... + n^2 = \frac{n(n+1)(2n+1)}{6}$$

que pode ser provada por indução.]

Ex. 28 *Dizemos que X segue a* **distribuição de Bernoulli de parâmetro** *p quando tem a seguinte distribuição*

x	0	1
$\Pr(X = x)$	$1-p$	p

Neste caso, calcule $E(X)$, $E(X^2)$ e $Var(X)$.

Ex. 29 *Num exame de múltipla escolha com 80 questões, cada questão tem 5 escolhas. O aluno recebe 4 pontos por resposta correta, mas perde 1 ponto por cada questão respondida incorretamente (se ele deixa a questão em branco, ele não ganha nem perde nada). Suponha que o aluno marca aleatoriamente as respostas das questões, sem pensar em nenhuma.*
a) Qual o valor esperado do número de pontos ganhos em cada questão? E a variância?
b) Qual o valor esperado do número de pontos no exame todo? E a variância?

Ex. 30 *Considere os seguintes dois jogos com dados justos:*
A) Você lança o dado 100 vezes e seu prêmio é a soma dos 100 dados.
B) Você lança um dado uma vez e seu prêmio é 100 vezes o número lançado.
Estes jogos são equivalentes? Calcule o seu prêmio esperado e a variância do prêmio em cada caso. Que jogo você prefere?

Ex. 31 *Uma loteria semanal tem 100 bilhetes e um prêmio fixo para o bilhete sorteado. Quem tem o maior valor esperado do prêmio total: quem compra 10 bilhetes numa semana ou quem compra 1 bilhete por semana durante 10 semanas? Que variância do prêmio total é maior?*

Ex. 32 *Sejam X_1, X_2, X_3 e X_4 os números obtidos nos quatro lançamentos sucessivos e independentes de um dado justo. Seja $S = X_1 + X_2 + X_3 + X_4$. Calcule $E(S)$ e $Var(S)$.*

Ex. 33 *Sejam X_1, X_2, ..., X_n os números obtidos em n lançamentos sucessivos e independentes de um dado justo. Seja $\bar{X} = \frac{X_1+X_2+...+X_n}{n}$. Calcule $E(\bar{X})$ e $Var(\bar{X})$.*

Ex. 34 *Sejam X_1, X_2, ..., X_n variáveis aleatórias independentes com o mesmo valor esperado μ e a mesma variância σ^2 (por exemplo, podem ser os resultados das n repetições de um mesmo experimento). Sejam $S = X_1 + X_2 + ... + X_n$ e $\bar{X} = S/n$. Calcule os valores esperados e variâncias de S e $\bar{X}$ em função de μ, σ e n.*

Ex. 35 *Seja X o número de caras obtidas ao lançarmos uma moeda justa 100 vezes (cada lançamento sendo independente de todos os anteriores). Calcule $E(X)$ e $Var(X)$. [Dica: seja $X_i = 1$ se o i-ésimo lançamento deu cara, e $X_i = 0$ caso tenha sido coroa. Note que $X = \sum X_i$.]*

Ex. 36 *A cada dia, o valor de uma ação pode subir \$3 (com probabilidade $\frac{1}{3}$) ou cair \$2 (com probabilidade $\frac{2}{3}$). Suponha que a valorização da ação em um dia é independente de todos os outrso. Você compra a ação e a vende N dias depois. Qual o valor esperado e a variância do seu lucro?*

Ex. 37 *a) Duas senhoras A e B deixam seus chapéus a e b (respectivamente) na recepção de um hotel. Mais tarde, o hotel devolve os chapéus às senhoras em ordem aleatória (isto é, as ordens (a,b) e (b,a) são igualmente prováveis). Seja X o número de senhoras que recebe o seu próprio chapéu. Calcule $E(X)$.*
b) Repita o problema anterior para três senhoras com três chapéus supondo que as 6 possíveis ordens de devolução são eqüiprováveis.

Ex. 38 *Um grupo de n amigos (numerados de 1 a n) resolve fazer um amigo oculto. Para tanto, os números de 1 a n são colocados numa caixa, e cada um deles retira, sucessivamente e sem reposição, um número da caixa sem olhar.*
a) Defina X_1, X_2,..., X_n da seguinte forma: $X_i = 1$ se o amigo i tira o seu próprio número, e $X_i = 0$ caso contrário. Mostre que $E(X_i) = \frac{1}{n}$.
b) Seja X o número total de amigos que retiraram o seu próprio número. Escreva X em função dos X_i e calcule $E(X)$ sem encontrar sua função de probabilidade.

Ex. 39 *Numa urna há 2 bolas pretas e 3 bolas brancas. Você retira bolas da urna sucessivamente, sem reposição. A cada bola preta retirada, você ganha \$1; a cada bola branca, você perde \$1.*
a) Se você retirar apenas 1 bola, qual o seu lucro esperado?
b) E se você retirar 2 bolas? E se retirar 3? 4? Todas?
c) Agora suponha que você pára de retirar as bolas assim que estiver lucrando, ou assim que retirar todas as bolas pretas, o que acontecer primeiro. Qual o seu lucro esperado?

Ex. 40 *Sejam μ, k e σ reais positivos fixos. Encontre uma distribuição de probabilidade que satisfaça a **igualdade***

$$\Pr(|X - \mu| \geq k\sigma) = \frac{1}{k^2}$$

[Dica: examine a demonstração da desigualdade de Chebyshev.]

Ex. 41 *Se $E(X) = 0$ e $Var(X) = 1$, encontre k tal que você possa garantir que $\Pr(|X| \leq k) \geq 99\%$.*

Ex. 42 *Se $E(X) = 5$ e $Var(X) = 9$, encontre a tal que você possa garantir que $\Pr(5 - a < X < 5 + a) \geq 75\%$.*

2.5 Covariância e Correlação

Definição 48 *A **covariância** entre duas variáveis X e Y é*

$$\boxed{Cov(X,Y) = E[(X - E(X))(Y - E(Y))]}$$

Proposição 49

$$\boxed{Cov(X,Y) = E(XY) - E(X)E(Y)}$$

Prova. Escreva $E(X) = \mu_X$ e $E(Y) = \mu_Y$. Então:

$$\begin{aligned} Cov(X,Y) &= E[(X - \mu_X)(Y - \mu_Y)] = E(XY - \mu_X Y - \mu_Y X + \mu_X \mu_Y) = \\ &= E(XY) - \mu_X E(Y) - \mu_Y E(X) + \mu_X \mu_Y = E(XY) - \mu_X \mu_Y \end{aligned}$$

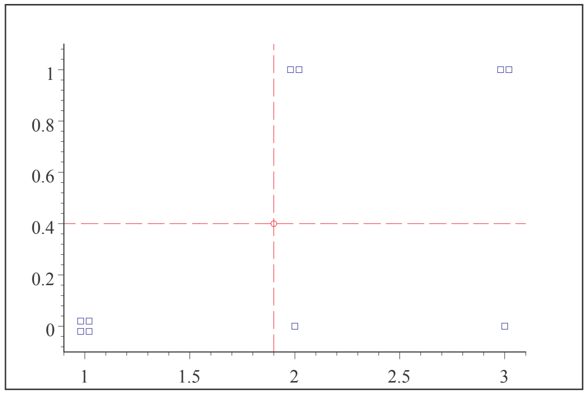
■

Intuição da Covariância: $X - E(X)$ são os resíduos de X com relação à média, isto é, $X - E(X)$ é positivo quando X está acima da média; o produto $(X - E(X))(Y - E(Y))$ será positivo sempre que ambos X e Y estiverem acima de suas médias, ou quando ambos estiverem abaixo de suas médias. Assim, $Cov(X,Y)$ será positivo se X e Y estiverem ambos acima ou abaixo de suas médias ao mesmo tempo mais frequentemente do que não. **Em suma, uma covariância positiva significa que, via de regra, X tem valor alto quando Y tem valor alto, e vice-versa.**

Exemplo 50 *Considere a seguinte distribuição conjunta*

$Y\backslash X$	1	2	3
0	0.4	0.1	0.1
1	0	0.2	0.2

cujo diagrama de dispersão é (o número de símbolos em cada ponto ilustra a probabilidade ali; as retas tracejadas são $X = E(X)$ *e* $Y = E(Y)$*):*



Esperamos que a covariância de X *e* Y *seja positiva – afinal, note como a maioria dos pontos (e os pontos mais "freqüentes") estão no primeiro e terceiro quadrantes com relação aos eixos com origem em* $(E(X), E(Y))$*. Isto indica que,* ***a grosso modo****, quando* X *está acima da sua média,* Y *também está acima da sua média (e quando* X *está abaixo,* Y *está abaixo). De fato, os cálculos mostram que*

$$\begin{aligned} E(X) &= 1.9;\ E(Y) = 0.4;\ E(XY) = (0.2)(2) + (0.2)(3) = 1.0 \\ Cov(X,Y) &= 1.0 - (1.9)(0.4) = 0.24 \end{aligned}$$

que mostra uma covariância positiva entre X *e* Y*.*
Uma interessante interpretação física: coloque uma massa igual à probabilidade $p(x, y)$ *em cada possível ponto* (x, y)*. O ponto* $(E(X), E(Y))$ *é o* ***centro de massa*** *deste sistema! A covariância será positiva se os pontos se encontrarem mais frequentemente (e fortemente) no primeiro e terceiro quadrante com relação a este novo par de eixos do que no segundo e quarto quadrantes.*

Proposição 51 $\boxed{\text{Se } X \text{ e } Y \text{ são independentes, então } Cov(X,Y) = E(XY) - E(X)E(Y) = 0.}$

Nota 52 *Não vale a volta: é possível ter* $Cov(X,Y) = 0$ *sem que* X *e* Y *sejam independentes! Por exemplo, considere a seguinte distribuição conjunta*

$Y\backslash X$	-1	0	1
-1		0.25	
0	0.25		0.25
1		0.25	

Note que $E(X) = E(Y) = E(XY) = 0$*, então* $Cov(X,Y) = 0$*, mas note que* X *e* Y *não são independentes!*

Assim, a covariância é uma extensão do conceito de independência. A covariância é positiva quando as variáveis estão "positivamente associadas"; é zero quando as variáveis **não são correlacionadas** (isto é, não se afetam **na média, a grosso modo**); e é negativa quando X e Y "se repelem".

Infelizmente, o valor numérico da covariância depende das unidades usadas para medir X e Y. Por exemplo, se uma troca de unidades multiplicar todos os valores de X por 10 a covariância também será multiplicada por 10:

$$Cov(10X, Y) = E(10XY) - E(10X)E(Y) = 10(E(XY) - E(X)E(Y)) = 10Cov(X,Y)$$

Definição 53 *Outra medida da "variação conjunta" de duas variáveis X e Y é a* ***correlação***

$$\boxed{\rho(X,Y) = \frac{Cov(X,Y)}{\sigma(X)\sigma(Y)}}$$

É fácil ver que $\rho(aX, bY) = \rho(X,Y)$ para quaisquer constantes positivas a e b. Assim, a correlação é independente da unidade usada para medir X e Y.

Exemplo 54 *Voltemos ao nosso exemplo anterior, onde $Cov(X,Y) = 0.24$. Para encontrar a correlação, precisamos também encontrar*

$$\begin{aligned} E\left(X^2\right) &= 4.3;\ Var(X) = 0.69;\ E\left(Y^2\right) = 0.4;\ Var(Y) = 0.24 \\ \rho(X,Y) &= \frac{0.24}{\sqrt{0.69}\sqrt{0.24}} = 0.5898 \end{aligned}$$

que mostra uma correlação positiva entre X e Y.

Note que $Cov(X,X) = E\left(X^2\right) - (E(X))^2 = Var(X)$.

2.5.1 Um pouco de Álgebra Linear

Uma outra maneira de interpretar valores esperados, variâncias e covariâncias requer um conhecimento básico de Álgebra Linear. Para entendê-la, voltemos ao exemplo anterior:

$Y \backslash X$	1	2	3
0	0.4	0.1	0.1
1	0	0.2	0.2

Escreva um vetor com os valores de X para todos os pontos possíveis de nosso espaço amostral, e outro com os correspondentes vetores de Y. Neste caso, teríamos:

$$\begin{aligned} v_X &= (1,2,2,3,3) \\ v_Y &= (0,0,1,0,1) \end{aligned}$$

Note que, como há 5 pontos possíveis, estes vetores pertencem a $\mathbb{R}^5$. Defina tabém o vetor $u = (1,1,1,1,1)$ (com 1 em todas as coordenadas). Usando que $E(X) = 1.9$ e $E(Y) = 0.4$, podemos também escrever os vetores dos **resíduos:**

$$\begin{aligned} r_X &= v_X - E(X)\,u = (-0.9, 0.1, 0.1, 1.1, 1.1) \\ r_Y &= v_Y - E(Y)\,u = (-0.4, -0.4, 0.6, -0.4, 0.6) \end{aligned}$$

Enfim, defina um produto interno em $\mathbb{R}^5$ usando as probabilidades da distribuição conjunta como pesos na ordem correta:

$$\langle (x_1,x_2,x_3,x_4,x_5), (y_1,y_2,y_3,y_4,y_5) \rangle = 0.4x_1y_1 + 0.1x_2y_2 + 0.2x_3y_3 + 0.1x_4y_4 + 0.2x_5y_5$$

No caso geral, o produto interno seria algo como

$$\langle x, y \rangle = \sum_{ij} x_{ij} y_{ij} \Pr(X = i; Y = j)$$

Não é difícil provar que, como todas as probabilidades são positivas, isto é de fato um produto interno válido. Todas as propriedades descritas daqui para a frente são verificáveis no caso geral.

Note que, com este produto interno, u é unitário, isto é, $\langle u, u \rangle = 1$, pois a soma das probabilidades deve ser 1. Com esta notação, temos:

$$\begin{aligned} E(X) &= \langle v_X, u \rangle;\ E(Y) = \langle v_Y, u \rangle; \\ E(XY) &= \langle v_X, v_Y \rangle \\ Var(X) &= \langle r_X, r_X \rangle;\ Var(Y) = \langle r_Y, r_Y \rangle; \\ \sigma(X) &= |r_X|;\ \sigma(Y) = |r_Y| \end{aligned}$$

e, em particular, $E(X - E(X)) = \langle r_X, u \rangle = 0$.[5]

Enfim, o sinal de
$$Cov(X,Y) = \langle r_X, r_Y \rangle$$
é uma maneira de verificar se os resíduos de X estão "alinhados" com os resíduos de Y. **Assim, a covariância é um produto interno, e tem todas as propriedades conhecidas dos produtos internos.** Por exemplo:

$$\boxed{Cov(X,Y) = Cov(Y,X)}$$
$$\boxed{Cov(aX,Y) = aCov(X,Y)}$$
$$\boxed{Cov(X+Y,Z) = Cov(X,Z) + Cov(Y,Z)}$$

Por outro lado, uma maneira ainda melhor de verificar o alinhamento de r_X com r_Y é encontrar o (cosseno do) ângulo entre eles:
$$\cos\theta = \frac{\langle r_X, r_Y \rangle}{|r_X|\,|r_Y|} = \frac{Cov(X,Y)}{\sigma(X)\,\sigma(Y)} = \rho(X,Y)$$
que independe das unidades usadas na determinação de X e Y. Este é o significado da correlação na linguagem da Álgebra Linear: **a correlação é o ângulo formado pelos vetores dos resíduos** (quando usamos um produto interno conveniente, ponderado pela distribuição de probabilidade).

Em particular, sabemos então que $\boxed{-1 \leq \rho(X,Y) \leq 1}$; e é fácil provar que:

$$\boxed{\text{Se } \rho(X,Y) = \pm 1 \text{ então } X \text{ e } Y \text{ obedecem perfeitamente a uma relação linear}}$$

De fato:

$$\begin{aligned} \rho(X,Y) &= \pm 1 \Rightarrow r_Y = \alpha r_X \Rightarrow Y_i - E(Y) = \alpha(X_i - E(X)) \text{ para cada ponto possível } (X_i, Y_i) \Rightarrow \\ &\Rightarrow \text{ todos os pontos } (X_i, Y_i) \text{ satisfazem } y - E(Y) = \alpha(x - E(X)) \end{aligned}$$

e esta última equação é uma reta de coeficiente angular α que passa por $(E(X), E(Y))$ – o sinal de α é o sinal da correlação.

Proposição 55 *Para quaisquer variáveis aleatórias X e Y:*

$$\boxed{Var(X+Y) = Var(X) + Var(Y) + 2Cov(X,Y)}$$

Em particular, re-encontramos a propriedade:

$$\boxed{\text{Se } X \text{ e } Y \text{ são independentes, então } Var(X+Y) = Var(X) + Var(Y)}$$

Prova. Enquanto é possível demonstrar esta propriedade usando muitos somatórios, a interpretação de variâncias e covariâncias como produtos internos facilita nossa vida imensamente:

$$Var(X+Y) = \langle r_{X+Y}, r_{X+Y} \rangle = \langle r_X + r_Y, r_X + r_Y \rangle = \langle r_X, r_X \rangle + \langle r_Y, r_Y \rangle + 2\langle r_X, r_Y \rangle = Var(X) + Var(Y) + 2Cov(X,Y)$$

que nada mais é do que uma versão da "Lei dos Cossenos". ■

Proposição 56
$$\boxed{Cov(aX+b, Y) = aCov(X,Y)}$$

---

[5] Assim, não é de se surpreender que algumas propriedades possam ser redemonstradas com esta notação de produto interno. Por exemplo

$$\begin{aligned} Var(X) &= \langle r_X, r_X \rangle = \langle v_X - E(X)u, v_X - E(X)u \rangle = \langle v_X, v_X \rangle - 2E(X)\langle v_X, u \rangle + (E(X))^2 \langle u, u \rangle = \\ &= \langle v_X, v_X \rangle - 2E(X)E(X) + (E(X))^2 = E(X^2) - (E(X))^2 \end{aligned}$$

Prova. De fato,

$$Cov\left(aX+b,Y\right)=\left\langle r_{aX+b},r_Y\right\rangle=\left\langle ar_X,r_Y\right\rangle=a\left\langle r_X,r_Y\right\rangle=aCov\left(X,Y\right)$$

■

Proposição 57

$$\boxed{\rho\left(aX+b,Y\right)=\begin{cases}\rho\left(X,Y\right) & se\ a>0\\ -\rho\left(X,Y\right) & se\ a<0\end{cases}}$$

Prova. De fato,

$$\rho\left(aX+b,Y\right)=\frac{Cov\left(aX+b,Y\right)}{\sigma\left(aX+b\right)\sigma\left(Y\right)}=\frac{aCov\left(X,Y\right)}{|a|\,\sigma\left(X\right)\sigma\left(Y\right)}=\frac{a}{|a|}\rho\left(X,Y\right).$$

■

Exemplo 58 *Uma urna tem cinco bolas numeradas 0, 0, 1, 2, 2; escolha sem reposição* ***três*** *delas. Defina X como a menor, Y como a maior. Vamos encontrar a distribuição conjunta de X e Y, $E\left(X\right)$, $Var\left(X\right)$, $\rho\left(X,Y\right)$ e $Var\left(X+Y\right)$.*
Em primeiro lugar, há $\binom{5}{3}=10$ maneiras de escolher as bolas, como na tabela abaixo (onde distinguimos as bolas 0_A, 0_B, 2_A e 2_B por clareza):

Escolha:	0_A0_B1	$0_A0_B2_A$	$0_A0_B2_B$	0_A12_A	0_A12_B	$0_A2_A2_B$	0_B12_A	0_B12_B	$0_B2_A2_B$	12_A2_B
X	0	0	0	0	0	0	0	0	0	1
Y	1	2	2	2	2	2	2	2	2	2

Como estas 10 escolhas são igualmente prováveis, temos:

$Y\backslash X$	0	1
1	10%	0%
2	80%	10%

Então

$$\begin{aligned}E\left(X\right)&=E\left(X^2\right)=0.1;\ Var\left(X\right)=0.1-0.01=0.09\ e\ \sigma\left(X\right)=0.3\\ E\left(Y\right)&=0.1+0.9\left(2\right)=1.9;\ Var\left(Y\right)=0.09\ e\ \sigma\left(X\right)=0.3\\ E\left(XY\right)&=0.2\Rightarrow Cov\left(X,Y\right)=0.2-\left(0.1\right)\left(1.9\right)=0.01\Rightarrow\rho\left(X,Y\right)=\frac{0.01}{\left(0.3\right)\left(0.3\right)}=\frac{1}{9}\\ Var\left(X+Y\right)&=Var\left(X\right)+Var\left(Y\right)+2Cov\left(X,Y\right)=0.09+0.09+0.02=0.2\end{aligned}$$

2.5.2 Exercícios

Ex. 43 *Jogue um dado duas vezes. Sejam S a soma dos dois dados, e D a sua diferença na ordem em que apareceram (D pode ser negativo). Mostre que $Cov\left(S,D\right)=0$, mas que S e D não são independentes.*

Ex. 44 *Se $E\left(X\right)=1$, $E\left(Y\right)=2$, $Var\left(X\right)=2$, $Var\left(Y\right)=1$ e $Cov\left(X,Y\right)=-1$, calcule $E\left(2X-3Y\right)$, $Var\left(2X-3Y\right)$ e $Cov\left(2X-3Y,X+Y\right)$.*

Ex. 45 *Se $E(X)=3$, $E(Y)=-2$, $Var(X)=4$, $Var(Y)=1$ e $Cov(X,Y)=-1$. Determine $E(X-2Y+1)$, $Var(X-2Y+1)$ e $Cov(X-2Y,2X-Y+1)$.*

Ex. 46 *Se $E\left(X\right)=3$, $E\left(Y\right)=2$, $Var\left(X\right)=1$, $Var\left(Y\right)=4$ e $Cov\left(X,Y\right)=-2$, encontre Y em função de X. [Dica: correlação.]*

Ex. 47 *Jogue um dado duas vezes. Sejam X o máximo dos dois, e Y o mínimo. Calcule $E\left(X+Y\right)$, $Var\left(X\right)$, $Var\left(Y\right)$, $E\left(XY\right)$, $Cov\left(X,Y\right)$, $Var\left(X+Y\right)$ e $\rho\left(X,Y\right)$.*

Ex. 48 *A tabela abaixo dá a distribuição conjunta de probabilidade de X e Y. Determine:*

$X\backslash Y$	0	1	2
-1	0.2	0	0.1
0	0.1	0.1	0.1
1	0.2	0.1	0.1

a) As distribuições marginais de X e de Y.
b) O valor esperado e a variância de X e de Y.
c) A covariância entre X e Y. Elas são independentes?
d) $E(X+Y)$ e $E(XY)$.

Ex. 49 *Dada a distribuição conjunta das variáveis X e Y, parcialmente conhecida, pela tabela abaixo, complete a tabela, supondo X e Y independentes. Então:*

$Y\backslash X$	0	1	2	$\Pr(Y=y)$
1				0.2
2	0.15		0.05	
3				
$\Pr(X=x)$	0.3			

a) Calcule $E(X)$, $E(Y)$, o desvio padrão de X e o desvio-padrão de Y.
b) Seja $Z = 2X - 4Y$. Construa a distribuição de probabilidade de Z.
c) Calcule $E(Z)$ e $Var(Z)$.

Ex. 50 *Considere a distribuição conjunta das variáveis X e Y:*

$Y\backslash X$	1	2	3
1	0	0.1	0.2
2	0.2	0.1	0.1
3	0.2	0	0.1

Seja $W = X + Y$. Calcule $E(W)$, $Var(W)$ e $\sigma(W)$.

2.6 Exercícios de Provas

Ex. 51 (A1 2003.2) *As variáveis aleatórias X e Y são tais que $Var(X) = 3$, $Var(Y) = 1$ e $Cov(X,Y) = -1$. Determine $Var(3X - 2Y)$ e $Cov(2X + Y, X - Y)$.*

Ex. 52 (A1 2003.2) *Um jogador aposta em um jogo honesto de cara-ou-coroa, de acordo com o seguinte sistema:*
i) aposta 1 real na primeira jogada, 2 reais na segunda, e assim sucessivamente, dobrando sempre;
ii) pára de jogar quando ganha;
iii) não joga mais de cinco partidas.
Determine a distribuição do seu ganho total e o valor esperado desse ganho.

Ex. 53 (A1 2004.2) *Um investidor está analisando três alternativas de investimento, para os quais as diversas probabilidades de retorno são dadas na tabela abaixo.*

Investimento X		*Investimento Y*		*Investimento Z*	
Retorno	*Prob.*	*Retorno*	*Prob.*	*Retorno*	*Prob.*
0	*0.2*	*-1*	*0.3*	*-1*	*0.1*
1	*0.4*	*1*	*0.4*	*1*	*0.1*
2	*0.4*	*5*	*0.3*	*2*	*0.8*

O investidor prefere o investimento de maior retorno médio. Entre dois investimentos com os mesmos retornos médios, ele prefere o de menor risco (ou seja, de menor variância). Analise os investimentos acima e diga qual deve ser o escolhido.

Ex. 54 (A2 2004.2) *Uma moeda honesta é lançada três vezes. Seja X o número de caras nos dois primeiros lançamentos e Y o número de caras nos dois últimos lançamentos.*
a) Construa a tabela de probabilidade conjunta de X e Y.
b) Encontre a distribuição marginal de X e a distribuição condicional de X dado que $Y = 1$. As variáveis X e Y são independentes?
c) Calcule a correlação entre X e Y.

Ex. 55 (T2 2005.2) *Cada pessoa de um casal escolhe independentemente um número do conjunto $\{-2, -1, 0, 1, 2\}$ (todos os números têm a mesma chance de serem escolhidos). Seja X o menor dos dois números e Y o maior (se forem iguais, então $X = Y$). A distribuição conjunta de X e Y é dada pela tabela*

Y↓ X→	-2	-1	0	1	2
-2	0.04	0	0	0	0
-1	0.08	0.04	0	0	0
0	0.08	0.08	0.04	0	0
1	0.08	0.08	0.08	0.04	0
2	0.08	0.08	0.08	0.08	0.04

a) Encontre a distribuição marginal, a esperança e a mediana de X.
b) Encontre a distribuição condicional de X dado que $Y = 0$.
c) Seja $Z = Y - X$. Encontre a distribuição de Z e calcule $E(Z)$.

Ex. 56 (T3 2005.2) *Cada pessoa de um casal escolhe independentemente um número do conjunto $\{-2, -1, 0, 1, 2\}$ (todos os números têm a mesma chance de serem escolhidos). Seja X o menor dos dois números e Y o maior (se forem iguais, então $X = Y$). A distribuição conjunta de X e Y é dada pela tabela*

Y↓ X→	-2	-1	0	1	2
-2	0.04	0	0	0	0
-1	0.08	0.04	0	0	0
0	0.08	0.08	0.04	0	0
1	0.08	0.08	0.08	0.04	0
2	0.08	0.08	0.08	0.08	0.04

donde se calculam $E(X) = -0.8$, $E(Y) = 0.8$ e $Var(X) = Var(Y) = 1.36$.
a) Calcule $Cov(X, Y)$ e $\rho(X, Y)$.
b) Calcule $Var(Y - X)$.
c) Calcule $Cov(Y - X, Y + X)$.
d) As variáveis $Y - X$ e $Y + X$ são independentes?

Ex. 57 (A1 2005.2) *A distribuição conjunta de X e Y é dada pela tabela*

Y↓ X→	-1	0	1
-1	0.17	0.04	0.09
0	0.02	0.20	0.18
1	0.11	0.16	0.03

a) Encontre as distribuições marginais de X e de Y e calcule $\Pr\left(X \leq 0 \mid Y^2 = 1\right)$.
b) Calcule $Cov(X, Y)$. As variáveis X e Y são independentes?
c) Calcule $\rho(X, 3X + 4Y)$.

Ex. 58 (A1 2005.2) *Em 1938, o físico Frank Benford observou que, em listas de números e estatísticas que ocorram de forma "natural", o **primeiro dígito** de tais números é 1 cerca de 30% do tempo (muito mais do que os $1/9 = 11.1\%$ esperados se a distribuição dos dígitos fosse uniforme). De fato, sendo X o primeiro dígito de um número destas listas, a distribuição de probabilidade de X é dada pela distribuição de Benford:*

$$\Pr(X = k) = c \ln\left(\frac{k+1}{k}\right) \text{ para } k = 1, 2, ...9$$

a) Encontre a constante c para que esta distribuição seja, de fato, uma distribuição de probabilidade.
b) Escreva uma ***fórmula simples*** *para a função de distribuição acumulada de X e calcule a sua mediana.*
c) Suponha que o primeiro dígito da população de cada município brasileiro segue de perto uma distribuição de Benford. Há 5564 *municípios do Brasil, dos quais* 1309 *têm entre* 5000 *e* 10000 *habitantes. Quantos municípios você estimaria terem entre* 5000 *e* 6000 *habitantes? Que hipótese adicional você fez para chegar a esta estimativa?*

Ex. 59 (T1 2006.1) *Numa urna há cinco bolas, numeradas 0, 0, 1, 2 e 2. Retire, sucessivamente e sem reposição, duas bolas desta urna, cujos números são X e Y, nesta ordem.*
a) Dado que $XY = 0$, qual a probabilidade de termos $X = 0$?
b) Encontre a função de probabilidade conjunta de X e Y.
c) Encontre a distribuição marginal de Y.
d) Encontre a função de probabilidade de $Z = XY$.

Ex. 60 (T2 2006.1) *Numa urna há cinco bolas, numeradas 0, 0, 1, 2 e 2. Retire, sucessivamente e sem reposição, duas bolas desta urna, cujos números são X e Y, nesta ordem. A distribuição conjunta de X e Y é dada pela tabela abaixo:*

$Y \backslash X$	0	1	2	*Marginal de Y*
0	0.1	0.1	0.2	0.4
1	0.1	0	0.1	0.2
2	0.2	0.1	0.1	0.4
Marginal de X	0.4	0.2	0.4	1

a) Calcule $E(X)$ e $Var(X)$.
b) Calcule $Cov(X,Y)$ e $\rho(X,Y)$. As variáveis X e Y são independentes?
c) Calcule $Var(X+Y)$ e $Cov(X-Y, 2X+Y)$.

Chapter 3

Principais Distribuições Discretas

3.1 Distribuição Uniforme

Seja X uma variável aleatória que assume os valores do conjunto $\{x_1, x_2, x_3, ..., x_n\}$ de maneira eqüiprovável, isto é, cada um deles tem probabilidade $\frac{1}{n}$. Dizemos que X tem uma **distribuição uniforme**. É fácil ver que

$$\begin{aligned} E(X) &= \frac{1}{n}\sum_{k=1}^{n} x_k \\ Var(X) &= \frac{1}{n}\sum_{k=1}^{n} x_k^2 - \frac{1}{n^2}\left(\sum_{k=1}^{n} x_k\right)^2 \end{aligned}$$

Por exemplo, se o conjunto for $\{1, 2, 3, ..., n\}$, então pode-se mostrar que (veja exercício 2.27):

$$E(X) = \frac{n+1}{2};\ E\left(X^2\right) = \frac{(n+1)(2n+1)}{6};\ Var(X) = \frac{n^2-1}{12}$$

3.2 Brevíssima Revisão de Análise Combinatória

3.2.1 Princípio Multiplicativo

Uma tarefa deve ser realizada em r estágios. Suponha que há n_1 maneiras de relizar o primeiro estágio; para cada uma destas n_1 maneiras, há n_2 maneiras de realizar o segundo; para cada uma destas, há n_3 maneiras de fazer o terceiro estágio e assim por diante. O **princípio multiplicativo** diz que o número de maneiras de realizar a tarefa toda será

$$N = n_1 n_2 n_3 ... n_r$$

3.2.2 Permutações

Seja A um conjunto com n elementos (distintos). Uma **permutação** em A é uma ordenação dos elementos de A.

Proposição 1 *Um conjunto A com n elementos tem um total de $n! = n(n-1)(n-2)...(3)(2)(1)$ permutações distintas*[1].

De fato, há n possibilidades para o primeiro elemento; para cada escolha deste, há $n-1$ possibilidades para o segundo objeto; para cada uma destas, há $n-2$ possibilidades para o terceiro; e assim por diante, até o penúltimo objeto da ordenação (2 possibilidades) e o último (que só terá 1 possibilidade). Utilizando o princípio multiplicativo, concluímos que há $n!$ maneiras de ordenar os n objetos distintos.

[1] O número $n!$ é denominado **fatorial de** n ou n**-fatorial**. Note que $n! = n.(n-1)!$. Por convenção, adota-se $0! = 1$.

3.2.3 Combinações

Sejam n e k números fixos. Seja A um conjunto com n elementos. A pergunta desta seção é: quantos subconjuntos de A têm k elementos (isto é, de quantas formas podemos escolher k elementos dentre os n elementos de A)?

Definição 2 *O **número binomial** $\binom{n}{k}$ é, por definição, o número de maneiras de escolher k elementos de um conjunto com n elementos; em outras palavras, se A é um conjunto com n elementos, A tem $\binom{n}{k}$ subconjuntos com k elementos cada*[2].

Exemplo 3 *O conjunto $A = \{1, 2, 3\}$ tem um total de $2^3 = 8$ subconjuntos. São: 1 subconjunto com 0 elementos (o vazio); 3 subconjuntos com 1 elemento; 3 subconjuntos com 2 elementos; e apenas 1 subconjunto com 3 elementos (o próprio A). Então*

$$\binom{3}{0} = 1;\ \binom{3}{1} = 3;\ \binom{3}{2} = 3;\ \binom{3}{3} = 1$$

Note que $\binom{3}{0} + \binom{3}{1} + \binom{3}{2} + \binom{3}{3} = 2^3$.

Proposição 4 *Os números binomiais são dados pela fórmula*

$$\boxed{\binom{n}{k} = \frac{n!}{k!(n-k)!}}$$

Prova. Suponha que A tem n elementos. Para escolher um subconjunto **ordenado** de A com k elementos, temos n escolhas para o primeiro elemento, $n-1$ para o segundo, e assim por diante, até $n-k+1$ escolhas para o k-ésimo elemento. Assim, temos um total de

$$n(n-1)(n-2)....(n-k+1) = n(n-1)(n-2)...(n-k+1)\frac{(n-k)(n-k-1)(n-k-2)...(2)(1)}{(n-k)(n-k-1)(n-k-2)...(2)(1)} = \frac{n!}{(n-k)!}$$

subconjuntos **ordenados**. Mas um subconjunto de A com k elementos tem $k!$ possíveis ordenações, isto é, cada subconjunto com k elementos de A foi contado $k!$ vezes dentre os subconjuntos ordenados. Assim, precisamos dividir o número acima por $k!$, obtendo a fórmula final. ■

Proposição 5 *Sejam n e k inteiros com $0 < k < n$. Então*[3]

$$\boxed{\binom{n}{k} = \binom{n-1}{k} + \binom{n-1}{k-1}}$$

Prova. Apresentamos duas demonstrações. Por álgebra:

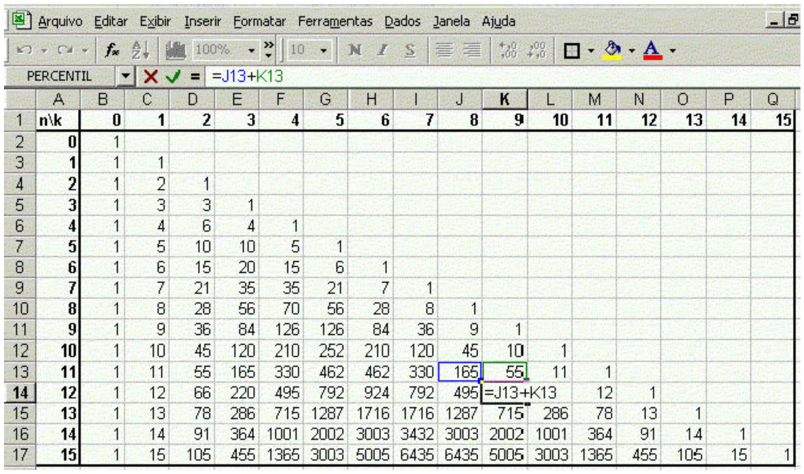
$$\begin{aligned}\binom{n-1}{k} + \binom{n-1}{k-1} &= \frac{(n-1)!}{k!(n-k-1)!} + \frac{(n-1)!}{(k-1)!(n-k)!} = \frac{(n-1)!}{k!(n-k)!}(n-k) + \frac{(n-1)!}{k!(n-k)!}k = \\ &= \frac{(n-1)!}{k!(n-k)!}(n-k+k) = \frac{(n-1)!n}{k!(n-k)!} = \frac{n!}{k!(n-k)!} = \binom{n}{k}\end{aligned}$$

Por combinatória: seja A um conjunto com n elementos. Seja x um elemento qualquer fixo de A. O número de subconjuntos com k elementos de A **que contêm** x é $\binom{n-1}{k-1}$ (pois x já está escolhido; dos outros $n-1$ elementos, devemos escolher $k-1$ para completar o subconjunto). O número de subconjuntos com k elementos de A **que não contêm** x é $\binom{n-1}{k}$ (pois x já está fora; dos outros $n-1$ elementos, temos de escolher k para fazer o subconjunto). Assim, separamos os $\binom{n}{k}$ subconjuntos de A em dois tipos disjuntos: os $\binom{n-1}{k-1}$ que contêm x e os $\binom{n-1}{k}$ que não contêm x. A soma destes últimos tem de ser o primeiro. ■

É fácil ver que $\binom{n}{0} = \binom{n}{n} = 1$. A partir deste fato e da fórmula recursiva acima, é fácil determinar todos os números binomiais e montar o famoso **triângulo de Pascal**. Por exemplo, usando uma planilha Excel:

[2] Este número também é denotado por C_n^k em alguns textos. Lê-se "combinação de n, k a k" ou simplesmente "n, k a k".

[3] É comum usar $\binom{n}{k} = 0$ sempre que $k < 0$ ou $k > n$. Neste caso, a restrição $0 < k < n$ desta proposição não é necessária.



Arquivo Editar Exibir Inserir Formatar Ferramentas Dados Janela Ajuda

PERCENTIL =J13+K13

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1	n\k	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
2	0	1															
3	1	1	1														
4	2	1	2	1													
5	3	1	3	3	1												
6	4	1	4	6	4	1											
7	5	1	5	10	10	5	1										
8	6	1	6	15	20	15	6	1									
9	7	1	7	21	35	35	21	7	1								
10	8	1	8	28	56	70	56	28	8	1							
11	9	1	9	36	84	126	126	84	36	9	1						
12	10	1	10	45	120	210	252	210	120	45	10	1					
13	11	1	11	55	165	330	462	462	330	165	55	11	1				
14	12	1	12	66	220	495	792	924	792	495	=J13+K13		12	1			
15	13	1	13	78	286	715	1287	1716	1716	1287	715	286	78	13	1		
16	14	1	14	91	364	1001	2002	3003	3432	3003	2002	1001	364	91	14	1	
17	15	1	15	105	455	1365	3003	5005	6435	6435	5005	3003	1365	455	105	15	1

onde o cursor destaca a fórmula utilizada em todas as células (exceto na coluna $k = 0$, onde copiamos o valor 1).

O leitor atento notará a relação entre os números da linha n acima e os coeficientes do desenvolvimento de $(x + y)^n$. De fato:

Proposição 6 (Binômio de Newton)

$$\boxed{(x + y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k}}$$

Prova. Ora, $(x + y)^n = (x + y)(x + y)...(x + y)$. Aplicando a propriedade distributiva, encontraremos vários termos da forma $x^k y^{n-k}$ para k variando de 0 a n. A única questão é: quantos termos da forma $x^k y^{n-k}$? Para encontrar um termo destes, basta escolher k binômios da forma $x + y$ donde os x virão. Como há n destas somas, há exatamente $\binom{n}{k}$ formas de fazer esta escolha. ■

Corolário 7 *Tomando $x = y = 1$, temos*

$$\boxed{\sum_{k=0}^{n} \binom{n}{k} = 2^n}$$

confirmando o que já sabíamos: um conjunto de n elementos tem 2^n subconjuntos.

3.3 Processo de Bernoulli

Definição 8 *Um **processo de Bernoulli** é uma seqüência de experimentos com as seguintes características:*
*a) Cada experimento tem apenas dois resultados possíveis, denominados **sucesso** e **falha**;*
b) Cada experimento tem a mesma probabilidade p de sucesso, e cada experimento é completamente independente de todos os outros.

Por exemplo, os seguintes experimentos podem ser modelados por processos de Bernoulli:
a) Jogue uma moeda justa $n = 100$ vezes (tome, por exemplo, sucesso=cara e $p = \frac{1}{2}$) e conte o número de caras;
b) Lance um dado até obter um 6 (tome sucesso="6" e $p = \frac{1}{6}$) e conte o número de lançamentos;
c) Jogue no vermelho da roleta em Las Vegas até sair no lucro ou perder \$200 (sucesso="vermelho" e $p = \frac{18}{38}$).

É comum usar a notação $q = 1 - p$ para a probabilidade de falha de cada experimento.

3.3.1 Distribuição de Bernoulli

Suponha que limitamos o processo acima a apenas 1 experimento. Seja X uma variável aleatória definida por $X = 1$ em caso de sucesso e $X = 0$ em caso de falha. Então a distribuição de X é simplesmente:

k	0	1
$\Pr(X = k)$	$q = 1 - p$	p

Como no exercício 2.28, é fácil calcular as principais propriedades de X, a saber

$$\begin{aligned} E(X) &= E\left(X^2\right) = p \\ Var(X) &= p - p^2 = p.q \end{aligned}$$

3.3.2 Distribuição Binomial

Definição e Função de Probabilidade

Suponha agora que o número de experimentos a serem feitos é determinado a priori – digamos, faremos n experimentos. Seja X a variável aleatória que representa o número de sucessos obtidos nestes n experimentos. Dizemos que X tem uma **distribuição binomial de parâmetros** n **e** p (e escrevemos $X \sim Bin(n, p)$).

Neste caso, chamaremos a função de probabilidade de X por um nome especial: BinomialDen, isto é

$$\Pr(X = k) = \text{BinomialDen}(k; n, p)$$

e chamaremos[4] a função acumulada de BinomialDist:

$$\Pr(X \leq k) = \text{BinomialDist}(k; n, p)$$

Como calcular esta misteriosa função *"BinomialDen"*? Considere o espaço amostral dado por todas as possíveis seqüências de Sucessos e Falhas do experimento. Por exemplo, para $n = 4$, temos 16 possíveis seqüências:

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

Note que todas as seqüências com um determinado número k de sucessos têm a mesma probabilidade, a saber

$$p^k q^{n-k}$$

pois cada um dos k sucessos tem probabilidade p de ocorrer, cada uma das $n-k$ falhas tem probabilidade $q = 1-p$ de ocorrer. Por exemplo, no caso $n = 4$ e $k = 2$, temos

$$\Pr(SSFF) = \Pr(SFSF) = \Pr(SFFS) = \Pr(FSSF) = \Pr(FSFS) = \Pr(FFSS) = p^2 q^2$$

Agora, quantas destas seqüências têm exatamente k letras S e $n-k$ letras F? Para determinar uma tal seqüência, basta escolher k posições para os S (dentre as n possíveis). Em suma, são $\binom{n}{k}$ seqüências com k sucessos, cada uma com probabilidade $p^k q^{n-k}$. Conclusão:

Proposição 9

$$\boxed{\text{BinomialDen}(k; n, p) = \Pr(X = k) = \tbinom{n}{k} p^k q^{n-k}}$$

Quanto à função acumulada, não há fórmula fechada simples para ela. No momento[5], o melhor que podemos fazer é:

$$\text{BinomialDist}(k; n, p) = \Pr(X \leq k) = \sum_{i=0}^{k} \binom{n}{k} p^k q^{n-k}$$

Exemplo 10 *Jogue uma moeda justa* 10 *vezes. Qual a chance de obter exatamente* 5 *caras? Exatamente* 2*? Qual a chance de obter menos de* 2 *caras?*
Seja X o número de caras obtidas em 10 *lançamentos da moeda justa. Então $X \sim Bin(10, 0.5)$. Portanto*

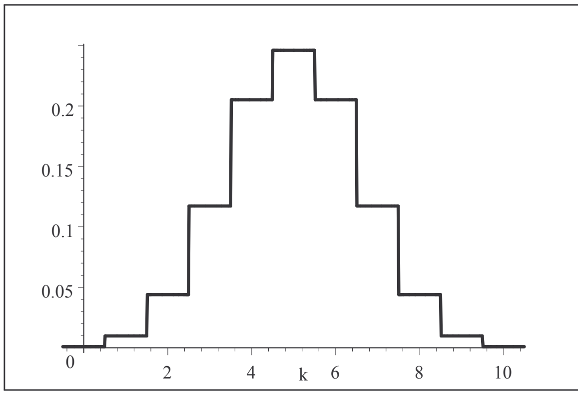
$$\begin{aligned} \Pr(X = 5) &= \text{BinomialDen}(5; 10, 0.5) = \binom{10}{5}\left(\frac{1}{2}\right)^5\left(\frac{1}{2}\right)^5 = \frac{252}{1024} = \frac{63}{256} = 24.609\% \\ \Pr(X = 2) &= \text{BinomialDen}(2; 10, 0.5) = \binom{10}{2}\left(\frac{1}{2}\right)^2\left(\frac{1}{2}\right)^8 = \frac{45}{1024} = 4.3945\% \\ \Pr(X \leq 1) &= \Pr(X = 0) + \Pr(X = 1) = \binom{10}{0}\left(\frac{1}{2}\right)^5\left(\frac{1}{2}\right)^5 + \binom{10}{1}\left(\frac{1}{2}\right)^5\left(\frac{1}{2}\right)^5 = 1.074\% \end{aligned}$$

[4] De fato, os nomes mais comuns na literatura para a função de probabilidade binomial são $Bin(k; n, p)$ ou $b(k; n, p)$. A notação que estamos usando coincide com a do Scientific Workplace e do Maple. Já o Excel usa as sintaxes *DISTRBINOM(k;n;p;0)* (*BINOMDIST(k;n;p;0)* em inglês) para a função de probabilidade e *DISTRBINOM(k;n;p;1)* (*BINOMDIST(k;n;p;1)*) para a função acumulada.

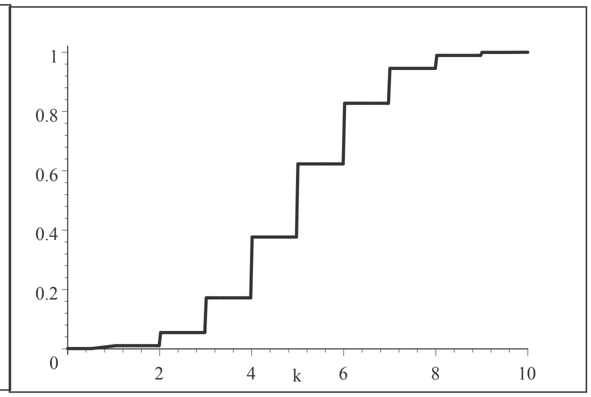
[5] Mais tarde veremos como **aproximar** BinomialDist por outras funções.

ou, se você tiver acesso a uma ferramenta computacional que calcule a função acumulada:

$$\Pr(X \le 1) = \mathrm{BinomialDist}(1; 10, 0.5) = 1.074\%$$



$\mathrm{BinomialDen}(k; 10, 0.5)$



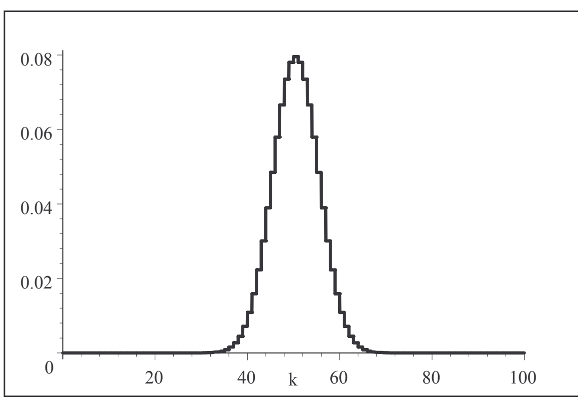
$\mathrm{BinomialDist}(k; 10, 0.5)$

Exemplo 11 *Jogue uma moeda justa* 100 *vezes. Qual a chance de obtermos exatamente* 50 *caras? Exatamente* 20*? Entre* 40 *e* 60 *caras?*
Seja X *o número de caras obtidas nos* 100 *lançamentos da moeda justa. Então* $X \sim Bin(100, 0.5)$*. Portanto*

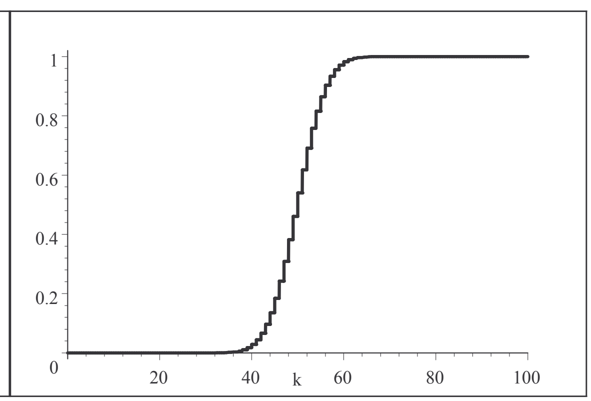
$$\begin{aligned}
\Pr(X = 50) &= \mathrm{BinomialDen}(50; 100, 0.5) = \binom{100}{50}\left(\frac{1}{2}\right)^{50}\left(\frac{1}{2}\right)^{50} = 7.959\% \\
\Pr(X = 20) &= \mathrm{BinomialDen}(20; 100, 0.5) = \binom{100}{20}\left(\frac{1}{2}\right)^{20}\left(\frac{1}{2}\right)^{80} = 4.228 \times 10^{-10} \\
\Pr(40 \le X \le 60) &= \sum_{k=40}^{60}\binom{100}{k}\left(\frac{1}{2}\right)^{k}\left(\frac{1}{2}\right)^{100-k}
\end{aligned}$$

Este somatório é bem difícil de calcular. Usando uma ferramenta computacional:

$$\Pr(40 \le X \le 60) = \mathrm{BinomialDist}(60; 100, 0.5) - \mathrm{BinomialDist}(39; 100, 0.5) = 96.48\%$$



$\mathrm{BinomialDen}(k; 100, 0.5)$



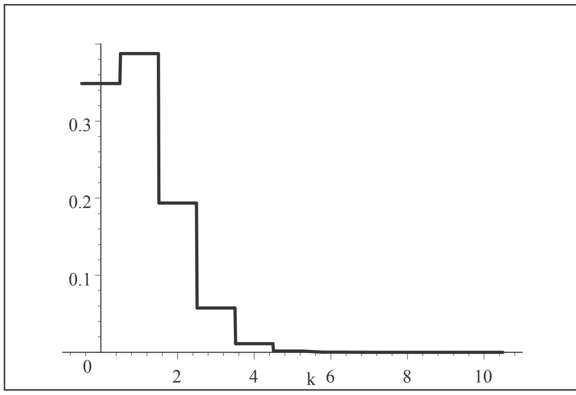
$\mathrm{BinomialDist}(k; 100, 0.5)$

Exemplo 12 *Role um dado* 6 *vezes. Qual a chance de obtermos exatamente um* 6*? Mais do que um* 6*?*
Seja X *o número de* 6 *rolados. Então* $X \sim Bin\left(6, \frac{1}{6}\right)$*. Queremos*

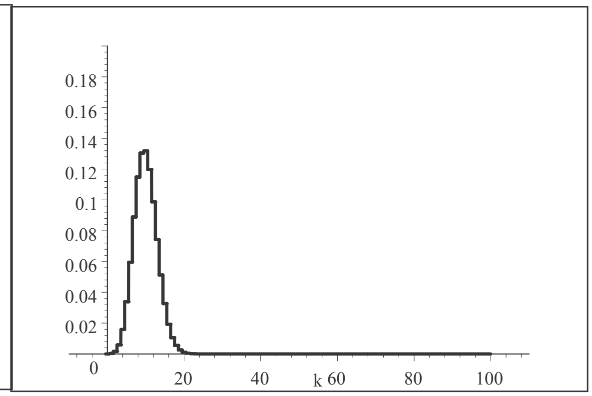
$$\begin{aligned}
\Pr(X = 1) &= \mathrm{BinomialDen}\left(1; 6, \frac{1}{6}\right) = \binom{6}{1}\left(\frac{1}{6}\right)^{1}\left(\frac{5}{6}\right)^{5} = \frac{3125}{7776} = 40.1878\% \\
\Pr(X > 1) &= 1 - \Pr(X \le 1) = 1 - \binom{6}{0}\left(\frac{1}{6}\right)^{0}\left(\frac{5}{6}\right)^{6} - \binom{6}{1}\left(\frac{1}{6}\right)^{1}\left(\frac{5}{6}\right)^{5} = 26.3224\%
\end{aligned}$$

Exemplo 13 *Sabe-se que 10% dos parafusos de uma fábrica têm defeito. Qual a probabilidade de termos exatamente 2 defeituosos em 10? E 20 defeituosos em 100?*
Solução:

$$\begin{aligned}
\textit{Número de defeituosos} &: X \sim Bin\,(10, 0.1) \;\; e\; Y \sim Bin\,(100, 0.1) \\
\Pr(X = 2) &= \mathrm{BinomialDen}\,(2; 10, 0.1) = \binom{10}{2} (0.1)^2 (0.9)^8 = 19.37\% \\
\Pr(Y = 20) &= \mathrm{BinomialDen}\,(20; 100, 0.1) = \binom{100}{20} (0.1)^{20} (0.9)^{80} = 0.1171\%
\end{aligned}$$



BinomialDen $(k; 10, 0.1)$



BinomialDen $(k; 100, 0.1)$

A planilha Excel *Discrete.xls* contém uma página com a distribuição binomial de parâmetros n e p (função de probabilidade e acumulada) que permite que você mude os valores de n e p e acompanhe como o gráfico da função de probabilidade muda.

Valor Esperado e Variância

Proposição 14 *Seja* $X \sim Bin\,(n, p)$. *Então*

$$\boxed{E\,(X) = np}$$

$$\boxed{Var\,(X) = npq}$$

Prova. Prova simples: Seja $X_i = 1$ se o i-ésimo experimento foi um sucesso e $X_i = 0$ caso contrário. Então X_i é uma prova de Bernoulli, e já calculamos

$$E\,(X_i) = p \text{ e } Var\,(X_i) = pq$$

Note que o número de sucessos total é $X = X_1 + X_2 + ... + X_n$. Então

$$\begin{aligned}
E\,(X) &= E\,(X_1) + E\,(X_2) + ... + E\,(X_n) = np \\
Var\,(X) &= Var\,(X_1) + Var\,(X_2) + ... + Var\,(X_n) = npq
\end{aligned}$$

já que as variáveis X_1, X_2, ..., X_n são todas independentes entre si.
Prova Algébrica Feia:

$$\begin{aligned}
E\,(X) &= \sum_{k=0}^{n} k.p\,(k) = \sum_{k=0}^{n} k.\frac{n!}{k!\,(n-k)!} p^k q^{n-k} = \sum_{k=1}^{n} n \frac{(n-1)!}{(k-1)!\,(n-k)!} p^k q^{n-k} = \\
&= np \sum_{k=1}^{n} \frac{(n-1)!}{(k-1)!\,(n-k)!} p^{k-1} q^{n-k} = np \sum_{j=0}^{n-1} \frac{(n-1)!}{j!\,(n-1-j)!} p^j q^{n-1-j} = \\
&= np \sum_{j=0}^{n-1} \binom{n-1}{j} p^j q^{n-1-j} = np\,(p+q)^{n-1} = np
\end{aligned}$$

onde usamos que: o termo em $k = 0$ pode ser descartado do somatório; a mudança de variáveis $j = k - 1$; o binômio de Newton para $(p+q)^{n-1}$; e o fato de que $p + q = 1$. Analogamente:

$$\begin{aligned}
E\left(X^2 - X\right) &= \sum_{k=0}^{n} \left(k^2 - k\right).p(k) = \sum_{k=2}^{n} k(k-1)\frac{n!}{k!(n-k)!}p^k q^{n-k} = \\
&= n(n-1)p^2 \sum_{k=2}^{n} \frac{(n-2)!}{(k-2)!(n-k)!}p^{k-2}q^{n-k} = \\
&= n(n-1)p^2 \sum_{j=0}^{n-2} \frac{(n-2)!}{j!(n-2-j)!}p^j q^{n-2-j} = n(n-1)p^2(p+q)^{n-2} = n(n-1)p^2
\end{aligned}$$

Portanto

$$\begin{aligned}
E\left(X^2\right) &= E\left(X^2 - X\right) + E(X) = n(n-1)p^2 + np = np((n-1)p+1) = np(np+q) \\
Var(X) &= E\left(X^2\right) - (E(X))^2 = np(np+q-np) = npq
\end{aligned}$$

■

3.3.3 Distribuição Geométrica

Suponha agora que realizamos um processo de Bernoulli com probabilidade de sucesso de cada prova $p > 0$. Seja X o número de tentativas feitas até o primeiro sucesso (inclusive). Dizemos que X tem uma **distribuição geométrica de parâmetro** p, isto é, $X \sim Geom(p)$.

Proposição 15 *Se $X \sim Geom(p)$, então*

$$\boxed{\Pr(X = k) = Geom(k; p) = q^{k-1}p}$$
$$\boxed{\Pr(X \leq k) = 1 - q^k}$$

Prova. De fato, para que o primeiro sucesso venha na tentativa k, precisamos de obter $k - 1$ falhas seguidas (cada uma com probabilidade q) e, enfim o sucesso (cuja probabilidade é p). Daí vem $\Pr(X = k) = q^{k-1}p$.
Por outro lado, $X > k$ significa que o primeiro sucesso não ocorreu nas primeiras k provas, isto é, as k primeiras provas foram falhas:

$$\Pr(X > k) = q^k$$

Portanto

$$\Pr(X \leq k) = 1 - q^k$$

Se você preferir uma prova algébrica usando a soma dos termos de uma P.G.:

$$\Pr(X \leq k) = \sum_{j=1}^{k} q^{j-1}p = p\left(1 + q + q^2 + ... + q^{k-1}\right) = p\frac{1-q^k}{1-q} = 1 - q^k$$

■

Proposição 16 *Se $X \sim Geom(p)$, então*

$$\boxed{E(X) = \tfrac{1}{p} \text{ e } Var(X) = \tfrac{q}{p^2}}$$

Prova. Sabemos que a série de potências

$$f(x) = \frac{1}{1-x} = 1 + x + x^2 + x^3 + ... = \sum_{k=0}^{\infty} x^k$$

converge para $|x| < 1$. Derivando termo a termo duas vezes:

$$f'(x) = \frac{1}{(1-x)^2} = 1 + 2x + 3x^2 + 4x^3 + ... = \sum_{k=0}^{\infty} kx^{k-1}$$
$$f'(x) = \frac{2}{(1-x)^3} = 2 + 6x + 12x^2 + 20x^4 + ... = \sum_{k=0}^{\infty} k(k-1)x^{k-2}$$

O que isto tem a ver com o valor esperado? Ora:

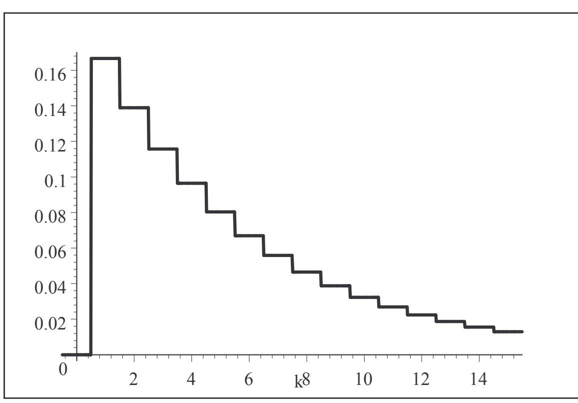
$$E(X) = \sum_{k=0}^{\infty} kq^{k-1}p = pf'(q) = \frac{p}{(1-q)^2} = \frac{1}{p}$$
$$E(X(X-1)) = \sum_{k=0}^{\infty} k(k-1)q^{k-1}p = qpf''(q) = \frac{2pq}{(1-q)^3} = \frac{2q}{p^2} \Rightarrow$$
$$\Rightarrow E\left(X^2\right) = \frac{2q}{p^2} + E(X) = \frac{2q}{p^2} + \frac{p}{p^2} = \frac{q+1}{p^2} \Rightarrow Var(X) = \frac{q+1}{p^2} - \frac{1}{p^2} = \frac{q}{p^2}$$

■

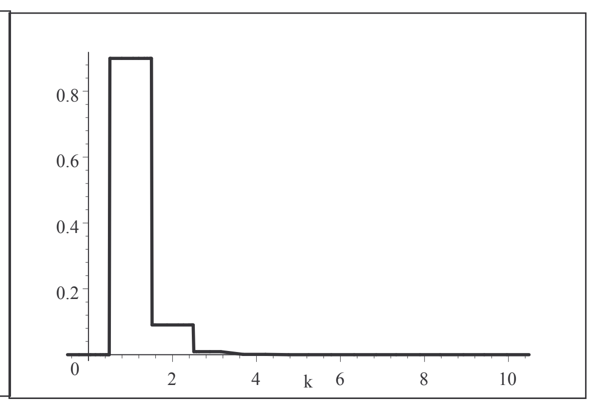
Exemplo 17 *Lança-se um dado não-tendencioso até a obtenção do primeiro* 6*. Seja* X *o número de lançamentos efetuados. Qual é a distribuição de* X*? Quanto valem* $\Pr(X=4)$, $\Pr(X>4)$, $\Pr(X<4)$ *e* $\Pr(X>2 \mid X<5)$*? Quanto valem* $E(X)$ *e* $Var(X)$*?*
Solução: X *tem distribuição geométrica com parâmetro* $p = \frac{1}{6}$*. As probabilidades são*

$$\Pr(X=4) = \left(\frac{5}{6}\right)^3\left(\frac{1}{6}\right) = \frac{125}{1296} = 9.645\%$$
$$\Pr(X>4) = \left(\frac{5}{6}\right)^4 = \frac{625}{1296} = 48.223\%$$
$$\Pr(X<4) = 1 - \left(\frac{5}{6}\right)^3 = \frac{91}{216} = 42.130\%$$
$$\Pr(X>2 \mid X<5) = \frac{\Pr(2<X<5)}{\Pr(X<5)} = \frac{\Pr(X=3)+\Pr(X=4)}{1-\Pr(X>4)} = \frac{q^2p+q^3p}{1-q^4} =$$
$$= \left(\frac{5}{6}\right)^2\left(\frac{1}{6}\right)\frac{1+\frac{5}{6}}{1-\left(\frac{5}{6}\right)^4} = \frac{25}{61} = 40.98\%$$

Em particular, $E(X) = \frac{1}{1/6} = 6$ *e* $Var(X) = \frac{5/6}{(1/6)^2} = 30$.



$Geom\left(k; \frac{1}{6}\right)$



$Geom(k; 0.9)$

Exemplo 18 *Lançar um satélite custa \$2b por lançamento, e dá um benefício de \$3b quando finalmente há sucesso (e então param-se os lançamentos). A probabilidade de sucesso é de 90% por lançamento. Qual o lucro* *esperado? Qual o menor valor da probabilidade que ainda faz o projeto valer a pena?*

$$\begin{aligned}
\textit{No de lançamentos} &: \quad X \sim Geom\,(0.9) \\
\textit{Lucro} &: \quad L = 3 - 2X \\
E\,(L) &= 3 - 2E\,(X) = 3 - \frac{2}{0.9} = \$777.78M \\
\textit{Break-even} &: \quad E\,(L) = 3 - \frac{2}{p} = 0 \Rightarrow p = \frac{2}{3}
\end{aligned}$$

3.3.4 Distribuição Binomial Negativa

Enfim, suponha que o processo de Bernoulli é repetido até obtermos o r-ésimo sucesso. Seja X o número de tentativas feitas (incluindo o último sucesso). Dizemos que X tem uma **distribuição binomial negativa de parâmetros** r **e** p, isto é, $X \sim NegBin\,(r,p)$. Note que a distribuição geométrica é um caso particular da binomial negativa quando $r = 1$.

Proposição 19 *Se $X \sim NegBin\,(r,p)$, então, para $k \geq r$ (k inteiro)*

$$\boxed{\Pr\,(X = k) = p.\,\mathrm{BinomialDen}\,(r-1; k-1, p) = \tbinom{k-1}{r-1} p^r q^{k-r}}$$

$$\boxed{E\,(X) = \tfrac{r}{p}} \ e \ \boxed{Var\,(X) = \tfrac{rq}{p^2}}$$

Prova. A demonstração é um dos exercícios da próxima subseção, tente usar as dicas ali presentes antes de ler esta prova. Depois, confira: seja Y o número de sucessos nos primeiros $k-1$ lançamentos e Z o número de sucessos no k-ésimo lançamento. Então $Y \sim Bin\,(k-1,p)$ e $Z \sim Be\,(p)$. Note que o r-ésimo sucesso acontece no k-ésimo lançamento se, e somente se, $Y = r-1$ e $Z = 1$, isto é

$$X = k \Leftrightarrow (Y = r-1 \text{ e } Z = 1)$$

Como Y e Z são independentes (pois tratam de lançamentos distintos):

$$\Pr\,(X = k) = \Pr\,(Y = r-1)\,.\Pr\,(Z = 1) = \binom{k-1}{r-1} p^{r-1} q^{k-r}.p = \binom{k-1}{r-1} p^r q^{k-r}$$

Enfim, seja X_1 o número de lançamentos até o primeiro sucesso, X_2 o número de lançamentos dali até o segundo sucesso (sem contar o lançamento do primeiro, mas contando o segundo), e assim por diante. Então note que

$$X = X_1 + X_2 + ... + X_r$$

e cada um dos X_i é uma variável com distribuição geométrica de parâmetro p. Então

$$\begin{aligned}
E\,(X) &= rE\,(X_1) = \frac{r}{p} \\
Var\,(X) &= rVar\,(X_i) = \frac{rq}{p^2}
\end{aligned}$$

onde, para a variância, usamos que os X_i são independentes entre si. ■

Exemplo 20 *Lança-se um dado justo até a obtenção do terceiro "seis". Seja X o número de lançamentos efetuados. Qual é a distribuição de X? Qual é a probabilidade de fazermos exatamente* 6 *lançamentos? Menos de* 6*? Mais de* 6*? Calcule $E\,(X)$ e $Var\,(X)$.*
Solução: A distribuição de X é Binomial Negativa, $X \sim NegBin\left(3, \frac{1}{6}\right)$. Então

$$\Pr\,(X = 6) = \frac{1}{6}\,\mathrm{BinomialDen}\left(2, 5, \frac{1}{6}\right) = 2.68\%$$

Para a próxima pergunta, seja Y o número de sucessos nos 6 *primeiros lançamentos (isto é, $Y \sim Bin\left(6, \frac{1}{6}\right)$). Se tivermos que esperar mais de* 6 *lançamentos pelo terceiro sucesso, é por que nos* 6 *primeiros lançamentos tivemos* 2 *ou menos sucessos, isto é:*

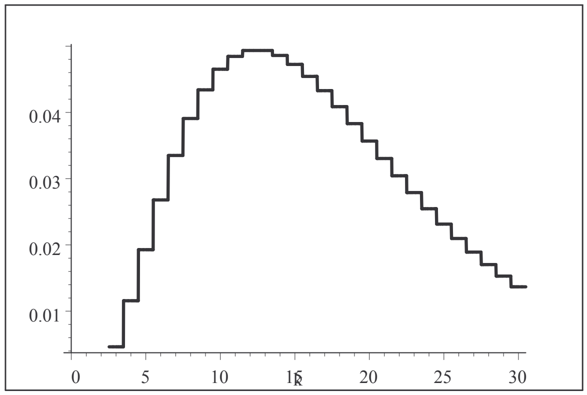
$$\Pr\,(X > 6) = \Pr\,(Y \leq 2) = \mathrm{BinomialDist}\left(2, 6, \frac{1}{6}\right) = 93.77\%$$

Então:

$$\Pr(X < 6) = 1 - \Pr(X = 6) - \Pr(X > 6) = 3.55\%$$

Enfim

$$E(X) = \frac{3}{\left(\frac{1}{6}\right)} = 18 \text{ e } Var(X) = \frac{3\left(\frac{5}{6}\right)}{\left(\frac{1}{6}\right)^2} = 90$$



$NegBin\left(k; 3, \frac{1}{6}\right) = \frac{1}{6}\,\text{BinomialDen}\left(2; k-1; \frac{1}{6}\right)$

3.3.5 Exercícios[6]

Ex. 1 *Suponha que a probabilidade de uma usina nuclear falhar num determinado ano é de* 0.1%. *Suponha que haja* 100 *usinas num determinado país. Qual a probabilidade de ao menos uma falhar neste ano?*

Ex. 2 *Role um dado* 30 *vezes. Qual a chance de obtermos exatamente cinco números* 6*?*

Ex. 3 *A probabilidade de um arqueiro acertar um alvo com uma flecha é de* 0.20. *Lançam-se* 5 *flechas no alvo. Qual é a probabilidade de que ele acerte exatamente* 4 *flechas no alvo? Pelo menos* 2*?*

Ex. 4 10% *dos parafusos produzidos por uma indústria são defeituosos. Numa amostra com* 5 *deles, qual a chance de termos* 0 *defeituosos?* 1*?* 2*? Não mais que* 2*?*

Ex. 5 *Minha chance de ganhar um set do Guga é* 30%. *Qual a chance de eu ganhar um jogo de* 3 *sets? E um de* 5 *sets? E* 7 *sets?*

Ex. 6 *Minhoca Gaúcho consegue fazer exatamente* 4 *finalizações por partida, e a cada uma delas ele tem* 30% *de chance de marcar um gol. Encontre a função de probabilidade do número de gols que ele faz por partida.*

Ex. 7 *Suponha que* $X \sim Bin(n, p)$. *Sabendo que* $E(X) = 30$ *e* $Var(X) = 20$, *calcule* n *e* p.

Ex. 8 *Sabe-se que* 4% *dos passageiros não aparecem nos seus vôos. Eu tenho* 98 *lugares no meu vôo, e reservei lugar para* 100 *passageiros. Qual a chance de eu ter lugar para todos que aparecerem?*

Ex. 9 *Um estudante marca ao acaso as respostas de um teste tipo* V *ou* F *com* 50 *questões.*
a) Qual a probabilidade de ele acertar 80% *(ou mais?) delas apenas adivinhando? E* 60% *ou mais?*
b) Se 100 *estudantes adivinham as questões ao acaso, qual a probabilidade de ao menos um tirar* 80% *ou mais?*
c) E se o teste for múltipla escolha com 5 *alternativas por questão, como mudam os itens anteriores?*

Ex. 10 *Um potencial paranormal tenta adivinhar* 10 *cartas escolhidas aleatoriamente, cada uma com* 5 *possibilidades eqüiprováveis. Se ele não tiver poder algum, qual a chance de adivinhar* 8 *cartas ou mais? E se* 1000 *pessoas comuns forem testadas, qual a chance de ao menos uma adivinhar* 8 *ou mais cartas por puro acaso?*

[6] Em alguns dos problemas desta seção, você vai precisar de uma calculadora ou computador que calcule a Distribuição Binomial Acumulada. Use a planilha EXCEL *Discrete.xls* para fazer os cálculos.

Ex. 11 *Aposte* 70 *vezes no número* 13 *da roleta em Monte Carlo. Qual a chance de você sair de lá com lucro? Qual é o lucro esperado?*

Ex. 12 *a) Mostre que* $\mathrm{BinomialDen}\,(k+1;n,p)=\frac{(n-k)p}{(k+1)(1-p)}\,\mathrm{BinomialDen}\,(k;n,p)$
b) Suponha que $X\sim Bin\,(10,p)$ *e que* $\Pr(X=5)=2\Pr(X=4)$. *Calcule* p.
c) Se $X\sim Bin(m,p)$, *qual o valor mais provável de* X? *Este número é chamado a* ***moda*** *de* X. *[Dica: quando é que* $k+1$ *é mais provável do que* k*?]*

Ex. 13 *Um estudante marca ao acaso as respostas de um teste de* 9 *questões de múltipla escolha com* 4 *alternativas por questão. Determine:*
a) o número esperado de acertos; a variância do número de acertos;
b) a nota esperada; a variância da nota;
c) a probabilidade de obter 4 acertos;
d) o número mais provável de acertos.

Ex. 14 *Numa loja há dois tipos de torta: maçã e chocolate. Todo dia entram* 10 *clientes nesta loja, e cada um deles pede uma torta – chocolate com* 40% *de chance, maçã com* 60% *de chance. Quantas tortas de cada tipo tem de estar estocadas por dia para que haja* 95% *de certeza de que todos os consumidores receberão as tortas que pediram?*

Ex. 15 *A e B disputam uma série de partidas e ganha a série quem primeiro alcançar* 10 *pontos. No momento, o jogo está* 6 *a* 4 *em favor de A. Em cada partida, a probabilidade de B ganhar é* 0.6 *e a de A,* 0.4. *Determine a probabilidade de B ganhar a série. [Dica: faça-os jogarem* 9 *partidas novas.]*

Ex. 16 *Uma urna contém* 3 *bolas brancas,* 3 *bolas pretas e* 4 *bolas vermelhas. Sacam-se, com reposição,* 5 *bolas. Determine a probabilidade de serem sacadas:*
a) exatamente 3 bolas brancas;
b) pelo menos 3 bolas brancas;
c) 2 brancas, 2 pretas e 1 vermelha.

Ex. 17 *Um laboratório é contratado para fornecer lotes de vacina a um distribuidor. Ocasionalmente, algumas vacinas se revelam ineficazes. Como não é possível testar todas (o teste inutiliza a vacina), o distribuidor adota o seguinte processo de seleção: extrai de cada lote uma amostra aleatória de* 10 *ampolas, contendo um número* X *de vacinas estéreis. Se* $X=0$, *o lote é aceito. Se* $X\geq 1$, *o lote é rejeitado*[7]. *Admitamos que o tamanho do lote seja suficientemente grande para quer a distribuição de* X *seja (aproximadamente) binomial, com* $n=10$ *e* $p=0.05$ *(a fração de vacinas estéreis em cada lote). Qual é a probabilidade de que o lote seja aceito?*

Ex. 18 *Se* $X\sim Bin(m,p)$ *e* $Y\sim Bin(n,p)$ *são independentes, qual é a distribuição de* $Z=X+Y$*?*

Ex. 19 *Distribua* 10000 *folhetos aleatoriamente por* 2000 *quadras em uma cidade. Qual a chance de a sua quadra não receber folheto algum? E de receber* 5 *folhetos? E* 10*?*

Ex. 20 *Um dado justo é rolado até que um* 6 *seja obtido. Seja* X *o número de lançamentos feitos. Calcule* $\Pr(T>6)$, $\Pr(T>4)$ *e* $\Pr(T>6\mid T>2)$.

Ex. 21 *Seja* $X\sim Geom\,(p)$. *Mostre que*

$$\Pr(X>a+b\mid X>a)=\Pr(X>b)$$

ou seja, a distribuição geométrica "não tem memória".

Ex. 22 *A probabilidade de uma tentativa ser bem sucedida é* 0.9. *Uma tentativa bem sucedida gera um lucro (ou seja, o custo já está descontado)* 100 *e uma tentativa mal sucedida tem um custo* 10. *Insistir-se-á até haver sucesso. Qual é o número esperado de tentativas? Qual é a probabilidade de serem feitas mais de duas tentativas? Qual é o lucro esperado?*

[7] Este processo é designado **plano de amostragem simples** com tamanho de amostra $n=10$ e o número de aceitação $c=0$.

Ex. 23 *A probabilidade de uma tentativa ser bem sucedida é* 0.4*. Uma tentativa bem sucedida gera um rendimento (ou seja, o custo da tentativa ainda não está descontado)* 100*. As duas primeiras tentativas têm um custo* 10 *e as demais, custo* 5*. Insistir-se-á até haver sucesso. Qual é o número esperado de tentativas? Qual é a probabilidade de serem feitas mais de três tentativas? Qual é o lucro esperado?*

Ex. 24 *Seja $X \sim Geom(p)$. Seja Y o número de fracassos antes do primeiro sucesso. Determine o conjunto de valores, a função de probabilidade, a esperança e a variância de Y.*

Ex. 25 *Num processo de Bernoulli, seja X o número de provas que têm de ser feitas até encontrarmos o r-ésimo sucesso (no caso $r = 1$, X teria uma distribuição geométrica).*
a) Mostre que

$$\Pr(X = k) = \binom{k-1}{r-1} p^r q^{k-r}$$

Esta distribuição é chamada de ***Binomial Negativa com parâmetros*** *r e p. [Dica: considere a probabilidade de obter $r - 1$ sucessos nos primeiros $k - 1$ lançamentos, e então 1 sucesso no próximo lançamento.]*
b) Mostre que $E(X) = \frac{r}{p}$ e $Var(X) = \frac{rq}{p^2}$. [Dica: seja X_1 o número de lançamentos até o primeiro sucesso, X_2 o número de lançamentos dali até o segundo sucesso, etc. Mostre que cada uma destas variáveis é geométrica e que $X = X_1 + X_2 + ... + X_k$.]

Ex. 26 *A probabilidade de o Vasco ganhar um campeonato é de* 0.2*. Qual é a probabilidade de, nos próximos* 10 *anos, o Vasco ganhar mais de um campeonato? E de o segundo campeonato ganho pelo Vasco ocorrer nos próximos 5 anos? E de o segundo campeonato ganho pelo Vasco só acontecer depois dos próximos* 10 *anos?*

Ex. 27 *Seja $X \sim NegBin(r; p)$. Seja Y o número de fracassos antes do r-ésimo sucesso. Determine o conjunto de valores, a função de probabilidade, a esperança e a variância de Y.*

Ex. 28 *Suponha que os sexos dos $n + 2$ filhos de um casal sejam independentes, exceto por um casal de gêmeos idênticos (portanto, de mesmo sexo). Suponha que cada filho tem probabilidade p de ser homem (inclusive os gêmeos). Mostre que a distribuição do número X de filhos homens deste casal é dada por*

$$\Pr(X = k) = p^{k-1} q^{n-k+1} \left(\binom{n}{k} p + \binom{n}{k-2} q \right)$$

para $k = 0, 1, ..., n + 2$. Calcule $E(X)$ e $Var(X)$.

3.4 Processo de Poisson

Queremos um modelo para o número de gols por partida no campeonato brasileiro (X). Suponha que sabemos que há 3 gols por jogo na média (isto é, 2 gols por hora). Podemos tentar:

- **Modelo 1:** Divida o jogo em 3 blocos de 30 minutos cada. Em cada bloco há um gol. Neste caso, $X = 3$ com 100% de chance (tecnicamente, $X \sim Bin(3, 1)$). A média está correta, mas o modelo não nos parece muito bom.

- **Modelo 2**: Divida o jogo em 6 blocos de 15 minutos cada. Em cada bloco há um gol com probabilidade 50%. Neste caso, $X \sim Bin(6, 0.5)$. A média será de 3 gols por jogo, e cada jogo pode ter de 0 a 6 gols. O modelo nos parece melhor, mas jamais gerará um daqueles raros 7×2 (como o Atlético-PR vs. Vasco de 2005).

- **Modelo 3:** Divida o jogo em 90 blocos de 1 minuto cada. Em cada bloco haverá um gol com probabilidade $\frac{1}{30}$. Assim, $X \sim Bin\left(90, \frac{1}{30}\right)$ tem o valor esperado correto, e até os raros jogos com mais de 8 gols (até 90 gols!) podem aparecer.

- **Modelo N:** Divida o jogo em N blocos, pode haver um gol por bloco com probabilidade $\frac{3}{N}$. Então $X \sim Bin\left(N, \frac{3}{N}\right)$.

O que acontece se tomarmos então $N \to \infty$? As distribuições binomiais de probabilidade se aproximam de uma nova distribuição, chamada **distribuição de Poisson.** Neste caso, dizemos que $X \sim Poi\,(3)$.

Em geral, $\lim_{n\to\infty} Bin\left(k; n, \frac{\mu}{n}\right) = Poi\,(k;\mu)$.

Proposição 21 *Se $X \sim Poi\,(\mu)$, tem-se*

$$\boxed{\Pr(X=k) = \frac{\mu^k}{k!}e^{-\mu}}$$

$$\boxed{E(X) = Var(X) = \mu}$$

Prova. Note que $p = \frac{\mu}{n}$ na binomial. Então:

$$\begin{aligned}
\lim_{n\to\infty} Bin\left(k; n, \frac{\mu}{n}\right) &= \lim_{n\to\infty} \frac{n!}{k!\,(n-k)!}\left(\frac{\mu}{n}\right)^k\left(1-\frac{\mu}{n}\right)^{n-k} = \frac{\mu^k}{k!}\lim_{n\to\infty}\frac{n(n-1)(n-2)\ldots(n-k+1)}{n^k}\frac{\left(1-\frac{\mu}{n}\right)^n}{\left(1-\frac{\mu}{n}\right)^k} = \\
&= \frac{\mu^k}{k!}\lim_{n\to\infty}\frac{n^k+\ldots}{n^k}\lim_{n\to\infty}\left(1-\frac{\mu}{n}\right)^n\lim_{n\to\infty}\left(1-\frac{\mu}{n}\right)^{-k} = \frac{\mu^k}{k!}.1.e^{-\mu}.1 = \frac{\mu^k}{k!}e^{-\mu} \\
E(X) &= \sum_{k=0}^{\infty} k\frac{\mu^k}{k!}e^{-\mu} = \mu e^{-\mu}\sum_{k=1}^{\infty}\frac{\mu^{k-1}}{(k-1)!} = \mu e^{-\mu}e^{\mu} = \mu \\
E\left(X^2\right) &= \sum_{k=0}^{\infty} k^2\frac{\mu^k}{k!}e^{-\mu} = \sum_{k=0}^{\infty}\left(k(k-1)+k\right)\frac{\mu^k}{k!}e^{-\mu} = \mu^2 e^{-\mu}\sum_{k=2}^{\infty}\frac{\mu^{k-2}}{(k-2)!} + \sum_{k=0}^{\infty}k\frac{\mu^k}{k!}e^{-\mu} = \mu^2+\mu \\
Var(X) &= E\left(X^2\right) - \left[E(X)\right]^2 = \mu
\end{aligned}$$

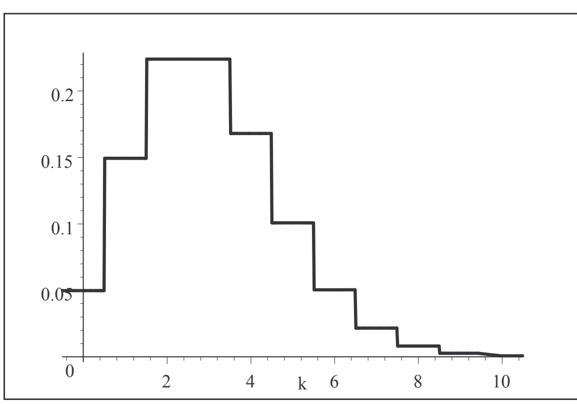
■

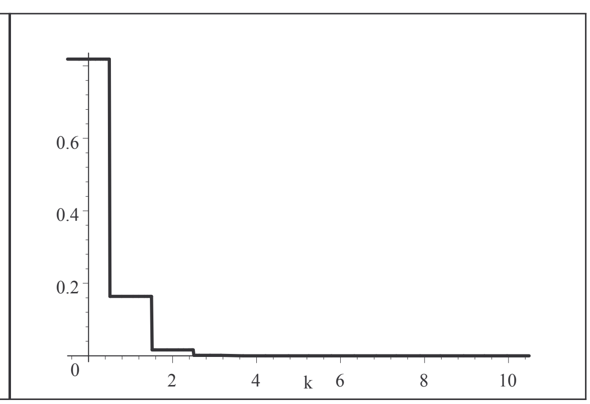
Exemplo 22 *Suponha que o número de gols numa partida segue uma distribuição de Poisson com média de 3 gols por partida. Qual a chance de haver 9 gols numa partida? Qual a chance de termos um 0×0?*

$$\begin{aligned}
\textit{No de gols} &: X \sim Poi\,(3) \\
\Pr(X=9) &= \frac{3^9}{9!}e^{-3} = 0.2701\% \\
\Pr(X=0) &= \frac{3^0}{0!}e^{-3} = e^{-3} = 4.979\%
\end{aligned}$$

Exemplo 23 *Um datilógrafo comete uma média de 0.2 erros por página. Assumindo uma distribuição de Poisson, qual a probabilidade de ele cometer 2 erros numa única página? Nenhum erro?*

$$\begin{aligned}
\textit{No de erros} &: X \sim Poi\,(0.2) \\
\Pr(X=2) &= \frac{0.2^2}{2!}e^{-0.2} = 1.637\% \\
\Pr(X=0) &= \frac{0.2^0}{0!}e^{-0.2} = e^{-0.2} = 81.87\%
\end{aligned}$$





$Poi\,(3)$ $Poi\,(0.2)$

Como a distribuição de Poisson é o limite de uma binomial onde o número de experimentos cresce (mas a probabilidade decresce mantndo a média fixa), é razoável Poisson para aproximar distribuições binomiais onde há vários experimentos com probabilidade de sucesso pequena. Revisitemos alguns exercícios da seção anterior para ilustrar esta aproximação:

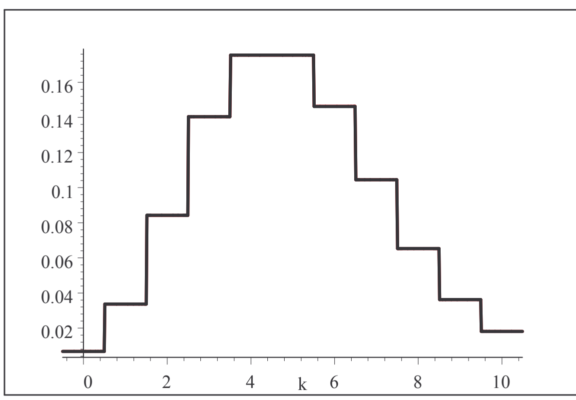
Exemplo 24 *Distribua* 10000 *folhetos aleatoriamente por* 2000 *quadras em uma cidade. Qual a chance de a sua quadra não receber folheto algum? E de receber* 5 *folhetos? E* 10*?*
Solução: seja X o número de folhetos recebidos na sua quadra. Então $X \sim Bin\left(10000, \frac{1}{2000}\right)$. Calcular a resposta exata dá bastante trabalho:

$$\begin{aligned}
\Pr(X=0) &= \text{BinomialDen}\left(0; 10000, \frac{1}{2000}\right) = \left(\frac{1999}{2000}\right)^{10000} = 0.6730\% \\
\Pr(X=5) &= \text{BinomialDen}\left(5, 10000, \frac{1}{2000}\right) = \binom{10000}{5}\left(\frac{1}{2000}\right)^{5}\left(\frac{1999}{2000}\right)^{9995} = 17.551\% \\
\Pr(X=10) &= \text{BinomialDen}\left(10, 10000, \frac{1}{2000}\right) = \binom{10000}{10}\left(\frac{1}{2000}\right)^{10}\left(\frac{1999}{2000}\right)^{9990} = 1.812\%
\end{aligned}$$

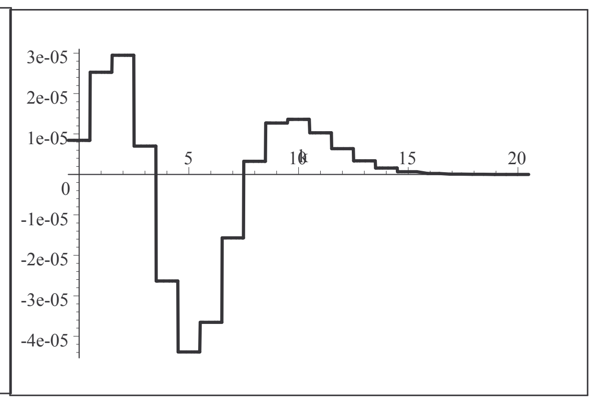
Como $n = 10000$ é grande e $p = \frac{1}{2000}$ é pequeno, talvez seja razoável trocar a binomial por uma distribuição de Poisson com parâmetro $\mu = np = 5$. De fato, para $Y \sim Poi(5)$:

$$\begin{aligned}
\Pr(Y=0) &= e^{-5}\frac{5^0}{0!} = 0.6738\% \\
\Pr(Y=5) &= e^{-5}\frac{5^5}{5!} = 17.547\% \\
\Pr(Y=10) &= e^{-5}\frac{5^{10}}{10!} = 1.813\%
\end{aligned}$$

que estão corretas até a quarta casa decimal (segunda da porcentagem)! Assim, neste caso vale a pena usar a distribuição de Poisson ao invés da binomial.



$Poi(5)$ e $BinDist\left(10000, \frac{1}{2000}\right)$: indistinguíveis



Diferença $Poi(5) - BinDist\left(10000, \frac{1}{2000}\right)$

Em geral, a distribuição de Poisson é usada sempre que tratamos de eventos “raros” mas que podem acontecer “a todo instante” – número de acidentes em uma estrada, número de casos de doença rara numa população, número de erros tipográficos, número de chamadas telefônicas recebidas numa empresa, etc.

Exemplo 25 *O número X de acidentes diários na ponte Rio-Niterói tem média* 1.8 *acidentes por dia. Supondo que X tenha distribuição de Poisson, determine a probabilidade de não haver acidentes amanhã e o número mais provável de acidentes amanhã. Se houve pelo menos um acidente num dia, qual a chance de ter havido exatamente dois acidentes neste dia?*
Solução: estamos supondo $X \sim Poi(1.8)$. Então

$$\Pr(X=0) = e^{-1.8}\frac{(1.8)^0}{0!} = e^{-1.8} = 16.53\%$$

Para a segunda pergunta, note que

$$\Pr(X = k+1) \geq \Pr(X = k) \Longleftrightarrow e^{-\mu}\frac{\mu^{k+1}}{(k+1)!} \geq e^{-\mu}\frac{\mu^k}{k!} \Leftrightarrow \mu \geq k+1$$

ou seja, a probabilidade vai aumentando enquanto tivermos $k \leq \mu - 1$, *depois começa a diminuir. Neste caso,*

$$\Pr(X = 0) \leq \Pr(X = 1) \geq \Pr(X = 2) \geq \Pr(X = 3) \geq ...$$

e, portanto, $X = 1$ *é o número mais provável, com probabilidade*

$$\Pr(X = 1) = e^{-1.8}\frac{(1.8)^1}{1!} = 29.75\%$$

(note que $\Pr(X = 2) = 26.78\%$*.apenas). Enfim, a última resposta é*

$$\Pr(X = 2 \mid X \geq 1) = \frac{\Pr(X = 2 \ e \ X \geq 1)}{\Pr(X \geq 1)} = \frac{\Pr(X = 2)}{1 - \Pr(X = 0)} = \frac{e^{-1.8}\frac{(1.8)^2}{2!}}{1 - e^{-1.8}} = 32.08\%$$

3.4.1 Exercícios

Ex. 29 *Determine a moda (isto é, o valor mais provável) de* $X \sim Poi(4)$.

Ex. 30 *Suponha os erros de digitação em um texto distribuídos segundo uma Poisson com média* 0.2 *erros por página. Determine a probabilidade de não haver erros nas* 10 *primeiras páginas.*

Ex. 31 *Uma empresa tem capacidade pra atender três clientes por dia. O número diário de clientes que chegam à empresa tem distribuição de Poisson com parâmetro* 2*. Chegando clientes em número superior à capacidade de atendimento, os excedentes são dispensados. Determine:*
a) a probabilidade de haver, em um dia, clientes não atendidos;
b) o número médio de clientes atendidos por dia;
c) para quanto deve ser ampliada a capacidade de atendimento para que a probabilidade de haver clientes não atendidos seja menor que 0.06*?*

Ex. 32 *No jogo da Sena, são sorteadas* 6 *dentre* 60 *dezenas, de modo que há* $\binom{60}{6} = 50\,063\,860$ *resultados possíveis. Cada apostador escolhe* 6 *dezenas e ganha a sena se suas* 6 *dezenas são sorteadas. Suponha que haja* 50 063 860 *apostadores que tenham escolhido suas dezenas ao acaso. Determine a probabilidade de haver* 0 *ganhadores,* 1 *ganhador e* 2 *ganhadores, respectivamente.*

Ex. 33 *Se a taxa média de mortalidade por afogamento acidental é de* 3 *por* 100000 *habitantes por ano, determine a probabilidade de que, em uma cidade de* 200000 *habitantes, se verifiquem em um ano mais de* 3 *mortes por afogamento. E menos de* 3*?*

Ex. 34 *Em uma excursão ao pantanal de Mato Grosso certa ave é avistada um número de vezes que é uma variável aleatória de Poisson com média* $\lambda = 0.8$*. Qual a chance de, numa excursão, não se avistar nenhuma daquelas aves? E de avistar mais de duas?*

Ex. 35 *Na revisão tipográfica de um livro encontrou-se em média* 1.5 *erros por página. Das* 800 *páginas do livro, estimar quantas não precisam ser modificadas por não apresentarem erros.*

Ex. 36 *Uma companhia de seguros observa que* 0.008% *das pessoas morrem em um tipo especial de acidentes a cada ano. Qual é a probabilidade da companhia ter que pagar para* 3 *ou mais vítimas das* 5000 *pessoas asseguradas contra tal tipo de acidente em um ano?*

Ex. 37 *Numa estrada há diariamente* 2 *acidentes para cada* 100 *km. Qual é a probabilidade de que em* 300 *km ocorram* 5 *acidentes num dia? E, em* 250 *km, pelo menos* 3 *acidentes?*

Ex. 38 (*) *Partículas radioativas são emitidas por uma fonte de modo que o número de partículas emitidas por segundo tem distribuição de Poisson com parâmetro* λ*. Um contador tem probabilidade* p *de detectar uma partícula emitida. Admitindo independência entre as detecções, qual é a distribuição do número de partículas detectadas por segundo?*

3.5 Distribuição Hipergeométrica

De uma caixa com r bolas "sucesso" e $N-r$ bolas "falha", extraímos sem reposição n bolas. Seja X o número de bolas de sucesso. Dizemos que X tem **distribuição hipergeométrica com parâmetros** n**,** r **e** N (isto é, $X \sim Hip\,(n, r, N)$).

Proposição 26 *Se* $X \sim Hip\,(n, r, N)$*, então*

$$\boxed{\Pr(X=k) = Hip\,(k;n,r,N) = \frac{\binom{r}{k}\binom{N-r}{n-k}}{\binom{N}{n}}}$$

$$\boxed{E(X) = np} \text{ e } \boxed{Var(X) = npq\tfrac{N-n}{N-1}}$$

onde $p = \frac{r}{N}$.

Prova. Há $\binom{N}{n}$ maneiras eqüiprováveis de escolher as n bolas das N disponíveis. Destas maneiras, quantas têm exatamente k sucessos e $n-k$ falhas? Basta escolher k dentre as r bolas sucesso, (são $\binom{r}{k}$ maneiras de faze isto) e então escolher $n-k$ dentre as $N-r$ bolas fracasso (há $\binom{N-r}{n-k}$ maneiras de escolher isto). Portanto, são $\binom{r}{k}\binom{N-r}{n-k}$ maneiras favoráveis dentre as $\binom{N}{n}$ maneiras eqüiprováveis, isto é

$$\Pr(X=k) = \frac{\binom{r}{k}\binom{N-r}{n-k}}{\binom{N}{n}}$$

Seja X_i o número de sucessos contados apenas olhando para a $i-$ésima bola retirada. Claramente, $X = X_1 + X_2 + .. + X_n$ e X_i é uma variável de Bernoulli com parâmetro $p = \frac{r}{N}$. Então

$$E(X) = E(X_1) + E(X_2) + ... + E(X_n) = np$$

No entanto, note que estas variáveis não são independentes! Assim, para calcular a variância, fazemos

$$Var(X) = Var(X_1) + Var(X_2) + ... + Var(X_n) + 2\sum_{i\neq j} Cov(X_i, X_j) = npq + 2\frac{n(n-1)}{2}V$$

onde $V = Cov(X_i, X_j)$ deve ser o mesmo para quaisquer duas variáveis distintas X_i e X_j Como calcular V? Ora, para $n = N$ devemos ter $Var(X) = 0$, já que neste caso todas as bolas são retiradas e então certamente $X = r$. Assim

$$0 = Npq + N(N-1)V \Rightarrow V = \frac{-pq}{N-1}$$

Portanto

$$Var(X) = npq - n(n-1)\frac{pq}{N-1} = npq\left(1 - \frac{n-1}{N-1}\right) = npq\frac{N-n}{N-1}$$

O fator $\frac{N-n}{N-1}$ é chamado de "fator de correção para populações finitas". Note como ele se aproxima de 1 à medida que $N \to \infty$ (com n fixo). ■

Exemplo 27 *Há* 6 *times paulistas num campeonato brasileiro com* 20 *clubes. Escolha* 4 *ao acaso (digamos, para rebaixamento). Qual a chance de nenhum ser paulista? E* 1 *ser paulista?* 2*?* 3*? Todos os* 4*?*

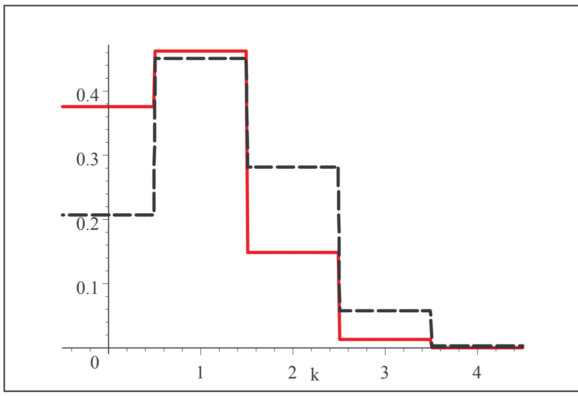
$$\begin{aligned}
&Paulistas &:& \quad X \sim Hip\,(4,6,20) \qquad Pr\,(X=0) = \frac{\binom{6}{0}\binom{14}{4}}{\binom{20}{4}} = \frac{1001}{4845} = 20.66\% \\
&\Pr(X=1) &=& \quad \frac{\binom{6}{1}\binom{14}{3}}{\binom{20}{4}} = 45.08\% \qquad \Pr(X=2) = \frac{\binom{6}{2}\binom{14}{2}}{\binom{20}{4}} = 28.17\% \\
&\Pr(X=3) &=& \quad \frac{\binom{6}{3}\binom{14}{1}}{\binom{20}{4}} = 5.78\% \qquad \Pr(X=4) = \frac{\binom{6}{4}\binom{14}{0}}{\binom{20}{4}} = \frac{1}{323} = 0.310\%
\end{aligned}$$

Exemplo 28 *E para os 4 cariocas?*

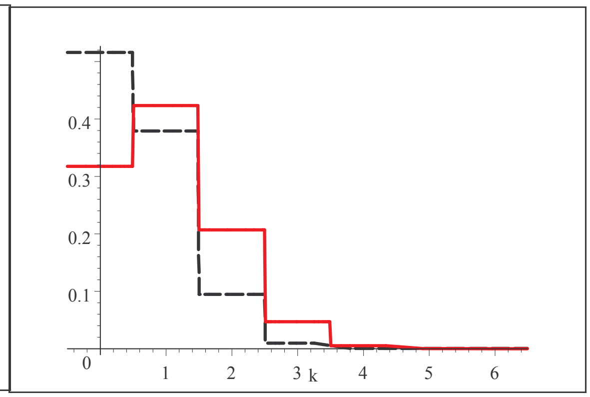
$$Cariocas \quad : \quad X \sim Hip\,(4,4,20) \qquad Pr\,(X=0) = \frac{\binom{4}{0}\binom{16}{4}}{\binom{20}{4}} = \frac{364}{969} = 37.56\%$$

$$\Pr\,(X=1) = \frac{\binom{4}{1}\binom{16}{3}}{\binom{20}{4}} = 46.23\% \qquad \Pr\,(X=2) = \frac{\binom{4}{2}\binom{16}{2}}{\binom{20}{4}} = 14.86\%$$

$$\Pr\,(X=3) = \frac{\binom{4}{3}\binom{16}{1}}{\binom{20}{4}} = 1.32\% \qquad \Pr\,(X=4) = \frac{\binom{4}{4}\binom{16}{0}}{\binom{20}{4}} = 0.0206\%$$



$Hip\,(k;4,6,20)$ e $Hip\,(k;4,4,20)$



Mega-sena: $Hip\,(k;6,6,60)$ e $Hip\,(k;6,10,60)$

Exemplo 29 *No jogo da sena são sorteadas 6 dentre 60 dezenas.Apostamos em 10 dezenas. Qual a probabilidade de acertarmos 4 dezenas? Qual o número esperado de testes em que devemos apostar para, pela primeira vez, acertar 4 dezenas?*
Solução: As 60 dezenas estão divididas em 10 apostadas (sucessos) e 50 não apostadas (falhas). Serão acertadas 4 dezenas se sortearem 4 das dezenas apostadas e 2 das não apostadas. Seja X o número de dezenas sorteadas de sucesso (isto é, que estão no meu jogo). Então $X \sim Hip\,(6,10,60)$. Portanto

$$\Pr\,(X=4) = \frac{\binom{10}{4}\binom{50}{2}}{\binom{60}{6}} = 0.5138\%$$

Seja Y o número de tentativas até obter uma quadra. Então $Y \sim Geom\,(0.5138\%)$ isto é

$$E\,(Y) = \frac{1}{0.005138} = \frac{\binom{60}{6}}{\binom{10}{4}\binom{50}{2}} = 194.61$$

Nota 30 *Note que os papéis de r e n são intercambiáveis. De fato,*

$$Hip\,(k;n,r,N) = \frac{n!r!\,(N-n)!\,(N-r)!}{k!\,(r-k)!\,(n-k)!N!}$$

é uma expressão que permanece invariante se trocarmos as posições de r e n. Em suma

$$Hip\,(k;n,r,N) = Hip\,(k;r,n,N)\,.$$

Também note que, se $n << N$, então $Hip\,(n,r,N) \cong Bin\,(n,p)$[8]*.*

3.5.1 Exercícios

Ex. 39 *Uma urna contém 5 bolas brancas e 3 bolas pretas. Retiram-se, sem reposição, duas bolas dessa urna. Seja X o número de bolas brancas sacadas. Determine a distribuição de X, $E\,(X)$ e $Var\,(X)$.*

[8] E os dois comentários juntos mostram também que, se $r << N$, então $Hip\,(n,r,N) \simeq Bin\,\left(r, \frac{n}{N}\right)$.

Ex. 40 *No jogo da mega-sena são sorteadas* 6 *dentre* 60 *dezenas. No ano de* 2005, *os prêmios médios foram de cerca de* $R\$18.000.000,00$ *para a sena,* $R\$18.000,00$ *para a quina e* $R\$240,00$ *para a quadra. Um volante básico com* 6 *dezenas custa* $R\$1,50$*; um volante com* 10 *dezenas custa* $\binom{10}{6}$ *vezes mais (isto é,* $\$315,00$*) e um volante com* 15 *dezenas custa* $\binom{15}{6}R\$1,50 = R\$7507,50$. *Usando os valores médios acima para os prêmios, descubra o valor esperado do lucro em cada uma destas apostas. Qual deveria ser o valor do prêmio da mega-sena para que a aposta básica começasse a valer a pena?*[9]

Ex. 41 *Uma lotérica recebe, para revender,* 1000 *bilhetes de uma loteria à qual concorrem* 100000 *bilhetes e são premiados* 10000 *bilhetes. Seja* X *o número de bilhetes premiados revendidos pela lotérica. Determine* $E(X)$ *e* $\sigma(X)$.

Ex. 42 *Em uma urna há* 10 *bolas pretas e* 5 *bolas brancas. Sacam-se, sem reposição,* 6 *bolas dessa urna. Determine o valor mais provável do número de bolas pretas sacadas. [Dica: encontre para que valores de* k *temos* $p(k+1) \geq p(k)$*].*

3.6 Exercícios de Provas

Ex. 43 (A1 2003.1) *Um sistema é formado por* 5 *componentes independentes e funciona se pelo menos três componentes funcionam. A probabilidade de falha de um componente é* 0.1. *Qual é a probabilidade de falha do sistema?*

Ex. 44 (A1 2003.1) *Suponha* X *com distribuição binomial negativa com parâmetros* $r = 6$ *e* $p = 0.8$. *Determine a moda de* X *e calcule* $\Pr(X < 10)$.

Ex. 45 (A1 2003.1) *Uma urna contém* b *bolas brancas e* p *bolas pretas. Sacam-se, sucessivamente e sem reposição, bolas dessa urna até que todas as bolas pretas tenham sido sacadas. Seja* X *o número de bolas sacadas. Determine a função de probabilidade de* X.

Ex. 46 (A1 2004.2) *Em um jogo de dados, o participante paga* $R\$25,00$ *para jogar os dados pela primeira vez e* $R\$15,00$ *a partir da segunda rodada. Ele joga até tirar um 1 ou um 6. Quando isto finalmente acontece, ele recebe* $R\$50,00$.
a) Qual é a probabilidade de que o jogador tenha prejuízo neste jogo?
b) Qual é o seu lucro (ou prejuízo) esperado?

Ex. 47 (A1 2004.2) *A ocorrência de pedidos de conserto em uma empresa de geração de energia é modelada como um processo de Poisson de taxa* 0.2 *por dia (isto significa que o número de pedidos em um período de* d *dias tem distribuição de Poisson com média* $0.2d$ *e que, além disso, as quantidades de acidentes em períodos disjuntos são independentes). O serviço funciona* 24 *horas por dia,* 7 *dias por semana.*
a) Qual é o número médio de pedidos de conserto em um período de um mês?
b) Qual é a probabilidade de que não haja pedidos de assistência técnica em uma dada semana?
c) Em uma semana, qual é o número médio de dias em que há pedidos de conserto?

Ex. 48 (AS 2004.2) *Vários lançamentos de um satélite serão tentados até que o primeiro deles tenha sucesso. Cada lançamento custa* $\$12$ *milhões, mas o benefício de colocar o satélite em órbita é estimado em* $\$75$ *milhões. Cada lançamento tem probabilidade de sucesso* p *e lançamentos distintos são considerados completamente independentes.*
a) Se $p = 0.3$, *qual o lucro esperado desta série de lançamentos? Qual a variância deste lucro?*
b) Qual o valor mínimo de p *para que está série de lançamentos tenha lucro esperado (positivo)?*

Ex. 49 (T2 2005.2) *Cada item abaixo define uma variável aleatória. Diga que distribuição discreta você usaria para cada variável, incluindo os parâmetros correspondentes.*
EXEMPLO: O número de caras obtidas em 90 *lançamentos independentes de uma moeda justa.*
Resposta: $X \sim Bin(90, 0.5)$ *onde* $n = 90$ *e* $p = 0.5$.

[9] Apenas uma vez na história até Março de 2006 a mega-sena pagou mais de $R\$52$ milhões; foi em 10 de Outubro de 1999, quando, após acumular 9 vezes seguidas, a mega-sena pagou $R\$64.905.517,65$ (já descontado o imposto de renda) a um grupo de juízes e advogados que haviam apostado na “Orixás Loterias” em Salvador. Veja http://www.caixa.gov.br/Loterias/.

a) O número de vezes que um aluno faz o vestibular da única faculdade que lhe interessa (a FGV), pressupondo que a cada vez ele tenha os mesmos 40% *de chance de passar.*
b) O número de clientes que procura Sherlock Holmes em um dia (estima-se que Sherlock receba, em média, 20 *clientes por ano).*
c) O número de questões que um aluno acerta num vestibular de 80 *questões com* 5 *opções cada, onde o aluno tenta adivinhar aleatoriamente* 70 *questões e deixa as outras* 10 *em branco.*
d) O número de dezenas que eu acertei no concurso passado da Sena (meu jogo tinha 10 *dezenas; a Sena escolhe* 6 *dezenas dentre* 60 *possíveis, sem repetição).*
e) O número de estudantes (dos 2141 *estudantes da FGV) que fazem aniversário hoje.*
f) O número de times cariocas a serem rebaixados no Campeonato Brasileiro, pressupondo que todos os times estejam em igualdade de condições (o campeonato tem 22 *times, dos quais* 4 *serão rebaixados; os cariocas são Fluminense, Botafogo, Flamengo e Vasco).*

Ex. 50 (A1 2005.2) *Acidentes ocorrem na ponte Rio-Niterói a uma taxa média de* 18 *acidentes por mês (*30 *dias). Um dia é* bom *quando não há acidente na ponte. Qual o valor esperado do número de dias* bons *por mês? [Dica: qual a probabilidade de um determinado dia não ter acidente algum?]*

Ex. 51 (A1 2005.2) *Na roleta em Monte Carlo, há números de* 0 *a* 36 *(a roleta é justa, então estes* 37 *números são equiprováveis). Você decide apostar seguidamente no número* 13 *somente até obter a primeira vitória. A cada derrota você perde* \$10*, mas uma vitória equivale a um lucro de* \$350*.*
a) Calcule o valor esperado e a variância do seu lucro (ou perda) nesta seqüência de apostas.
b) Calcule a probabilidade de "sair lucrando" desta seqüência de apostas (isto é, de ter lucro positivo).

Ex. 52 (AS 2005.2) *A cada dia de um período de* 90 *dias, uma determinada ação pode desvalorizar* \$4 *com probabilidade* 30%*, ou subir* \$2 *com probabilidade* 70% *(suponha que cada dia é independente dos demais). Seja* Z *o número de dias em que a ação subiu.*
a) Qual é a distribuição de probabilidade de Z*?*
b) Você compra a ação no início do período e a vende ao final dos 90 *dias. Qual o valor esperado do seu lucro?*
c) Estime a probabilidade do seu lucro ser de pelo menos \$50 *no período de* 90 *dias (juntando as valorizações e desvalorizações de cada dia).*

Ex. 53 (AS 2005.2) *Você trabalha no setor de Atendimento ao Consumidor que recebe* X *reclamações por dia, onde* X *tem distribuição de Poisson com parâmetro* $E(X) = \lambda = 7.1$ *reclamações por dia. Seu chefe disse que vai estourar aquela champagne importada no dia em que seu setor receber menos de* 2 *reclamações.*
a) Qual a probabilidade disto acontecer hoje?
b) Qual o valor esperado do número de dias que você tem de esperar para saborear a champagne?
c) Qual a probabilidade da comemoração ocorrer nos próximos 90 *dias?*

Ex. 54 (T2 2006.1) *Você vai prestar um exame oral onde seu professor lhe apresentará uma questão de cada vez. A cada questão, ele espera* 10 *minutos e apenas então você diz a sua resposta. Assim que você acertar uma questão, o teste acaba e você passa no curso. Se você não acertar questão alguma num total de* 60 *minutos (ou seja,* 6 *questões), o teste acaba e você está reprovado. Suponha que a probabilidade de você acertar cada questão é* 20%*, e que acertar cada questão é completamente independente de acertar as outras.*
a) Qual a probabilidade de você passar no curso?
b) Suponha que 60 *alunos com estas mesmas características fazem este exame. Supondo que a aprovação de cada aluno é independente da aprovação dos demais, calcule o valor esperado e o desvio-padrão do número de alunos aprovados.*
c) Na situação do item anterior, complete a seguinte frase com dois números: "a desigualdade de Chebyshev garante que a probabilidade do número de aprovados estar entre _______ e _______ é pelo menos $\frac{8}{9}$*".*
d) Seja Z *a duração (em minutos) do seu exame oral. Calcule a mediana e o valor esperado de* Z*.*

Chapter 4

Variáveis Aleatórias Contínuas

4.1 Distribuições Contínuas

4.1.1 Função de Distribuição Acumulada

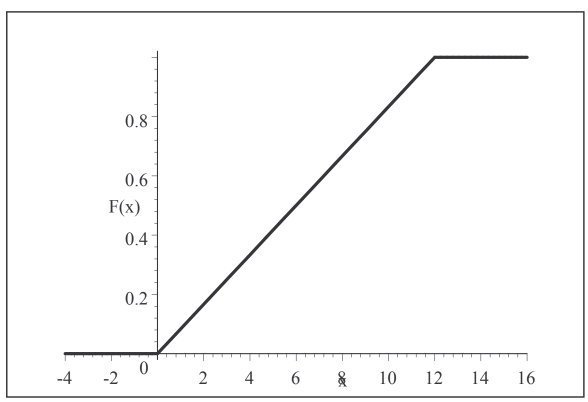
Gostaríamos de trabalhar com variáveis aleatórias contínuas, isto é, variáveis cujo espaço amostral é um "continuum" de valores (por exemplo, o instante exato em que uma ligação telefônica chega ou em que um gol acontece, o local de uma estrada onde um acidente ocorre). Para analisar uma variável aleatória contínua, precisamos definir a função de distribuição acumulada desta variável, de forma idêntica ao caso discreto:

Definição 1 *A* ***função de distribuição acumulada*** *(****função de distribuição; fda****) de uma variável aleatória* X *é*

$$\boxed{F_X(x) = \Pr(X \leq x)}$$

Exemplo 2 *Imagine uma roleta graduada de 0 a 12 como um relógio e um ponteiro. Dê um "peteleco" no ponteiro e deixe-o girar até parar numa posição aleatória (representada por um número* ***real*** X *de 0 a 12). Se a roleta é "justa", é razoável supor que não há "preferência" para pontos de um tipo ou de outro*[1]*. Que probabilidade você acha razoável designar para* $\Pr(X \leq 6) = F(6)$ *(isto é, qual a chance do ponteiro parar do lado direito do relógio)? Se não há preferência entre o lado direito e o esquerdo do relógio, é razoável supor que* $F(6) = \Pr(X \leq 6) = 50\%$*. Analogamente, como não há preferências entre intervalos (de mesmo tamanho), é razoável supor que* $F(7) = \Pr(X \leq 7) = \frac{7}{12}$*. Com efeito, é razoável supor que a f.d.a. desta variável* X *é*

$$F(x) = \Pr(X \leq x) = \begin{cases} 0, \text{ se } x < 0 \\ \frac{x}{12}, \text{ se } 0 \leq x \leq 12 \\ 1, \text{ se } x > 12 \end{cases}$$



A partir desta função, é fácil calcular a probabilidade de X *estar num intervalo qualquer* $(a, b]$*. No nosso exemplo, para* $a, b \in [0, 12]$*, teríamos*

$$\Pr(a < X \leq b) = F(b) - F(a) = \frac{b}{12} - \frac{a}{12} = \frac{b-a}{12}$$

[1] Note que, neste ponto, não sabemos exatamente o que significam "aleatório" ou "preferência". Por enquanto, use a sua intuição – mais tarde seremos capazes de dar uma definição mais precisa destas idéias, usando a linguagem que estamos desenvolvendo.

ou seja, a probabilidade de X *estar num intervalo de tamanho* Δx *contido em* $[0, 12]$ *é simplesmente* $\Delta x/12$*. Este modelo é chamado de* ***distribuição uniforme*** *no intervalo* $[0, 12]$*.*
Qual a probabilidade de obter $X = a$*? Estranhamente, temos* $\Pr(X = a) \leq \Pr(X \in [a, a + \varepsilon]) = \frac{\varepsilon}{12}$ *para qualquer* ε*. Assim, a única probabilidade razoável seria*

$$\Pr(X = a) = 0$$

Repetimos: neste modelo contínuo, a probabilidade de obtermos $X = a$ *é ZERO (qualquer que seja* a*)! O leitor atento perguntará – mas escolheremos um certo número* x*, não? Qual era a probabilidade de este número* x *ser escolhido? Respondemos: antes de realizar o experimento,* $\Pr(X = x) = 0$*. Não há contradição aqui – simplesmente, probabilidade ZERO não significa IMPOSSÍVEL! Pense em probabilidade ZERO como um evento muito muito raro; se repetirmos nosso experimento* N *vezes e colocarmos* n *como o número de vezes em que* $X = x$*, teríamos* $n/N \to 0$ *quando* $N \to \infty$ *(mesmo que* $n \neq 0$*).*

Repetimos: probabilidade ZERO não implica IMPOSSÍVEL[2]. Analogamente, $\Pr(X \neq x) = 1$, mas o evento $X \neq x$ não acontece necessariamente SEMPRE.

A propriedade a seguir é completamente análoga ao caso discreto – pense no seu significado:

Proposição 3 *Se* $F(x)$ *é a f.d.a de uma variável aleatória real* X*, então*

$$\boxed{F \text{ é não-decrescente}}$$

$$\boxed{F(-\infty) = \Pr(X \leq -\infty) = 0 \text{ e } F(+\infty) = \Pr(X \in \mathbb{R}) = 1}$$

$$\boxed{\Pr(a \leq X \leq b) = F(b) - F(a)}$$

Note que no caso contínuo $\Pr(a < X \leq b) = \Pr(a \leq X \leq b)$, pois $\Pr(X = a) = 0$.

4.1.2 Quantis

Analogamente ao caso discreto, temos:

Definição 4 *O* $q-$*quantil de uma variável aleatória contínua* X *é qualquer valor* x_q *onde a função acumulada "acerta"* q*. Formalmente*

$$\boxed{F(x_q) = q}$$

De fato, esta definição é mais simples do que a do caso discreto, pois agora a f.d.a não dá "saltos". Novamente, é comum chamar $x_{0.25}$, $x_{0.5}$ e $x_{0.75}$ de **primeiro, segundo e terceiro quartis** da variável aleatória X. O segundo quartil também é comumente chamado de **mediana**.

Exemplo 5 *Se* X *tem distribuição uniforme em* $[0, 12]$*, seus quartis são:*

$$\begin{aligned} F(x) &= 0.25 \Longleftrightarrow \frac{x}{12} = 0.25 \Longleftrightarrow x = 3 \\ F(x) &= 0.5 \Longleftrightarrow \frac{x}{12} = 0.5 \Longleftrightarrow x = 6 \\ F(x) &= 0.75 \Longleftrightarrow \frac{x}{12} = 0.75 \Longleftrightarrow x = 9 \end{aligned}$$

ou seja, os quartis da distribuição uniforme em $[0, 12]$ *são* 3*,* 6 *e* 9 *respectivamente.*

[2] Ainda vale que $\Pr(\emptyset) = 0$, isto é, IMPOSSÍVEL implica PROBABILIDADE ZERO. Mas a volta não vale, nunca valeu, nem no caso discreto! Por exemplo, A e B são eventos mutuamente excludentes se, e somente se, $A \cap B = \emptyset$, e isto implica $\Pr(A \cap B) = 0$. No entanto, podemos ter $\Pr(A \cap B) = 0$ sem ter $A \cap B = \emptyset$!

4.1.3 Função Densidade de Probabilidade

Se dividirmos a probabilidade de X estar num intervalo $[x, x+\Delta x]$ pelo tamanho do intervalo, temos a densidade média de probabilidade[3]

$$\frac{\Pr(x \leq X \leq x+\Delta x)}{\Delta x} = \frac{F(x+\Delta x) - F(x)}{\Delta x}$$

Para calcular a densidade de probabilidade **em um ponto**, devemos tomar $\Delta x \to 0$. Então[4]:

Definição 6 *A* ***função densidade de probabilidade (fdp)*** *de X é a derivada da função acumulada*[5]*:*

$$\boxed{f_X(x) = \lim_{\Delta x \to 0} \frac{F(x+\Delta x)-F(x)}{\Delta x} = F'_X(x)}$$

Exemplo 7 *Tomávamos X de maneira uniforme em $[0,12]$. Se $a, b \in [0,12]$, tínhamos:*

$$\frac{\Pr(a \leq X \leq b)}{\Delta x} = \frac{b-a}{12(b-a)} = \frac{1}{12}$$

então a densidade de probabilidade é

$$f_X(x) = F'(x) = \lim_{\Delta x \to 0} \frac{\Pr(x \leq X \leq x+\Delta x)}{\Delta x} = \frac{1}{12} \text{ (para } 0 < x < 12\text{)}$$

Para ser completo, temos

$$f_X(x) = \begin{cases} \frac{1}{12}, \text{ para } 0 < x < 12 \\ 0, \text{ caso contrário} \end{cases}$$

Pelo Teorema Fundamental do Cálculo, é imediato notar que:

Proposição 8 *Dada a f.d.p de uma variável aleatória contínua X, encontramos probabilidades pela fórmula*

$$\boxed{\Pr(a \leq X \leq b) = F(b) - F(a) = \int_a^b f(t)\, dt}$$

Em particular, como $F(-\infty) = 0$, note que

$$\boxed{F(x) = \Pr(X \leq x) = \int_{-\infty}^x f(t)\, dt}$$

Repetimos: a f.d.p. NÃO mostra probabilidades – *sua INTEGRAL DEFINIDA (isto é, a ÁREA sob o gráfico da f.d.p.) é que é uma probabilidade*[6]. Intuitivamente, a f.d.p. é maior nos pontos onde há mais "probabilidade" – mas este paralelo não funciona bem pois, em cada ponto, a probabilidade é ZERO. Talvez seja melhor pensar que $\Pr(X = x)$ seja "$f(x)\,dx$" (ao invés de $f(x)$).

A propriedade a seguir é novamente análoga à do caso discreto; pense no seu significado.

---

[3] A densidade linear média **de massa** de um arame é a massa total deste arame dividida pelo seu comprimento total, com unidade Kg/m. Densidade média de **carga** elétrica seria carga total sobre comprimento total, com unidade C/m. Como probabilidade não tem unidade, a unidade de densidade de probabilidade é $1/$(unidade de X).

[4] Se o arame da nota acima não fosse homogêneo, falaríamos na densidade linear de massa **em um ponto,** a saber, $\lim_{\Delta L \to 0} \frac{\Delta M}{\Delta L}$, onde ΔM é a massa e ΔL é o comprimento de um pedacinho de arame em volta daquele ponto.

[5] Pelo menos onde esta derivada existir. Nos pontos isolados onde $F'(x)$ não existe, você pode definir $f(x)$ da maneira que quiser, pois isto não afetará em nada o modelo probabilístico.

[6] Uma interpretação da f.d.p. vem da aproximação linear:

$$\Pr(x \leq X \leq x+\Delta x) = F(x+\Delta x) - F(x) \approx f(x).\Delta x$$

isto é, "*a quantidade de probabilidade no intervalo $[x, x+\Delta x]$ é proporcional a $f(x)$ e Δx para Δx pequeno*".

Se você preferir, como $f(x) = F'(x)$, temos que "*a f.d.p. diz quanto a f.d.a. está crescendo no ponto a*". Compare com o caso discreto – naquele caso, os "saltos" da f.d.a. eram exatamente os valores da função de probabilidade. Neste sentido, a f.d.p no caso contínuo é *parecida* com a função de probabilidade do caso discreto.

Outro paralelo: faça um gráfico de barras para a função de probabilidade de uma variável discreta X, onde a *área* (ao invés da altura) da barra sobre $X = k$ é a probabilidade $\Pr(X = k)$. À medida que o espaço amostral de X aumenta para incluir mais e mais valores possíveis de k, este gráfico de barras se aproximará da função densidade $f(x)$ (da mesma forma que as "somas de Riemann" de uma função se aproximam de sua integral definida)!

Proposição 9 *Se $f(x)$ é a f.d.p. de uma variável aleatória real X, então para todo x real:*

$$\boxed{0 \leq f(x)}$$

$$\boxed{\int_{-\infty}^{\infty} f(t)\, dt = 1}$$

Definição 10 *A **moda** de uma variável aleatória é o valor x onde a densidade $f(x)$ é máxima.*

Exemplo 11 *Dizemos que a variável contínua X tem **distribuição de Cauchy** quando sua f.d.p. é dada por*

$$f(x) = c\frac{1}{1+x^2}$$

Calcule o valor da constante de c e a f.d.a. de X. Calcule então $\Pr(0 \leq X \leq 1)$, os quartis e a moda de X.
Solução: devemos ter

$$\int_{-\infty}^{\infty} \frac{c}{1+x^2} dx = 1 \Rightarrow c(\arctan\infty - \arctan(-\infty)) = 1 \Rightarrow c = \frac{1}{\pi}$$

pois

$$\int \frac{1}{1+x^2} dx = \arctan x + C \qquad e \qquad \arctan(\pm\infty) = \pm\frac{\pi}{2}$$

A f.d.a. será

$$F(x) = \int_{-\infty}^{x} \frac{1}{\pi}\frac{1}{1+t^2} dt = \frac{1}{\pi}(\arctan t)_{-\infty}^{x} = \frac{1}{\pi}\left(\arctan x + \frac{\pi}{2}\right) = \frac{\arctan x}{\pi} + \frac{1}{2}$$

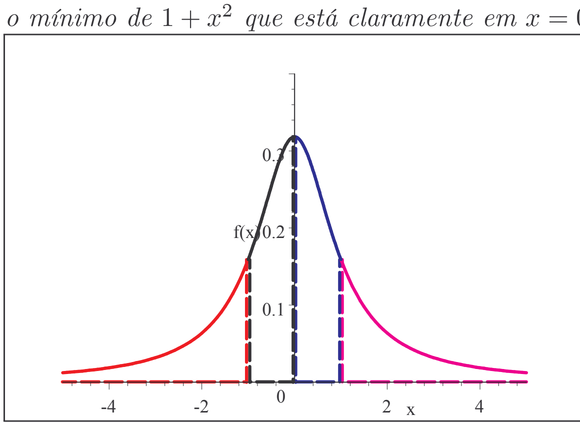
Assim,

$$\Pr(0 \leq X \leq 1) = F(1) - F(0) = \frac{\arctan 1 - \arctan 0}{\pi} = \frac{\frac{\pi}{4} - 0}{\pi} = \frac{1}{4} = 25\%$$

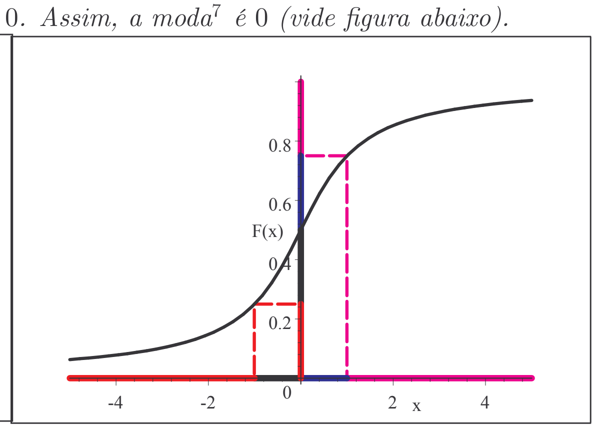
Os quartis vêm de

$$\begin{aligned}
F(x) &= \frac{1}{4} \Leftrightarrow \frac{\arctan x}{\pi} = \frac{-1}{4} \Leftrightarrow x = \tan\left(-\frac{\pi}{4}\right) = -1 \\
F(x) &= \frac{1}{2} \Leftrightarrow \frac{\arctan x}{\pi} = 0 \Leftrightarrow x = \tan(0) = 0 \\
F(x) &= \frac{3}{4} \Leftrightarrow \frac{\arctan x}{\pi} = \frac{1}{4} \Leftrightarrow x = \tan\left(\frac{\pi}{4}\right) = 1
\end{aligned}$$

ou seja, os quartis são -1, 0 e 1. Enfim, para encontrar a moda precisamos encontrar o máximo de $\frac{1}{1+x^2}$, isto é, o mínimo de $1+x^2$ que está claramente em $x=0$. Assim, a moda[7] *é 0 (vide figura abaixo).*



F.d.p da distribuição de Cauchy



F.d.a. da distribuição de Cauchy

Note como a probabilidade do intervalo $[0,1]$ (isto é, $\Pr(0 \leq X \leq 1)$) pode ser vista como a área sob o gráfico da f.d.p entre $x=0$ e $x=1$ (região em azul no gráfico da esquerda) ou como o tamanho da imagem do intervalo $[0,1]$ por F, isto é, $F(1) - F(0)$ (segmento vertical em azul no gráfico da direita).

[7] Você pode derivar $f(x) = \frac{1}{1+x^2}$ e igualar a 0, mas neste caso nosso argumento evita contas chatas.

4.1.4 Funções de Variáveis Aleatórias Contínuas

Seja X uma variável aleatória contínua cujas f.d.p e f.d.a. são $f(x)$ e $F(x)$, respectivamente. Seja $Y = h(X)$. Como encontrar as f.d.p. $g(y)$ e f.d.a. $G(y)$ desta nova variável?

A solução é procurar construir eventos equivalentes usando X e Y. Afinal:

$$Y \leq y \Longleftrightarrow h(X) \leq y$$

Resolvendo a desigualdade $h(X) \leq y$ encontramos $X \in I$ para algum subconjunto I da reta real. Então

$$G(y) = \Pr(Y \leq y) = \Pr(X \in I)$$

e esta probabilidade pode ser calculada a partir de $F(x)$ (a f.d.a. de X). Tendo $G(y)$, é fácil calcular $g(y) = G'(y)$.

Exemplo 12 *Suponha que X tem distribuição uniforme em $[0,1]$. Qual é a f.d.p. da variável de $Y = \sqrt{X}$? Solução: seja $G(y)$ a f.d.a. de Y. Como é certo que $0 \leq Y \leq 1$, então:*

$$\begin{aligned} G(y) &= 0 \text{ para } y < 0 \\ G(y) &= 1 \text{ para } y > 1 \end{aligned}$$

Agora, para $0 < y < 1$, temos

$$Y \leq y \Leftrightarrow \sqrt{X} \leq y \Leftrightarrow X \leq y^2$$

Como a distribuição de X é $F_X(x) = x$ para $0 < x < 1$:

$$G(y) = \Pr(Y \leq y) = \Pr\left(X \leq y^2\right) = F_X\left(y^2\right) = y^2$$

Derivando esta expressão, encontramos a f.d.p. de Y

$$g(y) = 2y \text{ para } 0 \leq y \leq 1$$

Juntando tudo

$$g(y) = \begin{cases} 2y \text{ se } 0 \leq y \leq 1 \\ 0 \text{ caso contrário} \end{cases}$$

Intuitivamente, isto significa que, se tomarmos números aleatórios entre 0 e 1 de maneira uniforme e então extrairmos suas raízes quadradas, estes novos números não se espalham uniformemente *entre 0 e 1, mas "se concentram mais" perto de 1 (onde está a moda de Y).*

Proposição 13 *Seja X uma variável aleatória de densidade $f(x)$ e seja $Y = h(X)$ onde h é uma função crescente. Então a densidade $g(y)$ da variável Y satisfaz*

$$\boxed{g(y) = \frac{f(x)}{h'(x)} = \frac{f(x)}{\frac{dy}{dx}}}$$

onde $x = h^{-1}(y)$, ou seja,

$$\boxed{g(y)\,dy = f(x)\,dx}$$

Prova. Seja $y = h(x)$. Então

$$G(y) = \Pr(Y \leq y) = \Pr(h(X) \leq h(x))$$

Como h é crescente

$$\Pr(h(X) \leq h(x)) = \Pr(X \leq x) = F(x)$$

ou seja

$$G(h(x)) = F(x)$$

Derivando com relação a x (e usando a regra da cadeia):

$$h'(x)\,G'(h(x)) = F'(x) \Rightarrow g(h(x)) = \frac{f(x)}{h'(x)} \Rightarrow g(y) = \frac{f(x)}{h'(x)}$$

■

Note que, se h é decrescente, vale a mesma fórmula trocando $h'(x)$ por $|h'(x)| = -h'(x)$.

Exemplo 14 *Voltemos ao exemplo anterior: X é uniforme em $[0,1]$ e $Y = h(X) = \sqrt{X}$; podemos usar a fórmula acima pois h é crescente no intervalo $[0,1]$. Então:*

$$g(y) = \frac{f(x)}{h'(x)} = \frac{f(x)}{\frac{1}{2\sqrt{x}}} = 2\sqrt{x} f(x)$$

onde $y = h(x) = \sqrt{x}$. Substituindo x e usando que $f(x) = 1$ para $0 \le x \le 1$ e $f(x) = 0$ caso contrário, temos

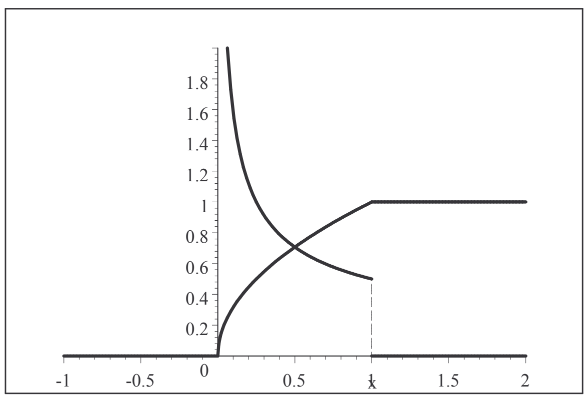
$$g(y) = \begin{cases} 2y \text{ se } 0 \le y \le 1 \\ 0 \text{ caso contrário} \end{cases}$$

Exemplo 15 *Seja X uniforme em $[-1,1]$ e sejam $Y = X^2$ e $W = |X|$. Encontre as densidades de Y e W. Em primeiro lugar, não podemos usar nossa fórmula pois as funções acima não são monótonas. Por outro lado, se $0 \le y \le 1$, temos*

$$\Pr(Y \le y) = \Pr\left(X^2 \le y\right) = \Pr(-\sqrt{y} \le X \le \sqrt{y}) = \int_{-\sqrt{y}}^{\sqrt{y}} \frac{1}{2} dx = \sqrt{y} \Rightarrow f_Y(y) = \frac{1}{2\sqrt{y}}$$

Para $y < 0$ é claro que $\Pr(Y \le y) = 0$ e, para $y > 1$, temos $\Pr(Y \le y) = 1$.
Para $0 \le w \le 1$, temos

$$\Pr(W \le w) = \Pr(|X| \le w) = \int_{-w}^{w} \frac{1}{2} dx = w \Rightarrow f_W(w) = 1 \Rightarrow W \text{ é uniforme em } [0,1]$$



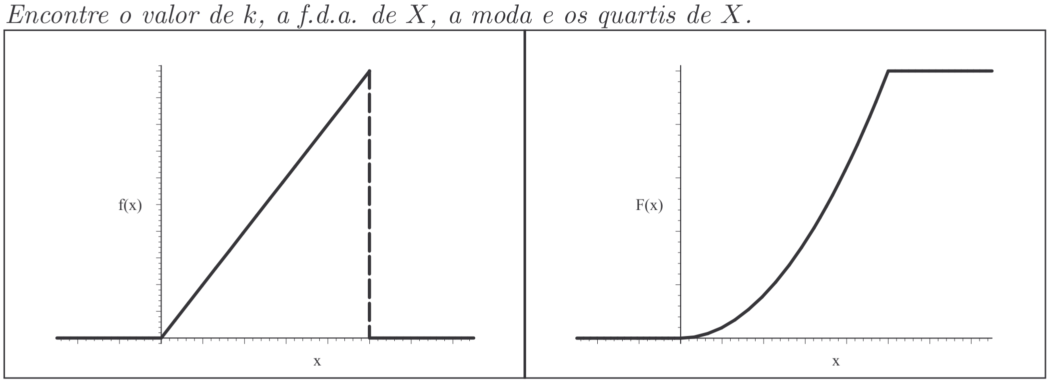
F.d.p e f.d.a de Y

4.1.5 Exercícios ilustrados

Ex. 1 *A densidade de uma variável aleatória X é*

$$f(x) = \begin{cases} kx, \text{ se } x \in [0,1] \\ 0, \text{ caso contrário} \end{cases}$$

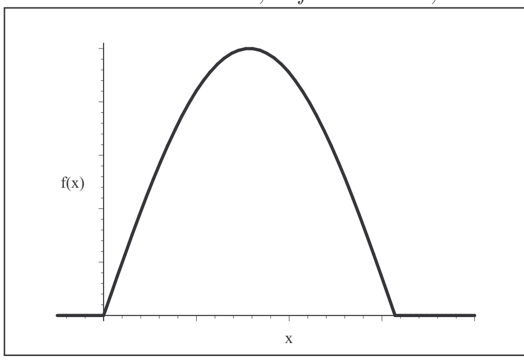
Encontre o valor de k, a f.d.a. de X, a moda e os quartis de X.

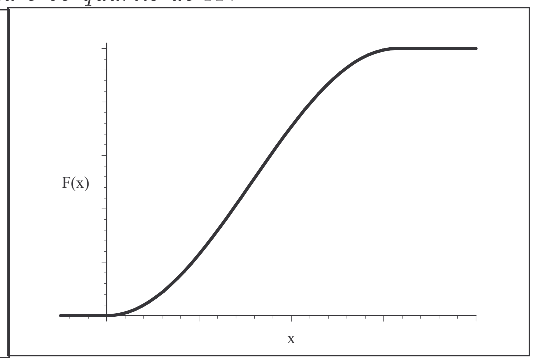


Ex. 2 *A densidade de uma variável aleatória X é*

$$f(x) = \begin{cases} k \sin x, & se\ x \in [0, \pi] \\ 0, & caso\ contrário \end{cases}$$

Encontre o valor de k, a f.d.a. de X, a moda e os quartis de X.





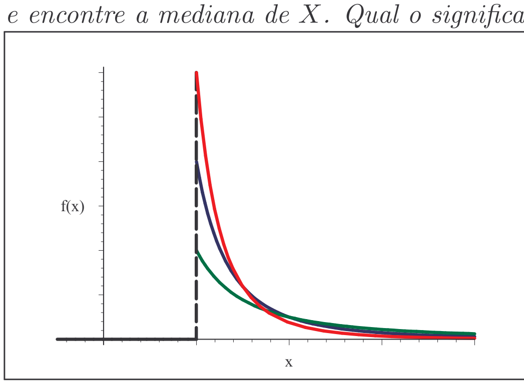
Ex. 3 *É comum usar a* ***distribuição de Pareto***[8] ***de parâmetros*** *$A > 0$* ***e*** *$\alpha > 0$ para descrever a distribuição de riquezas X entre indivíduos de uma população. Sua densidade é dada por*

$$f(x) = \begin{cases} kx^{-\alpha-1}, & se\ x \geq A \\ 0 & caso\ contrário \end{cases}$$

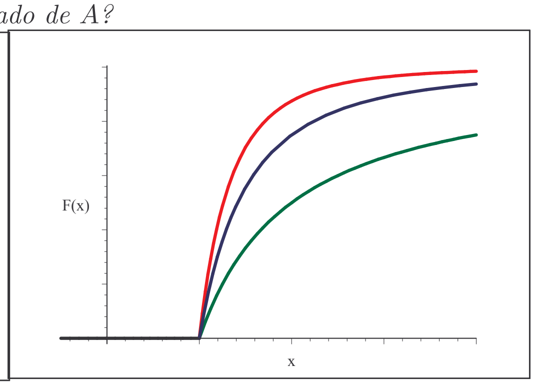
Encontre o valor de k em função de A e α. Em seguida, mostre que a f.d.a. de X é dada por

$$F(x) = \begin{cases} 1 - \left(\frac{A}{x}\right)^{\alpha} & se\ x \geq A \\ 0, & caso\ contrário \end{cases}$$

e encontre a mediana de X. Qual o significado de A?



F.d.p. Pareto para $A = 1$ e $\alpha = 1, 2, 3$



F.d.a. Pareto para $A = 1$ e $\alpha = 1, 2, 3$

Ex. 4 *Se X tem distribuição uniforme em $[0, 12]$, encontre as funções de distribuição e, a partir destas, as funções de densidade, das seguintes variáveis:*
a) $Y = \frac{X}{12}$.
b) $Z = X^2$.
c) $W = (X - 6)^2$.

Ex. 5 *Se X tem distribuição uniforme em $[0, 2\pi]$, calcule as funções de distribuição e, a partir destas, as funções de densidade das variáveis $Y = \cos X$ e $Z = \sin X$.*

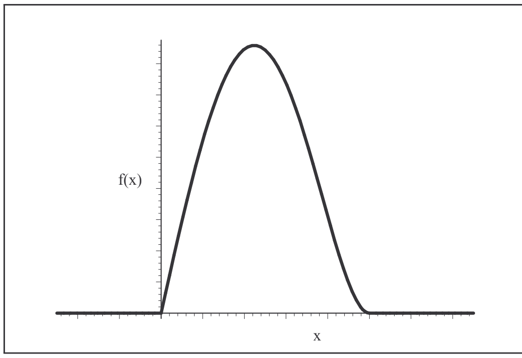
Ex. 6 *a) Seja X uma variável aleatória com função de distribuição $F(x) = \frac{1}{\pi} \frac{1}{1+x^2}$. Seja $Y = \frac{1}{\pi} \frac{1}{1+X^2}$. Qual é a distribuição de Y?*
b) Em geral, seja X uma variável aleatória cuja função de distribuição é $F(x)$. Seja $Y = F(X)$ (a mesma F). Qual é a distribuição de Y?

[8] Os parâmetros A e α não parecem ter nome universal. De acordo com a Wikipedia, a distribuição de Pareto também aproxima outros fenômenos, como "tamanho de vilas/cidades", "tamanho de arquivos transmitidos via Internet", "volume de óleo em reservas naturais", "tamanho de meteoritos", etc.

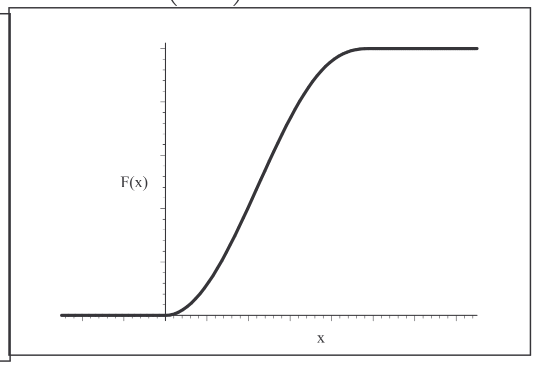
Ex. 7 *Em hidrologia, usa-se a* ***distribuição de Kumaraswamy*** *cuja f.d.a. é dada por*

$$F_X(x) = \begin{cases} 0, \text{ se } x < 0 \\ \left[1 - (1 - x^a)^b\right], \text{ se } 0 \leq x \leq 1 \\ 1, \text{ se } x > 1 \end{cases}$$

Encontre seus quartis e sua f.d.p. e mostre que a moda é $\left(\frac{a-1}{ab-1}\right)^{1/a}$.



F.d.p. Kumaraswamy para $a = 2$ *e* $b = 3$

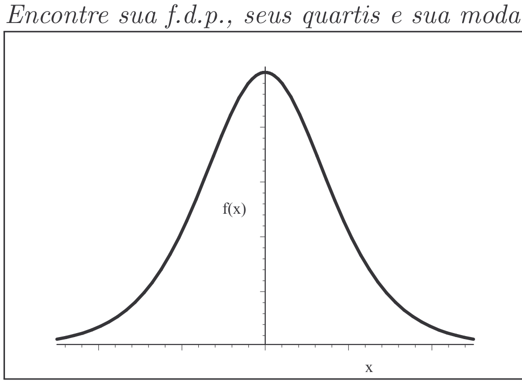


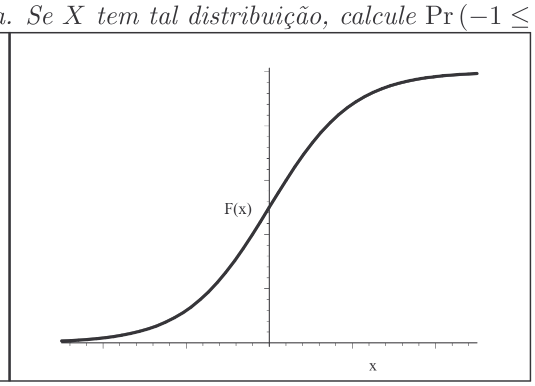
F.d.a. Kumaraswamy para $a = 2$ *e* $b = 3$

Ex. 8 *A distribuição* ***logística*** *(normalizada) é baseada no modelo logístico e tem f.d.a. dada por*

$$F(x) = \frac{1}{1 + e^{-x}}$$

Encontre sua f.d.p., seus quartis e sua moda. Se X *tem tal distribuição, calcule* $\Pr(-1 \leq X \leq 1)$.





4.2 Valor Esperado e Variância

4.2.1 Valor Esperado

Lebremos que, no caso discreto, tínhamos

$$\begin{aligned} \mu &= E(X) = \sum_{k=-\infty}^{\infty} k.\Pr(X = k) \\ Var(X) &= E\left((X - \mu)^2\right) \end{aligned}$$

No caso contínuo, o somatório vira uma integral, e $\Pr(X = t) \approx f(t)\,dt$, então:

Definição 16 *Se* X *é uma variável aleatória com densidade* $f(x)$*, definimos seu* ***valor esperado (valor médio, esperança)****, por*

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

Todas as propriedades obtidas no caso discreto continuam valendo trocando somatórios por integrais:

Proposição 17 *Se $Y = h(X)$, então*

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

Prova. Provemos esta propriedade apenas no caso em que h é crescente. Neste caso, sabemos que $g(y) dy = f(x) dx$. Assim, usando a substituição $y = h(x)$:

$$E(Y) = \int_{-\infty}^{\infty} yg(y) dy = \int_{-\infty}^{\infty} h(x) f(x) dx$$

■

Proposição 18 *Sejam a e b constantes quaisquer. Então*

$$\boxed{E(aX + b) = aE(X) + b}$$

Em particular:

$$\begin{aligned} E(b) &= b \\ E(aX) &= aE(X) \\ E(X - \mu_X) &= 0 \end{aligned}$$

Prova. Seja $f(x)$ a densidade de X. Então:

$$E(aX + b) = \int_{-\infty}^{\infty} (ax + b) f(x) dx = a \int_{-\infty}^{\infty} xf(x) dx + b \int_{-\infty}^{\infty} f(x) dx = aE(X) + b.1 = aE(X) + b$$

■

Mais tarde, mostraremos também que, se X e Y são duas variáveis aleatórias, então

$$\boxed{E(X + Y) = E(X) + E(Y)}$$

4.2.2 Variância

Pelos mesmos motivos do caso discreto, uma boa medida de dispersão de uma variável aleatória contínua X é a sua variância:

Definição 19 *A **variância** e o **desvio-padrão** de uma variável aleatória X com densidade $f(x)$ e média $E(X) = \mu$ são*

$$\boxed{Var(X) = E\left((X - \mu)^2\right) = \int_{-\infty}^{\infty} (x - \mu)^2 f(x) dx}$$

$$\boxed{\sigma(X) = \sqrt{Var(X)}}$$

Proposição 20 *Temos*

$$\boxed{Var(aX + b) = a^2 Var(X)}$$

$$\boxed{Var(X) = E\left(X^2\right) - (E(X))^2}$$

Prova. A demonstração do primeiro fato é 100% idêntica à do caso discreto! Copiando e colando, sendo $Y = aX + b$, temos:

$$\begin{aligned} Var(Y) &= E\left((Y - \mu_Y)^2\right) = E\left(((aX + b) - (a\mu_X + b))^2\right) = \\ &= E\left(a^2 (X - \mu_X)^2\right) = a^2 E\left((X - \mu_X)^2\right) = a^2 Var(X) \end{aligned}$$

A segunda também é análoga ao caso discreto, trocando somatórios por integrais:

$$Var(X) = \int_{-\infty}^{\infty} \left(x^2 - 2\mu x + \mu^2\right) f(x)\,dx = \int_{-\infty}^{\infty} x^2 f(x)\,dx - 2\mu \int_{-\infty}^{\infty} x f(x)\,dt + \mu^2 = E\left(X^2\right) - \mu^2$$

■

Mais tarde, mostraremos que, $\boxed{\text{se } X \text{ e } Y \text{ são independentes, } Var(X+Y) = Var(X) + Var(Y)}$

Para terminar nosso repeteco do caso discreto, aqui está a versão "contínua" da desigualdade de Chebyshev:

Teorema 21 (Desigualdade de Chebyshev) *Seja X uma variável aleatória com valor esperado $\mu = E(X)$ e desvio-padrão $\sigma = \sigma(X)$. Então*

$$\boxed{\Pr(\mu - k\sigma < X < \mu + k\sigma) \geq 1 - \tfrac{1}{k^2}}$$

ou seja

$$\boxed{\Pr(|X - \mu| > k\sigma) \leq \tfrac{1}{k^2}}$$

Prova. É o mesmo truque do caso discreto, trocando somatórios por integrais. De fato:

$$\begin{aligned}
Var(X) &= \int_{-\infty}^{\infty} (x-\mu)^2 f(x)\,dx = \\
&= \int_{-\infty}^{\mu-k\sigma} (x-\mu)^2 f(x)\,dx + \int_{\mu-k\sigma}^{\mu+k\sigma} (x-\mu)^2 f(x)\,dx + \int_{\mu+k\sigma}^{\infty} (x-\mu)^2 f(x)\,dx \geq \\
&\geq \int_{-\infty}^{\mu-k\sigma} (k\sigma)^2 f(x)\,dx + 0 + \int_{\mu+k\sigma}^{\infty} (k\sigma)^2 f(x)\,dx = \\
&= k^2\sigma^2 \left(\int_{-\infty}^{\mu-k\sigma} f(x)\,dx + \int_{\mu+k\sigma}^{\infty} f(x)\,dx \right) = k^2\sigma^2 \Pr(|X-\mu| > k\sigma)
\end{aligned}$$

Então, substituindo $Var(X) = \sigma^2$:

$$\Pr(|X-\mu| > k\sigma) \leq \frac{1}{k^2}$$

■

Exemplo 22 *Dizemos que X tem* ***distribuição uniforme*** *no intervalo $[\alpha, \beta]$ (representado por $X \sim U(\alpha, \beta)$) quando a densidade de X é dada por*

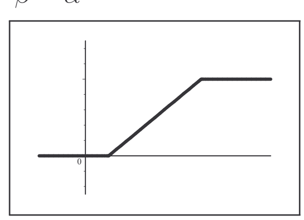
$$f(x) = \begin{cases} c, & \text{para } x \in [\alpha, \beta] \\ 0, & \text{caso contrário} \end{cases}$$

Calcule c, $E(X)$, $Var(X)$, $\sigma(X)$ e a f.d.a. $F(x)$.

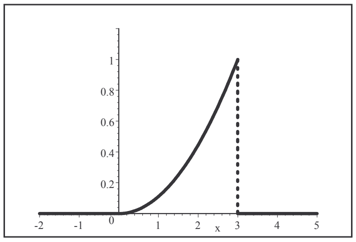
Solução: Como a área sob a curva $f(x)$ tem de ser 1, é claro que $c = \frac{1}{\beta-\alpha}$. Então:

$$\begin{aligned}
E(X) &= \int_{-\infty}^{\infty} t f(t)\,dt = \int_{\alpha}^{\beta} \frac{t}{\beta-\alpha}dt = \frac{\beta^2-\alpha^2}{\beta-\alpha}\frac{1}{2} = \frac{\beta+\alpha}{2} \\
E\left(X^2\right) &= \int_{-\infty}^{\infty} t^2 f(t)\,dt = \int_{\alpha}^{\beta} \frac{t^2}{\beta-\alpha}dt = \frac{\beta^3-\alpha^3}{\beta-\alpha}\frac{1}{3} = \frac{\beta^2+\alpha\beta+\alpha^2}{3} \\
Var(X) &= \frac{\beta^2+\alpha\beta+\alpha^2}{3} - \left(\frac{\beta+\alpha}{2}\right)^2 = \frac{\beta^2+\alpha^2-2\alpha\beta}{12} = \frac{(\beta-\alpha)^2}{12} \\
\sigma(X) &= \frac{\beta-\alpha}{\sqrt{12}} \\
F(x) &= \int_{-\infty}^{x} f(t)\,dt = \int_{\alpha}^{x} f(t)\,dt = \frac{x-\alpha}{\beta-\alpha} \text{ para } x \in [\alpha, \beta]\,. \text{ Portanto}
\end{aligned}$$

$$F(x) = \begin{cases} 0, & \text{se } x < \alpha \\ \frac{x-\alpha}{\beta-\alpha}, & \text{se } \alpha \leq x \leq \beta \\ 1, & \text{se } \beta < x \end{cases}$$



Exemplo 23 *Seja* $f(x) = \begin{cases} \frac{x^2}{9}, \text{ se } x \in [0,3] \\ 0, \text{ caso contrário} \end{cases}$. *Calcule* $E(X)$ *e* $Var(X)$.



$$E(X) = \int_{-\infty}^{\infty} tf(t)\,dt = \int_0^3 t\frac{t^2}{9}dt = \left.\frac{t^4}{36}\right]_0^3 = \frac{9}{4} = 2.25$$
$$E(X^2) = \int_0^3 \frac{t^4}{9}dt = \left.\frac{t^5}{45}\right]_0^3 = \frac{27}{5} = 5.4$$
$$Var(X) = E(X^2) - [E(X)]^2 = \frac{27}{5} - \left(\frac{9}{4}\right)^2 = \frac{27}{80} = 0.3375$$

Exemplo 24 *Seja* $f(x) = \begin{cases} 0, \text{ se } x < 0 \\ ce^{-x}, \text{ se } x \geq 0 \end{cases}$. *Calcule* c, $F(x)$, $E(X)$ *e* $\sigma(X)$.

$$\begin{aligned}
\int_0^{\infty} ce^{-x}dx &= \left.-ce^{-x}\right]_0^{\infty} = 0 + ce^{-0} = c \Rightarrow c = 1 \\
F(x) &= \int_{-\infty}^{x} f(t)\,dt = \int_0^x e^{-t}dt = \left.-e^{-t}\right]_0^x = 1 - e^{-x} \text{ para } x > 0 \\
E(X) &= \int_{-\infty}^{\infty} tf(t)\,dt = \int_0^{\infty} \underbrace{t}_{u}\underbrace{e^{-t}dt}_{dv} = \left.-te^{-t}\right]_0^{\infty} - \int_0^{\infty} -e^{-t}.1.dt = \int_0^{\infty} e^{-t}dt = 1 \\
E(X^2) &= \int_{-\infty}^{\infty} t^2 f(t)\,dt = \int_0^{\infty} t^2 e^{-t}dt = \left(\left.-t^2e^{-t}\right]_0^{\infty} - \int_0^{\infty}\left(-e^{-t}\right)2tdt = 2\int_0^{\infty} te^{-t}dt = 2 \\
Var(X) &= 2 - 1^2 = 1 \Rightarrow \sigma(X) = 1
\end{aligned}$$

4.2.3 Exercícios

Ex. 9 *A densidade de uma variável aleatória X é*

$$f(x) = \begin{cases} 2x, \text{ se } x \in [0,1] \\ 0, \text{ caso contrário} \end{cases}$$

Mostre que $E(X) = \frac{2}{3}$ *e* $Var(X) = \frac{1}{18}$.

Ex. 10 *Calcule* $E(X)$ *e* $Var(X)$ *onde X é uma variável aleatória com densidade dada por*

$$f(x) = \begin{cases} \frac{\sin x}{2}, \text{ se } x \in [0,\pi] \\ 0, \text{ caso contrário} \end{cases}$$

Ex. 11 *Mostre que a média e a variância da distribuição de Pareto de parâmetros* $A > 0$ *e* $\alpha > 0$, *de densidade*

$$f(x) = \begin{cases} \alpha A^{\alpha} x^{-\alpha-1}, \text{ se } x \geq A \\ 0 \text{ caso contrário} \end{cases}$$

são, respectivamente,

$$\begin{aligned}
E(X) &= \frac{\alpha A}{\alpha - 1} \text{ (desde que } \alpha > 1) \\
Var(X) &= \frac{\alpha A^2}{(\alpha-1)^2(\alpha-2)} \text{ (desde que } \alpha > 2)
\end{aligned}$$

Ex. 12 *Mostre que a distribuição de Cauchy, cuja densidade é dada por*

$$f(x) = \frac{1}{\pi}\frac{1}{1+x^2}$$

não tem *valor esperado (pois as integrais correspondentes divergem).*

Ex. 13 *Seja X uma variável aleatória e considere a função $f(t) = E\left((X-t)^2\right)$. Mostre que $f(t)$ tem um mínimo para $t = E(X)$.*

Ex. 14 *Seja X uma variável aleatória e considere a função $g(t) = E(|X-t|)$. Mostre que $g(t)$ tem um mínimo quando t é a mediana de X.*

Ex. 15 *Use integrais duplas (trocando a ordem de integração) para mostrar que, se X é uma variável não-negativa, então*

$$E(X) = \int_0^\infty \Pr(X > t)\ dt$$

Ex. 16 *O **coeficiente de curtose** de uma variável aleatória X com valor esperado $\mu = E(X)$ e desvio-padrão $\sigma = \sigma(X)$ é definido por*

$$\frac{E\left((X-\mu)^4\right)}{\sigma^4}$$

a) Use que $Var(Y) = E\left(Y^2\right) - (E(Y))^2 \geq 0$ com um Y apropriado para mostrar que o coeficiente de curtose é sempre maior ou igual a 1.
b) Mostre que o coeficiente de curtose da distribuição uniforme no intervalo $[a,b]$ é $\frac{9}{5}$.

4.3 Exercícios de Provas

Ex. 17 (A1 2004.2) *A renda de um indivíduo escolhido ao acaso em uma população tem uma função de distribuição acumulada dada por*

$$F(x) = \begin{cases} 0, \text{ se } y < 100 \\ 1 - \left(\frac{100}{y}\right)^2, \text{ se } y \geq 100 \end{cases}$$

Este modelo probabilístico é conhecido como modelo de Pareto, com parâmetros $\alpha = 2$ e $\beta = 100$.
a) Qual é a densidade desta distribuição de probabilidade?
b) Qual é a probabilidade de que um indivíduo escolhido ao acaso tenha renda maior que 200?
c) Qual é o valor médio da renda na população?

Chapter 5

Principais Distribuições Contínuas

5.1 Distribuição Uniforme

Definição 1 *Dizemos que a variável aleatória X tem* ***distribuição uniforme no intervalo*** $[a,b]$ *(denotada por $X \sim U[a,b]$) quando sua densidade é dada por*

$$\boxed{f(x) = \begin{cases} \frac{1}{b-a}, & se\ a \leq x \leq b \\ 0, & caso\ contrário \end{cases}}$$

ou seja, sua função de distribuição acumulada é

$$\boxed{F(x) = \begin{cases} 0, & se\ x < a \\ \frac{x-a}{b-a}, & se\ a \leq x \leq b \\ 1, & se\ b < x \end{cases}}$$

No capítulo anterior, mostramos que:

Proposição 2 *Se $X \sim U[a,b]$ então*

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

Note que, tomando $U = \frac{X-a}{b-a}$, temos $U \sim U[0,1]$, e, em particular, $E(U) = \frac{1}{2}$ e $Var(U) = \frac{1}{12}$. Conversamente, se $U \sim U[0,1]$, tome $X = (b-a)U + a$ e obteremos $X \sim U[a,b]$[1].

5.2 Distribuição Exponencial

Suponha que eventos ocorram de acordo com um processo de Poisson à taxa média de λ eventos por unidade de tempo (digamos, por hora). Dado um certo intervalo $[0,t]$, seja X o número de eventos ocorridos neste intervalo. Sabe-se que $X \sim Poi(\lambda t)$, isto é,

$$\Pr(X = k) = \frac{(\lambda t)^k}{k!} e^{-\lambda t}$$

Por outro lado, seja T o **tempo de ocorrência do primeiro evento**. Note que $T > t$ é equivalente a $X = 0$ (isto é, o primeiro evento ocorrer depois de t horas é equivalente ao número de eventos em t horas ser zero). Assim

$$\Pr(T \leq t) = 1 - \Pr(T > t) = 1 - \Pr(X = 0) = 1 - e^{-\lambda t}$$

[1] Em Excel, a função *ALEATÒRIO()* retorna um número aleatório U cuja distribuição é uniforme em $[0,1]$. Usando a propriedade acima, é fácil simular distribuiçãoes uniformes em outros intervalos. Por exemplo, se queremos que X venha da distribuição uniforme em $[12,60]$, em Excel basta fazer

$$= 48 * ALEATÓRIO() + 12$$

Assim, T segue uma distribuição cujas funções acumulada e densidade são, respectivamente

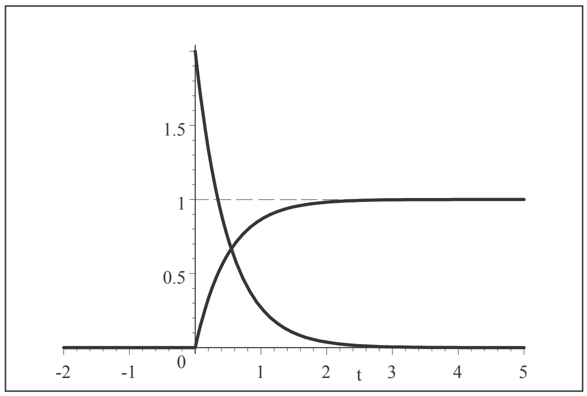
$$\begin{aligned} F(t) &= \Pr(T \le t) = 1 - e^{-\lambda t} \text{ (para } t \ge 0) \\ f(t) &= F'(t) = \lambda e^{-\lambda t} \text{ (para } t \ge 0) \end{aligned}$$

Definição 3 *Dizemos que a variável aleatória T tem* ***distribuição exponencial de parâmetro*** *λ (denotada $T \sim Exp(\lambda)$) se sua densidade é dada por*

$$\boxed{f(t) = \begin{cases} \lambda e^{-\lambda t}, & se\ t \ge 0 \\ 0, & caso\ contrário \end{cases}}$$

ou seja, sua função de distribuição acumulada é

$$\boxed{F(t) = \begin{cases} 0, & se\ t < 0 \\ 1 - e^{-\lambda t}, & se\ t \ge 0 \end{cases}}$$



F.D.P e F.D.A. exponencial com $\lambda = 2$

Proposição 4 *Se $T \sim Exp(\lambda)$, então*

$$\boxed{E(T) = \tfrac{1}{\lambda} \qquad Var(T) = \tfrac{1}{\lambda^2} \qquad Med(T) = \tfrac{\ln 2}{\lambda}}$$

Prova.

$$\begin{aligned} E(T) &= \int_0^\infty t\lambda e^{-\lambda t} dt = \left(-te^{-\lambda t} - \frac{e^{-\lambda t}}{\lambda}\right]_{t=0}^{t=\infty} = \frac{1}{\lambda} \\ E\left(T^2\right) &= \int_0^\infty t^2 \lambda e^{-\lambda t} dt = \left(-t^2 e^{-\lambda t}\right]_{t=0}^{t=\infty} + 2\int_0^\infty te^{-\lambda t} dt = (0-0) + \frac{2}{\lambda} E(T) = \frac{2}{\lambda^2} \\ Var(T) &= E\left(T^2\right) - (E(T))^2 = \frac{1}{\lambda^2} \Rightarrow \sigma(Z) = \frac{1}{\lambda} \\ F(t) &= 0.5 \Rightarrow 1 - e^{-\lambda t} = 0.5 \Rightarrow t = \frac{\ln 2}{\lambda} \end{aligned}$$

■

Note o paralelo entre todas as fórmulas obtidas para a distribuição discreta $Geom(p)$ (quantos experimentos fazer até o primeiro sucesso) e a distribuição contínua $Exp(\lambda)$ (quanto tempo esperar até o primeiro "sucesso"):

Variável	F.D.A.	1-F.D.A.	F.D.P.	$E(X)$	$Var(X)$	$Med(X)$
$Geom(p)$	$\Pr(X \le k) = 1 - q^k$	$\Pr(X > k) = q^k$	$\Pr(X = k) = p(1-p)^{k-1}$	$\frac{1}{p}$	$\frac{1}{p^2}$	$\frac{\ln 2}{-\ln q}$
$Exp(\lambda)$	$\Pr(T \le t) = 1 - \left(e^{-\lambda}\right)^t$	$\Pr(T > t) = e^{-\lambda t}$	$f(t) = \lambda\left(e^{-\lambda}\right)^t$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$	$\frac{\ln 2}{\lambda}$

Exemplo 5 *Sua empresa recebe* 20 *ligações telefônicas por dia (em média), no horário comercial de* 8*h às* 18*h. Usando um modelo de Poisson para estas ligações, qual a chance de a primeira ligação do dia chegar antes das*

9h? E entre 10h e 12h?
Solução: Seja T o tempo de chegada da primeira ligação, em horas, a partir das 8h (isto é,$T = 0$ às 8h da manhã). A distribuição de T será uma distribuição exponencial de parâmetro $\lambda = 20/10 = 2$ ligações por hora. Então as probabilidades pedidas são

$$\begin{aligned}
\Pr(T \leq 1) &= F(1) = 1 - e^{-2} = 86.466\% \\
\Pr(2 \leq T \leq 4) &= F(4) - F(2) = e^{-4} - e^{-8} = 1.798\%
\end{aligned}$$

Exemplo 6 *Suponha que um equipamento funciona à taxa de 0.5 falhas por hora, isto é, o tempo T da próxima falha satisfaz $T \sim Exp(0.5)$. Calculemos o tempo esperado da primeira falha, a probabilidade do equipamento não falhar em 4 horas e a probabilidade de haver no máximo 1 falha em 4 horas. Temos*

$$\begin{aligned}
E(T) &= 1/0.5 = 2 \text{ horas} \\
F(T) &= 1 - e^{-0.5t} \Rightarrow \Pr(T > 4) = 1 - F(4) = e^{-2} = 13.53\%
\end{aligned}$$

Enfim, se X é o número de falhas nas próximas 4 horas, sabemos que $X \sim Poi(2)$. Então

$$\Pr(X \leq 1) = \Pr(X = 0) + \Pr(X = 1) = e^{-2}\left(\frac{2^0}{0!} + \frac{2^1}{1!}\right) = 3e^{-2} = 40.60\%$$

Exemplo 7 *O tempo de falha em horas de um equipamento é $T \sim Exp(0.25)$ – isto é, o valor esperado do tempo de falha é de 4 horas. Se o equipamento falhar em 4 horas ou menos, você perde \$3000, mas, se ele durar mais do que 4 horas, você ganha \$3000. Qual o lucro esperado de usar este equipamento?*
Solução: seja L o lucro em reais. Então $L = \begin{cases} -3000 \text{ se } T \leq 4 \\ 3000 \text{ se } T > 4 \end{cases}$. Portanto:

$$\begin{aligned}
E(L) &= -3000\Pr(T \leq 4) + 3000\Pr(T > 4) = -3000F(4) + 3000(1 - F(4)) = 3000 - 6000F(4) = \\
&= 3000 - 6000\left(1 - e^{-4(0.25)}\right) = 6000e^{-1} - 3000 = -\$792.72
\end{aligned}$$

Proposição 8 *Se $T \sim Exp(\lambda)$ e $Y = aT$, então $Y \sim Exp(\lambda/a)$.*

Prova. De fato,

$$f_Y(y) = \frac{f_T(t)}{a} = \frac{\lambda e^{-\lambda t}}{a} = \left(\frac{\lambda}{a}\right)e^{-\left(\frac{\lambda}{a}\right)y}$$

que corresponde à distribuição exponencial de parâmetro λ/a. ■

Corolário 9 $\boxed{\textit{Se } T \sim Exp(\lambda) \textit{ então } \lambda T \sim Exp(1)}$.

5.2.1 Exercícios

Ex. 1 *Seja $U \sim U[0,1]$. Encontre as f.d.a., as f.d.p. e os valores esperados de:*
a) $A = U + 2$ b) $B = U^3$ c) $C = \frac{1}{U+1}$
d) $D = \ln(U + 1)$ e) $E = \left|U - \frac{1}{2}\right|$ f) $F = \left(U - \frac{1}{2}\right)^2$

Ex. 2 *Suponha que $U \sim U[0,1]$. Qual a probabilidade da equação $Ux^2 - 5Ux + 4 = 0$ ter raízes reais?*

Ex. 3 *Seja $X \sim U[a,b]$. Para que escolhas de c e d tem-se que $U = cX + d$ tem distribuição uniforme em $[0,1]$? [Dica: há duas possibilidades!]*

Ex. 4 *Suponha que $T \sim Exp(5)$ e $Y = 3T + 4$. Calcule $E(Y)$, $Var(Y)$, a f.d.a. e a f.d.p. de Y.*

Ex. 5 *Seja T o tempo (em dias) da próxima falha do seu ar condicionado. Suponha que $T \sim Exp(2)$. Determine (e tente interpretar) $E(T)$, $\sigma(T)$, a f.d.a. de T, a **função de sobrevivência de** T (isto é, $R(t) = \Pr(T > t)$) e os quartis de T.*

Ex. 6 *A vida de um equipamento tem distribuição exponencial com média* 1000 *horas. Qual a garantia máxima que pode ser dada pelo fabricante para que pelo menos* 99% *dos equipamentos fabricados não falhem durante a garantia?*

Ex. 7 *A vida de um equipamento tem distribuição exponencial com média* 1000 *horas. Um equipamento com vida superior a* 800 *horas dá lucro de* \$1000 *mas um equipamento com vida inferior a* 800 *horas dá prejuízo de* \$500*. Qual o lucro esperado deste equipamento?*

Ex. 8 *Suponha que* $T \sim Exp(1)$*. Seja* $f(n) = E(T^n)$ *onde* $n = 0, 1, 2, \ldots$*. Calcule* $f(n+1)$ *em função de* $f(n)$ *e, a partir daí, deduza uma fórmula para* $E(T^n)$*.*

Ex. 9 *a) Prove que a distribuição exponencial não tem memória, isto é, se* $T \sim Exp(\lambda)$*, então*

$$\Pr(T > r + s \mid T > r) = \Pr(T > s)$$

para quaisquer r*,* s *reais positivos. Compare esta propriedade com a propriedade semelhante da distribuição geométrica.*
b) A vida de uma bateria tem distribuição exponencial com média 500 *horas. Você precisa que a bateria funcione sem falhas durante* 24 *horas, mas, embora ela esteja funcionando neste instante, você não sabe por quanto tempo ela já foi usada. É possível determinar a probabilidade de ela pifar nas próximas* 24 *horas? Em caso positivo, qual é esta probabilidade?*

Ex. 10 *Suponha que o tempo* T *(em minutos) de espera até um ônibus passar pelo ponto onde você está tem distribuição* $Exp\left(\frac{1}{30}\right)$ *(isto é,* $E(T) = 30$*). Você aguardou* 60 *minutos e o ônibus ainda não passou. Qual o valor esperado do tempo que você ainda tem que aguardar? O que este exemplo te diz sobre a aplicabilidade deste modelo para o tempo de espera por transporte urbano?*

Ex. 11 *Seja* $Y \sim U[0,1]$*. Encontre a distribuição de* $Z = -\ln Y$*. Em EXCEL, simule* 1000 *amostras desta variável* Z *e calcule a média destas amostras. Esta média está longe do valor que você esperava?*

Ex. 12 *Você compra duas lâmpadas para colocar na sua sala. Os tempos de vida destas lâmpadas são independentes e dados, em meses, por* $T_1 \sim Exp\left(\frac{1}{4}\right)$ *e* $T_2 \sim Exp\left(\frac{1}{3}\right)$ *(isto é,* $E(T_1) = 4$ *e* $E(T_2) = 3$*). Qual a probabilidade de ambas as lâmpadas ainda estarem funcionando daqui a um mês? E daqui a* 3 *meses? Seja* T *o tempo durante o qual ambas as lâmpadas funcionam ao mesmo tempo. Qual é a distribuição de* T*?*

Ex. 13 *Suponha que* $U \sim U[0,1]$ *e* $V \sim U[0,1]$ *são independentes. Seja* $X = \max(U, V)$*. Encontre a f.d.a., a f.d.p. e o valor esperado de* X*.*

Ex. 14 *Suponha que* $T_1 \sim Exp(\lambda_1)$ *e* $T_2 \sim Exp(\lambda_2)$ *onde* T_1 *e* T_2 *são independentes. Seja* $T = \min(T_1, T_2)$*. Mostre que* $T \sim Exp(\lambda_1 + \lambda_2)$*.*

Ex. 15 *Considere um processo de Poisson com uma taxa de* λ *eventos por hora. Seja* X *o número de eventos ocorridos num intervalo de tempo* $[0, t]$*. Então* $X \sim Poi(\lambda t)$*.*
a) Calcule $\Pr(X \geq 2)$*.*
b) Sendo T *o instante em que o segundo evento acontece, explique porque*

$$\Pr(T \leq t) = \Pr(X \geq 2)$$

c) Mostre que a f.d.p. de T *é* $f(t) = \begin{cases} \lambda^2 t e^{-\lambda t}, \text{ para } t \geq 0 \\ 0, \text{ caso contrário} \end{cases}$
Esta distribuição é um caso particular da distribuição gama a ser estudada a seguir.

5.3 Distribuição Gama

5.3.1 A função Gama

Façamos uma breve pausa para introduzir as principais propriedades da função Γ.

Definição 10 *Para $\alpha > 0$, definimos a função $\Gamma(\alpha)$ como*

$$\boxed{\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx}$$

Proposição 11 *Sempre que as integrais acima convergirem, vale*

$$\boxed{\Gamma(p+1) = p\Gamma(p)}$$

Prova.

$$\Gamma(p+1) = \int_0^\infty x^p e^{-x} dx = \left(-x^p e^{-x}\right]_0^\infty + p\int_0^\infty x^{p-1} e^{-x} dx = (0-0) + p\Gamma(p)$$

■

Proposição 12 *Para n natural positivo, vale*

$$\boxed{\Gamma(n) = (n-1)!}$$

Prova. Por indução: vale para $n = 1$ pois $\Gamma(1) = \int_0^\infty e^{-x} dx = 1 = 0!$; e o passo de indução é:

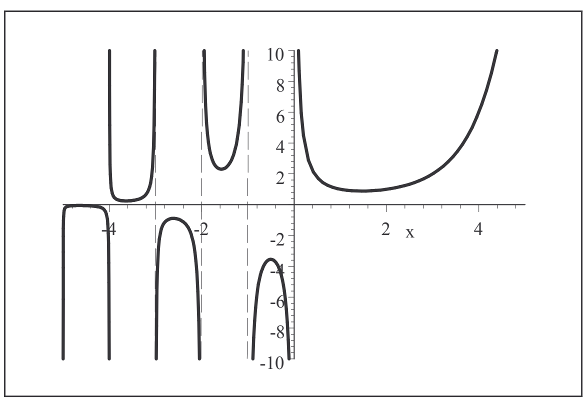
$$\text{se } \Gamma(k) = (k-1)!, \text{ então } \Gamma(k+1) = k\Gamma(k) = k((k-1)!) = k!$$

■

Nota 13 *Como a integral que define $\Gamma(x)$ só converge para $x > 0$, usa-se a propriedade $\Gamma(p) = \Gamma(p+1)/p$ recursivamente para definir $\Gamma(x)$ quando $x < 0$, isto é*

$$\Gamma(x) = \frac{\Gamma(x+n)}{x(x+1)(x+2)(x+3)\ldots(x+n-1)}$$

onde n é escolhido de forma que $x + n > 0$. Com esta definição, o gráfico de $\Gamma(x)$ fica assim



Em outras palavras, a função $\Gamma(x)$ é uma generalização da função fatorial para $x \in \mathbb{R}$ e, portanto, cresce extremamente rápido para $x > 2$.

Proposição 14 *Temos*

$$\boxed{\int_0^\infty e^{-x^2} dx = \frac{\sqrt{\pi}}{2}}$$

$$\boxed{\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}}$$

Prova. Seja I a integral imprópria acima. Então

$$\begin{aligned} I^2 &= \left(\int_0^\infty e^{-x^2}dx\right)\left(\int_0^\infty e^{-y^2}dy\right) = \int_{R^2} e^{-x^2-y^2}dA = \int_0^{\frac{\pi}{2}}\int_0^\infty e^{-r^2}rdrd\theta = \\ &= \int_0^{\frac{\pi}{2}}\left(-\frac{e^{-r^2}}{2}\right]_0^\infty d\theta = \frac{\pi}{2}\left(0+\frac{1}{2}\right) = \frac{\pi}{4} \Rightarrow I = \frac{\sqrt{\pi}}{2} \end{aligned}$$

Portanto

$$\Gamma\left(\frac{1}{2}\right) = \int_0^\infty \frac{e^{-x}}{\sqrt{x}}dx = \int_0^\infty \frac{1}{u}e^{-u^2}(2udu) = \int_0^\infty 2e^{-u^2}du = 2\frac{\sqrt{\pi}}{2} = \sqrt{\pi}$$

■

5.3.2 Distribuição Gama

Mantendo o processo de Poisson da seção anterior, seja agora Z o **tempo de ocorrência do n-ésimo evento** e X o número de ocorrências no intervalo $[0,t]$. Teremos

$$\begin{aligned} \Pr(Z>t) &= \Pr(X<n) = \sum_{k=0}^{n-1}\Pr(X=k) = e^{-\lambda t}\sum_{k=0}^{n-1}\frac{(\lambda t)^k}{k!} \Rightarrow \\ &\Rightarrow \Pr(Z\le t) = 1-\Pr(X<n) = 1-e^{-\lambda t}\sum_{k=0}^{n-1}\frac{(\lambda t)^k}{k!} \end{aligned}$$

Assim, a f.d.a e a f.d.p de Z seriam

$$\begin{aligned} F(t) &= 1-e^{-\lambda t}\sum_{k=0}^{n-1}\frac{(\lambda t)^k}{k!} \\ f(t) &= F'(t) = \lambda e^{-\lambda t}\sum_{k=0}^{n-1}\frac{(\lambda t)^k}{k!} - e^{-\lambda t}\sum_{k=0}^{n-1}\frac{k(\lambda t)^{k-1}\lambda}{k!} = \\ &= \lambda e^{-\lambda t}\left(\sum_{k=0}^{n-1}\frac{(\lambda t)^k}{k!} - \sum_{k=1}^{n-1}\frac{(\lambda t)^{k-1}}{(k-1)!}\right) = \lambda e^{-\lambda t}\frac{(\lambda t)^{n-1}}{(n-1)!} = \frac{\lambda^n}{(n-1)!}t^{n-1}e^{-\lambda t} \end{aligned}$$

Definição 15 *Sejam $\alpha,\lambda>0$. Dizemos que a variável aleatória Z tem* ***distribuição Gama de parâmetros*** *α e β (denotada $Z\sim\Gamma(\alpha,\lambda)$) se sua densidade é dada por*

$$\boxed{\mathrm{GammaDen}(t) = \begin{cases} \frac{\lambda^\alpha}{\Gamma(\alpha)}t^{\alpha-1}e^{-\lambda t}, \textit{ se } t\ge 0 \\ 0, \textit{ caso contrário} \end{cases}}$$

ou, no caso específico em que $\alpha=n$ é natural,

$$\boxed{\mathrm{GammaDen}(t) = \begin{cases} \frac{\lambda^n}{(n-1)!}t^{n-1}e^{-\lambda t}, \textit{ se } t\ge 0 \\ 0, \textit{ caso contrário} \end{cases}}$$

Note que esta é, de fato, uma f.d.p, já que

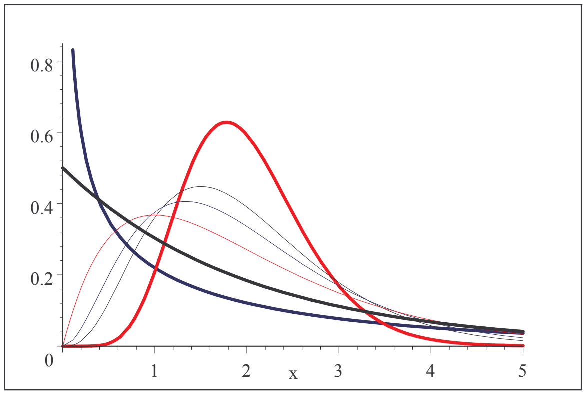
$$\int_{-\infty}^\infty f(t)\,dt = \frac{\lambda}{\Gamma(\alpha)}\int_0^\infty (\lambda t)^{\alpha-1}e^{-\lambda t}dt = \frac{\lambda}{\Gamma(\alpha)}\int_0^\infty x^{\alpha-1}e^{-x}\frac{dx}{\lambda} = \frac{1}{\Gamma(\alpha)}\Gamma(\alpha) = 1$$

onde tomamos $x=\lambda t$.

Também note que, tomando $\alpha=1$ na distribuição Gama, naturalmente ficamos com

$$f(t) = \lambda e^{-\lambda t}$$

uma distribuição exponencial de parâmetro λ (que é a f.d.p. do tempo de ocorrência do **primeiro evento** de um Processo de Poisson).



Família de distribuições Gama com $E(Z) = \alpha/\lambda = 2$

Proposição 16 $\boxed{\textit{Se } Z \sim Gamma(\alpha,\lambda), \textit{ então } E(Z) = \alpha/\lambda,\ Var(Z) = \alpha/\lambda^2 \textit{ e } Moda(Z) = \alpha/\lambda - 1/\lambda.}$

Prova.

$$E(Z) = \int_0^\infty t.\frac{\lambda^\alpha}{\Gamma(\alpha)} t^{\alpha-1} e^{-\lambda t} dt = \frac{1}{\Gamma(\alpha)} \int_0^\infty (\lambda t)^\alpha e^{-\lambda t} dt$$

Fazendo $x = \lambda t$

$$E(Z) = \frac{1}{\Gamma(\alpha)} \int_0^\infty x^\alpha e^{-x} \frac{dx}{\lambda} = \frac{1}{\lambda\Gamma(\alpha)} \Gamma(\alpha+1) = \frac{\alpha}{\lambda}$$

Analogamente

$$\begin{aligned} E\left(Z^2\right) &= \frac{1}{\lambda\Gamma(\alpha)} \int_0^\infty (\lambda t)^{\alpha+1} e^{-\lambda t} dt = \frac{1}{\lambda^2\Gamma(\alpha)} \int_0^\infty x^{\alpha+1} e^{-x} dx = \frac{1}{\lambda^2\Gamma(\alpha)} \Gamma(\alpha+2) = \frac{(\alpha+1)\alpha}{\lambda^2} \\ Var(Z) &= E\left(Z^2\right) - (E(Z))^2 = \frac{\alpha^2+\alpha}{\lambda^2} - \frac{\alpha^2}{\lambda^2} = \frac{\alpha}{\lambda^2} \Rightarrow \sigma(z) = \frac{\sqrt{\alpha}}{\lambda} \end{aligned}$$

Enfim, derivando a densidade e igualando a 0

$$\begin{aligned} f'(t) &= \frac{\lambda^\alpha}{\Gamma(\alpha)} \left((\alpha-1) t^{\alpha-2} e^{-\lambda t} - \lambda t^{\alpha-1} e^{-\lambda t}\right) = 0 \Rightarrow \\ &\Rightarrow \alpha - 1 - \lambda t = 0 \Rightarrow t = \frac{\alpha-1}{\lambda} \text{ para } \alpha > 1 \end{aligned}$$

■

Note os paralelos entre $Y \sim NegBin(n,p)$ (que é o número de experimentos de um Processo de Bernoulli até o n-ésimo sucesso) e $Z \sim Gamma(n,\lambda)$ (tempo de ocorrência do n-ésimo sucesso):

Variável	1-F.D.A.	$E(X)$	$Var(X)$
$NegBin(n,p)$	$\Pr(Y > k) = \Pr(X < n)$ onde $X \sim Bin(k,p)$	$\frac{n}{p}$	$\frac{n}{p^2}$
$Gamma(n,\lambda)$	$\Pr(Z > t) = \Pr(X < n)$ onde $X \sim Poi(\lambda t)$	$\frac{n}{\lambda}$	$\frac{n}{\lambda^2}$

Proposição 17 $\boxed{\textit{Se } X \sim Gamma(a,\lambda) \textit{ e } Z = \lambda X, \textit{ então } Z \sim Gamma(a,1).}$

Prova. De fato,

$$f_Z(z) = \frac{f_X(x)}{\lambda} = \frac{1}{\lambda}\frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x} = \frac{(\lambda x)^{\alpha-1} e^{-(\lambda x)}}{\Gamma(\alpha)} = \frac{z^{\alpha-1}}{\Gamma(\alpha)} e^{-z}$$

que é a f.d.p. da distribuição $Gamma(a,1)$. Por este motivo, diz-se que o parâmetro λ é apenas um **parâmetro de escala desta distribuição**. ■

5.4 Distribuição Normal

É a distribuições mais importante da Estatística, devido ao Teorema Central do Limite (que veremos mais tarde). Foi usada por De Moivre em 1734 como aproximação da Distribuição Binomial para n grande (que veremos mais tarde); por volta de 1800 Gauss utilizou-a para fazer análise de erros (motivo pelo qual esta distribuição é também chamada de Distribuição Gaussiana).

Definição 18 *Dizemos que uma variável aleatória X tem distribuição normal com parâmetros μ e σ^2 (denotado por $X \sim N\left(\mu, \sigma^2\right)$) quando sua f.d.p é dada por*

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

Note que $f(\mu + h) = f(\mu - h)$, isto é, a densidade é simétrica com relação a $x = \mu$. Note também que a função acima é, de fato, uma densidade de probabilidade, já que, fazendo $z = \frac{x-\mu}{\sigma}$, temos

$$\int_{-\infty}^{\infty} f(x)\,dx = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left(-\frac{z^2}{2}\right) dz$$

Agora, tomando $z = \sqrt{2}t$:

$$\int_{-\infty}^{\infty} f(x)\,dx = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} e^{-t^2} dt = 1$$

pela última proposição da seção sobre a função Gama.

Proposição 19 $\boxed{\textit{Se } X \sim N\left(\mu, \sigma^2\right) \textit{ então } Z = \frac{X-\mu}{\sigma} \sim N(0,1).}$

Prova. De fato, seja $h(X) = \frac{X-\mu}{\sigma}$ (que é crescente). Então:

$$f_Z(z) = \frac{f_X(x)}{h'(X)} = \sigma \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}$$

que é a f.d.p. de uma distribuição normal de parâmetros $\mu = 0$ e $\sigma = 1$. ■

Proposição 20 $\boxed{\textit{Se } X \sim N\left(\mu, \sigma^2\right), \textit{ então } E(X) = Moda(X) = Med(X) = \mu \textit{ e } Var(X) = \sigma^2.}$

Prova. Tomando $Z = \frac{X-\mu}{\sigma}$ de novo, temos $E(X) = \sigma E(Z) + \mu$. Mas:

$$E(Z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} z \exp\left(-\frac{z^2}{2}\right) dz = \frac{1}{\sqrt{2\pi}} \left(-e^{-z^2/2}\right]_{-\infty}^{\infty} = 0$$

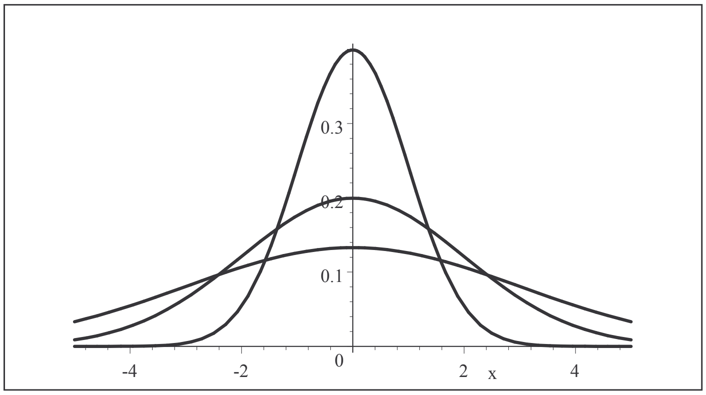
e, portanto, $E(X) = \mu$. Note também que a f.d.p. de Z é uma função par, portanto $Med(Z) = 0$ e $Med(X) = \sigma.0 + \mu = \mu$. Enfim, note que

$$f'(x) = \frac{1}{\sqrt{2\pi}\sigma} \frac{-2(x-\mu)}{2\sigma^2} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) = \frac{x-\mu}{\sqrt{2\pi}\sigma^3} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

é negativo para $x < \mu$ e positivo para $x > \mu$, então seu máximo é alcançado em $Moda(X) = \mu$. Enfim

$$\begin{aligned}
Var(X) &= \sigma^2 Var(Z) = \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{\infty} z^2 \exp\left(-\frac{z^2}{2}\right) dz = \\
&= \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{\infty} -z d \exp\left(-\frac{z^2}{2}\right) = \frac{\sigma^2}{\sqrt{2\pi}} \left(\left(-z \exp\left(-\frac{z^2}{2}\right)\right]_{-\infty}^{\infty} - \int_{-\infty}^{\infty} -\exp\left(-\frac{z^2}{2}\right) dz\right) = \\
&= \frac{\sigma^2}{\sqrt{2\pi}} \left((0-0) + \sqrt{2\pi}\right) = \sigma^2
\end{aligned}$$

■



Distribuições Normais com $\mu = 0$ e $\sigma = 1, 2, 3$

Infelizmente, a função de distribuição acumulada da distribuição normal não pode ser calculada usando apenas as funções matemáticas elementares. No entanto, ela é tão comum que várias calculadoras e pacotes computacionais são capazes de calculá-la numericamente[2]. Neste texto, usaremos a notação

$$F(z) = \Pr(Z \leq z) = \text{NormalDist}(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-x^2/2} dx$$

para a f.d.a. de uma variável Z com distribuição normal padrão $N(0,1)$, e NormalInv para a sua função inversa. No apêndice o leitor encontrará uma tabela para esta função acumulada. **Aprenda a usá-la para resolver problemas**. Alguns valores notáveis são:

$$\begin{aligned}
\Pr(Z \leq 0) &= \text{NormalDist}(0) = 0.5 \text{ (por simetria)} \\
\Pr(-1 \leq Z \leq 1) &= \frac{1}{\sqrt{2\pi}} \int_{-1}^{1} e^{-z^2/2} dz = \text{NormalDist}(1) - \text{NormalDist}(-1) = 0.682\,689\,49 \\
\Pr(-2 \leq Z \leq 2) &= \frac{1}{\sqrt{2\pi}} \int_{-2}^{2} e^{-z^2/2} dz = \text{NormalDist}(2) - \text{NormalDist}(-2) = 0.954\,499\,74 \\
\Pr(-3 \leq Z \leq 3) &= \frac{1}{\sqrt{2\pi}} \int_{-3}^{3} e^{-z^2/2} dz = \text{NormalDist}(3) - \text{NormalDist}(-3) = 0.997\,300\,2 \\
\Pr(-4 \leq Z \leq 4) &= \frac{1}{\sqrt{2\pi}} \int_{-4}^{4} e^{-z^2/2} dz = \text{NormalDist}(4) - \text{NormalDist}(-4) = 0.999\,936\,66
\end{aligned}$$

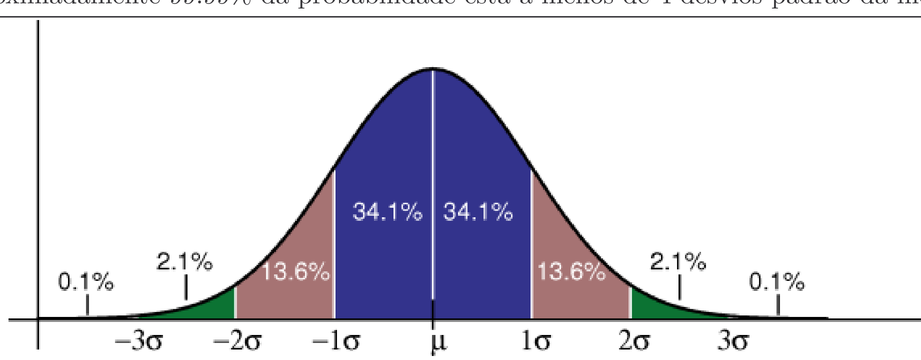
Isto significa que, **em qualquer distribuição normal:**

Aproximadamente 68.27% da probabilidade está a menos de 1 desvio-padrão da média

Aproximadamente 95.45% da probabilidade está a menos de 2 desvios-padrão da média

Aproximadamente 99.73% da probabilidade está a menos de 3 desvios-padrão da média

Aproximadamente 99.99% da probabilidade está a menos de 4 desvios-padrão da média[3]



Probabilidades notáveis na distribuição normal (Fonte: Wikipedia)

[2]O Excel, por exemplo, usa *DIST.NORMP* para a f.d.a. $F(z)$ da normal padrão e *INV.NORMP* para a função inversa $F^{-1}(z)$.
[3]E, pela simetria, metade destas probabilidades estarão nos intervalos $[0,1]$, $[0,2]$, $[0,3]$ e $[0,4]$ respectivamente.

Exemplo 21 *Suponha que a altura de um homem adulto em centímetros é* $X \sim N\left(168, 6^2\right)$. *Qual a probabilidade de um homem adulto ter mais de* $1.80m$*?*
Solução: tome

$$Z = \frac{X - 168}{6}$$

Então

$$X > 180 \Leftrightarrow Z > \frac{180 - 168}{6} = 2$$

Assim

$$\Pr(X > 180) = \Pr(X > 2) = 1 - \text{NormalDist}(2) = 2.2750\%$$

Exemplo 22 *Alguns testes de Q.I. são criados de forma que a pontuação de uma pessoa escolhida ao acaso seja* $X \sim N\left(100, 15^2\right)$. *Supondo que este modelo seja válido, calcule* $\Pr(X \le 70)$, $\Pr(X > 115)$, $\Pr(X \ge 190)$, $\Pr(|X - 100| < 10)$ *e encontre* a *tal que* $\Pr(|X - 100| \le a) = 0.95$.
Solução: tomando

$$Z = \frac{X - 100}{15}$$

sabemos que $Z \sim N(0, 1)$. *Então consultamos uma tabela (ou usamos um pacote computacional) para encontrar*

$$\begin{aligned}
\Pr(X \le 70) &= \Pr(Z \le -2) = \text{NormalDist}(-2) = 2.2750\% \\
\Pr(X > 115) &= \Pr(Z > 1) = 1 - \text{NormalDist}(1) = 15.8655\% \\
\Pr(X \ge 190) &= \Pr(Z \ge 6) = 1 - \text{NormalDist}(6) = 9.865 \times 10^{-10} \\
\Pr(|X - 100| < 10) &= \Pr\left(|Z| < \frac{2}{3}\right) = \text{NormalDist}\left(\frac{2}{3}\right) - \text{NormalDist}\left(-\frac{2}{3}\right) = 49.5015\% \\
\Pr(|X - 100| \le a) &= \Pr\left(|Z| \le \frac{a}{15}\right) = 95\% \Rightarrow \frac{a}{15} = \text{NormalInv}(0.975) \approx 1.9600 \Rightarrow a \approx 29.4
\end{aligned}$$

Exemplo 23 *Seja* $X \sim N\left(300, 5^2\right)$. *Calcule* $\Pr(X < 290)$ *e* $\Pr(|X - \mu| < 2\sigma)$. *Compare esta última com* $\Pr(|Y - \mu_Y| < 2\sigma_Y)$ *onde* $Y \sim N(1311, 309.85)$.
Solução: seja $Z = \frac{X-300}{5}$. *Então*

$$\begin{aligned}
\Pr(X < 290) &= \Pr(Z < -2) = \text{NormalDist}(-2) = 2.2750\% \\
\Pr(|Y - \mu_Y| < 2\sigma_Y) &= \Pr(|X - \mu| < 2\sigma) = \Pr(|Z| < 2) = \text{NormalDist}(2) - \text{NormalDist}(-2) = 95.4500\%
\end{aligned}$$

5.4.1 Exercícios

Ex. 16 *Calcule* $\Gamma(1.5)$ *e* $\Gamma(2.5)$.

Ex. 17 *Calcule* $\Gamma(3.4)$ *em função de* $\Gamma(0.4) = 2.218$. *Confira sua resposta com o auxílio de uma calculadora ou computador.*

Ex. 18 *No campeonato de futebol da sua empresa, os gols acontecem como num processo de Poisson com uma média de* 2.89 *gols por jogo (um jogo tem exatamente* 90 *minutos; não há acréscimos). No bolão da sua empresa, cada participante escolhe um possível minuto de jogo para o segundo gol daquele jogo. Em que minuto você apostaria? E se fosse para apostar no terceiro gol do jogo?*

Ex. 19 *Seja* $Z \sim N(0, 1)$. *Um livro mostra três tabelas distintas de probabilidades para três funções* F, A *e* R*:*

$$\begin{aligned}
F(z) &= \text{NormalDist}(z) = \Pr(Z \le z) \\
R(z) &= \Pr(Z > z) \\
A(z) &= \Pr(0 \le Z \le z)
\end{aligned}$$

Qual é a relação entre F *e* R*? Entre* R *e* A*? E entre* F *e* A*?*

Ex. 20 *Seja* $X \sim N(10, 16)$ *(isto é, o desvio-padrão é* 4*). Usando uma tabela ou computador, calcule:*
a) $\Pr(X \le 10)$ *b)* $\Pr(X \le 18)$ *c)* $\Pr(X > 13)$ *d)* $\Pr(13 \le X \le 18)$
e) $\Pr(6 \le X \le 14)$ *f)* $\Pr(X \le 0)$ *g)* $\Pr(9 \le X \le 11)$ *h)* $\Pr(X > -4)$

Ex. 21 *Seja $Z \sim N(0,1)$. Use uma tabela ou um computador para encontrar os valores de a tais que:*
a) $\Pr(-a < Z < a) = 0.9$ *b)* $\Pr(-a < Z < a) = 0.95$ *c)* $\Pr(-a < Z < a) = 0.99$
d) $\Pr(Z > a) = 0.75$ *e)* $\Pr(Z \leq a) = 0.81$

Ex. 22 *As alturas dos 1000 estudantes da FGV são normalmente distribuídas com média $1.72m$ e desvio-padrão $5cm$. Qual o valor esperado do número de estudantes com mais de $1.80m$? Qual a probabilidade de algum deles ter mais de $2m$ de altura? [Usar tabela.]*

Ex. 23 *Um equipamento militar vem em duas versões: tipo A, com tempo de vida em horas dado por $X_A \sim N(30, 6^2)$, e tipo B, com tempo de vida em horas $X_B \sim N(34, 3^2)$. Que equipamento seria preferível para uma missão de 34 horas? E para uma de 40 horas?*

Ex. 24 *Suponha que os graus numéricos de uma turma grande de alunos de Probabilidade sejam aproximadamente normalmente distribuídos, com média 6 e desvio-padrão 2. Você quer dar notas de A a D aos alunos de forma que 15% da turma receba A, 40% receba B, 25% receba C e 20% receba nota D. Quais devem ser os cortes numéricos que separam estas notas? [Usar tabela]*

Ex. 25 *Encontre os quartis da distribuição $N(\mu, \sigma^2)$ em função de μ e σ. [Usar tabela]*

Ex. 26 *Mostre que os pontos de inflexão da densidade normal estão a um desvio-padrão da média.*

Ex. 27 *Dizemos que X tem distribuição* lognormal *com parâmetros μ e σ^2 quando $Y = \ln X$ tem distribuição normal de parâmetros μ e σ^2.*
a) Encontre a função densidade da variável lognormal X.
b) Se X tem distribuição lognormal com parâmetros $\mu = 1.5$ e $\sigma^2 = 1$, calcule $\Pr(X > 2)$.

5.5 Taxa de Falhas

Definição 24 *Seja T o tempo de vida de um equipamento, isto é, o instante da sua primeira falha, cuja f.d.a é $F(t)$. A **confiabilidade** deste equipamento é*

$$\boxed{R(t) = \Pr(T > t) = 1 - F(t)}$$

Exemplo 25 *Se a falha for um processo de Poisson de parâmetro λ (em falhas por hora), então $T \sim Exp(\lambda)$, isto é*

$$\begin{aligned} E(T) &= Var(T) = 1/\lambda \\ \Pr(T \leq t) &= 1 - e^{-\lambda t} \\ R(t) &= e^{-\lambda t} \end{aligned}$$

Definição 26 *A **taxa média de falhas** de um equipamento em um intervalo $[t, t + \Delta t]$ é a chance de ele falhar nos próximos Δt dado que ele ainda não falhou, dividido por Δt, isto é*

$$TMF = \frac{\Pr(T \leq t + \Delta t \mid T > t)}{\Delta t} = \frac{F(t + \Delta t) - F(t)}{(1 - F(t))\,\Delta t} = -\frac{R(t + \Delta t) - R(t)}{R(t)\,.\Delta t}$$

*Tome o limite quando $\Delta t \to 0$ e temos então a **taxa instantânea de falhas***

$$\boxed{\lambda(t) = \lim_{\Delta t \to 0} TMF = \frac{F'(t)}{1 - F(t)} = -\frac{R'(t)}{R(t)}}$$

Exemplo 27 *Se a falha for um processo de Poisson de parâmetro λ (em falhas por hora), então*

$$\begin{aligned} TMF &= -\frac{e^{-\lambda(t+\Delta t)} - e^{-\lambda t}}{e^{-\lambda t}.\Delta t} = \frac{1 - e^{-\lambda \Delta t}}{\Delta t} \\ \lambda(t) &= \frac{\lambda e^{-\lambda t}}{e^{-\lambda t}} = \lambda \end{aligned}$$

isto é, no processo de Poisson, a taxa de falhas é constante e igual a λ falhas por hora. Reciprocamente, se a taxa de falhas é constante, mostra-se que o tempo de falha segue uma distribuição exponencial.

Exemplo 28 *Considere 2 componentes independentes em paralelo, onde* $T_1 \sim Exp(\lambda_1)$ *é o tempo de falha do primeiro componente e* $T_2 \sim Exp(\lambda_2)$ *é o tempo de falha do segundo. Seja* T *o tempo de falha do sistema. Calcule* $E(T) = MTBF$, $R(t)$ *e* $\lambda(t)$.
Solução:

$$\begin{aligned}
R(t) &= \Pr(T > t) = \Pr(T_1 > t \text{ ou } T_2 > t) = \Pr(T_1 > t) + \Pr(T_2 > t) - \Pr(T_1 > t) . \Pr(T_2 > t) = \\
&= e^{-\lambda_1 t} + e^{-\lambda_2 t} - e^{-(\lambda_1+\lambda_2)t} \\
F(t) &= 1 - R(t) = 1 - e^{-\lambda_1 t} - e^{-\lambda_2 t} + e^{-(\lambda_1+\lambda_2)t} \Rightarrow f(t) = \lambda_1 e^{-\lambda_1 t} + \lambda_2 e^{-\lambda_2 t} - (\lambda_1 + \lambda_2) e^{-(\lambda_1+\lambda_2)t} \\
E(T) &= \int_0^\infty t f(t)\, dt = \frac{1}{\lambda_1} + \frac{1}{\lambda_2} - \frac{1}{\lambda_1 + \lambda_2}
\end{aligned}$$

5.6 Exercícios de Provas

Ex. 28 (A2 2004.2) *Uma distribuição de probabilidade muito utilizada para modelar o tempo de vida de equipamentos é a distribuição de Rayleigh, cuja densidade é dada por* $f(x) = 2axe^{-ax^2}$ *, para* $x > 0$*, onde* a *é um parâmetro.*
a) Verifique que se X *tem distribuição de Rayleigh, então* $Y = X^2$ *tem distribuição exponencial.*
Nos demais itens, suponha que um certo equipamento tem um tempo de vida, em meses, dado por uma distribuição de Rayleigh com parâmetro $a = 0.01$.
b) Calcule a probabilidade de que o equipamento dure mais do que 20 *meses.*
c) Verifique que a função de taxa de falhas do equipamento é dada por

$$\lambda(x) = \frac{f(x)}{1 - F(x)} = 0,02x$$

d) Em que dia é mais provável que o equipamento falhe: no primeiro dia do décimo mês, ou no primeiro dia do vigésimo mês? Justifique.
e) Se você tiver a opção de usar, por um dia, um equipamento com 10 *meses de uso ou um com* 20 *meses de uso (ambos funcionando), o que você prefere? Justifique.*

Ex. 29 (AS 2004.2) *Seja* T *o tempo em que a primeira falha de um disco rígido de uma certa marca ocorre desde o momento em que ele é novo. Usualmente, o valor esperado de* T *é de* 30 *meses, mas você tem usado um disco rígido desta marca por* 24 *meses (desde quando ele era novo) sem problema algum.*
a) Suponha que a distribuição de T *é exponencial. Qual é o valor esperado do tempo de ocorrência da próxima falha* ***do seu disco****?*
b) Repita o item anterior supondo que a distribuição de T *é uniforme (iniciando em* $T = 0$*).*
c) Suponha agora que T *tem uma distribuição (aproximadamente) normal de desvio-padrão* 10 *meses. Qual a probabilidade de seu disco rígido durar* ***mais*** 16 *meses sem falhar?*

Ex. 30 (T3 2005.2) *Suponha que seu processo de resoluções de questões é um processo de Poisson com uma taxa* ***média*** *de* 1 *questão resolvida a cada* 30 *minutos. Seja* X *o número de questões que você resolve em* t *horas, e* T *o tempo necessário em horas para você resolver uma prova de* 2 *questões.*
a) Calcule $\Pr(X \geq 2)$ *e* $\Pr(T \leq t)$.
b) Mostre que a f.d.p. de T *é da forma* $f(t) = \begin{cases} cte^{-2t} & \text{para } t \geq 0 \\ 0 & \text{para } t < 0 \end{cases}$ *onde* c *é uma constante.*
c) Qual o valor esperado e a variância de T*?*
d) Se a turma for composta de alunos assim, que percentagem espera-se terminar a prova em 90 *minutos?*

Ex. 31 (A1 2005.2) *Duas máquinas de encher pacotes de açúcar estão reguladas de maneiras ligeiramente diferentes. A máquina "A" enche pacotes com massa* $X_A \sim N\left(991g, (4g)^2\right)$*. A máquina "B" enche pacotes com massa* $X_B \sim N\left(988g, (14g)^2\right)$*. Todo pacote cujo peso seja menor que* $995g$ *é sempre rejeitado pelo controle de qualidade.*
a) Qual máquina produz a menor percentagem de pacotes rejeitados? Que percentagem é esta?
b) Os pacotes que passam pelo controle de qualidade são vendidos com lucro de \$0,40 *cada, mas os rejeitados são*

re-enchidos manualmente até o peso correto de $1Kg$ *por um custo adicional de* $\$0,15$ *por pacote. Qual o lucro esperado (por pacote) da máquina "B"?*
c) O fabricante decide regular a máquina "A" para que 67% *dos pacotes passem pelo controle de qualidade. Qual deve ser o valor da nova média (mantendo o desvio-padrão antigo)?*

Ex. 32 (A1 2005.2) *Seja* T *uma variável aleatória com distribuição exponencial de parâmetro* λ*, isto é, de densidade dada por*

$$f(t) = \begin{cases} \lambda e^{-\lambda t}, \text{ para } t \geq 0 \\ 0, \text{ caso contrário} \end{cases}$$

a) Seja $X = e^{-\lambda T}$*. Encontre a distribuição de* X *e calcule* $E(X)$*.*
b) Calcule $E(T^n)$ *em função de* λ *e* n *(onde* n *é um número natural).*
c) Um ***coeficiente de assimetria*** *(*skewness*) bastante usado para distribuições em geral é*

$$Sk(T) = \frac{E(T^3) - 3E(T)E(T^2) + 2(E(T))^3}{(\sigma(T))^3}$$

(Em geral, $Sk > 0$ *indica que* $Moda < Mediana < Média$ *e a distribuição tem "rabo mais comprido do lado direito do que do lado esquerdo"). Calcule o coeficente de assimetria da distribuição exponencial.*

Chapter 6

Variáveis Aleatórias Contínuas Bidimensionais

6.1 Função de Densidade Conjunta

Para uma variável contínua, tínhamos uma *Função Densidade de Probabilidade* $f(x)$, cuja principal propriedade era

$$\Pr(X \in I) = \int_I f(x)\,dx$$

onde I é um subconjunto da reta real (tipicamente, I é uma união de intervalos). Como conseqüência deste fato, tínhamos

$$\begin{aligned} f(x) &\geq 0 \text{ para todo } x \in \mathbb{R} \\ \int_{-\infty}^{\infty} f(x)\,dx &= 1 \end{aligned}$$

Além disso, definíamos

$$\begin{aligned} E(X) &= \int_{-\infty}^{\infty} x f(x)\ dx \\ Var(X) &= E\left[(X - E(X))^2\right] = E\left(X^2\right) - [E(X)]^2 \end{aligned}$$

Agora para duas variáveis aleatórias contínuas X e Y (assumindo valores reais), temos:

Definição 1 *Uma* ***Função de Densidade Conjunta*** $f(x, y)$ *das variáveis* X *e* Y *é uma função com a seguinte propriedade:*

$$\boxed{\Pr((X, Y) \in R) = \int\int_R f(x, y)\,dA}$$

onde R *é um subconjunto qualquer do plano* XY *(isto é, uma região dentro do plano* $\mathbb{R}^2$*). Conseqüentemente, uma função de densidade conjunta tem de satisfazer as seguintes propriedades básicas:*

$$\boxed{f(x, y) \geq 0 \text{ para todo } (x, y) \in \mathbb{R}^2}$$

$$\boxed{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x, y)\,dA = 1}$$

Assassinando violentamente o rigor matemático, diríamos que "$f(x, y)\,dA$ é a probabilidade no ponto (x, y)".

Definição 2 *Dada uma função de densidade conjunta $f(x,y)$, definimos o* ***Valor Esperado****,* ***Variância*** *e* ***Desvio-Padrão*** *da variável X como a seguir:*

$$\boxed{E(X) = \int\int_{\mathbb{R}^2} x f(x,y)\, dA}$$

$$\boxed{Var(X) = E\left[(X - E(X))^2\right] = E\left(X^2\right) - [E(X)]^2}$$

$$\boxed{\sigma(X) = \sqrt{Var(X)}}$$

Aliás, em geral, o valor esperado de uma função $g(X,Y)$ será:

$$\boxed{E(g(X,Y)) = \int\int_{\mathbb{R}^2} g(x,y) f(x,y)\, dA}$$

Exemplo 3 (A) *Escolha um ponto (x,y) "aleatoriamente" dentro do quadrado $Q = [0,2] \times [0,2]$. Uma possível densidade para este experimento é tomar*

$$f(x,y) = \begin{cases} \frac{1}{4} \text{ se } (x,y) \in Q \\ 0, \text{ caso contrário} \end{cases}$$

Note que $\int\int_{\mathbb{R}^2} f(x,y)\, dA = 1$. Esta densidade escolhe pontos de coordenadas "pequenas" (perto de 0) tão freqüentemente como pontos de coordenadas "grandes" (perto de 2). Por exemplo, note que

$$\Pr(X + Y > 2) = \int\int_T \frac{1}{4} dA = \frac{1}{4} \acute{A}rea(T) = \frac{1}{2}$$

onde T é o triângulo de vértices $(2,0), (0,2)$ e $(2,2)$.

Definição 4 *Quando a função de densidade conjunta é constante dentro de um certo conjunto R (e 0 fora dele), dizemos que a variável (x,y) é* ***distribuída uniformemente*** *em R. Neste caso, temos*

$$\boxed{f(x,y) = \begin{cases} \frac{1}{\acute{A}rea(R)} \text{ se } (x,y) \in R \\ 0, \text{ caso contrário} \end{cases}}$$

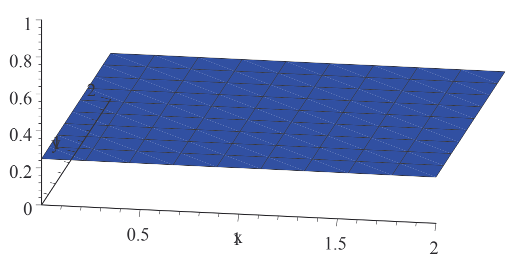
Exemplo 5 (B) *Agora, escolha um ponto (x,y) dentro de Q com a densidade conjunta dada por*

$$g(x,y) = \begin{cases} \frac{xy}{4} \text{ se } (x,y) \in Q \\ 0, \text{ caso contrário} \end{cases}$$

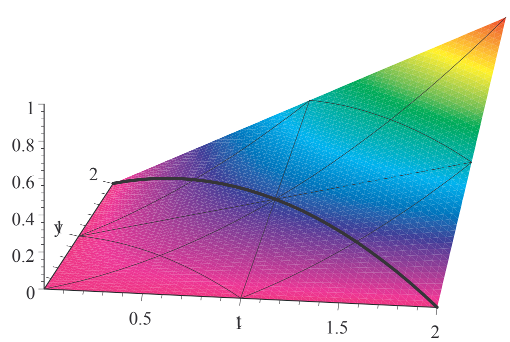
Novamente, verifique que $\int\int_Q \frac{xy}{4} dA = 1$. Esta densidade escolhe pontos de coordenadas "grandes" mais freqüentemente do que pontos de coordendas "pequenas". Intuitivamente, uma pequena região próxima de $(2,2)$ é escolhida 4 vezes mais freqüentemente do que uma região próxima de $(1,1)$ que tenha a mesma área. Note que

$$\Pr(X + Y > 2) = \int\int_T \frac{xy}{4} dA = \int_0^2 \int_{2-x}^2 \frac{xy}{4}\, dy\, dx = \int_0^2 \frac{x}{8}\left(2^2 - (2-x)^2\right) dx = \frac{5}{6}$$

onde T ainda é o triângulo de vértices $(2,0), (0,2)$ e $(2,2)$.



A) $f(x,y) = 1/4$



B) $f(x,y) = xy/4$ (destacada a linha $X + Y = 2$)

6.2 Distribuições Marginais e Condicionais; Covariância e Correlação

Dadas duas variáveis discretas X e Y e sua distribuição conjunta $p(x,y) = \Pr(X = x \text{ e } Y = y)$, tínhamos o conceito de Covariância e Correlação:

$$\begin{aligned} Cov(X,Y) &= E[(X - E(X))(Y - E(Y))] = E(XY) - E(X)E(Y) \\ \rho(X,Y) &= \frac{Cov(X,Y)}{\sigma(X)\sigma(Y)} \end{aligned}$$

Também tínhamos distribuições marginais e condicionais

$$\begin{aligned} p_X(x) &= \Pr(X = x) = \sum_y p(x,y) \\ p_{X|Y}(x|y) &= \Pr(X = x|Y = y) = \frac{p(x,y)}{p_Y(y)} \end{aligned}$$

Agora, dada uma Densidade Conjunta $f(x,y)$ para V.A. Contínuas X e Y, temos as mesmíssimas definições de Covariância e Correlação (e as mesmas propriedades!), apenas trocando os somatórios de dentro dos valores esperados pelas integrais correspondentes.

Definição 6 *A **Distribuição Marginal** de X será*

$$\boxed{f_X(x) = \int_{-\infty}^{\infty} f(x,y)\,dy}$$

*e a **Distribuição Condicional** de X dado Y será*

$$\boxed{f_{X|Y}(x|y) = \frac{f(x,y)}{f_Y(y)}}$$

Definição 7 *A **Covariância** e a **Correlação** entre duas variáveis X e Y são respectivamente*

$$\boxed{Cov(X,Y) = E((X - E(X))(Y - E(Y))) = E(XY) - E(X)E(Y)}$$
$$\boxed{\rho(X,Y) = \frac{Cov(X,Y)}{\sigma(X)\sigma(Y)}}$$

Definição 8 *A **Esperança Condicional** de X na certeza de que $Y = y$ é*

$$\boxed{E[X \mid Y = y] = \int_{-\infty}^{\infty} x f_{X|Y}(x|y)\,dx}$$

(às vezes denotada simplesmente $E[X|y]$).

Exemplo 9 (A) *Voltando ao exemplo A acima, temos as distribuições marginais*

$$\begin{aligned} f_X(x) &= \int_0^2 \frac{1}{4}dy = \frac{1}{2} \text{ (para } 0 < x < 2\text{)} \\ f_Y(y) &= \int_0^2 \frac{1}{4}dx = \frac{1}{2} \text{ (para } 0 < y < 2\text{)} \end{aligned}$$

ou seja, X e Y ambos têm distribuições uniformes em $[0,2]$. Também, dado y entre 0 e 2, tem-se

$$f_{X|Y}(x|y) = \frac{1/4}{1/2} = \frac{1}{2} \text{ (para } 0 < x < 2\text{)}$$

isto é, mesmo que o valor de $Y = y$ seja dado, a distribuição de X não se altera. Como

$$E(XY) = \int_0^2 \int_0^2 xy.\frac{1}{4}\,dy\,dx = \int_0^2 \frac{x}{2}dx \int_0^2 \frac{y}{2}dy = E(X)E(Y)$$

temos $Cov(X,Y) = 0$. Enfim, note que $E(X|y) = E(X) = 1$ para $0 < y < 2$.

Exemplo 10 (B) *Voltando ao exemplo B, temos as marginais*

$$\begin{aligned} f_X(x) &= \int_0^2 \frac{xy}{4} dy = \frac{x}{2} \text{ (para } 0 < x < 2\text{)} \\ f_Y(y) &= \int_0^2 \frac{xy}{4} dx = \frac{y}{2} \text{ (para } 0 < y < 2\text{)} \end{aligned}$$

e, dado y entre 0 e 2, temos a condicional

$$f_{X|Y}(x|y) = \frac{f(x,y)}{f_Y(y)} = \frac{xy/4}{y/2} = \frac{x}{2} \text{ para } 0 < x < 2$$

Para calcular $Cov(X,Y)$ notamos que:

$$E(XY) = \int_0^2 \int_0^2 xy\frac{xy}{4}\, dy\, dx = \left(\int_0^2 x\frac{x}{2}dx\right)\left(\int_0^2 y\frac{y}{2}dy\right) = E(X)E(Y)$$

e, portanto, novamente, $Cov(X,Y) = 0$. Enfim, note que $E(X|y) = E(X) = \frac{8}{6}$ para $0 < y < 2$.

Exemplo 11 (C) *Escolha um ponto agora uniformemente no círculo C de centro $(0,0)$ e raio 1. Então*

$$f(x,y) = \begin{cases} \frac{1}{\pi} & \text{se } (x,y) \in C \\ 0, & \text{caso contrário} \end{cases}$$

Por simetria, vê-se que

$$\begin{aligned} E(X) &= \int\int_C \frac{x}{\pi} dA = 0 \\ E(Y) &= \int\int_C \frac{y}{\pi} dA = 0 \\ E(XY) &= \int\int_C \frac{xy}{\pi} dA = 0 \end{aligned}$$

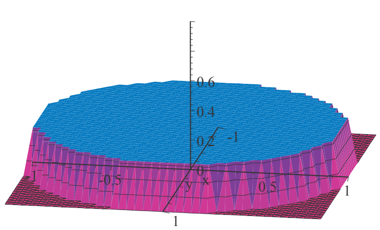
e, assim, $Cov(X,Y) = \rho(X,Y) = 0$ de novo. No entanto

$$\begin{aligned} f_X(x) &= \int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} \frac{1}{\pi} dy = \frac{2}{\pi}\sqrt{1-x^2} \text{ para } -1 < x < 1 \\ f_Y(y) &= \int_{-\sqrt{1-y^2}}^{\sqrt{1-y^2}} \frac{1}{\pi} dx = \frac{2}{\pi}\sqrt{1-y^2} \text{ para } -1 < y < 1 \end{aligned}$$

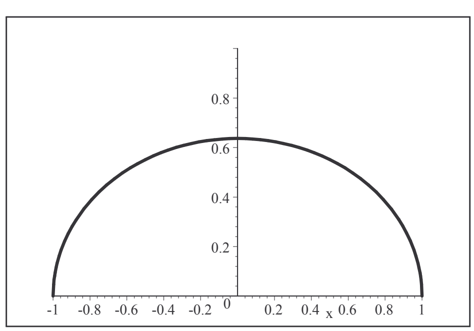
enquanto, dado $-1 < y < 1$:

$$f_{X|Y}(x|y) = \frac{1/\pi}{\frac{2}{\pi}\sqrt{1-y^2}} = \frac{1}{2\sqrt{1-y^2}} \text{ para } -\sqrt{1-y^2} < x < \sqrt{1-y^2}$$

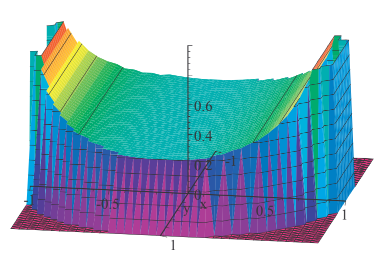
ou seja, dado $Y = y$, a distribuição de X é uniforme, a saber, $X \sim U\left[-\sqrt{1-y^2}, \sqrt{1-y^2}\right]$! Portanto, $E[X|y] = 0$ qualquer que seja $y \in [-1,1]$.



C) $f(x,y) = 1/\pi$ em C



C) $f_X(x) = \frac{2}{\pi}\sqrt{1-x^2}$



C) $f_{X|Y}(x|y) = \frac{1}{2\sqrt{1-y^2}}$ em C

Definição 12 *As variáveis X e Y são* ***independentes*** *quando a sua densidade conjunta é o produto das marginais, isto é,*

$$\boxed{f(x,y) = f_X(x)\, f_Y(y)}$$

Nos exemplos A e B acima, X e Y são independentes, mas em C, X e Y não o são. De fato, é fácil ver que:

Proposição 13 *X e Y são independentes se, e somente se, $f(x,y)$ é da forma $g(x)\,h(y)$ num retângulo da forma $[a,b]\times[c,d]$ (e 0 caso contrário; note que este retângulo pode ser "infinito").*

Prova. De fato, se $f(x,y) = g(x)\,h(y)$ em $[a,b]\times[c,d]$, temos

$$\begin{aligned} \int\int g(x)\,h(y)\,dA &= 1 \Rightarrow \left(\int_a^b g(x)\,dx\right)\left(\int_c^d h(y)\,dy\right) = 1 \\ f_X(x) &= \int_c^d g(x)\,h(y)\,dy = g(x)\int_c^d h(y)\,dy \\ f_Y(y) &= \int_a^b g(x)\,h(y)\,dx = h(y)\int_a^b g(x)\,dx \end{aligned}$$

Então

$$f_X(x)\,f_Y(y) = g(x)\,h(y)\left(\int_c^d h(y)\,dy\right)\left(\int_a^b g(x)\,dx\right) = g(x)\,h(y) = f(x,y)$$

e as variáveis são independentes. Por outro lado, se X e Y são independentes, basta tomar $g(x) = f_X(x)$ e $h(y) = f_Y(y)$ e acabou. ■

Proposição 14 *Se X e Y são independentes, $Cov(X,Y) = 0$ e $E[X|y] = E(X)$ para valores válidos de y (isto é, sempre que $f_Y(y) \neq 0$).*

Prova. De fato, neste caso

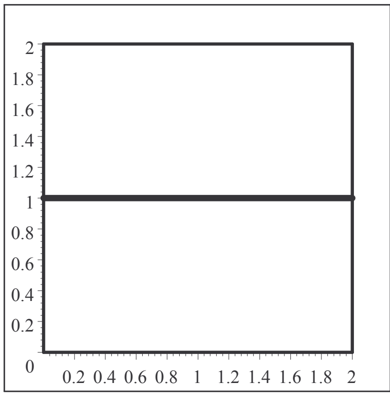
$$\begin{aligned} E(XY) &= \int\int xyf(x,y)\,dA = \int_a^b\int_c^d xyf_X(x)\,f_Y(y)\;dy\,dx = \\ &= \left(\int_a^b xf_X(x)\,dx\right)\left(\int_c^d yf_Y(y)\,dy\right) = E(X)\,E(Y) \\ E(X|y) &= \int_{-\infty}^{\infty} xf_{X|Y}(x|y)\,dx = \int_{-\infty}^{\infty} x\frac{f(x,y)}{f_Y(y)}dx = \int_{-\infty}^{\infty} xf_X(x)\,dx = E(X) \end{aligned}$$

■

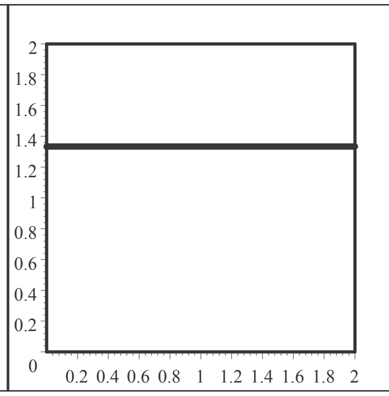
Definição 15 *Os gráficos de $E[X|y]$ (uma função de y) e $E[Y|x]$ (uma função de x) são chamados de* ***curvas de regressão*** *(de X sobre y e vice-versa, respectivamente).*

Exemplo 16 *Nos 3 exemplos acima, temos que $E[X|y]$ e $E[Y|x]$ são constantes (nos suportes das respectivas densidades):*

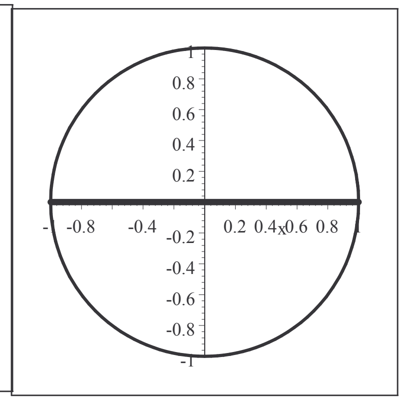
$$\begin{aligned} A &: E[X|y] = E[Y|x] = 1 \\ B &: E[X|y] = E[Y|x] = \frac{4}{3} \\ C &: E[X|y] = E[Y|x] = 0 \end{aligned}$$



A: $E[Y|x] = 1$.



B: $E[Y|x] = \frac{4}{3}$.



C: $E[Y|x] = 0$.

Exemplo 17 *Suponha que a densidade conjunta de X e Y é*

$$f(x,y) = \begin{cases} \frac{x+y}{8} & \text{para } (x,y) \in [0,2] \times [0,2] \\ 0, & \text{caso contrário} \end{cases}$$

Note que X e Y não são independentes pois f não é o produto de duas funções em x e y. Mais especificamente, note que as marginais são:

$$\begin{aligned} f_X(x) &= \int_0^2 \frac{x+y}{8} dy = \frac{x+1}{4} \text{ para } x \in [0,2] \text{ (e 0, caso contrário)} \\ f_Y(y) &= \int_0^2 \frac{x+y}{8} dx = \frac{y+1}{4} \text{ para } y \in [0,2] \text{ (e 0, caso contrário)} \end{aligned}$$

enquanto a condicional de X dado $Y = y$ é

$$f_{X|Y}(x|y) = \frac{f(x,y)}{f_Y(y)} = \frac{x+y}{8} \frac{4}{y+1} = \frac{x+y}{2(1+y)}$$

Note que o valor esperado de X dado $Y = y$ é

$$E[X|y] = \int_0^2 x \frac{x+y}{2(1+y)} dx = \frac{4+3y}{3(1+y)}$$

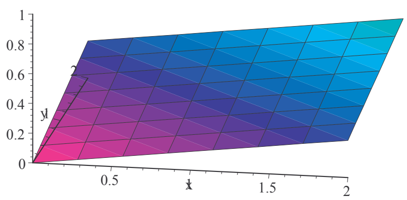
que é uma função decrescente de y! Assim, não é surpresa que X e Y estejam negativamente correlacionadas:

$$\begin{aligned} E(X) &= \int_0^2 \int_0^2 x \frac{x+y}{8} \, dx \, dy = \frac{7}{6} \\ E(X^2) &= \int_0^2 \int_0^2 x^2 \frac{x+y}{8} \, dx \, dy = \frac{5}{3} \Rightarrow Var(X) = \frac{5}{3} - \left(\frac{7}{6}\right)^2 = \frac{11}{36} \\ E(XY) &= \int_0^2 \int_0^2 xy \frac{x+y}{8} \, dx \, dy = \frac{4}{3} \\ Cov(X,Y) &= E(XY) - E(X)E(Y) = \frac{4}{3} - \frac{7}{6}\frac{7}{6} = -\frac{1}{36} \\ \rho(X,Y) &= \frac{Cov(X,Y)}{\sigma(X)\sigma(Y)} = \frac{-1/36}{11/36} = -\frac{1}{11} \end{aligned}$$

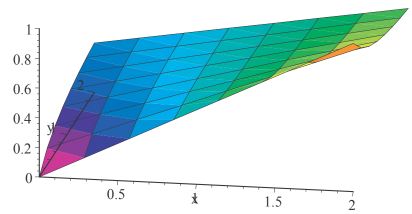
Enfim, note que $E[X|Y] = \frac{4+3Y}{3(1+Y)}$ é uma nova variável aleatória, cujo valor esperado é

$$E(E[X|Y]) = \int_{-\infty}^{\infty} E(X|y) f_Y(y) \, dy = \int_0^2 \frac{4+3y}{3(1+y)} \frac{y+1}{4} dy = \int_0^2 \frac{4+3y}{12} dy = \frac{7}{6}$$

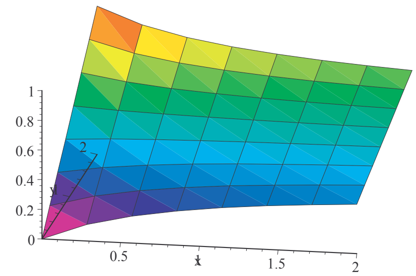
ou seja, $E(E[X|Y]) = E(X)$! Coincidência? Não! É um teorema (veja exercícios).



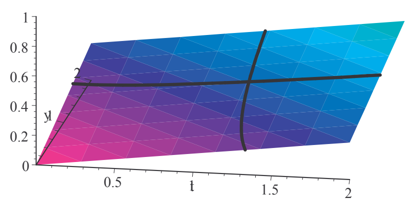
$f(x,y) = \frac{x+y}{8}$



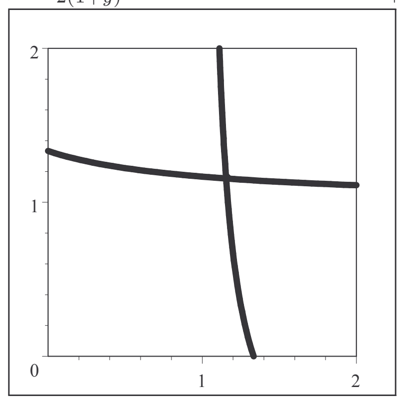
$f_{X|Y}(x|y) = \frac{x+y}{2(1+y)}$



$f_{Y|X}(y|x) = \frac{x+y}{2(1+x)}$



$E[X|y]$ e $E[Y|x]$ no espaço...



...e no plano XY.

6.2.1 Exercícios Ilustrados

Ex. 1 *A densidade conjunta de* (X,Y) *é*

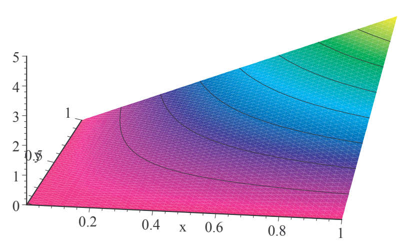
$$f(x,y) = \begin{cases} kxy, \text{ se } 0 \le x \le 1 \text{ e } 0 \le y \le 1 \\ 0, \text{ caso contrário} \end{cases}$$

a) Determine o valor da constante k.
b) Determine as densidades marginais de X *e* Y.
c) X *e* Y *são independentes? Calcule* $Cov(X,Y)$.
d) Calcule $\Pr(X+Y \le 1)$ *e* $\Pr\left(X+Y \le \frac{1}{2}\right)$.
e) Determine a densidade condicional $f_{Y|X}$ *e a esperança condicional* $E[Y \mid X = x]$.

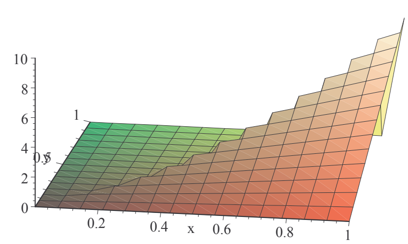
Ex. 2 *A densidade conjunta de* (X,Y) *é dada por*

$$f(x,y) = \begin{cases} kxy, \text{ se } 0 \le y \le x \le 1 \\ 0, \text{ caso contrário} \end{cases}$$

Repita os itens a-e do exercício anterior.



1) $f(x,y) = kxy$



2) $f(x,y) = kxy$ $(0 \le y \le x \le 1)$

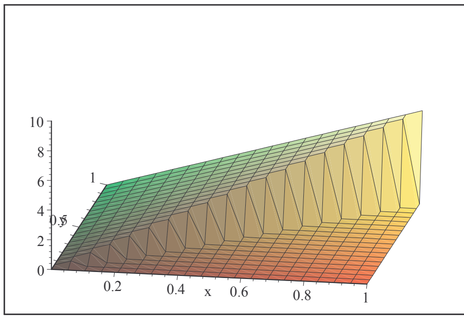
Ex. 3 *Sejam X e Y variáveis aleatórias independentes distribuídas uniformemente em $[0,1]$ Calcule $E\left(X^Y\right)$.*

Ex. 4 *A densidade conjunta de X e Y é*

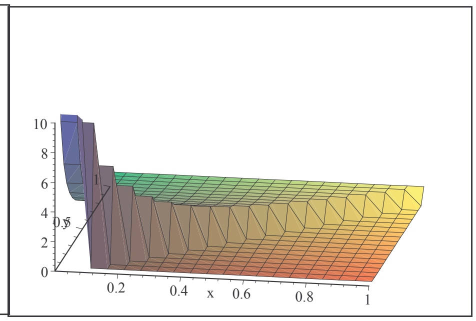
$$f(x,y) = \begin{cases} 6x, \text{ se } 0 \le x \le y \le 1 \\ 0, \text{ caso contrário} \end{cases}$$

Determine $E[X \mid Y = y]$ (onde $0 < y < 1$).

Ex. 5 *Y tem distribuição $U(0;1)$ e, na certeza de $Y = y$, X tem distribuição $U(0;y)$. Determine $Pr(X > 0.4)$ e calcule $Cov(X,Y)$.*



4) $f(x,y) = 6x$ $(0 \le x \le y \le 1)$



5) $f(x,y) = ??$ $(0 \le x \le y \le 1)$

Ex. 6 *Se $X \sim Exp(\lambda)$ e $\alpha > 0$, determine $E[X \mid X > \alpha]$.*

Ex. 7 *Note que $E[X|y]$ é uma função de y, digamos, $g(y)$. Então $E[X|Y] = g(Y)$ é uma variável aleatória que depende de Y (e não de X). Podemos então falar do valor médio desta variável aleatória, a saber, $E(E[X|Y])$. Mostre que*

$$E(E[X|Y]) = E(X)$$

6.3 Funções de Variáveis Contínuas

Dadas variáveis aleatórias contínuas X e Y com densidade conjunta $f_{X,Y}(x,y)$ e dadas $W = g(X,Y)$ e $Z = h(X,Y)$, como obter a distribuição conjunta de W e Z? O teorema a seguir resume o resultado:

Teorema 18 *Sejam X e Y variáveis aleatórias com densidade $f_{X,Y}(x,y)$. Sejam $W = g(X,Y)$ e $Z = h(X,Y)$ duas novas variáveis. Então a densidade conjunta de W e Z é dada por*[1]

$$\boxed{f_{W,Z}(w,z) = f_{X,Y}(x,y) \cdot |J|}$$

onde J é o Jacobiano da transformação $T : (x,y) \mapsto (g(x,y), h(x,y))$, a saber:

$$\boxed{J = \frac{\partial(x,y)}{\partial(w,z)} = \begin{vmatrix} \frac{\partial x}{\partial w} & \frac{\partial x}{\partial z} \\ \frac{\partial y}{\partial w} & \frac{\partial y}{\partial z} \end{vmatrix}}$$

[1] Se você não consegue lembrar se tem que usar as derivadas de x e y com relação a z e w ou vice-versa, este mnemônico vai te ajudar: no caso unidimensional, era

$$f(x)\,dx = g(y)\,dy$$

No caso bidimensional, é

$$f(x,y)\,\partial(x,y) = g(w,z)\,\partial(w,z)$$

Tecnicamente, isto não faz sentido algum, mas jogando um dos ∂ para o outro lado e lembrando de colocar um módulo para consertar tudo:

$$f(x,y) = g(w,z)\left|\frac{\partial(w,z)}{\partial(x,y)}\right|$$

ou, equivalentemente,

$$g(w,z) = f(x,y)\left|\frac{\partial(x,y)}{\partial(w,z)}\right|$$

(Tecnicamente, precisamos das seguintes hipóteses adicionais: J nunca se anula na região suporte de $f_{X,Y}(x,y)$; T é uma transformação diferenciável bijetiva, pelo menos entre o suporte de $f_{X,Y}(x,y)$ e o interior do suporte de $f_{W,Z}(w,z)$; $f_{X,Y}(x,y)$ é contínua em seu suporte).

Prova. O argumento a seguir não é uma demonstração formal, mas é convincente (e tem o espírito correto). Considere uma região R no plano XY cuja imagem $T(R)$ está no plano WZ. Como T é bijetiva, os eventos $(X,Y) \in R$ e $(W,Z) \in T(R)$ são equivalentes, e portanto:

$$\Pr((X,Y) \in R) = \Pr((W,Z) \in T(R))$$

Sendo $f_{W,Z}(w,z)$ a nova densidade conjunta, a equação acima se escreve

$$\int\int_R f_{X,Y}(x,y)\,dxdy = \int\int_{T(R)} f_{W,Z}(w,z)\,dwdz$$

Mas, do Cálculo, sabemos que, dentro das hipóteses acima,

$$\int\int_R f_{X,Y}(x,y)\,dxdy = \int\int_{T(R)} f_{X,Y}(x,y)\left|\frac{\partial(x,y)}{\partial(w,z)}\right| dwdz$$

onde $w = g(x,y)$, $z = h(x,y)$ e

$$\left|\frac{\partial(x,y)}{\partial(w,z)}\right| = \left|\begin{vmatrix} \frac{\partial x}{\partial w} & \frac{\partial x}{\partial z} \\ \frac{\partial y}{\partial w} & \frac{\partial y}{\partial z} \end{vmatrix}\right| = \left|\frac{\partial x}{\partial w}\frac{\partial y}{\partial z} - \frac{\partial x}{\partial z}\frac{\partial y}{\partial w}\right|$$

Assim, é de se esperar que

$$f_{W,Z}(w,z) = f_{X,Y}(x,y)\left|\frac{\partial(x,y)}{\partial(w,z)}\right|$$

■

Exemplo 19 (A) *Voltando ao exemplo A, escolha um ponto (X,Y) uniformemente no quadrado $Q = [0,2] \times [0,2]$. Qual é a distribuição conjunta de $Z = X^2$ e $W = Y^3$? Calculemos o determinante Jacobiano de (W,Z) com relação a (X,Y)*

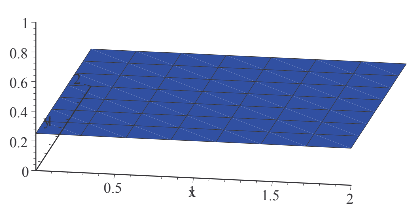
$$\frac{\partial(W,Z)}{\partial(X,Y)} = \begin{vmatrix} \frac{\partial W}{\partial X} & \frac{\partial W}{\partial Y} \\ \frac{\partial Z}{\partial X} & \frac{\partial Z}{\partial Y} \end{vmatrix} = \begin{vmatrix} 0 & 3Y^2 \\ 2X & 0 \end{vmatrix} = -6XY^2$$

Infelizmente, precisamos exatamente do Jacobiano inverso (de (X,Y) com relação a (W,Z))! Mas o coeficiente de correção será o inverso, isto é

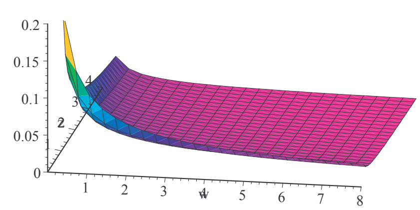
$$f_{Z,W}(z,w) = f_{X,Y}(x,y)\,|J| = \frac{1}{4}\frac{1}{|-6xy^2|} = \frac{1}{24}\frac{1}{\sqrt{z}}\frac{1}{\sqrt[3]{w^2}} \text{ para } 0 < x,y < 2$$

ou seja, a densidade conjunta de Z e W é (não se esqueça de calcular o novo suporte com z e w!):

$$f(z,w) = \begin{cases} \frac{1}{24} z^{-1/2} w^{-2/3}, \text{ para } (z,w) \in [0,4] \times [0,8] \\ 0, \text{ caso contrário} \end{cases}$$



$f(x,y) = \frac{1}{4}$



$f(z,w) = \frac{1}{24} z^{-1/2} w^{-2/3}$

Se quisermos encontrar a distribuição de apenas uma variável nova $Z = g(X, Y)$, há duas alternativas:

- Encontre a acumulada de Z fazendo a integral correspondente no plano XY. Afinal:

$$\Pr(Z \leq z) = \Pr(g(X, Y) \leq z) = \int\int_R f(x, y)\, dA$$

  onde R é a região definida por $g(X, Y) \leq z$.

- Invente uma variável qualquer W (por exemplo, $W = X$ ou $W = Y$) para acompanhar Z e encontre a nova densidade conjunta de Z e W. A partir daqui, integre com relação a W para encontrar a marginal de Z, que é a sua densidade.

Exemplo 20 *Suponha que a f.d.p. de X e Y é dada por*

$$f(x, y) = \begin{cases} e^{-x-y}, \text{ se } x \geq 0 \text{ e } y \geq 0 \\ 0, \text{ caso contrário} \end{cases}$$

Qual é a distribuição de $Z = X + Y$? Usando o primeiro método, basta calcular a acumulada de Z:

$$\Pr(Z \leq z) = \Pr(X + Y \leq z)$$

Para $z < 0$, esta probabilidade é claramente 0. Para z positivo, $X + Y \leq z$ é (no plano XY) o triângulo T de vértices $(0,0)$, $(0,z)$ e $(z,0)$. Então:

$$\begin{aligned} \Pr(X + Y \leq z) &= \int\int_T e^{-x-y} dA = \int_0^z \int_0^{x-z} e^{-x-y}\, dy\, dx = \\ &= \int_0^z e^{-x}\left(-e^{-y}\right]_{y=0}^{y=x-z} dx = \int_0^z e^{-x}\left(1 - e^{x-z}\right) dx = \\ &= \int_0^z e^{-x} - e^{-z} dx = \left(-e^{-x}\right]_{x=0}^{x=z} - ze^{-z} = 1 - e^{-z} - ze^{-z} \end{aligned}$$

Daqui é fácil obter a densidade de Z. Para $z > 0$:

$$f(z) = \frac{d\left(1 - e^{-z} - ze^{-z}\right)}{dz} = ze^{-z}$$

Ou seja

$$f(z) = \begin{cases} ze^{-z}, \text{ para } z \geq 0 \\ 0, \text{ caso contrário} \end{cases}$$

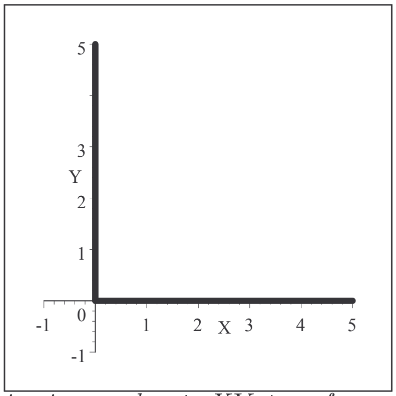
Exemplo 21 *Refaremos o exemplo acima usando o segundo método. Seja, por exemplo, $W = X$. Então*

$$\frac{\partial(w, z)}{\partial(x, y)} = \begin{vmatrix} \frac{\partial w}{\partial x} & \frac{\partial w}{\partial y} \\ \frac{\partial z}{\partial x} & \frac{\partial z}{\partial y} \end{vmatrix} = \begin{vmatrix} 1 & 0 \\ 1 & 1 \end{vmatrix} = 1$$

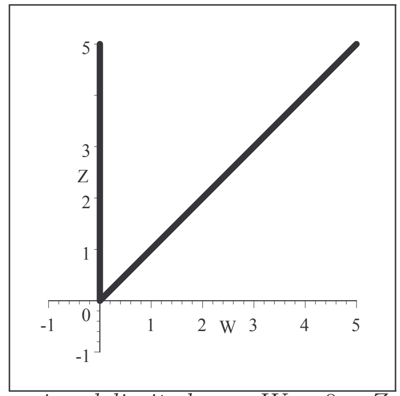
Portanto,

$$f(w, z) = \frac{f(x, y)}{1} = e^{-x-y} = e^{-z} \text{ para } x > 0 \text{ e } y > 0$$

Precisamos colocar a região suporte em termos de w e z! A reta $x = 0$ se transforma em $w = 0$; a reta $y = 0$ se *transforma em $z = x + 0 = w$:*



$T(x,y) \underset{\leadsto}{=} (x, x+y)$



O primeiro quadrante XY transforma-se... *...na regiao delimitada por $W = 0$ e $Z = W$.*

Assim, a nova densidade é

$$f(w,z) = \begin{cases} e^{-z} \text{ se } 0 \le w \le z \\ 0, \text{ caso contrário} \end{cases}$$

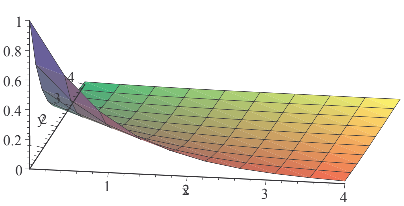
A marginal de Z é então

$$f_Z(z) = \int_0^z e^{-z} dw = z e^{-z} \text{ para } z \ge 0$$

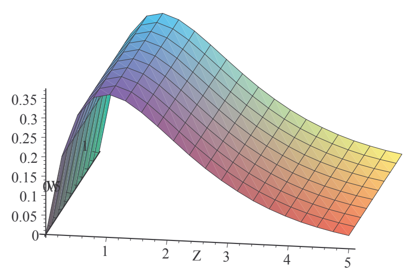
e 0 caso contrário.

6.3.1 Exercícios Ilustrados

Ex. 8 *X e Y são independentes com distribuição $Exp(1)$. Se $Z = X + Y$ e $W = \frac{X}{X+Y}$:*
a) Determine a densidade conjunta de Z e W.
b) Determine as densidades marginais de Z e W e identifique essas distribuições marginais.
c) Verifique se Z e W são independentes.



$(z,w) \underset{\leadsto}{=} \left(x+y, \frac{x}{x+y}\right)$



$f(x,y) = ??$ $f(z,w) = ??$

Ex. 9 *A densidade conjunta de (X,Y) é*

$$f(x,y) = \begin{cases} 4xy, \text{ se } 0 \le x \le 1 \text{ e } 0 \le y \le 1 \\ 0, \text{ caso contrário} \end{cases}$$

a) Encontre a densidade de $Z = X + Y$.
b) Encontre a densidade de $W = X - Y$.
c) Calcule a correlação entre Z e W. As variáveis W e Z são independentes?

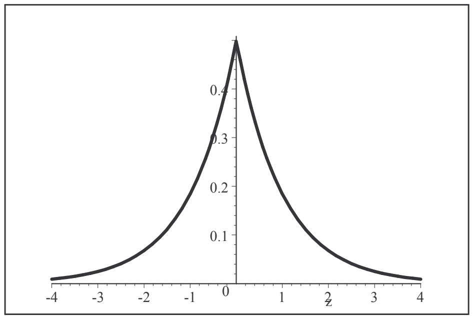
Ex. 10 *Repita os itens do exercício anterior para a densidade*

$$f(x,y) = \begin{cases} 8xy, \text{ se } 0 \le y \le x \le 1 \\ 0, \text{ caso contrário} \end{cases}$$

Ex. 11 *X e Y são independentes com distribuição $Exp(\lambda)$. Mostre que $Z = X - Y$ tem* ***distribuição de Laplace de parâmetros*** *$\mu = 0$ e $b = \frac{1}{\lambda}$*[2]*, isto é, sua densidade é dada por*

$$f(z) = \frac{\lambda e^{\lambda|z|}}{2}$$

Qual seria a densidade de Z se $X \sim Exp(\lambda_1)$ e $Y \sim Exp(\lambda_2)$?



Laplace de parâmetros $\mu = 0$ e $b = 1$.

Ex. 12 *Sejam X e Y variáveis independentes com densidades dadas por $f(x)$ e $g(y)$, respectivamente. Seja $Z = X + Y$. Mostre que a densidade de Z pode ser escrita da forma*

$$h(z) = \int_{-\infty}^{\infty} f(t)\, g(z-t)\, dt$$

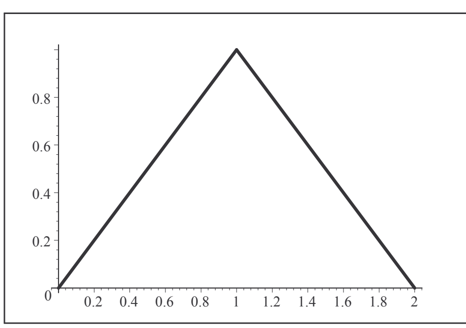
desde que esta integral seja convergente. Esta nova função h é chamada de ***convolução*** *de f com g e é representada por $h = f * g$.*

Ex. 13 *Se X e Y são variáveis independentes com distribuição $U[0,1]$ mostre que $Z = X + Y$ tem densidade dada por*

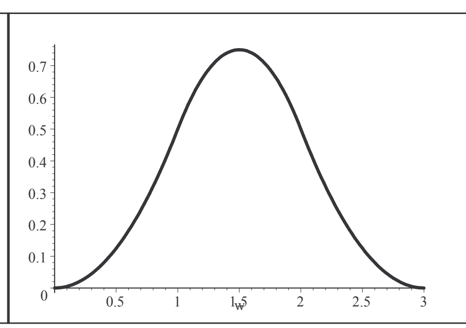
$$f(z) = \begin{cases} z, \text{ se } 0 \le z \le 1 \\ 2 - z, \text{ se } 1 \le z \le 2 \\ 0, \text{ caso contrário} \end{cases}$$

Ex. 14 *X, Y e Z são independentes e têm distribuição $U[0,1]$. Mostre que a densidade de $W = X + Y + Z$ é*

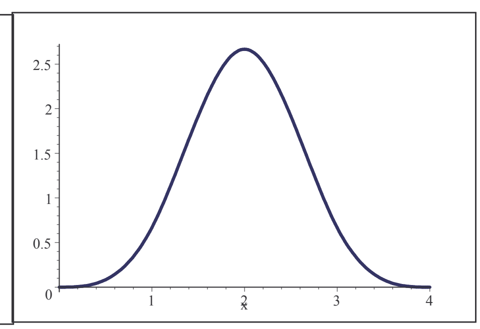
$$f(w) = \begin{cases} \frac{w^2}{2}, \text{ se } 0 < w < 1 \\ -w^2 + 3w - \frac{3}{2}, \text{ se } 1 < w < 2 \\ \frac{w^2}{2} - 3w + \frac{9}{2}, \text{ se } 2 < w < 3 \\ 0, \text{ caso contrário} \end{cases}$$



Soma de 2 var. indep. $U[0,1]$.



Soma de 3 var. indep. $U[0,1]$.



Soma de 4 var. indep. $U[0,1]$.

[2] Por definição, a distribuição de Laplace de parâmetros μ e b tem densidade dada por

$$f(x) = \frac{1}{2b} e^{-\frac{|x-\mu|}{b}}$$

Ela é uma variante da distribuição normal com $E(X) = \mu$ e $Var(X) = 2b^2$. Em termos de $\sigma = \sqrt{2}b$, teríamos

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

Ex. 15 *X e Y são independentes com distribuição $Exp(\lambda_1)$ e $Exp(\lambda_2)$, respectivamente. Mostre que, se $\lambda_1 \neq \lambda_2$, então $Z = X + Y$ tem densidade dada por*

$$f(z) = \begin{cases} \frac{\lambda_1\lambda_2}{\lambda_2-\lambda_1}\left(e^{-\lambda_1 z} - e^{-\lambda_2 z}\right), & \text{se } z \geq 0 \\ 0, & \text{caso contrário} \end{cases}$$

O que acontece quando λ_1 e λ_2 se aproximam de um valor comum λ? Qual é o nome desta distribuição?

Ex. 16 *Sejam X e Y independentes com distribuições $N(0,1)$ cada. Mostre que, as variáveis $W = aX + bY$ e $Z = bX - aY$ são independentes e determine as suas distribuições marginais.*

Ex. 17 *X e Y são independentes e têm distribuição $N(0,1)$. Determine a densidade de $W = \frac{Y}{X}$.*

Ex. 18 *A **distribuição normal bidimensional centrada em** $(0,0)$ é a distribuição cuja densidade é dada por*

$$f(x,y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2(1-\rho^2)}\left(\frac{x^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2} - \frac{2\rho xy}{\sigma_x\sigma_y}\right)\right)$$

onde σ_x, σ_y são constantes positivas e $-1 \leq \rho \leq 1$.
a) Mostre que, se $\rho = 0$, então X e Y são variáveis independentes com distribuições marginais $N\left(0,\sigma_x^2\right)$ e $N\left(0,\sigma_y^2\right)$, respectivamente.
b) No caso geral $\rho \neq 0$, encontre a densidade conjunta de X e $Z = \frac{Y}{\sigma_Y} - \rho\frac{X}{\sigma_X}$ e mostre que estas variáveis são independentes. Em particular, conclua que a marginal de X ainda é $N\left(0,\sigma_x^2\right)$ (analogamente, mostra-se que a marginal de Y ainda é $N\left(0,\sigma_y^2\right)$).
c) Mostre que a correlação entre X e Y é exatamente ρ. [Dica: $Cov(X,Z) = 0$, então...]
d) Mostre que as curvas de nível $f(x,y) = c$ são elipses (ou, no caso $\sigma_x = \sigma_y$, circunferências).

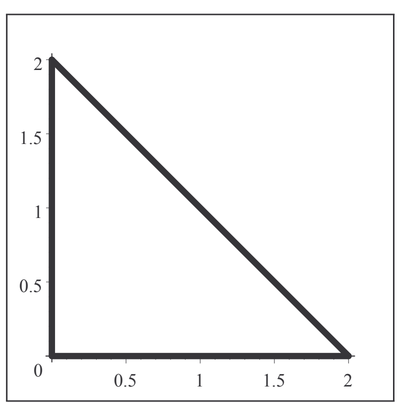
Ex. 19 *Suponha que a distribuição de X e Y é normal bidimensional com médias 0, desvios-padrão 1 e correlação ρ, isto é*

$$f(x,y) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2(1-\rho^2)}\left(x^2 + y^2 - 2\rho xy\right)\right)$$

*a) Mostre que $E[Y \mid X = x] = \rho x$. [Dica: sabemos que X e $Z = Y - \rho X$ são independentes; use $Y = Z + \rho X$.]. Em particular, se $0 < \rho < 1$, note que, sempre que $x > \mu_X = 0$, tem–se $E[Y \mid X = x] = \rho x < x$. Este fenômeno é chamado **retorno à média**. Por exemplo, X pode ser a altura de um pai e Y a altura de seu filho. Apesar de haver uma correlação positiva entre X e Y, filhos de pais que são mais altos do que a média tendem a ser mais baixos que os pais (mas ainda mais altos do que a média). Outro exemplo: estudantes que tiraram notas altas na A1 terão, em média, notas mais baixas na A2 (e vice-versa: os estudantes que tiraram notas baixas na A1 tendem a tirar notas mais altas na A2). Isto não é sinal de cansaço, excesso de estudo ou sugestão de que não vale a pena estudar para a A1 – é simplesmente o retorno à média em ação.*

6.4 Exercícios de Provas

Ex. 20 (A2 2004.2) *A distribuição conjunta de duas variáveis aleatórias X e Y é uniforme (isto é, tem densidade constante e igual a k) na região da figura.*



a) Determine o valor de k
b) Encontre a distribuição marginal de X.
c) Calcule a covariância de X e Y.
d) Qual é a probabilidade condicional de que X seja maior que $\frac{1}{2}$ dado que $Y = 1$?

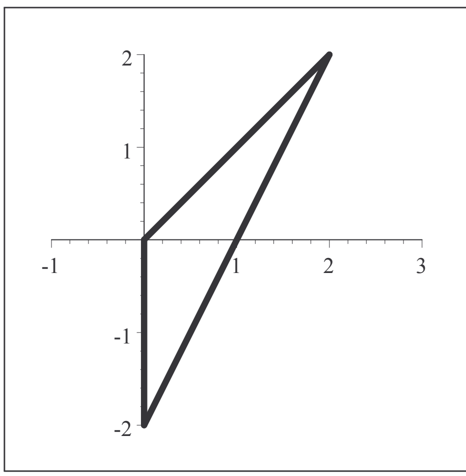
Ex. 21 (AS 2004.2) *Dois componentes de um sistema eletrônico têm tempos de vida dados, respectivamente, por X e Y (em meses). Suponha que a distribuição conjunta de X e Y é dada pela distribuição:*

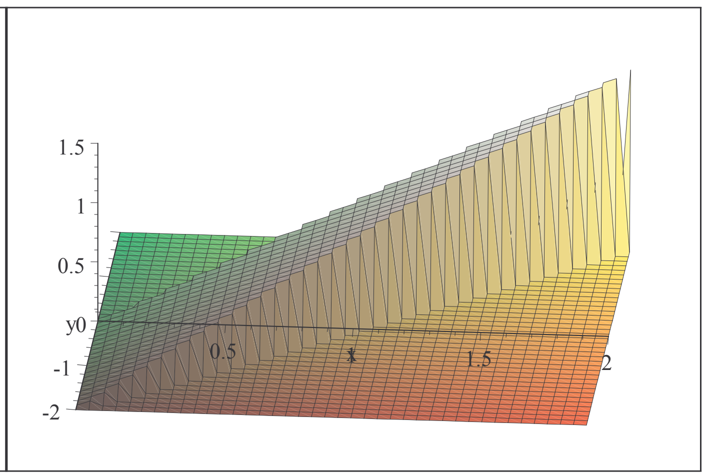
$$f(x,y) = \begin{cases} 3e^{-x-3y}, & se\ x, y > 0 \\ 0, & caso\ contrário \end{cases}$$

a) Calcule a densidade marginal de X.
b) Calcule a densidade condicional de X dado que $Y = y$ e determine $Cov(X,Y)$.
c) Qual a probabilidade do componente de tempo de vida X durar mais que o componente de tempo de vida Y?
d) Para que o sistema funcione, é necessário que pelo menos um dos dois componentes esteja funcionando. Qual a probabilidade do sistema funcionar pelo menos 1 mês?

Ex. 22 (T5 2005.2) *Considere o triângulo T de vértices $(0,0)$, $(0,-2)$ e $(2,2)$ (veja figura). X e Y são duas variáveis aleatórias cuja densidade conjunta é dada por*

$$f(x,y) = \begin{cases} \frac{3x}{4}, & se\ (x,y) \in T \\ 0, & caso\ contrário \end{cases}$$



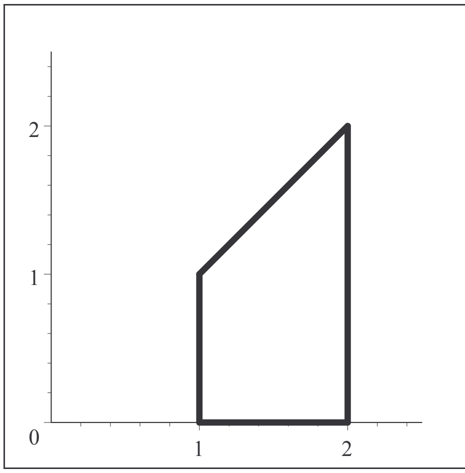


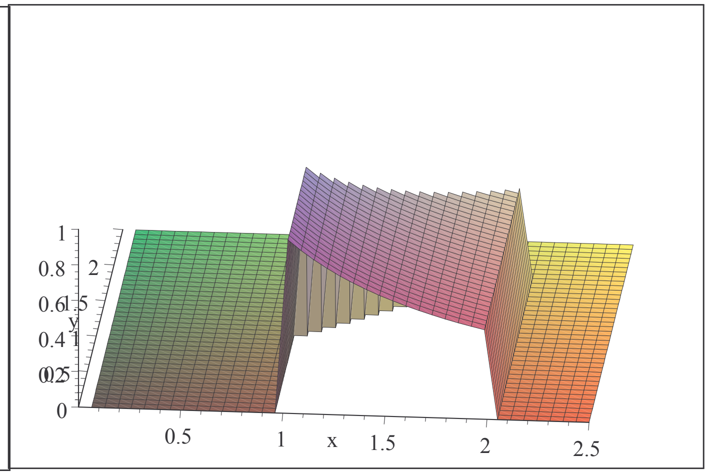
a) Mostre que f é, de fato, uma densidade conjunta de probabilidade e encontre a densidade marginal de X.
b) Calcule $\Pr(Y \leq 0 \mid X \leq 1)$.
c) Encontre $E[Y|x]$ e desenhe a curva de regressão correspondente.
d) Sabe-se que $E(XY) = E(Y^2) = \frac{4}{5}$, $E(X^2) = \frac{6}{5}$ e $E(Y) = \frac{1}{2}$. Calcule $\rho(X,Y)$. As variáveis X e Y são independentes?
e) Seja $Z = Y - \frac{3X}{2} + 1$. Encontre a densidade conjunta de X e Z e calcule $Cov(X,Z)$.

Ex. 23 (A2 2005.2) *As variáveis aleatórias X e Y tem densidade conjunta dada por*

$$f_{X,Y}(x,y) = \begin{cases} \frac{1}{x}, & se\ (x,y) \in R \\ 0, & caso\ contrário \end{cases}$$

onde R é o trapézio de vértices $(1,0)$, $(2,0)$, $(2,2)$ e $(1,1)$.



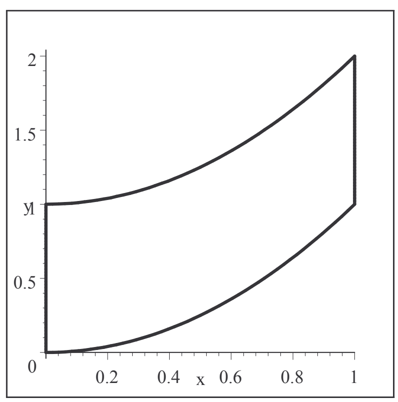


a) Encontre e esboce o gráfico da densidade marginal de Y.
b) Encontre a densidade conjunta de X e $Z = \frac{Y}{X}$. Mostre que X e Z são independentes. Que distribuição marginal têm X e Z? Brevemente, determine $E(X)$, $E(Z)$, $E(X^2)$ e $E(Z^2)$.
c) Calcule $E(Y)$ e $Var(Y)$. [Dica: você pode usar $Y = XZ$.]
d) Escolhem-se 100 pontos (X, Y) independentemente de acordo com esta distribuição. Seja W o número de pontos escolhidos que ficam dentro da região S definida por $Y \geq 1$. Estime $\Pr(W \geq 40)$ (Usar o Excel).
e) Calcule $\Pr\left(X \geq \frac{3}{2} \mid Y \geq 1\right)$.
f) Encontre $E[Y|x]$ e desenhe a curva de regressão correspondente.

Ex. 24 (AS 2005.2) *As variáveis X e Y tem densidade conjunta dada por*

$$f(x,y) = \begin{cases} \frac{1}{2\sqrt{y-x^2}}, & se\ (x,y) \in R \\ 0, & caso\ contrário \end{cases}$$

onde R é a região definida por $x^2 < y < x^2 + 1$ e $0 < x < 1$.



a) Mostre que X e $Z = Y - X^2$ são independentes.
b) Encontre a distribuição marginal de X e mostre que $E(X^n) = \frac{1}{n+1}$ para $n > 0$.
c) Calcule $Cov(X, Y)$. [Sugestão: use $Y = Z + X^2$ e os itens anteriores.]
d) Encontre $E[Y|X = x]$ e esboce a curva de regressão correspondente. [Dica: use Z de novo!]

Chapter 7

Somas e Médias de Variáveis Aleatórias

7.1 Motivação: Inferência Estatística

Até aqui, temos nos preocupado com o seguinte problema: dada uma variável aleatória X com uma distribuição conhecida (digamos, $N(30, 400)$), encontre probabilidades (por exemplo, $\Pr(X \geq 60)$) e propriedades (como valor esperado, desvio-padrão, etc.) relacionadas à variável X.

Infelizmente, quando tentamos aplicar estes modelos à realidade, confrontamo-nos com o seguinte dilema: que modelo usar para cada situação? As "estórias" que vêm por trás de cada distribuição ajudam a escolher os modelos, mas, mesmo assim, ainda pode ser difícil descobrir os parâmetros a serem usados (os p's, λ's, μ's e σ's corretos).

Exemplo 1 *Suponha que queremos saber se um determinado referendo terá como resultado "sim" ou "não". Temos então que criar um modelo para uma **POPULAÇÃO COM** N **ELEITORES**, cada um deles votando "sim" ou "não" (digamos por agora que estas são as únicas duas opções de cada eleitor). Um **MODELO** razoável para esta população seria imaginar que, tomando um eleitor aleatoriamente desta população, ele dirá "sim" com probabilidade* p *e "não" com probabilidade* $1-p$*. Assim, se* X *representa a resposta de um eleitor escolhido ao acaso (*$X=1$ *significa "sim" e* $X=0$ *significa "não"),* X *satisfaz o modelo de Bernoulli, e só falta descobrir o **PARÂMETRO*** p*... Como? Uma idéia é tomar uma **AMOSTRA** de* n *eleitores e perguntar a eles o que eles votariam – ou seja, supomos que realizamos* n *provas de Bernoulli. Desta amostra, tomamos a seguinte **ESTATÍSTICA:** o número* S_n *de eleitores (na amostra) que disse "sim". A partir desta estatística* S_n*, queremos fazer inferências sobre possíveis valores de* p*: será que* $p \geq 0.5$ *é uma hipótese razoável? Será que* $p \approx \frac{S_n}{n}$ *(isto é, a proporção na população seria parecida com a proporção da amostra)?*

A idéia da Estatística é tentar estimar os modelos (incluindo parâmetros) de determinadas variáveis aleatórias. O exemplo acima inclui os termos e perguntas básicas da estatística, a saber:

POPULAÇÃO: todos os elementos/resultados sob investigação; freqüentemente, confundimos a população em si com um **MODELO** de probabilidade a ela associado – afinal de contas, é o **modelo** que estamos tentando estimar! Este modelo terá **PARÂMETROS** a serem estimados/aceitados/rejeitados no processo de inferência estatística.

AMOSTRA: qualquer subconjunto da população.

ESTATÍSTICA: qualquer função das amostras (de preferência, uma função que seja útil de alguma forma para descobrir/aceitar/rejeitar os parâmetros do modelo da população!).

Assim, as perguntas básicas a serem entendidas no processo de inferência estatística são:

a) Qual a população a ser amostrada?

Você tem que pensar com cuidado qual é a população que você QUER estudar; freqüentemente, é possível esboçar um modelo inicial razoável neste estágio, com alguns parâmetros a serem verificados ou estimados.

b) Como obter os dados (a amostra)?

No que se segue, suporemos que as amostragens serão do tipo **AAS (Amostragem Aleatória Simples)**, isto é, sorteiam-se elementos da população aleatoriamente (de maneira que todos sejam equiprováveis) independentemente uns dos outros (portanto, com reposição). Assim, no caso em que a população é representada por uma variável X contínua, teríamos que a distribuição conjunta da amostra seria uma função de n variáveis (!):

$$f(x_1, x_2, ..., x_n) = f(x_1) f(x_2) ....f(x_n)$$

onde $f(x)$ é a f.d.p. de X. Note-se que fazer uma amostragem que realmente tenha estas características (eqüiprobabilidade e independência) pode ser operacionalmente bastante complicado!

c) Que informações pertinentes (estatísticas) serão retiradas da amostra?

No exemplo acima decidimos contar apenas o número S_n de "sim" dentro da amostra. Este S_n foi a estatística utilizada.

d) Como se comportam as estatísticas quando esta amostragem é utilizada em populações conhecidas?

Por exemplo, se $X \sim Bernoulli(p)$, qual é a real distribuição de S_n? Resposta: usando AAS, $S_n \sim Bin(n, p)$. Precisamos estudar este problema de Probabilidade (a partir dos modelos, entender como se comportam as estatísticas) para poder entender melhor o problema de Estatística (a partir das estatísticas, inferir os modelos; isto é, a partir da estatística S_n, inferir algo sobre o parâmetro p). Tais distribuições das estatísticas são chamadas *Distribuições Amostrais*. Neste capítulo, estudaremos as Distribuições Amostrais da Soma (e da Média) de variáveis aleatórias.

7.2 Somas das Principais Distribuições Aleatórias

Definição 2 *Seja $(X_1, X_2, ..., X_n)$ uma AAS tomada a partir de uma distribuição de uma v.a. X. Definimos a* ***soma da amostra*** *e a* ***média amostral*** *respectivamente por*

$$\boxed{S_n = X_1 + X_2 + ... + X_n}$$

$$\boxed{\bar{X} = \frac{S_n}{n} = \frac{X_1+X_2+...+X_n}{n}}$$

que, claramente, também são variáveis aleatórias.

Nem sempre é fácil encontrar a distribuição de S_n (e de $\bar{X}$). Há alguns casos que já conhecemos (lembre-se: em cada caso, supõe-se que todos os X_i têm a mesma distribuição de X e são independentes entre si, ou seja, estamos fazendo amostragens aleatórias simples):

- Se $X \sim Be(p)$, então $S_n \sim Bin(n, p)$;
- Se $X \sim Bin(m, p)$, então $S_n \sim Bin(mn, p)$;
- Se $X \sim Geom(p)$, então $S_n \sim NegBin(n, p)$;
- Se $X \sim NegBin(m, p)$, então $S_n \sim NegBin(mn, p)$;
- Se $X \sim Poi(\lambda)$ então $S_n \sim Poi(n\lambda)$;
- Se $X \sim Exp(\lambda)$ então $S_n \sim Gamma(n, \lambda)$;
- Se $X \sim Gamma(m, \lambda)$ então $S_n \sim Gamma(mn, \lambda)$;

Todas as afirmações acima são facilmente justificáveis diretamente a partir das definições (as "estórias") que estão por trás de cada distribuição (ou, se você preferir, por cálculos usando os métodos das seções 2.1.3 e 6.3).

Já se X é uniforme a distribuição de S_n é surpreendentemente complicada:

Exemplo 3 *Suponha que X é uniforme em $[0, 1]$. Então a distribuição de $\bar{X}$ dependerá do valor de n. Por exemplo, para $n = 2$, a f.d.p. de $\bar{X}$ é fácil de calcular (ex. 6.13):*

$$f_2(x) = \begin{cases} 4x, \text{ se } 0 < x < \frac{1}{2} \\ 4(1-x), \text{ se } \frac{1}{2} < x < 1 \\ 0, \text{ caso contrário} \end{cases}$$

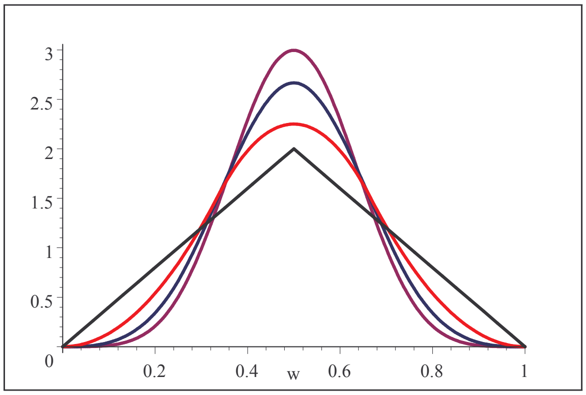
e, para $n=3$ e $n=4$ temos um pouco mais de trabalho (ex. 6.14):

$$f_3(x)=\begin{cases}\frac{27x^2}{2}, \text{ se } 0<x<\frac{1}{3}\\ -27\left(x-\frac{1}{2}\right)^2+\frac{9}{4}, \text{ se } \frac{1}{3}<x<\frac{2}{3}\\ \frac{27(1-x)^2}{2}, \text{ se } \frac{2}{3}<x<1\\ 0, \text{ caso contrário}\end{cases} \quad ; \quad f_4(x)=\begin{cases}\frac{128x^3}{3}, \text{ se } 0<x<\frac{1}{4}\\ -128x\left(x-\frac{1}{2}\right)^2+\frac{8}{3}, \text{ se } \frac{1}{4}<x<\frac{1}{2}\\ f_4(1-x), \text{ se } \frac{1}{2}<x<1\\ 0, \text{ caso contrário}\end{cases}$$

Como você pode ver, é complicado escrever uma fórmula geral que valha para todo n. É possível mostrar que

$$f_n(x)=\begin{cases}\sum_{j=0}^{k}(-1)^j\frac{n^2}{(n-j)!j!}(nx-j)^{n-1}, \text{ para } \frac{k}{n}\le x\le\frac{k+1}{n}\\ 0, \text{ caso contrário}\end{cases}$$

Os gráficos de f_2 a f_5 estão abaixo:



Para completar nossa biblioteca de "somas conhecidas", aqui vai a proposição básica que leva ao caso normal (cuja prova é essencialmente o exercício 6.16):

Proposição 4 *Se $X_1\sim N\left(\mu_1,\sigma_1^2\right)$ e $X_2\sim N\left(\mu_2,\sigma_2^2\right)$ são independentes, então qualquer combinação linear não-nula $X=aX_1+bX_2$ também terá distribuição normal, a saber, $X\sim N\left(\mu,\sigma^2\right)$ onde $\mu=a\mu_1+b\mu_2$ e $\sigma^2=a^2\sigma_1^2+b^2\sigma_2^2$. Em suma:*

$$\boxed{\left.\begin{array}{c}X_1\sim N\left(\mu_1,\sigma_1^2\right)\\ X_2\sim N\left(\mu_2,\sigma_2^2\right)\\ X_1 \text{ e } X_2 \text{ independentes}\end{array}\right\}\Rightarrow aX_1+bX_2\sim N\left(a\mu_1+b\mu_2,a^2\sigma_1^2+b^2\sigma_2^2\right)}$$

Prova. Inicialmente, analisemos o caso em que $Z_1\sim N(0,1)$, $Z_2\sim N(0,1)$ são normais-padrão independentes. As variáveis Z_1 e Z_2 teriam distribuição conjunta dada por

$$f_{Z_1,Z_2}(z_1,z_2)=\frac{1}{2\pi}\exp\left(\frac{-z_1^2-z_2^2}{2}\right)$$

Considere as novas variáveis $Z=aZ_1+bZ_2$ e $W=bZ_1-aZ_2$. Note que

$$Z^2+W^2=\left(a^2+b^2\right)\left(Z_1^2+Z_2^2\right)$$

enquanto o determinante Jacobiano para perfazer a mudança de variáveis é

$$J=\left|\frac{\partial(Z,W)}{\partial(Z_1,Z_2)}\right|=\begin{vmatrix}a & b\\ b & -a\end{vmatrix}=a^2+b^2$$

ou seja

$$\begin{aligned}f_{Z,W}(z,w) &= \frac{f_{Z_1,Z_2}(z_1,z_2)}{J}=\frac{1}{2\pi\left(a^2+b^2\right)}\exp\left(-\frac{z^2+w^2}{2\left(a^2+b^2\right)}\right)=\\ &= \left(\frac{1}{2\pi\sqrt{a^2+b^2}}\exp\left(-\frac{z^2}{2\left(a^2+b^2\right)}\right)\right)\left(\frac{1}{2\pi\sqrt{a^2+b^2}}\exp\left(-\frac{w^2}{2\left(a^2+b^2\right)}\right)\right)\end{aligned}$$

Daqui, vemos claramente que Z e W são independentes, e a distribuição marginal de Z deve ser

$$f_Z(z) = \frac{1}{2\pi\sqrt{a^2+b^2}} \exp\left(-\frac{z^2}{2(a^2+b^2)}\right)$$

ou seja, $Z \sim N\left(0, a^2+b^2\right)$.
No caso geral em que $X_1 \sim N\left(\mu_1, \sigma_1^2\right)$ e $X_2 \sim N\left(\mu_2, \sigma_2^2\right)$, basta tomar $\mu = a\mu_1 + b\mu_2$ e escrever

$$X = aX_1 + bX_2 \Leftrightarrow X - \mu = (a\sigma_1)\frac{X_1-\mu_1}{\sigma_1} + (b\sigma_2)\frac{X_2-\mu_2}{\sigma_2}$$

ou seja, $Z = X - \mu$ é uma combinação linear (de coeficientes $a\sigma_1$ e $b\sigma_2$) de $Z_1 = \frac{X_1-\mu_1}{\sigma_1}$ e $Z_2 = \frac{X_2-\mu_2}{\sigma_2}$ (ambas normais padrão!). Portanto, recaímos no caso anterior e concluímos que $Z = X - \mu \sim N\left(0, (a\sigma_1)^2 + (b\sigma_2)^2\right)$, isto é, $X \sim N\left(\mu, a^2\sigma_1^2 + b^2\sigma_2^2\right)$. ■

Em particular, tomando $a = b = 1$, vamos que se $X_1 \sim N\left(\mu_1, \sigma_1^2\right)$ e $X_2 \sim N\left(\mu_2, \sigma_2^2\right)$ são independentes, então $X_1 + X_2 \sim N\left(\mu_1+\mu_2, \sigma_1^2+\sigma_2^2\right)$. Por indução, é fácil então ver que:

- $\boxed{\text{Se } X_i \sim N\left(\mu,\sigma^2\right) \text{ são independentes, então } S_n \sim N\left(n\mu, n\sigma^2\right) \text{ e } \bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)}$

Este fato é muito importante e merece ser repetido num formato "normal padrão":

Teorema 5 *Se $X_1, X_2, ..., X_n$ são amostras independentes da distribuição $N\left(\mu,\sigma^2\right)$ então*

$$\boxed{\frac{S_n - n\mu}{\sqrt{n}\sigma} \sim N(0,1)}$$

$$\boxed{\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)}$$

7.3 Lei dos Grandes Números

E se a distribuição de X_i não for uma destas clássicas? Por exemplo, digamos que X_1, X_2, ..., X_n são uma amostragem aleatória simples de uma variável X com $E(X) = \mu$ e $\sigma(X) = \sigma$. O que podemos afirmar sobre S_n e $\bar{X}$? Pelo menos a média e variância destas variáveis é fácil de achar:

Proposição 6

$$\boxed{E(S_n) = n\mu; \quad Var(S_n) = n\sigma^2; \quad \sigma(S_n) = \sqrt{n}\sigma}$$

$$\boxed{E\left(\bar{X}\right) = \mu; \quad Var\left(\bar{X}\right) = \frac{\sigma^2}{n}; \quad \sigma(X) = \frac{\sigma}{\sqrt{n}}}$$

Prova. Basicamente, este é o exercício 2.33. Em suma

$$\begin{aligned} E\left(\sum X_i\right) &= \sum E(X_i) = n\mu;\ Var\left(\sum X_i\right) = \sum Var(X_i) = n\sigma^2 \\ E\left(\bar{X}\right) &= E\left(\frac{S_n}{n}\right) = \frac{n\mu}{n} = \mu; \quad Var\left(\bar{X}\right) = Var\left(\frac{S_n}{n}\right) = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n} \end{aligned}$$

onde **para as variâncias** utilizamos que X_1, X_2, ..., X_n são independentes[1]. ■

Em particular, observe o que acontece com $\bar{X}$: à medida que n cresce, $E\left(\bar{X}\right)$ permanece constante mas $Var\left(\bar{X}\right)$ se aproxima de 0. Parece então que a distribuição de $\bar{X}$ fica cada vez mais concentrada em $E(X)$ (veja

[1] Se X_1, X_2, X_3, ..., X_n fosse uma amostragem aleatória **sem reposição** a partir de uma população com N elementos, então

$$\begin{aligned} Var(S_n) &= \sum_i Var(X_i) + 2\sum_{i,j} Cov(X_i, X_j) = \\ &= n\sigma^2 + n(n-1)\,Cov(X_1, X_2) \end{aligned}$$

no exemplo 3 acima como a distribuição de $\bar{X}$ vai se concentrando em volta de $E(X) = 0.5$). Como formalizar esta intuição? Quem traduz "variância pequena" no fato "probabilidade concentrada" é a **desigualdade de Chebyshev**: se $E(X) = \mu$ e $\sigma(X) = \sigma$, então para qualquer k positivo

$$\Pr(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2}$$

Agora é fácil provar a versão fraca da **Lei dos Grandes Números (LGN)** ou "Lei da Médias":

Teorema 7 (Lei dos Grandes Números, Bernoulli, 1713) *Suponha que X tem uma distribuição tal que $\mu = E(X)$ e $\sigma^2 = Var(X)$ são finitos. Seja $\bar{X} = \frac{X_1+X_2+...+X_n}{n}$ onde X_1, X_2, X_3, ..., X_n são provas independentes com a distribuição de X. Então, para qualquer $\varepsilon > 0$ fixo, temos:*

$$\Pr\left(\left|\bar{X} - \mu\right| \geq \varepsilon\right) \to 0$$

quando $n \to \infty$. Conseqüentemente, quando $n \to \infty$

$$\boxed{\Pr\left(\left|\bar{X} - \mu\right| < \varepsilon\right) \to 1}$$

Prova. Colocando $\bar{X}$ na desigualdade de Chebyshev e lembrando que $E\left(\bar{X}\right) = E(X) = \mu$ e $\sigma\left(\bar{X}\right) = \sigma/\sqrt{n}$:

$$0 \leq \Pr\left(\left|\bar{X} - \mu\right| \geq \frac{k\sigma}{\sqrt{n}}\right) \leq \frac{1}{k^2}$$

O k pode ser qualquer. Escolha $k = \varepsilon\sqrt{n}/\sigma$:

$$0 \leq \Pr\left(\left|\bar{X} - \mu\right| \geq \varepsilon\right) \leq \frac{\sigma^2}{n\varepsilon^2}$$

Pelo Teorema do Sanduíche (Confronto), temos que

$$\lim_{n\to\infty} \Pr\left(\left|\bar{X} - \mu\right| \geq \varepsilon\right) = 0$$

■

Vale a pena destacar que este teorema vale **independentemente da distribuição original** de X (desde que esta tenha $E(X)$ e $Var(X)$ finitos).

Em particular, suponha que X é uma prova de Bernoulli com probabilidade p de sucesso (quando $X = 1$). Assim, $\bar{X}$ será simplesmente a proporção $\hat{p}$ de sucessos a longo prazo. A Lei acima indica que, para qualquer $\varepsilon \geq 0$, esta proporção satisfará

$$\Pr(|\hat{p} - p| \geq \varepsilon) \to 0$$

quando $n \to \infty$, ou seja

$$\boxed{\Pr(|\hat{p} - p| < \varepsilon) \to 1}$$

Isto significa que **finalmente temos alguma base para justificar a tal "interpretação freqüentista" das probabilidades citada no início do livro: a Lei dos Grandes Números diz que a proporção $\hat{p}$ de sucessos a longo prazo (muito muito provavelmente) ficará próxima da probabilidade p do evento**[2]**.**

Como para $n = N$ deveríamos ter $Var(S_n) = 0$, concluímos que $Cov(X_1, X_2) = \frac{-\sigma^2}{N-1}$. Assim:

$$Var(S_n) = n\sigma^2\left(1 - \frac{n-1}{N-1}\right) = n\sigma^2\left(\frac{N-n}{N-1}\right)$$

e, portanto

$$Var(\bar{X}) = \frac{\sigma^2}{n}\left(\frac{N-n}{N-1}\right)$$

O fator $\frac{N-n}{N-1}$ é chamado **fator de correção para populações finitas**.

[2] Existe uma versão mais forte da lei dos Grandes Números que basicamente diz que

$$\Pr(\hat{p} \to p) = 1$$

quando $n \to \infty$, mas para tanto precisaríamos definir o significado de "a probabilidade de uma seqüência $\hat{p}$ tender para p"...

Para a estatística, esta lei é muito útil: por exemplo, se quisermos estimar a proporção de eleitores p que vai votar "sim" num referendo (como no exemplo inicial deste capítulo), podemos tomar uma amostra de n eleitores e calcular $\bar{X} = \tilde{p}$ (a proporção de eleitores **da amostra** que disse "sim"). A Lei dos Grandes Números garante que, à medida que n cresce, a probabilidade de $\tilde{p}$ estar próximo de p se aproxima de 1! Nos exercícios e nas próximas seções veremos maneiras de estimar o mínimo n necessário de forma que possamos ter uma boa certeza de que p já está bem próximo de $\tilde{p}$.

7.3.1 Exercícios

Ex. 1 *Sejam* $X_1 \sim U[-1,1]$ *e* $X_2 \sim U[0,1]$ *independentes. Encontre a distribuição de* $S = X_1 + X_2$.

Ex. 2 *Sejam* $X_1 \sim Exp(\lambda)$ *e* $X_2 \sim U[0,1]$ *independentes. Encontre a distribuição de* $S = X_1 + X_2$.

Ex. 3 *Qual é a relação entre a densidade* $f(x)$ *da variável* S_n *e a densidade* $g(y)$ *da variável* $\bar{X}$ *no caso contínuo?*

Ex. 4 *Numa urna há 4 bolas, numeradas 1, 2, 3 e 3. Retire, com reposição, bolas* $X_1, X_2, ..., X_n$ *desta urna.*
a) Encontre a distribuição de $S_2 = X_1 + X_2$ *e de* $S_2/2$.
b) Encontre a distribuição de $S_3 = S_2 + X_3 = X_1 + X_2 + X_3$ *e de* $S_3/3$.
c) Encontre a distribuição de S_4 *e de* $S_4/4$.
d) Faça os histogramas das distribuições de $S_1 = X_1$, S_2, S_3 *e* S_4. *Que tipo de curva os histogramas parecem formar à medida que n cresce?*

Ex. 5 *O suporte de uma função* $f(x)$ *é o conjunto dos valores de* x *onde* $f(x) > 0$. *Sejam* X *e* Y *duas variáveis aleatórias independentes tais que o suporte da densidade* $f_X(x)$ *é* $[a,b]$ *e o suporte da densidade* $f_Y(y)$ *é* $[c,d]$. *Encontre o suporte da densidade da variável aleatória* $Z = X + Y$. *A resposta pode mudar caso* X *e* Y *não sejam independentes?*

Ex. 6 *Seja* $X \sim N(\mu, 1)$ *e considere uma AAS de tamanho 16 de* X. *Calcule* $\Pr(|X - \mu| < 1)$ *e compare-a com* $\Pr\left(\left|\bar{X} - \mu\right| < 1\right)$.

Ex. 7 *Se* $X \sim N(3,4)$ *e* $Y \sim N(7,1)$ *são independentes, qual é a distribuição de:*
a) $5X$ *b)* $X + Y$ *c)* $-X$ *d)* $X - Y$ *e)* $2Y - 3X$

Ex. 8 *Suponha que* $X_1, X_2, ..., X_{10} \sim N\left(100, 10^2\right)$. *Calcule* $\Pr(95 \le X \le 105)$ *e compare-a com* $\Pr\left(95 \le \bar{X} \le 105\right)$. *Quantas amostras (ao invés de 10) deveríamos tomar para que* $\Pr\left(95 \le \bar{X} \le 105\right)$ *fosse 95% ou mais?*

Ex. 9 *Figos têm massa normalmente distribuídas com média* $60g$ *e desvio-padrão* $8g$. *Determine a probabilidade da massa de uma dúzia de figos ser superior a* $750g$.

Ex. 10 *As notas de um aluno em 3 testes são variáveis aleatórias* $P_1 \sim N(7,1)$, $P_2 \sim N(6,2)$ *e* $P_3 \sim N(5, 2.76)$. *Qual a distribuição da média* $\frac{P_1+P_2+P_3}{3}$ *deste aluno? Se a nota mínima para passar é 6, qual a chance deste aluno passar? E se a nota mínima fosse 7? Qual a chance de* P_3 *ser pior que a média das outras duas? [Dica: faça* $X = P_1 + P_2 - 2P_3$*]*

Ex. 11 *O peso de um passageiro de elevador é supostamente* $X \sim N(80, 100)$. *Suponha que a capacidade máxima de um elevador é de* $800kg$.
a) Qual a probabilidade de 7 passageiros ultrapassarem a capacidade do elevador? E 8? E 9?
b) Qual o número máximo de passageiros para que haja 99% de chance do elevador suportar a carga?
c) Para quanto devemos aumentar a capacidade do elevador para que haja 99% de chance dele suportar 9 passageiros?

Ex. 12 *Uma empacotadora de café alega que o peso do café em gramas de cada um dos pacotes é uma variável aleatória* $X \sim N\left(1000, 5^2\right)$. *Se isto for verdade, qual a probabilidade da média dos pesos de 10 pacotes de café pesar menos que* $995g$*?*

Ex. 13 *Há duas populações* $X \sim N\left(\mu_1, \sigma_1^2\right)$ *e* $Y \sim N\left(\mu_2, \sigma_2^2\right)$. *Faça duas amostragens aleatórias simples de tamanhos m e n (respectivamente) destas duas populações e obtenha assim duas médias:* $\bar{X}$ *e* $\bar{Y}$. *Qual será a distribuição de* $D = \bar{X} - \bar{Y}$*?*

Ex. 14 *Numa determinada função de uma empresa, o salário dos funcionários do sexo masculino é* $H \sim N\left(600, 10^2\right)$ *enquanto o salário das funcionárias do sexo feminino é ligeiramente menor:* $M \sim N\left(590, 10^2\right)$*. Selecione uma amostra com* 25 *homens e* 25 *mulheres nesta função desta empresa, cujas médias salariais são respectivamente* $\bar{H}$ *e* $\bar{M}$*.*
a) Calcule a probabilidade de um funcionário que recebe mais de 610 *ser do sexo masculino.*
b) Calcule $\Pr\left(\bar{H} - \bar{M} \geq a\right)$ *para os valores* $a = 0$, 2 *e* 10.
c) Refaça os itens anteriores se a média salarial das funcionárias fosse igual à dos homens (isto é, se $M \sim N\left(600, 10^2\right)$*).*

Ex. 15 *Lance uma moeda* 100 *vezes. Seja* S_{100} *o número de caras obtidas. Sabemos que a distribuição de* S_{100} *é* $Bin\,(100, 0.5)$*, com média* 50 *e variância* $100\,(0.5)\,(0.5) = 25$*. Assim, o desvio-padrão é* 5*. O que a Desigualdade de Chebyshev diz sobre a probabilidade de o número de caras estar a menos de* 3 *desvios-padrão da média, isto é,* $\Pr\left(|S_{100} - 50| \leq 15\right)$*? Compare esta desigualdade com o valor real desta probabilidade calculada via binomial acumulada.*

Ex. 16 *Escolha* X_1, X_2,...X_n *uniformemente em* $[0, 1]$*. Mostre que*

$$\Pr\left(\left|\bar{X} - \frac{1}{2}\right| \geq \varepsilon\right) \leq \frac{1}{12n\varepsilon^2}$$

Em particular, calcule a probabilidade do "erro" $\left|\bar{X} - \frac{1}{2}\right|$ *ser maior que* 0.1 *para* $n = 100$, $n = 1000$ *e* $n = 10000$ *de acordo com esta estimativa.*

Ex. 17 *Escolha* X_1, X_2,...X_n *uniformemente em* $[\mu - h, \mu + h]$*. Use a desigualdade de Chebyshev para completar a expressão (em função de* n, ε *e* h*):*

$$\Pr\left(\left|\bar{X} - \mu\right| \geq \varepsilon\right) \leq \_\_\_\_\_\_\_\_$$

Em particular, calcule a probabilidade do "erro" $\left|\bar{X} - \mu\right|$ *ser maior que* 0.1 *para* $n = 100$, $n = 1000$ *e* $n = 10000$ *de acordo com esta estimativa.*

Ex. 18 *Suponha que* $X \sim U\,[\mu - 1, \mu + 1]$ *mas você não conhece* μ*. Estime o menor* n *tal que*

$$\Pr\left(\left|\bar{X} - \mu\right| \geq 0.1\right) \leq 5\%$$

Isto significa que, se você tomar n *amostras desta distribuição uniforme, você tem* 95% *de confiança de que a média* $\bar{X}$ *da amostra que você tem em mãos aproxima o* μ *desconhecido em uma casas decimais. Qual seria o* n *necessário para que haja* 95% *de confiança de que o erro é menr que* 0.01*?*

Ex. 19 *Se* $E\,(X) = 0$ *e* $Var\,(X) = 1$*, encontre* k *tal que você possa garantir que* $\Pr\left(|X| \leq k\right) \geq 99\%$*. Que valor de* k *você usaria se* X *fosse normal?*

Ex. 20 *Seja* $S_n \sim Bin\,(n, p)$ *e* $\bar{X} = \frac{S_n}{n}$*. Mostre que*

$$\Pr\left(\left|\bar{X} - p\right| \geq \varepsilon\right) \leq \frac{p\,(1-p)}{n\varepsilon^2} \leq \frac{1}{4n\varepsilon^2}$$

Se você quiser ter 95% *de certeza de que* $\bar{X}$ *está a menos de* $\varepsilon = 0.1$ *do valor real de* p*, que* n *você usaria?*

Ex. 21 *Sejam* X_1 *e* X_2 *variáveis independentes com distribuição de Cauchy:*

$$f\,(x) = \frac{1}{\pi\,(1 + x^2)}$$

Mostre que $\bar{X} = \frac{X_1 + X_2}{2}$ *também tem uma distribuição de Cauchy (use um sistema algébrico computacional para ajudar com as contas). Conclua que a Lei dos Grandes Números não vale para a média de* n *variáveis aleatórias com distribuição de Cauchy.*

7.4 Teorema Central do Limite (TCL)

7.4.1 TCL para Distribuições Binomiais

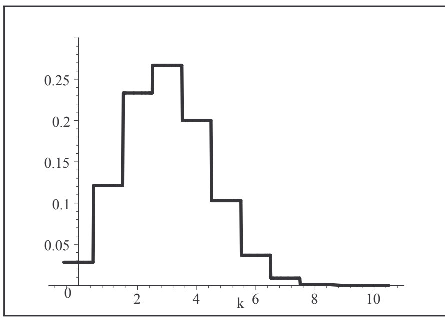
Seja X uma variável aleatória com distribuição binomial de parâmetros n e p (note que $X = S_n$ é a soma de n variáveis aleatórias de Bernoulli). Sabemos que $E(X) = np$ e $\sigma(X) = \sqrt{npq}$. Assim, consideremos a seguinte normalização de X:

$$A_n = \frac{X - np}{\sqrt{npq}}$$

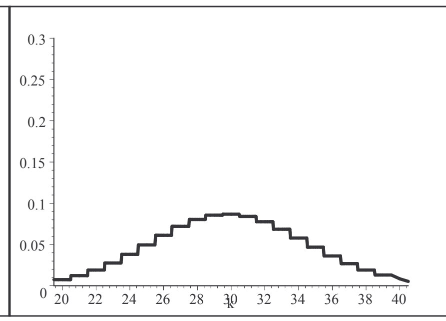
Claramente, $E(A_n) = 0$ e $\sigma(A_n) = 1$. A distribuição de A_n terá o mesmo formato do histograma da distribuição binomial (apenas re-graduando o eixo de X).

Considerando as distribuições de A_n para n cada vez maior, pode-se verificar que o formato destas distribuições parece cada vez mais se aproximar de uma distribuição normal (com os re-escalamentos e translações de eixos necessários).

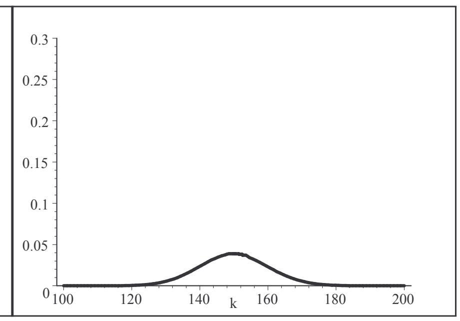
Exemplo 8 *Os histogramas abaixo mostram as funções de probabilidade de $X \sim Bin(n, 0.3)$ para diversos valores de n. Note como à medida que n cresce o formato da função de probabilidade parece se aproximar do formato de uma distribuição normal. Fazendo uma translação e um reescalamento em X, encontramos as distribuições para $A_n = \frac{X-np}{\sqrt{npq}}$. No entanto, como o eixo X foi re-escalado de um fator $\sqrt{npq}$, este fator deve ser recolocado nas alturas do histograma se quisermos que* ***as áreas*** *ainda somem 1. Usando estes histogramas* ***por áreas****, note como eles se aproximam de uma distribuição normal padrão.*



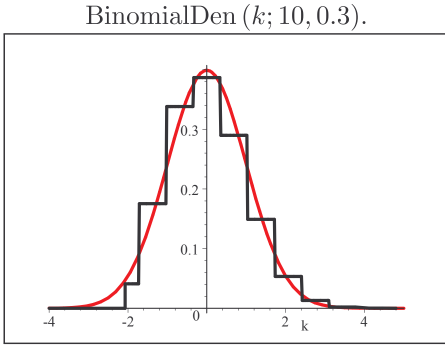
$\text{BinomialDen}(k; 10, 0.3)$.



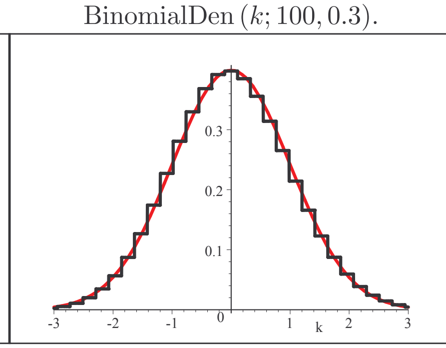
$\text{BinomialDen}(k; 100, 0.3)$.



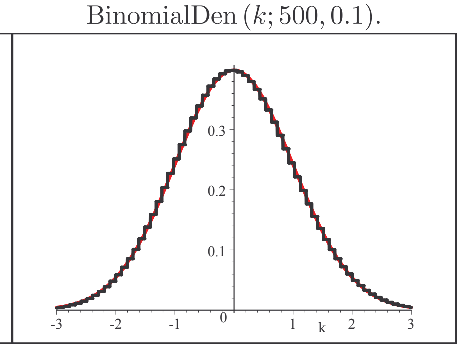
$\text{BinomialDen}(k; 500, 0.1)$.



$\sqrt{2.1} \times$ dist. de $A_{10} = \frac{X-3}{\sqrt{10(0.21)}}$



$\sqrt{21}\times$ dist. de $A_{100} = \frac{X-30}{\sqrt{100(0.21)}}$



$\sqrt{105} \times$ dist. de $A_{500} = \frac{X-150}{\sqrt{500(0.21)}}$

Em suma, a idéia é que

$$\boxed{\text{BinomialDen}(x; n, p) \approx \frac{1}{\sqrt{npq}}\phi\left(\frac{x-np}{\sqrt{npq}}\right) \text{ para } n \text{ grande}}$$

onde $\phi(\alpha) = \frac{1}{\sqrt{2\pi}} e^{-\alpha^2/2}$ é a densidade da normal-padrão.

Formalizar esta idéia infelizmente nos leva a alguns detalhes técnicos (como assim "n grande"?). A princípio, gostaríamos de dizer que

$$\lim_{n\to\infty} \sqrt{npq}\, \text{BinomialDen}(x; n, p) = \phi\left(\frac{x-np}{\sqrt{npq}}\right)$$

mas esta expressão não faz sentido pois, à medida que n cresce, o lado direito também muda e se aproxima de 0 (o que não é útil)! Façamos ao contrário: fixemos $a = \frac{x-np}{\sqrt{npq}}$ na distribuição normal do lado direito e procuremos o x da distribuição binomial que corresponda a este a. Como x deve ser inteiro, a opção mais próxima é:

$$x = [np + a\sqrt{npq}]$$

onde o símbolo $[\alpha]$ significa "o inteiro mais próximo de α". Agora o lado direito está fixo em $\phi(a)$, e podemos finalmente enunciar o teorema de forma precisa:

Teorema 9 (Aproximação Normal Pontual à Distribuição Binomial) *Seja $X \sim Bin(n,p)$. Então*

$$\lim_{n\to\infty} \sqrt{npq}\,\mathrm{BinomialDen}\left([np + a\sqrt{npq}]\,; n, p\right) = \phi(a)$$

onde $\phi(\alpha) = \mathrm{NormalDen}(\alpha)$ é a densidade normal padronizada, isto é,

$$\phi(\alpha) = \frac{1}{\sqrt{2\pi}} e^{-\alpha^2/2}$$

Prova. Vamos provar este teorema apenas no caso em que $a = 0$ e np é inteiro. Para tanto, precisamos da Fórmula de Stirling

$$\lim_{n\to\infty} \frac{n!}{\sqrt{2\pi n}\, n^n e^{-n}} = 1$$

Neste caso

$$\begin{aligned} \sqrt{npq}\,\mathrm{BinomialDen}(np; n, p) &= \sqrt{npq}\frac{n!}{(np)!\,(nq)!}p^{np}q^{nq} \simeq \\ &\simeq \sqrt{npq}\frac{\sqrt{2\pi n}\, n^n e^{-n}}{\sqrt{2\pi np}\,(np)^{np}\, e^{-np}\sqrt{2\pi nq}\,(nq)^{nq}\, e^{-nq}}p^{np}q^{nq} = \frac{1}{\sqrt{2\pi}} = \phi(0) \end{aligned}$$

Um método análogo pode ser utilizado nos outros casos, mas foge ao escopo deste livro. ■

A fórmula

$$\mathrm{BinomialDen}(x; n, p) \approx \frac{1}{\sqrt{npq}}\phi\left(\frac{x-np}{\sqrt{npq}}\right)$$

é freqüentemente denominada de **Aproximação Normal Pontual à Distribuição Binomial**.

Exemplo 10 *Lancemos uma moeda 100 vezes. Qual a probabilidade de obter 55 caras? Ora, a probabilidade exata é*

$$\mathrm{BinomialDen}(55; 100, 0.5) = \binom{100}{55}(0.5)^{100} = 4.847\%$$

mas a aproximação normal pode ser usada sem necessidade nem de uma tabela

$$\mathrm{BinomialDen}(55; 100, 0.5) \simeq \frac{\mathrm{NormalDen}\left(\frac{55-50}{5}\right)}{5} = \frac{\phi(1)}{5} = \frac{1}{5}\frac{1}{\sqrt{2\pi}}e^{-0.5} = 4.839415\%$$

O leitor atento notará que, dado $X = x$ fixo, há vários valores de $A_n = a$ que satisfazem o enunciado do TCL, pois

$$[np + a\sqrt{npq}] = x \Longleftrightarrow x - \frac{1}{2} \le np + a\sqrt{npq} \le x + \frac{1}{2} \Leftrightarrow \frac{x - \frac{1}{2} - np}{\sqrt{npq}} \le a \le \frac{x + \frac{1}{2} - np}{\sqrt{npq}}$$

A escolha que fizemos de usar $a = \frac{x-np}{\sqrt{npq}}$ parece então um tanto arbitrária – porque não utilizar algum outro dos valores de a no intervalo onde o teorema é válido?

Exemplo 11 *No exemplo anterior, poderíamos calcular* $\mathrm{NormalDen}(a)$ *para outros valores de a no intervalo definido por* $\left[\frac{54.5-50}{5}, \frac{55.5-50}{5}\right] = [0.9, 1.1]$ *e todos eles dariam aproximações razoáveis. De fato*

$$\begin{aligned} \frac{\mathrm{NormalDen}(0.9)}{5} &= \frac{\phi(0.9)}{5} = \frac{1}{5}\frac{1}{\sqrt{2\pi}}e^{-0.45} = 5.087\% \\ \frac{\mathrm{NormalDen}(1.1)}{5} &= \frac{\phi(1.1)}{5} = \frac{1}{5}\frac{1}{\sqrt{2\pi}}e^{-0.55} = 4.603\% \end{aligned}$$

também não estão muito longe dos corretos 4.847%. *Aliás, qualquer número entre* 0.9 *e* 1.1 *podia entrar ali... Que tal fazer uma média?*

$$\begin{aligned} \Pr(X=55) &\approx \frac{1}{1.1-0.9}\int_{0.9}^{1.1}\frac{\phi(t)}{5}dt = \int_{0.9}^{1.1}\phi(t)\,dt = \\ &= \text{NormalDist}(1.1) - \text{NormalDist}(0.9) = 4.8394064\% \end{aligned}$$

também dá um resultado muito bom que pode ser obtido de uma tabela da acumulada da normal.

Em outras palavras, o TCL acima diz que, se $X \sim Bin(n,p)$, então para x inteiro temos:

$$\Pr(X=x) = \text{BinomialDen}(x;n,p) \cong \frac{\phi(t)}{\sqrt{npq}}$$

para todo $t \in [x_1, x_2]$ onde

$$x_1 = \frac{x-\frac{1}{2}-np}{\sqrt{npq}} \text{ e } x_2 = \frac{x+\frac{1}{2}-np}{\sqrt{npq}}$$

Portanto, $\Pr(X=x)$ também estará bem próximo do valor médio de $\phi(t)/\sqrt{npq}$ neste intervalo:

$$\Pr(X=x) \cong \frac{1}{x_2-x_1}\int_{x_1}^{x_2}\frac{\phi(t)}{\sqrt{npq}}dt = \int_{x_1}^{x_2}\phi(t)\,dt = \text{NormalDist}(x_2) - \text{NormalDist}(x_1)$$

Esta nova formulação é muito útil pois pode ser usada para intervalos da distribuição binomial, como no teorema a seguir!

Teorema 12 (Aproximação Normal à Binomial) *Seja $X \sim Bin(n,p)$, e sejam a e b números inteiros fixos. Defina*

$$\boxed{a^* = \frac{a-np-\frac{1}{2}}{\sqrt{npq}} \quad \text{e } b^* = \frac{b-np+\frac{1}{2}}{\sqrt{npq}}}$$

Então

$$\boxed{\lim_{n\to\infty}\Pr(a \le X \le b) = \int_{a^*}^{b^*}\phi(t)\,dt}$$

ou seja, a distribuição de X pode ser bem aproximada por ***áreas*** *sob $N(np, npq)$ para n grande.*

Prova. (Esboço) O argumento dos últimos parágrafos mostra que

$$\Pr(X=x) \cong \frac{1}{x_2-x_1}\int_{x_1}^{x_2}\frac{\phi(x)}{\sqrt{npq}}dx = \int_{x_1}^{x_2}\phi(x)\,dx$$

pois $x_2 - x_1 = \frac{1}{\sqrt{npq}}$. Tomando agora o somatório para x variando de a até b, obtemos o resultado desejado! ■

Assim, podemos aproximar probabilidades binomiais do tipo $\Pr(a \le X \le b)$ utilizando tabelas da distribuição normal padrão sempre que n for "razoavelmente grande" (na prática, usa-se $n \ge 30$):

$$\boxed{X \sim Bin(n,p) \Rightarrow \Pr(a \le X \le b) \simeq \text{NormalDist}\left(\frac{b+\frac{1}{2}-np}{\sqrt{npq}}\right) - \text{NormalDist}\left(\frac{a-\frac{1}{2}-np}{\sqrt{npq}}\right)}$$

Exemplo 13 *Jogue uma moeda* 100 *vezes. Qual a probabilidade de obter de* 40 *a* 60 *caras? A resposta exata é terrível de calcular*

$$\text{BinomialDist}(60;100,0.5) - \text{BinomialDist}(39;100,0.5) = \sum_{k=40}^{60}\binom{100}{k}\left(\frac{1}{2}\right)^{100} = 96.480\%$$

mas podemos usar uma aproximação normal, tomando $a^* = \frac{39.5-50}{5} = -2.1$ e $b^* = \frac{60.5-50}{5} = 2.1$. Assim, usando uma tabela, temos

$$\Pr(40 \le X \le 60) \cong \text{NormalDist}(2.1) - \text{NormalDist}(-2.1) = 96.427\%$$

Analogamente, qual a probabilidade de obtermos entre 35 *e* 65 *caras? A resposta exata é* 99.821% *enquanto a aproximação normal nos daria*

$$\Pr(35 \le X \le 65) \cong \text{NormalDist}(3.1) - \text{NormalDist}(-3.1) = 99.806\%$$

Exemplo 14 *Na faculdade GHX, um determinado curso tem* 50 *vagas. Estima-se que há* 40% *de chance de um aluno que passa no vestibular se matricular de fato na GHX. Se a faculdade GHX chamar* 100 *alunos, qual a probabilidade de não haver vagas para estes alunos todos? Se cada aluno for uma prova de Bernoulli, o número de alunos que se matriculam é* $X \sim Bin(100, 0.4)$. *A probabilidade de haver problemas é*

$$\Pr(X \geq 51) = 1 - \text{BinomialDist}(50; 100, 0.4)$$

que pode ser calculada com o auxílio de um computador como sendo 1.676%. *Se não tivermos um computador, usamos a seguinte aproximação normal*

$$\Pr(X \geq 51) \cong 1 - \text{NormalDist}\left(\frac{50.5 - 40}{\sqrt{100(0.4)(0.6)}}\right) = 1 - \text{NormalDist}(2.1433) = 1.604\%$$

ou seja, a faculdade tem uma política de "overbooking" relativamente segura.

7.4.2 TCL (Caso Geral)

De fato, o Teorema Central do Limite vale para **qualquer** densidade de probabilidade que tenha média e variância finitas (sejam elas discretas[3] ou contínuas!), não só para Bernoulli/Binomial:

Teorema 15 (Teorema Central do Limite) *Sejam* X_1, X_2, ... X_n *amostras independentes de uma densidade com valor esperado* μ *e variância* σ^2. *Seja*

$$\boxed{X^* = \sqrt{n}\frac{\bar{X}-\mu}{\sigma} = \frac{S_n - n\mu}{\sigma\sqrt{n}}}$$

Então

$$\boxed{\lim_{n\to\infty} \Pr(a \leq X^* \leq b) = \int_a^b \phi(x)\, dx}$$

isto é, a distribuição de X^* *se aproxima (em áreas) da distribuição normal padrão.*

Provar que $E(X^*) = 0$ e $Var(X^*) = 1$ é um exercício simples – difícil é mostrar que a distribuição de X^* se aproxima de alguma forma de uma distribuição normal (mesmo que sua distribuição seja discreta para n fixo)! Não provaremos este fato aqui – o leitor interessado pode consultar *"Introduction to Probability Theory and its Applications"*, por W. Feller[4].

Em particular, este teorema indica que $NegBin(n, p)$ (soma de n variáveis tipo $Geom(p)$), $Poi(n\lambda)$ (soma de n variáveis tipo $Poi(\lambda)$), $Gamma(n, \lambda)$ (soma de n variáveis $Exp(\lambda)$), também se aproximam de uma distribuição normal à medida que n cresce! Assim:

$$\boxed{\text{Se } n \text{ é grande e } X \sim NegBin(n,p)\text{, então } X \approx N\left(\frac{n}{p}, \frac{nq}{p^2}\right)}$$

$$\boxed{\text{Se } \lambda \text{ é grande e } X \sim Poi(\lambda)\text{, então } X \approx N(\lambda, \lambda)}$$

$$\boxed{\text{Se } n \text{ é grande e } X \sim Gamma(n,\lambda)\text{, então } X \approx N\left(\frac{n}{\lambda}, \frac{n}{\lambda^2}\right)}$$

Lembre que, se $X_1, X_2, ..., X_n$ forem normais desde o princípio, então X^* **é normal padrão e o limite não é necessário** (ou seja, n não precisa ser "grande"). Uma ilustração do Teorema Central do Limite é o exemplo 3, onde a média de várias distribuições uniformes se aproxima de uma normal (apesar da distribuição original uniforme não se parecer nem um pouco com uma normal!).

---

[3] Para ser exato, no caso discreto é necessário assumir também que os possíveis valores da variável X sejam *comensuráveis*.

[4] De fato, este texto prova uma versão ainda mais geral do TCL: **mesmo que as distribuições de X_1, X_2, ..., X_n sejam diferentes,** a distribuição da soma S_n e da média $\bar{X}$ se aproximarão de distribuições normais (desde que algumas hipóteses adicionais sejam feitas sobre as distribuições dos X_i).

7.5 Aplicação à Estatística: Distribuição Amostral de uma Proporção

Como no exemplo do início deste capítulo, seja p a proporção de eleitores que vai votar "sim" num referendo (ou que vai votar num certo candidato, ou que torce para um certo time). Ao entrevistarmos um eleitor ao acaso, é razoável supor que sua resposta X é uma variável aleatória com distribuição de Bernoulli de parâmetro p. Sabemos que $E(X) = p$ e $Var(X) = p(1-p)$.

Seja $S_n = X_1 + X_2 + ... + X_n$ o número de "sim" encontrados ao entrevistarmos n eleitores escolhidos aleatoriamente e independentemente[5]. Sabemos que $S_n \sim Bin(n,p)$. A proporção de sucessos nesta amostra é

$$\hat{p} = \frac{S_n}{n} = \bar{X}$$

Mas, pela seção anterior, sabemos que a distribuição de S_n pode ser bem aproximada pela normal $N(np, npq)$ para n grande, isto é

$$\boxed{\hat{p} = \bar{X} = \frac{S_n}{n} \approx N\left(p, \frac{pq}{n}\right)}$$

Em particular, note que

$$\begin{aligned} \Pr\left(p - k\sqrt{\frac{pq}{n}} \leq \hat{p} \leq p + k\sqrt{\frac{pq}{n}}\right) &= F(k) - F(-k) = 2F(k) - 1 \Rightarrow \\ &\Rightarrow \Pr\left(|p - \hat{p}| \leq k\sqrt{\frac{pq}{n}}\right) = 2F(k) - 1 \end{aligned}$$

onde $F(k) = \text{NormalDist}(k)$ é a f.d.a. da normal-padrão. Como $pq = p(1-p)$ assume o valor máximo de $\frac{1}{4}$ quando $p = \frac{1}{2}$, podemos afirmar que

$$\Pr\left(|p - \hat{p}| \leq \frac{k}{2\sqrt{n}}\right) \geq \Pr\left(|p - \hat{p}| \leq k\sqrt{\frac{pq}{n}}\right) = 2F(k) - 1$$

Escreva $\frac{k}{2\sqrt{n}} = \varepsilon$ e obtenha

$$\Pr(|p - \hat{p}| \leq \varepsilon) \geq 2F\left(2\varepsilon\sqrt{n}\right) - 1$$

Assim, mesmo sem saber p, podemos escolher n de forma a garantir que $\hat{p}$ esteja a menos de ε unidades de p com uma certeza de pelo menos $2F(2\varepsilon\sqrt{n}) - 1$!

Exemplo 16 *Numa pesquisa de opinião pública sobre torcidas[6], o IBGE entrevistou 2000 pessoas maiores de 16 anos e descobriu que 15% destas declaravam ser torcedores do Flamengo. Qual a confiança de que a proporção real de torcedores do Flamengo esteja entre 14% e 16% (ou seja, do erro ao se tomar $\hat{p}$ por p ser de 1% ou menos)? Entre 10% e 20% (erro de 5%)? Que margem de erro ε você deveria usar para ter 95% de confiança de que $|p - \hat{p}| \leq \varepsilon$?*
Solução:

$$\begin{aligned} \Pr(|\tilde{p} - p| \leq 0.01) &\geq 2\,\text{NormalDist}\left(0.02\sqrt{2000}\right) - 1 = 62.89\% \\ \Pr(|\tilde{p} - p| \leq 0.05) &\geq 2\,\text{NormalDist}\left(0.1\sqrt{2000}\right) - 1 = 99.999226\% \\ 2\,\text{NormalDist}\left(2\varepsilon\sqrt{n}\right) - 1 &= 0.95 \Rightarrow 2\varepsilon\sqrt{n} = \text{NormalInv}(0.975) = 1.96 \Rightarrow \varepsilon = \frac{1.96}{2\sqrt{n}} = 2.19\% \end{aligned}$$

Em outras palavras, temos 63% de confiança de que o $\hat{p}$ encontrado ficará a menos de 1% do p real; temos 99.9992% de confiança de que o erro é menor do que 5%; e temos 95% de confiança de que a margem de erro é menor que 2.2 pontos percentuais.

[5] Note que "independentemente" significa "com reposição", ou seja, seria teoricamente possível escolher o mesmo eleitor duas vezes para indicar a sua resposta. Na prática, sendo N o número total de eleitores na população, é costume tomar $n << N$, e praticamente não há diferença entre "com reposição" e "sem reposição". Isto dito, não é tão difícil fazer os cálculos num modelo sem reposição – basta notar que $S_n \sim Hip(n, pN, N)$.

[6] http://www.ibope.com.br/opp/pesquisa/opiniaopublica/download/imprensa_torcidas_1_mencao.pdf

Note-se que estes "95% de confiança" têm de ser interpretados corretamente: não é que a variável aleatória p tem 95% de chance de estar entre $\hat{p}-\varepsilon$ e $\hat{p}+\varepsilon$ – afinal, p **não é uma variável aleatória**, p é um parâmetro real e fixo da população, mas que infelizmente desconhecemos! A variável aleatória é $\hat{p}=\bar{X}$! Estes 95% têm de ser interpretados assim: "**se fizermos esta rodada de entrevistas várias vezes, encontraremos a cada vez um $\hat{p}$ diferente; existe** 95% **de chance deste $\hat{p}$ estar a menos de ε do p verdadeiro**".

Um fato espantoso é que os cálculos do exemplo acima absolutamente não dependem do número de torcedores no Brasil! Diga-se de passagem, a imensa maioria das pesquisas estatísticas trabalha com 95% de confiança; quando uma pesquisa de opinião diz algo como "a margem de erro é de b pontos percentuais", ela realmente quer dizer "temos pelo menos 95% de confiança de que a nossa estimativa da proporção e o valor real da proporção estão a menos de b pontos percentuais um do outro". Afinal, é sempre possível que o instituto de pesquisa tenha escolhido exatamente os únicos 2000 torcedores do "Fim-de-Mundo Futebol Clube" em sua pesquisa aleatória, obtendo um $\hat{p}$ patológico **por puro azar**.

De qualquer forma, se usarmos os usuais 95% de confiança, temos:

$$2\,\text{NormalDist}\left(2\varepsilon\sqrt{n}\right)-1=0.95\Rightarrow 2\varepsilon\sqrt{n}=1.9600\Rightarrow \boxed{\varepsilon\sqrt{n}=0.98}$$

independentemente do tamanho da população!

Exemplo 17 *Gostaríamos de entrevistar n eleitores sobre sua opinião a respeito de um referendo do tipo "sim" ou "não". Quantos eleitores devemos entrevistar para que a margem de erro entre a proporção p de "sim" na população e a proporção $\hat{p}$ da amostra seja 3% (ou menos)?*
Solução: Note que é impossível ***garantir*** *que $|p-\hat{p}|<3\%$ a menos que entrevistemos praticamente a população toda! Salvo algo seja dito explicitamente em contrário, a "margem de erro" usada em pesquisas costuma se referir a 95% de confiança. Assim $\varepsilon=0.03$ e:*

$$n=\left(\frac{0.98}{\varepsilon}\right)^2=\left(\frac{0.98}{0.03}\right)^2\approx 1067 \text{ entrevistados}$$

Novamente, note que o número de entrevistados independe da população deste país!

Exemplo 18 *Se cada vestibulando classificado tem 40% de chance de ir para a GHX, qual a chance de mais do que 50% dos 100 classificados irem para a GHX? Aproximadamente, temos*

$$\hat{p}\sim N\left(0.4,\frac{(0.4)\,(0.6)}{100}\right)=N\,(0.4,0.0024)$$

Assim, estimamos

$$\Pr(\hat{p}>0.5)\cong 1-\text{NormalDist}\left(\frac{0.5-0.4}{\sqrt{0.0024}}\right)=1-\text{NormalDist}\,(2.04124)=2.0613\%$$

Note que, se usássemos 51% ou mais:

$$\Pr(\hat{p}\geq 0.51)\cong 1-\text{NormalDist}\left(\frac{0.51-0.4}{\sqrt{0.0024}}\right)=1-\text{NormalDist}\,(2.24536)=1.2372\%$$

A resposta exata (1.676%) está bem perto destas duas.

Exemplo 19 *Suponha que 10% dos itens de uma linha de produção são defeituosos. Tomando 20 itens ao acaso, qual a chance de que 15% ou mais destes sejam defeituosos? O cálculo exato pede o uso da função acumulada da distribuição binomial $X\sim Bin\,(20,0.1)$:*

$$\Pr(\hat{p}\geq 0.15)=\Pr(X\geq 3)=1-\text{BinomialDist}\,(2;20,0.1)=32.307\%$$

mas arriscando uma aproximação normal $\hat{p}\approx N\left(0.1,\frac{(0.1)(0.9)}{20}\right)=N\,(0.1,0.0045)$:

$$\Pr(\hat{p}\geq 0.15)=1-\text{NormalDist}\left(\frac{0.15-0.1}{\sqrt{0.0045}}\right)=1-\text{NormalDist}\,(0.74536)=22.803\%$$

Note que a aproximação não é das melhores, pois $n=20$ não é grande o suficiente.

Exemplo 20 *Se $p = 30\%$ dos estudantes de uma escola são mulheres, retirando uma AAS com 10 estudantes, qual a chance de obter $\hat{p}$ a menos de 0.01 de p? A aproximação normal nos dá*

$$\begin{aligned} \hat{p} &\approx N\left(p, \frac{pq}{n}\right) = N(0.3, 0.021) \Rightarrow \Pr(|\hat{p} - p| < 0.01) = \\ &= \Pr\left(|Z| < \frac{0.01}{\sqrt{0.021}}\right) = \text{NormalDist}\left(\frac{0.01}{\sqrt{0.021}}\right) - \text{NormalDist}\left(-\frac{0.01}{\sqrt{0.021}}\right) = 5.5016\% \end{aligned}$$

Vale comparar isto com a resposta exata; afinal, $|\hat{p} - p| < 1\%$ se e somente se 3 estudantes da amostra exatamente são mulheres (se 2 ou 4 forem mulheres, $|\tilde{p} - p| = 10\%$)! Assim, tomando $X \sim Bin(10, 0.3)$:

$$\Pr(|\hat{p} - p| < 0.01) = \Pr(X = 3) = \binom{10}{3} (0.3)^3 (0.7)^7 = 26.683\%$$

*que é **muito longe** da aproximação normal pois o n é muito pequeno!*

7.5.1 Exercícios

Vários destes exercícios já apareceram em seções anteriores - a idéia é usar agora aproximações normais e compará-las com as respostas antigas!

Ex. 22 *No exercício 15 definimos S_{100} como o número de caras em 100 lançamentos de uma moeda. O Teorema de Chebyshev mostrava apenas que*

$$\Pr(35 \leq S_{100} \leq 65) \geq \frac{1}{3^2} = 11.11\%$$

Estime esta mesma probabilidade usando agora uma aproximação normal à distribuição binomial $Bin(100, 0.5)$ e compare-a com a probabilidade exata obtida da distribuição binomial acumulada.

Ex. 23 *No exercício 16 definimos $\bar{X}$ como a média de 100 variáveis aleatórias, cada uma distribuída uniformemente em $[0, 1]$. Naquele exercício usamos a desigualdade de Chebyshev para mostrar que*

$$\Pr\left(\left|\bar{X} - 0.5\right| \geq 0.1\right) \leq 8.333...\%$$

Estime esta probabilidade usando agora uma aproximação normal para a distribuição de $\bar{X}$. Compare-a com o valor exato[7] da probabilidade, a saber:

$$\Pr\left(\left|\bar{X} - 0.5\right| \geq 0.1\right) = \Pr(S_n \leq 40 \text{ ou } S_n \geq 60) = 5.0131246 \times 10^{-4}$$

Ex. 24 *a) No exercício 18 vimos que, quando $X \sim U[\mu - 1, \mu + 1]$, $n = 667$ amostras são necessárias para garantir que $\bar{X}$ está a menos de 0.1 unidades de μ (com 95% de confiança). Supondo que a distribuição de $\bar{X}$ seja praticamente normal, mostre que cerca de 129 amostras seriam suficientes.*
b) Supondo que $\bar{X}$ é praticamente normal, mostre que o número de amostras encontrado naquele exercício para que $\left|\bar{X} - \mu\right| \leq 0.01$ com 95% de confiança (que era $n = 66667$) pode ser reduzido para aproximadamente $n = 12805$.

Ex. 25 *Um instituto realiza uma pesquisa eleitoral com 1500 eleitores que vão votar em um de apenas dois candidatos (suponha que todos os entrevistados escolhem um dos candidatos).*
a) Qual a margem de erro da pesquisa em pontos percentuais?
b) Suponha agora que nesta amostra de eleitores, 49% declararam que vão votar no candidato A e 51% declararam que vão votar no candidato B. Com a margem de erro do item (a), diríamos que os candidatos estão em "empate técnico". Por este motivo, o instituto decidiu fazer uma segunda pesquisa com uma margem de erro de apenas 0.5 ponto percentual. Quantos eleitores o instituto deverá entrevistar?

[7] Bom, o valor **exato** mesmo é

$$\Pr(0.4 \leq \bar{X} \leq 0.6) = 1 - \frac{(660000)(227)P}{100!}$$

onde P é um número primo de 150 algarismos, a saber

$$P = 622\,609\,997\,659\,051\,599\,722\,950\,244\,464\,645\,718\,071\,677\,183\,516\,621\,629\,450\,601\,297\,296\,841\,554\,960 \\ 217\,389\,903\,579\,462\,227\,539\,317\,464\,835\,342\,588\,500\,249\,887\,137\,669\,481\,946\,430\,446\,686\,029\,506\,871$$

Com sete casas decimais, dá 5.0131246×10^{-4}.

Ex. 26 *Um instituto faz pesquisas sobre um referendo do tipo sim/não em* 26 *estados brasileiros mais o Distrito Federal. Em cada estado, ele divulga um intervalo de confiança para a porcentagem de eleitores "sim" (com* 95% *de confiança). Qual a probabilidade da proporção real não se encaixar no intervalo divulgado em* 3 *ou mais deste locais? E em* 5 *ou mais?*

Ex. 27 *Refaça o seguinte problema (comparando com a resposta exata obtida anteriormente):*
Role um dado 30 *vezes. Usa a aproximação normal à binomial para estimar a probabilidade de obtermos exatamente cinco números* 6*.*

Ex. 28 *Refaça o seguinte problema (comparando com a resposta exata obtida anteriormente):*
Um estudante marca ao acaso as respostas de um teste tipo V ou F com 50 *questões. Usando a aproximação normal à binomial:*
a) Qual a probabilidade de ele acertar 80% *(ou mais?) delas apenas adivinhando? E* 60% *ou mais?*
b) Se 100 *estudantes adivinham as questões ao acaso, qual a probabilidade de ao menos um tirar* 80% *ou mais?*
c) E se o teste for múltipla escolha com 5 *alternativas por questão, como mudam os itens anteriores?*

Ex. 29 *Refaça o seguinte problema (comparando com a resposta exata obtida anteriormente):*
Um potencial paranormal tenta adivinhar 10 *cartas escolhidas aleatoriamente, cada uma com* 5 *possibilidades eqüiprováveis. Se ele não tiver poder algum, qual a chance de adivinhar* 8 *cartas ou mais? E se* 1000 *pessoas comuns forem testadas, qual a chance de ao menos uma adivinhar* 8 *ou mais cartas por puro acaso?*

Ex. 30 *Refaça o seguinte problema (comparando com a resposta exata obtida anteriormente):*
Distribua 10000 *folhetos aleatoriamente por* 2000 *quadras em uma cidade. Qual a chance de a sua quadra não receber folheto algum? E de receber* 5 *folhetos? E* 10*?*

7.6 Exercícios de Provas

Ex. 31 (A2 2004.2) *Sua empacotadora produz automaticamente pacotes de um certo produto. O peso do pacote é regulável. No entanto, o peso do pacote sofre flutuações, de modo que ao se regular a máquina para um certo peso p, são produzidos pacotes cujo peso tem distribuição normal com média p e desvio padrão* 10 *gramas.*
a) Se a máquina for regulada para $p = 500g$, qual é a probabilidade de que o peso do pacote fique abaixo de $490g$?
b) Um fiscal amostrará um certo número n de pacotes e, caso a média dos pesos destes pacotes fique abaixo de $495g$, ele multará a sua fábrica. Qual o mínimo valor de n para que você tenha pelo menos 99% *de chance de não receber a multa?*
c) Qual o valor de p para que a probabilidade de que o peso de um pacote fique abaixo de $490g$ seja igual a 1%*?*

Ex. 32 (A2 2004.2) *Um jogador participa de um programa de prêmios em que ele gira* 24 *vezes uma roleta graduada (continuamente) de* 0 *a* 100 *reais. O prêmio final é a soma dos prêmios de cada giro da roleta.*
a) Calcule a média e a variância do prêmio final ganho pelo candidato.
b) Calcule a probabilidade aproximada de que o prêmio ganho seja maior do que 1400 *reais.*

Ex. 33 (AS 2004.2) *Uma roleta sorteia um número uniformemente entre* 0 *e* 10*. Você vai girá-la n vezes e somar os resultados de cada rodada. Você ganha um prêmio de* \$100 *se a soma ficar entre* 240 *e* 270*.*
a) Calcule a média e a variância da soma dos resultados em função de n.
b) O que é melhor: rodar a roleta 48 *ou* 54 *vezes? Justifique a sua resposta.*
c) E se cada rodada da roleta custar \$0.50*, qual opção é melhor? Explique.*

Ex. 34 (AS 2005.2) *Refaça o seguinte problema (comparando com a resposta exata obtida anteriormente):*
A cada dia de um período de 90 *dias, uma determinada ação pode desvalorizar* \$4 *com probabilidade* 30%*, ou subir* \$2 *com probabilidade* 70% *(suponha que cada dia é independente dos demais). Seja Z o número de dias em que a ação subiu.*
a) Qual é a distribuição de probabilidade de Z?
b) Você compra a ação no início do período e a vende ao final dos 90 *dias. Qual o valor esperado do seu lucro?*
c) Estime a probabilidade do seu lucro ser de pelo menos \$50 *no período de* 90 *dias (juntando as valorizações e desvalorizações de cada dia).*

Chapter 8

Outras Distribuições Amostrais

8.1 Estimação de Parâmetros

A idéia de usar $\bar{X}$ para tentar estimar μ do capítulo anterior pode ser generalizada. Assim, seja X uma variável aleatória com uma certa distribuição (é comum pensar que a distribuição de X representa uma certa **população**). Seja θ um **parâmetro** desta distribuição (por exemplo, θ pode ser a taxa λ da distribuição exponencial, ou o p da distribuição binomial). A partir de uma Amostragem Aleatória Simples X_1, X_2, ..., X_n de X gostaríamos de inventar alguma **estatística** $\hat{\theta}$ que representasse θ de alguma forma.

Definição 1 *Um **estimador** $\hat{\theta}$ de um parâmetro θ (da distribuição de X, isto é, da população) é uma função das observações da amostra, isto é*

$$\hat{\theta} = \hat{\theta}\left(X_1, X_2, ..., X_n\right)$$

Note que $\hat{\theta}$ é, de fato, uma variável aleatória, com a sua própria distribuição de probabilidade. Por exemplo, nas últimas aulas, vimos como usar o estimador $\hat{\mu} = \bar{X}$. Vimos que

$$E\left(\bar{X}\right) = \mu \qquad Var\left(\bar{X}\right) = \frac{\sigma^2}{n}$$

onde $\sigma^2 = Var\left(X\right)$. No caso particular em que X tem distribuição normal, vimos que $\bar{X}$ também tem distribuição normal. Enfim, o Teorema Central do Limite diz que, mesmo que X não tenha distribuição normal, a distribuição de $\bar{X}$ se aproximará de uma normal quando $n \to \infty$ (desde que $E\left(X\right), Var\left(X\right)$ sejam finitos).

Definição 2 *Um estimador $\hat{\theta}$ é dito **não-viesado** (**não-viciado**, **não-tendencioso**) quando*

$$\boxed{E\left(\hat{\theta}\right) = \theta}$$

*O **viés** de um estimador é*

$$\boxed{Viés\left(\hat{\theta}\right) = E\left(\hat{\theta}\right) - \theta}$$

Definição 3 *Um estimador $\hat{\theta}_1$ é dito **mais eficiente** do que $\hat{\theta}_2$ quando*

$$Var\left(\hat{\theta}_1\right) < Var\left(\hat{\theta}_2\right)$$

Definição 4 *Um(a seqüência de) estimador(es) $\hat{\theta}$ é dito **consistente** quando, para todo $\varepsilon > 0$ fixo,*

$$\boxed{\lim_{n\to\infty} \Pr\left(\left|\hat{\theta} - \theta\right| > \varepsilon\right) = 0}$$

Por exemplo, a Lei dos Grandes Números diz exatamente que $\bar{X}$ é um estimador consistente de μ.

Proposição 5 *Se*

$$\boxed{\lim_{n\to\infty} Viés\left(\hat{\theta}\right) = 0 \quad e \quad \lim_{n\to\infty} Var\left(\hat{\theta}\right) = 0}$$

então $\hat{\theta}$ é consistente.

Prova. A demonstração é análoga à da Lei dos Grandes Números. Para começar, façamos o caso em que $\hat{\theta}$ é não-tendencioso, isto é, $E\left(\hat{\theta}\right) = \theta$. Então, pela desigualdade de Chebyshev:

$$0 \leq \Pr\left(\left|\hat{\theta} - E\left(\hat{\theta}\right)\right| > \varepsilon\right) \leq \frac{Var\left(\hat{\theta}\right)}{\varepsilon^2}$$

Como ε é fixo e $\lim_{n\to\infty} Var\left(\hat{\theta}\right) = 0$, pelo Teorema do Confronto

$$\lim_{n\to\infty} \Pr\left(\left|\hat{\theta} - E\left(\hat{\theta}\right)\right| > \varepsilon\right) = 0$$

No caso em que $\hat{\theta}$ é tendencioso, devemos fazer uma pequena alteração. Pela desigualdade triangular:

$$\left|\hat{\theta} - \theta\right| \leq \left|\hat{\theta} - E\left(\hat{\theta}\right)\right| + \left|E\left(\hat{\theta}\right) - \theta\right| = \left|\hat{\theta} - E\left(\hat{\theta}\right)\right| + \left|Viés\left(\hat{\theta}\right)\right|$$

Portanto

$$\varepsilon < \left|\hat{\theta} - \theta\right| \Rightarrow \varepsilon < \left|\hat{\theta} - E\left(\hat{\theta}\right)\right| + \left|Viés\left(\hat{\theta}\right)\right| \Rightarrow \left|\hat{\theta} - E\left(\hat{\theta}\right)\right| > \varepsilon - \left|Viés\left(\hat{\theta}\right)\right|$$

e assim

$$0 \leq \Pr\left(\left|\hat{\theta} - \theta\right| > \varepsilon\right) \leq \Pr\left(\left|\hat{\theta} - E\left(\hat{\theta}\right)\right| > \varepsilon - \left|Viés\left(\hat{\theta}\right)\right|\right)$$

Como o viés se aproxima de 0, é possível tomar n grande o suficiente para que $\left|Viés\left(\hat{\theta}\right)\right| < \frac{\varepsilon}{2}$. Assim

$$0 \leq \Pr\left(\left|\hat{\theta} - \theta\right| > \varepsilon\right) \leq \Pr\left(\left|\hat{\theta} - E\left(\hat{\theta}\right)\right| > \frac{\varepsilon}{2}\right)$$

Mas a probabilidade da direita se aproxima de 0 (como vimos acima, bastando trocar ε por $\frac{\varepsilon}{2}$). Assim,

$$\lim_{n\to\infty} \Pr\left(\left|\hat{\theta} - \theta\right| > \varepsilon\right) = 0$$

■

Em particular, note que se $\hat{\theta}$ é não-tendencioso e $Var\left(\hat{\theta}\right) \to 0$, então $\hat{\theta}$ será consistente.

8.2 Estimadores pontuais da variância

8.2.1 Média Conhecida

Suponha que a média $\mu = E(X)$ é conhecida, mas não sabemos o valor de $\sigma^2 = Var(X)$. Um possível estimador para $\sigma^2 = E\left((X-\mu)^2\right)$ é

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^n (X_i - \mu)^2}{n}$$

Este estimador é não-viesado. De fato

$$E\left(\hat{\sigma}^2\right) = E\left(\frac{\sum_{i=1}^n (X_i - \mu)^2}{n}\right) = \frac{\sum_{i=1}^n E(X_i - \mu)^2}{n} = \frac{\sum_{i=1}^n Var(X_i)}{n} = \frac{nVar(X)}{n} = Var(X)$$

pois para cada X_i temos $E(X_i) = \mu$.

Na prática, raramente sabemos o valor de μ sem saber o valor de σ^2, então este estimador praticamente não é utilizado.

8.2.2 Média Desconhecida

Se não soubermos μ, como estimar a variância da população σ^2? Uma primeira idéia é tentar:

Definição 6 *A **variância verdadeira da amostra** é o estimador*

$$\boxed{\hat{\sigma}^2 = \frac{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2}{n}}$$

A palavra "verdadeira" é essencial para distinguir este estimador da "variância da amostra" que definiremos em breve. Infelizmente, esta "variância verdadeira" é viesada! De fato:

Proposição 7 *Para a variância verdadeira da amostra, tem-se*

$$\boxed{E\left(\hat{\sigma}^2\right) = \frac{n-1}{n}\sigma^2}$$

$$\boxed{Viés\left(\hat{\sigma}^2\right) = -\frac{\sigma^2}{n}}$$

Prova. Em primeiro lugar, lembremos que

$$E\left(X^2\right) = Var\left(X\right) + \left(E\left(X\right)\right)^2$$

Assim

$$\begin{aligned} E\left(\left(X_1 - \bar{X}\right)^2\right) &= Var\left(X_1 - \bar{X}\right) + \left(E\left(X_1 - \bar{X}\right)\right)^2 = \\ &= Var\left(\frac{(n-1)X_1 - X_2 - X_3 - ... - X_n}{n}\right) + 0 \end{aligned}$$

pois $E\left(X_1\right) = E\left(\bar{X}\right) = \mu$. Agora, como X_1, X_2, ..., X_n são independentes

$$\begin{aligned} Var\left(\frac{(n-1)X_1 - X_2 - X_3 - ... - X_n}{n}\right) &= \frac{(n-1)^2 Var\left(X_1\right) + Var\left(X_2\right) + ... + Var\left(X_n\right)}{n^2} = \\ &= \frac{(n-1)^2 + (n-1)}{n^2}\sigma^2 = \frac{n-1}{n}\sigma^2 \end{aligned}$$

Juntando tudo:

$$E\left(\hat{\sigma}^2\right) = E\left(\frac{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2}{n}\right) = \frac{\sum_{i=1}^{n} E\left(\left(X_i - \bar{X}\right)^2\right)}{n} = \frac{n-1}{n}\sigma^2$$

Ou seja, $\hat{\sigma}^2$ tem um viés de

$$Viés\left(\hat{\sigma}^2\right) = E\left(\hat{\sigma}^2\right) - \sigma^2 = -\frac{\sigma^2}{n}$$

■

Como $\hat{\sigma}^2$ é viesado, muitos preferem usar o seguinte outro estimador para a variância de X:

Definição 8 *A **variância da amostra** é o estimador*

$$\boxed{S^2 = \frac{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2}{n-1}}$$

obtido trocando o denominador n por $n-1$ na variância verdadeira da amostra.

Note-se que, enquanto $\hat{\sigma}^2$ a princípio representa um estimador qualquer para σ^2, a notação S^2 é reservada para o estimador acima. Por outro lado, vamos utilizar $\hat{\sigma}^2 =$ (Variância Verdadeira da Amostra) no resto desta seção. Note que

$$\boxed{S^2 = \frac{n}{n-1}.\hat{\sigma}^2}$$

Proposição 9 *S^2 é um estimador não viesado de σ^2, isto é*

$$\boxed{E\left(S^2\right)=\sigma^2}$$

Prova.

$$E\left(S^2\right)=\frac{n}{n-1}E\left(\hat{\sigma}^2\right)=\frac{n}{n-1}\frac{n-1}{n}\sigma^2=\sigma^2$$

■

Agora, para verificar se S^2 é consistente, falta descobrir a variância de S^2! Temos:

Proposição 10

$$\boxed{Var\left(S^2\right)=\frac{(n-1)\mu_4-(n-3)\sigma^4}{n(n-1)}}$$

onde

$$\mu_4=E\left((X-\mu)^4\right)$$

Prova. Façamos apenas o caso $n=2$. Se $\mu=0$, então $\sigma^2=E\left(X^2\right)$, e:

$$\begin{aligned} S^2 &= \left(X_1-\bar{X}\right)^2+\left(X_2-\bar{X}\right)^2=\frac{\left(X_1-X_2\right)^2}{2}\Rightarrow S^4=\frac{\left(X_1-X_2\right)^4}{4}\Rightarrow \\ \Rightarrow & \quad E\left(S^4\right)=\frac{E\left(X^4\right)}{2}-2\mu E\left(X^3\right)+\frac{3}{2}\left(E\left(X^2\right)\right)^2=\frac{\mu_4+3\sigma^4}{2} \end{aligned}$$

Então

$$Var\left(S^2\right)=E\left(S^4\right)-\left(E\left(S^2\right)\right)^2=\frac{\mu_4+3\sigma^4}{2}-\sigma^4=\frac{\mu_4+\sigma^4}{2}$$

Caso $\mu\neq 0$, basta notar que S^2 não se altera ao tomarmos $Y=X-\mu$ (pois teríamos $\bar{Y}=\bar{X}-\mu$ também). O caso geral consiste em abrir somatórios enormes e analisar os termos todos um a um... ■

Proposição 11 $\boxed{\text{Se } \mu_4<\infty, \text{ então } S^2 \text{ e a variância verdadeira } \hat{\sigma}^2 \text{ são ambos estimadores consistentes de } \sigma^2.}$

Prova. De fato, é fácil ver que $E\left(S^2\right)=\sigma^2$ e $Var\left(S^2\right)\to 0$ quando $n\to\infty$. Para $\hat{\sigma}^2$, basta notar que

$$\begin{aligned} E\left(\hat{\sigma}^2\right) &= \frac{n-1}{n}\sigma^2\to\sigma^2 \\ Var\left(\hat{\sigma}^2\right) &= \left(\frac{n-1}{n}\right)^2 Var\left(S^2\right)\to 0 \end{aligned}$$

também quando $n\to\infty$. ■

8.3 Erro Quadrático Médio

Como vimos acima, ambos S^2 e a variância verdadeira (que continuaremos chamando de $\hat{\sigma}^2$) são estimadores consistentes de σ^2; enquanto S^2 é não-viesado, $\hat{\sigma}^2$ tem a vantagem de ser mais eficiente. Qual é melhor? Um possível critério e calcular o **erro quadrático médio** de cada um deles, como definido a seguir:

Definição 12 *O erro quadrático médio de um estimador T de um parâmetro θ é*

$$EQM\left(T;\theta\right)=E\left((T-\theta)^2\right)$$

Proposição 13

$$EQM\left(T;\theta\right)=Var\left(T\right)+\left(Viés\left(T\right)\right)^2$$

Prova. De fato

$$E\left((T-\theta)^2\right) = Var\,(T-\theta) + (E\,(T-\theta))^2 = Var\,(T) + (Viés\,(T))^2$$

■

Note assim que

$$\begin{aligned} EQM\left(\hat{\sigma}^2;\sigma^2\right) &= (n-1)\frac{(n-1)\,\mu_4-(n-3)\,\sigma^4}{n^3}+\left(-\frac{\sigma^2}{n}\right)^2=\frac{(n-1)^2\,\mu_4-\left(n^2-5n+3\right)\sigma^4}{n^3} \\ EQM\left(S^2;\sigma^2\right) &= \frac{(n-1)\,\mu_4-(n-3)\,\sigma^4}{n\,(n-1)} \end{aligned}$$

Em outras palavras, $\hat{\sigma}^2$ é melhor se tivermos

$$\begin{aligned} EQM\left(\hat{\sigma}^2;\sigma^2\right) &< EQM\left(S^2;\sigma^2\right) \iff \\ &\iff \beta=\frac{\mu_4}{\sigma^4}>\frac{3n^2-8n+3}{(n-1)\,(2n-1)} \end{aligned}$$

onde β é o chamado **coeficiente de curtose** da distribuição. No caso específico da distribuição normal, temos

$$EQM\left(\hat{\sigma}^2;\sigma^2\right)=\frac{2n-1}{n^2}\sigma^4<\frac{2}{n-1}\sigma^4=EQM\left(S^2;\sigma^2\right)$$

o que indica que, por este critério, o estimador viciado $\hat{\sigma}^2$ é melhor do que S^2 para estimar σ^2!

8.3.1 Exercícios

Ex. 1 *Seja $X\sim Be\,(p)$ e seja $X_1,\,X_2,\,X_3,\,...,\,X_n$ uma AAS de X. Considere os seguintes estimadores de p:*
a) $\hat{p}_1 = X_3$
b) $\hat{p}_2 = \bar{X}$
c) $\hat{p}_3 = \frac{\max(X_1,X_2,...,X_n)+\min(X_1,X_2,...,X_n)}{2}$
d) $\hat{p}_4 = \frac{\left(\sum_{i=1}^n X_i\right)-\max(X_1,X_2,...,X_n)-\min(X_1,X_2,...,X_n)}{n-2}$ *(isto é, a média dos $n-2$ termos centrais).*
Quais destes estimadores são viesados? Qual é mais eficiente? Qual tem o menor EQM? Algumas de suas respostas podem depender de p e n.

Ex. 2 *Você quer estimar um parâmetro θ e, para tanto, criou dois estimadores que acabam por ter as seguintes distribuições amostrais*

$$\begin{aligned} \hat{\theta}_1 &\sim N\left(\theta,5^2\right) \\ \hat{\theta}_2 &\sim N\left(\theta-1,2^2\right) \end{aligned}$$

Na sua opinião, qual é melhor? Que critério você usou?

Ex. 3 *Considere dois estimadores $\hat{\theta}_1$ e $\hat{\theta}_2$, ambos não-viesados, isto é*

$$E\left(\hat{\theta}_1\right)=E\left(\hat{\theta}_2\right)=\theta$$

Suponha que $Var\left(\hat{\theta}_1\right)=2Var\left(\hat{\theta}_2\right)$. Queremos criar um novo estimador de θ que seja uma combinação linear destes dois, isto é

$$T=a\hat{\theta}_1+b\hat{\theta}_2$$

a) Que condição a e b têm de satisfazer para que T seja não-viesado?
b) Sendo T não-viesado, como escolher a e b para que T tenha a menor variância possível?

Ex. 4 *Seja* $X \sim Poi(\lambda)$*. Pode-se mostrar que* $\mu_4(X) = \lambda + 3\lambda^2$*.*
a) Sabemos que $\mu = \lambda$*; isto sugere o uso do estimador* $\bar{X}$ *para* λ*. Calcule o seu EQM.*
b) Sabemos que $\sigma^2 = \lambda$*; isto sugere o uso de* S^2 *para estimar* λ*. Calcule o seu EQM.*
c) Qual dos estimadores acima lhe parece melhor?

Ex. 5 *Seja* X *uma variável aleatória com densidade dada por*

$$f(x) = \begin{cases} ax + \frac{1}{2} \text{ se } -1 \le x \le 1 \\ 0, \text{ caso contrário} \end{cases}$$

Calcule $\mu = E(X)$ *em função de* a *e, a partir daí, crie um estimador não-viesado para* a *baseado numa AAS* X_1*,* X_2*, ...,* X_n *(dica: use* $\bar{X}$*). Qual é a variância deste estimador?*

Ex. 6 *Suponha que* $X \sim U[0, 2a]$*, onde gostaríamos de estimar o valor de* $\mu = a$*. Considere o estimador* $\bar{X}$*. Mostre que* $\bar{X}$ *é não-viesado, consistente e calcule seu EQM.*

Ex. 7 *Suponha que* $X \sim U[0, 2a]$*, onde gostaríamos de estimar o valor de* $\mu = a$*.*
a) Considere o estimador $M = \max(X_1, X_2, ..., X_n)/2$*. Encontre a densidade de* M *(dica: use a acumulada!) e mostre que*

$$E(M) = \frac{n}{n+1}a$$

Mostre que M *é consistente e calcule seu EQM.*
b) A partir de M*, construa um novo estimador* M_2 *não-viesado para* a *e calcule o seu EQM.* M_2 *é consistente?*
c) Que estimador você prefere: $\bar{X}$ *(do problema anterior),* M *ou* M_2*? Por quê?*

Ex. 8 *a) Seja* $X \sim NegBin(r, p)$*. Tome apenas uma amostra* X *desta distribuição. Mostre que*

$$\hat{p} = \frac{r-1}{X-1}$$

é um estimador não-viesado de p*.*
b) Você lança uma moeda (possivelmente viciada) até obter a quinta cara (K*). Se a seqüência de lançamentos foi* $CCKCKCCCCKCKK$*, qual seria a sua estimativa da probabilidade* p *desta moeda dar cara? E se a seqüência fosse* $CCKCKKCCKK$*? Escreva uma seqüência que levaria à estimativa* $p = 50\%$*.*

Ex. 9 *Seja* $X \sim Exp(\lambda)$*. Tome uma AAS* X_1*,* X_2 *desta distribuição. Considere os estimadores* $\bar{X}$ *e* $G = \sqrt{X_1 X_2}$ *para a média* $\mu = \frac{1}{\lambda}$*.Calcule o viés de cada um deles.Qual deles tem o menor erro quadrático?*

Ex. 10 *Suponha que* $\mu = E(X)$ *é conhecido. Considere o estimador*

$$Y = \frac{\sum_{i=1}^{n} (X_i - \mu)^2}{n}$$

de σ^2*. Mostre que* Y *é não-viesado e que calcule seu erro quadrático médio em função de* $\mu_4 = E(X-\mu)^4$*. Em particular, se* X *é normal, quanto vale* $EQM(Y; \sigma^2)$*? {Dica: suponha* $\mu = 0$ *para começar.]*

8.4 Distribuição Qui-quadrado

Dada a distribuição de uma variável aleatória X, descobrimos no capítulo anterior o valor esperado e a variância de $\bar{X}$. Mais ainda, no caso em que X tem distribuição normal, fomos capazes de descobrir a distribuição exata de $\bar{X}$ (que também era normal).

Neste capítulo, já descobrimos o valor esperado e variância de S^2 (pelo menos em função de $\sigma^2 = Var(X)$ e de $\mu_4(X)$). Mas qual será a distribuição exata de S^2? Novamente, a resposta depende da distribuição de X e é complicada no caso geral.

No entanto, se X é normal, podemos resumir o que sabemos sobre S^2 e descobrir sua distribuição como a seguir:

Proposição 14 *Se* $Z \sim N(0, 1)$ *então* $\mu_4(Z) = 3$*.*

Prova. Tomando $z^2 = 2u$, temos:

$$\begin{aligned}\mu_4 &= E\left((Z-0)^4\right) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} z^4 e^{-z^2/2} dz = \frac{2}{\sqrt{2\pi}} \int_0^{\infty} 4u^2 e^{-u} \frac{du}{\sqrt{2u}} = \frac{4}{\sqrt{\pi}} \int_0^{\infty} u^{3/2} e^{-u} du = \\ &= \frac{4}{\sqrt{\pi}} \Gamma\left(\frac{5}{2}\right) = \frac{4}{\sqrt{\pi}} \frac{3}{2} \frac{1}{2} \Gamma\left(\frac{1}{2}\right) = 3\end{aligned}$$

■

Proposição 15 *Se X é normal, então*

$$\boxed{Var\left(S^2\right) = \frac{2\sigma^4}{n-1} \quad e \quad Var\left(\hat{\sigma}^2\right) = \frac{2(n-1)}{n^2}\sigma^4}$$

Prova. Como S^2 não depende de μ, podemos supor $\mu = 0$. Mas então $X = \sigma Z$ onde Z é normal padrão, isto é

$$\mu_4(X) = \sigma^4 \mu_4(Z) = 3\sigma^4$$

Portanto

$$Var\left(S^2\right) = \frac{(n-1)\mu_4 - (n-3)\sigma^4}{n(n-1)} = \frac{3(n-1) - (n-3)}{n(n-1)} \sigma^4 = \frac{2\sigma^4}{n-1}$$

e, como $\hat{\sigma}^2 = \frac{n-1}{n} S^2$, a outra afirmação segue de imediato. ■

Os resultados acima foram aplicações imediatas dos resultados das últimas seções. Agora estamos prontos para resolver uma série de "exercícios" que levam à definição da distribuição qui-quadrado (que é a distribuição de S^2 no caso em que X é normal).

Ex. 11 *Seja $X \sim N(0,1)$. Qual é a distribuição de $Y = X^2$, seu valor esperado e variância?*

Solução 16 *As f.d.a. de X e Y satisfazem (para $y \geq 0$)*

$$F_Y(y) = \Pr(Y \leq y) = \Pr\left(X^2 \leq y\right) = \Pr\left(-\sqrt{y} \leq X \leq \sqrt{y}\right) = F_X\left(\sqrt{y}\right) - F_X\left(-\sqrt{y}\right)$$

Derivando dos dois lados:

$$f_Y(y) = \frac{1}{2\sqrt{y}}\left(f_X\left(\sqrt{y}\right) - f_X\left(-\sqrt{y}\right)\right)$$

No caso da distribuição normal, temos (para $y \geq 0$)

$$f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \Rightarrow f(y) = \frac{1}{2\sqrt{y}}\left(2\frac{e^{-y/2}}{\sqrt{2\pi}}\right) = \frac{(1/2)^{1/2}}{\sqrt{\pi}} y^{-1/2} e^{-y/2}$$

Comparando com a distribuição Gama, vemos que

$$Y \sim \Gamma\left(\frac{1}{2}, \frac{1}{2}\right)$$

e portanto

$$E(Y) = 1; \ Var(Y) = 2$$

Ex. 12 *Sejam $X_1, X_2, ..., X_n \sim N(0,1)$ e independentes. Qual é a distribuição de $Y = \sum X_i^2$?*

Solução 17 *Como cada X_i^2 tem distribuição $\Gamma\left(\frac{1}{2}, \frac{1}{2}\right)$, basta somá-las:*

$$Y \sim \Gamma\left(\frac{1}{2} + \frac{1}{2} + ... + \frac{1}{2}, \frac{1}{2}\right) = \Gamma\left(\frac{n}{2}, \frac{1}{2}\right)$$

e, portanto, $E(Y) = n$ e $Var(Y) = 2n$.

Esta distribuição é importante o suficiente para ganhar um nome à parte:

Definição 18 *A distribuição* $\Gamma\left(\frac{n}{2}, \frac{1}{2}\right)$ *é também chamada de* ***qui-quadrado com*** n ***graus de liberdade****, cuja notação é*

$$Y \sim \chi^2(n)$$

Em particular, $E(Y) = n$ *e* $Var(Y) = 2n$*. A fórmula exata da densidade é*

$$f(y) = \begin{cases} \frac{1}{\Gamma\left(\frac{n}{2}\right)2^{n/2}} y^{n/2-1} e^{-y/2}, \text{ se } y > 0 \\ 0, \text{ caso contrário} \end{cases}$$

Proposição 19 *A distribuição de* S^2 *é determinada por*

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$$

Prova. Façamos apenas o caso $n = 2$. Então

$$\frac{S^2}{\sigma^2} = \frac{(X_1 - X_2)^2}{2\sigma^2} = \left(\frac{X_1 - X_2}{\sqrt{2}\sigma}\right)^2$$

Como $X_1, X_2 \sim N(\mu, \sigma^2)$, temos que $X_1 - X_2 \sim N(0, 2\sigma^2)$, isto é, $Z = \frac{X_1 - X_2}{\sqrt{2}\sigma} \sim N(0,1)$. Assim, $\frac{S^2}{\sigma^2} = Z \sim N(0,1)$. ■

Na prática, usam-se tabelas para calcular probabilidades associadas à distribuição qui-quadrado com até 30 graus de liberdade. Depois disto, usa-se a seguinte aproximação que não demonstraremos:

Proposição 20 *Se* $Y \sim \chi^2(n)$ *então* $Z = \sqrt{2Y} - \sqrt{2n-1}$ *é aproximadamente* $N(0,1)$ *(especialmente para* n *grande).*

8.4.1 Exercícios

Ex. 13 *Seja* $X \sim N(0, \sigma^2)$ *e considere uma AAS de tamanho* 2 *de* X*. Calcule* $\Pr\left(X_1^2 + X_2^2 \leq \sigma^2\right)$.

Ex. 14 *Suponha que* $X \sim \chi^2(5)$*. Determine* $\Pr(2 < X < 4)$.

Ex. 15 *Suponha que* $X \sim N(1, 9)$*. Retire uma amostra* $X_1, X_2, \ldots, X_{10}$ *e calcule* S^2 *para esta amostra. Use uma tabela para estimar a probabilidade* $\Pr\left(S^2 < 10\right)$ *e o valor de* a *tal que* $\Pr\left(S^2 < a\right) = 5\%$.

Ex. 16 *Encontre os percentis* P_5 *e* P_{95} *(isto é, os* $0,05$ *e* $0,95$ *quantis) da distribuição* $\chi^2(10)$.

Ex. 17 *Seja* $Y \sim \chi^2(41)$*. Use a aproximação normal para estimar* $\Pr(Y \geq 50)$ *e* $\Pr(Y \leq 18)$*. Para que valor de* a *teríamos* $\Pr(Y \geq a) = 5\%$*?*

8.5 Distribuição t de Student

O químico e estatístico inglês William Sealy Gosset (1876–1937) trabalhava na cervejaria Guiness em Dublin e aplicava seus métodos estatísticos (que freqüentemente se utilizavam de amostras "pequenas") para selecionar as melhores variedades de cevada. Como a cervejaria proibia que seus funcionários publicassem artigos científicos para proteger seus segredos industriais, Gosset publicou seus resultados sob o pseudônimo "Student". Um de seus resultados foi a determinação da distribuição t (que ficou conhecida como "t de Student"), publicada em 1908.

Definição 21 *Sejam* $Z \sim N(0,1)$ *e* $X \sim \chi^2(n)$ *independentes. A distribuição da variável*

$$\boxed{T = \frac{Z}{\sqrt{X/n}}}$$

é chamda de ***t de Student com*** n ***graus de liberdade*** *(notação:* $T \sim t(n)$*).*

Proposição 22 *A densidade de $T \sim t(n)$ é*

$$\boxed{f(t) = \text{TDen}(t;n) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)\sqrt{\pi n}}\left(1+\frac{t^2}{n}\right)^{-(n+1)/2}}$$

Prova. A demonstração é simplesmente uma aplicação dos métodos de mudança de variáveis, mas as contas ficam um bocado longas... A distribuição conjunta de X e Z é o produto das marginais

$$f(x,z) = \frac{x^{(n-2)/2}e^{-x/2}}{2^{n/2}\Gamma(n/2)}\frac{1}{\sqrt{2\pi}}e^{-z^2/2} \text{ (para } x \geq 0)$$

Façamos a mudança de variáveis para X e $T = \frac{Z}{\sqrt{X/n}}$. O Jacobiano é

$$\frac{\partial(X,T)}{\partial(X,Z)} = \begin{vmatrix} 1 & 0 \\ \frac{\sqrt{n}ZX^{-3/2}}{2} & \frac{1}{\sqrt{X/n}} \end{vmatrix} = \frac{\sqrt{n}}{\sqrt{X}}$$

Então a nova densidade conjunta é obtida susbtituindo $z = t\sqrt{x/n}$ e usando este Jacobiano:

$$f(x,t) = f\left(x, t\sqrt{x}/n\right)\frac{\sqrt{x}}{\sqrt{n}} = \frac{x^{n/2}e^{-x/2}}{\sqrt{n}2^{n/2}\Gamma(n/2)}\frac{1}{\sqrt{2\pi}}e^{-t^2x/(2n)} \text{ para } x \geq 0$$

Integrando com relação a x encontramos a marginal de t:

$$f(t) = \frac{1}{\sqrt{\pi n}2^{(n+1)/2}\Gamma(n/2)}\int_0^\infty x^{n/2}e^{-\left(1+t^2/n\right)x/2}dx$$

Tomando $u = \frac{1}{2}\left(1+\frac{t^2}{n^2}\right)x$, temos $du = \frac{1}{2}\left(1+\frac{t^2}{n^2}\right)dx$ e

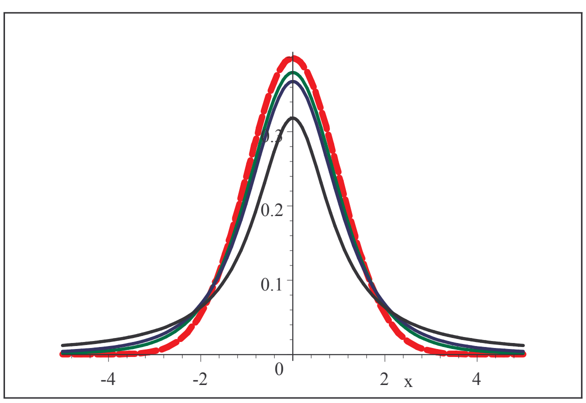
$$\begin{aligned} f(t) &= \frac{1}{\sqrt{\pi n}2^{(n+1)/2}\Gamma(n/2)}\int_0^\infty \frac{u^{n/2}}{\left(\frac{1}{2}\left(1+\frac{t^2}{n}\right)\right)^{n/2}}e^{-u}\frac{du}{\frac{1}{2}\left(1+\frac{t^2}{n}\right)} = \\ &= \frac{2^{(n+1)/2}}{\sqrt{\pi n}2^{(n+1)/2}\Gamma(n/2)}\left(1+\frac{t^2}{n}\right)^{-(n+1)/2}\int_0^\infty u^{n/2}e^{-u}du \end{aligned}$$

Enfim, usando a definição da função Gama:

$$f(t) = \frac{1}{\sqrt{\pi n}\Gamma(n/2)}\left(1+\frac{t^2}{n}\right)^{-(n+1)/2}\Gamma\left(\frac{n+1}{2}\right)$$

■

Note que $f(t)$ é uma função par; de fato, o gráfico de $f(t)$ é bastante assemelhado ao de uma distribuição normal, mas ligeiramente menos "concentrada" na origem (isto é, a distribuição t tem "caudas mais grossas").



$t(1)$, $t(3)$ e $t(5)$. Em tracejado, $N(0,1)$.

Proposição 23 *Se $T \sim t(n)$, então*

$$\boxed{E(T)=0} \text{ (desde que } n \geq 2\text{)}$$
$$\boxed{Var(T)=\tfrac{n}{n-2}} \text{ (desde que } n \geq 3\text{)}$$
$$\boxed{Med(T)=0} \text{ (qualquer } n \geq 1\text{)}$$

Nos outros casos, o valor esperado e variância de T não existem.

Prova. É fácil ver que a mediana é 0 pois $f(t)$ é par. Quanto ao valor esperado:

$$E(T)=k\int_{-\infty}^{\infty} t\left(1+\frac{t^2}{n}\right)^{-(n+1)/2} dt$$

onde k é uma constante que não depende de t. Por simetria, se esta integral existir, será 0; para ver se a integral converge, basta tomar $u=1+\frac{t^2}{n}$ e verificar a convergência de

$$\int_0^{\infty} t\left(1+\frac{t^2}{n}\right)^{-(n+1)/2} dt=\int_1^{\infty} u^{-(n+1)/2}\frac{ndu}{2}=\left(\frac{u^{(1-n)/2}}{(1-n)/2}\frac{n}{2}\right]_{u=1}^{u=\infty}$$

Esta integral converge desde que $1-n<0$, isto é, $n>1$.
Enfim, para calcular $Var(T)=E\left(T^2\right)$, façamos:

$$\begin{aligned}
E\left(T^2\right) &= 2\int_0^{\infty} t^2 f(t)\,dt=\frac{2\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{\pi n}\Gamma\left(\frac{n}{2}\right)}\int_0^{\infty} t^2\left(1+\frac{t^2}{n}\right)^{-(n+1)/2} dt= \\
&= \frac{2\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{\pi n}\Gamma\left(\frac{n}{2}\right)}\int_0^{\infty}\frac{nt}{1-n}d\left(\left(1+\frac{t^2}{n}\right)^{-(n-1)/2}\right)= \\
&= \frac{2\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{\pi n}\Gamma\left(\frac{n}{2}\right)}\left\{\left(\frac{nt}{1-n}\left(1+\frac{t^2}{n}\right)^{-(n-1)/2}\right]_{t=0}^{t=\infty}-\int_0^{\infty}\frac{n}{1-n}\left(1+\frac{t^2}{n}\right)^{-(n-1)/2}dt\right\}
\end{aligned}$$

O primeiro termo se anula desde que $n>1$. Assim:

$$E\left(T^2\right)=\frac{2\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{\pi n}\Gamma\left(\frac{n}{2}\right)}\frac{n}{n-1}\int_0^{\infty}\left(1+\frac{t^2}{n}\right)^{-(n-1)/2}dt$$

Fazendo $u=t\sqrt{\frac{n-2}{n}}$, temos:

$$E\left(T^2\right)=\frac{2\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{\pi n}\Gamma\left(\frac{n}{2}\right)}\frac{n}{n-1}\int_0^{\infty}\left(1+\frac{u^2}{n-2}\right)^{-(n-1)/2}\sqrt{\frac{n}{n-2}}du$$

Mas, se $n>2$, podemos criar a variável $Y\sim t(n-2)$ cuja densidade será:

$$f(y)=\frac{\Gamma\left(\frac{n-1}{2}\right)}{\Gamma\left(\frac{n-2}{2}\right)\sqrt{\pi(n-2)}}\left(1+\frac{y^2}{n-2}\right)^{-(n-1)/2}$$

Como a integral de $f(y)$ tem de ser 1, concluímos que

$$\frac{\Gamma\left(\frac{n-1}{2}\right)}{\Gamma\left(\frac{n-2}{2}\right)\sqrt{\pi(n-2)}}\int_0^{\infty}\left(1+\frac{y^2}{n-2}\right)^{-(n-1)/2}dy=\frac{1}{2}$$

Esta integral é exatamente a integral que precisávamos para terminar o cálculo de $E\left(T^2\right)$. De fato:

$$\begin{aligned}
E\left(T^2\right) &= \frac{2\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{\pi n}\Gamma\left(\frac{n}{2}\right)}\frac{n}{n-1}\sqrt{\frac{n}{n-2}}\frac{1}{2}\frac{\Gamma\left(\frac{n-2}{2}\right)\sqrt{\pi(n-2)}}{\Gamma\left(\frac{n-1}{2}\right)}= \\
&= \frac{n}{n-1}\frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n-1}{2}\right)}\frac{\Gamma\left(\frac{n-2}{2}\right)}{\Gamma\left(\frac{n}{2}\right)}=\frac{n}{n-1}\frac{n-1}{2}\frac{2}{n-2}=\frac{n}{n-2}
\end{aligned}$$

Portanto, $Var(T) = \frac{n}{n-2}$. ■

Em particular, $Var(T) = \frac{n}{n-2} \geq 1$ confirma que a distribuição t de Student "se espalha mais" do que $N(0,1)$ (que tem variância 1), ou seja, tem "caudas mais grossas".

Os casos $n = 1$ e $n = 2$ podem ser analisados separadamente:

- A distribuição $t(1)$ tem densidade dada por

$$\text{TDen}(t;1) = \frac{\Gamma(1)}{\Gamma\left(\frac{1}{2}\right)\sqrt{\pi}}\left(1+\frac{t^2}{2}\right)^{-1/2} = \frac{1}{\pi}\frac{1}{1+t^2}$$

  que é exatamente a **distribuição de Cauchy** com f.d.a. dada por

$$\text{TDist}(t;1) = \frac{1}{2} + \frac{\arctan t}{\pi}$$

  (já havíamos visto que esta distribuição não tem média nem variância).

- A distribuição $t(2)$ tem densidade

$$\text{TDen}(t;2) = \frac{\Gamma\left(\frac{3}{2}\right)}{\Gamma(1)\sqrt{2\pi}}\left(1+\frac{t^2}{2}\right)^{-3/2} = \frac{1}{(2+t^2)^{3/2}}$$

  cuja distribuição acumulada também pode ser encontrada analiticamente

$$\text{TDist}(t;2) = \frac{1}{2}\left(1+\frac{t}{\sqrt{2+t^2}}\right)$$

8.5.1 Para que serve?

Teorema 24 *Seja X_1, X_2, ..., X_n uma AAS da variável aleatória normal X. Então $\bar{X}$ e S^2 são independentes, e*

$$T = \frac{\bar{X}-\mu}{S/\sqrt{n}} \sim t(n-1)$$

Prova. Vamos mostrar que $\bar{X}$ e S^2 são independentes apenas no caso $n = 2$. De fato, sabemos que $X_1 + X_2$ e $X_1 - X_2$ são independentes (veja a demonstração da proposição 7.4 com $a = 1$ e $b = -1$). Portanto, $\bar{X} = \frac{X_1+X_2}{2}$ e $S^2 = \frac{(X_1-X_2)^2}{2}$ serão independentes.
Para a segunda afirmação, sabemos que

$$Z = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1) \text{ e } Y = \frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$$

Como $\bar{X}$ e S^2 são independentes para qualquer n, Z e Y também o serão. Pela definição da distribuição t, temos:

$$T = \frac{Z}{\sqrt{Y/(n-1)}} \sim t(n-1)$$

Expandindo o lado esquerdo:

$$T = \frac{Z}{\sqrt{Y/(n-1)}} = \frac{\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}}{S/\sigma} = \frac{\bar{X}-\mu}{S/\sqrt{n}}$$

concluindo a demonstração. ■

Exemplo 25 *Seja $X \sim N(\mu,\sigma^2)$. Tome uma AAS de X com 8 amostras. Calcule e compare a) $\Pr\left(\bar{X}-\mu > \sigma/2\right)$ com b) $\Pr\left(\bar{X}-\mu > S/2\right)$:*
a) Como $Z = \frac{\bar{X}-\mu}{\sigma/\sqrt{8}} \sim N(0,1)$, temos:

$$\Pr\left(\bar{X}-\mu > \sigma/2\right) = \Pr\left(\frac{\bar{X}-\mu}{\sigma/\sqrt{8}} > \frac{\sqrt{8}}{2}\right) = \Pr\left(Z > \sqrt{2}\right) = 1 - \text{NormalDist}\left(\sqrt{2}\right) = 7.865\%$$

b) Como $T = \frac{\bar{X}-\mu}{S/\sqrt{8}} \sim t(7)$, *temos:*

$$\Pr\left(\bar{X} - \mu > \frac{S}{2}\right) = \Pr\left(\frac{\bar{X}-\mu}{S/\sqrt{8}} > \frac{\sqrt{8}}{2}\right) = \Pr\left(T > \sqrt{2}\right) = 1 - \text{TDist}\left(\sqrt{2}; 7\right) = 10.01\%$$

Note como as duas probabilidades são siginificativamente diferentes. No entanto, se σ *é desconhecido, é impossível verificar se o evento em (a) acontece ou não – enquanto o evento em (b) não depende de* σ *para absolutamente nada!*

Em suma, a distribuição t é utilizada para estimar ou testar hipóteses sobre a distribuição de μ quando σ é desconhecido e n é pequeno.

Exemplo 26 *Vamos encontrar um intervalo de confiança para* μ *baseado no* $\bar{X}$ *e* S^2 *de uma amostra de* 8 *elementos. Para começar, encontre os* 0.025 *e* 0.975 *percentis da distribuição* $t(7)$*:*

$$\text{TInv}(0.975; 7) = 2.365 \text{ e } \text{TInv}(0.025; 7) = -2.365$$

Em outras palavras:

$$\Pr(|T| \leq 2.365) = 95\% \Rightarrow \Pr\left(\left|\bar{X} - \mu\right| \leq \frac{2.365}{\sqrt{8}} S\right) = 95\% \Rightarrow \Pr\left(\left|\bar{X} - \mu\right| \leq 0.836S\right) = 95\%$$

o que nos dá o intervalo para μ

$$\mu \in \left[\bar{X} - 0.836S, \bar{X} + 0.836S\right]$$

com 95% *de confiança! Este processo é facilmente adaptável para qualquer outro número* n *de amostras*[1].

8.6 A distribuição F de Snedecor

O nome desta distribuição vem de George W. Snedecor (1881 - 1974), um matemático e físico americano especialmente interessado em experimentos biológicos e agrícolas.

Definição 27 *Sejam* $X \sim \chi^2(n)$ *e* $Y \sim \chi^2(m)$ *variáveis independentes. A distribuição da variável*

$$\boxed{W = \frac{Y/m}{X/n}}$$

é chamada de F ***(de Snedecor) com*** m ***graus de liberdade no numerador e*** n ***graus de liberdade no denominador*** *(notação:* $W \sim F(m, n)$*).*

Proposição 28 *Se* $W \sim F(m, n)$, *então* $\frac{1}{W} \sim F(n, m)$.

Prova. Claro, pois

$$W = \frac{Y/m}{X/n} \Rightarrow \frac{1}{W} = \frac{X/n}{Y/m}$$

e esta também é a razão de duas distribuições qui-quadrado independentes, e portanto satisfaz a definição da distribuição $F(n, m)$. ■

[1] Se soubéssemos σ usaríamos $Z = \sqrt{8}\frac{\bar{X}-\mu}{\sigma} \sim N(0, 1)$, e então:

$$\begin{aligned} \text{NormalInv}(0.975) &= 1.960 \Rightarrow \\ &\Rightarrow \Pr(|Z| < 1.960) = 95\% \Rightarrow \Pr\left(\left|\bar{X} - \mu\right| \leq \frac{1.960}{\sqrt{8}}\sigma\right) = 95\% \Rightarrow \Pr\left(\left|\bar{X} - \mu\right| \leq 0.693\sigma\right) = 95\% \\ &\Rightarrow \mu \in \left[\bar{X} - 0.693\sigma, \bar{X} + 0.693\sigma\right] \text{ com 95\% de confiança.} \end{aligned}$$

Proposição 29 *A densidade de $W \sim F(m,n)$ é dada por*

$$f(w) = \mathrm{FDen}(w;m,n) = \frac{\Gamma\left(\frac{m+n}{2}\right) m^{m/2} n^{n/2}}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)} \frac{w^{(m/2)-1}}{(n+mw)^{(m+n)/2}} \text{ se } w \geq 0$$

e $f(w) = 0$ caso $w < 0$.

Prova. De novo, é só fazer um monte de contas. Afinal, se $X \sim \chi^2(n)$ e $Y \sim \chi^2(m)$ são independentes, sua distribuição conjunta é (para $x, y \geq 0$):

$$f(x,y) = \frac{1}{\Gamma\left(\frac{n}{2}\right) 2^{n/2}} x^{n/2-1} e^{-x/2} \frac{1}{\Gamma\left(\frac{m}{2}\right) 2^{m/2}} y^{m/2-1} e^{-y/2}$$

Fazendo a transformação $(X,Y) \to (W,Y)$ temos

$$X = \frac{nY}{mW} \Rightarrow \left|\frac{\partial(X,Y)}{\partial(W,Y)}\right| = \left\| \begin{array}{cc} -\frac{nY}{mW^2} & \frac{n}{mW} \\ 0 & 1 \end{array} \right\| = \frac{nY}{mW^2}$$

Portanto, a distribuição conjunta de W e Y será

$$\begin{aligned} f(w,y) &= \left(\frac{1}{\Gamma\left(\frac{n}{2}\right) 2^{n/2}} \left(\frac{ny}{mw}\right)^{n/2-1} e^{-ny/(2mw)}\right)\left(\frac{1}{\Gamma\left(\frac{m}{2}\right) 2^{m/2}} y^{m/2-1} e^{-y/2}\right)\frac{ny}{mw^2} = \\ &= \frac{n}{mw^2} \frac{\left(\frac{n}{mw}\right)^{n/2-1}}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right) 2^{(m+n)/2}} y^{(m+n)/2-1} \exp\left(-\frac{y}{2}\left(1+\frac{n}{mw}\right)\right) \end{aligned}$$

Note que esta fórmula vale para $w, y \geq 0$ (caso contrário, $f = 0$). Integrando com relação a y encontraremos a densidade marginal de W:

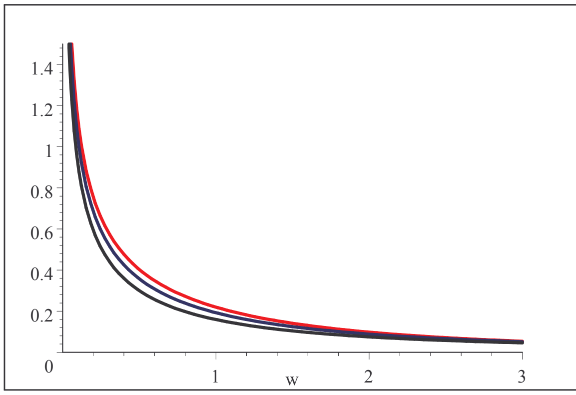
$$f(w) = \frac{1}{w} \frac{\left(\frac{n}{mw}\right)^{n/2}}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right) 2^{(m+n)/2}} \int_0^\infty y^{(m+n)/2-1} \exp\left(-\frac{y}{2}\left(1+\frac{n}{mw}\right)\right) dy$$

Fazendo $u = \left(1+\frac{n}{mw}\right)\frac{y}{2} = \frac{mw+n}{mw}\frac{y}{2}$, obtemos

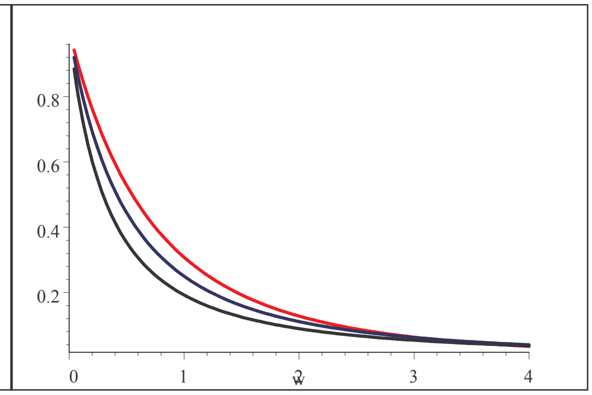
$$\begin{aligned} f(w) &= \frac{1}{w} \frac{\left(\frac{n}{mw}\right)^{n/2}}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right) 2^{(m+n)/2}} \int_0^\infty \left(\frac{2mw}{mw+n}\right)^{(m+n)/2-1} u^{(m+n)/2-1} e^{-u} \frac{2mwdu}{mw+n} = \\ &= \frac{1}{w} \frac{\left(\frac{n}{mw}\right)^{n/2}}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right) 2^{(m+n)/2}} \left(\frac{2mw}{mw+n}\right)^{(m+n)/2} \int_0^\infty u^{(m+n)/2-1} e^{-u} du = \\ &= \frac{1}{w} \frac{\left(\frac{n}{mw}\right)^{n/2}}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right) 2^{(m+n)/2}} \left(\frac{2mw}{mw+n}\right)^{(m+n)/2} \Gamma\left(\frac{m+n}{2}\right) = \\ &= \frac{\Gamma\left(\frac{m+n}{2}\right) m^{m/2} n^{n/2}}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)} \frac{w^{(m/2)-1}}{(n+mw)^{(m+n)/2}} \end{aligned}$$

■

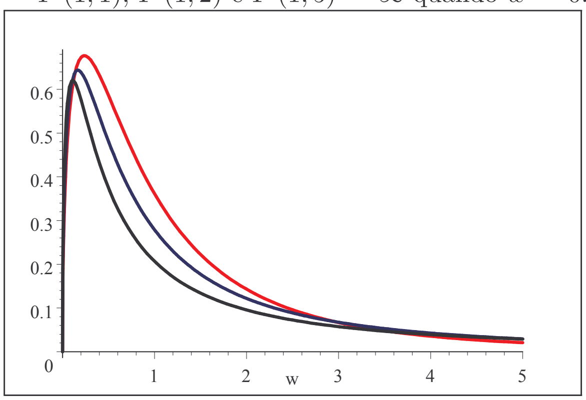
O formato da distribuição $F(m,n)$ varia com m e n. Para $m = 1$ e $m = 2$, a densidade é decrescente, enquanto para $m \geq 3$ a densidade é unimodal. Para $m = 1$, a densidade tem o eixo vertical como assíntota. Veja alguns exemplos abaixo (se sua impressão for colorida, $n = 1$ é preto, $n = 2$ é azul e $n = 5$ é vermelho):



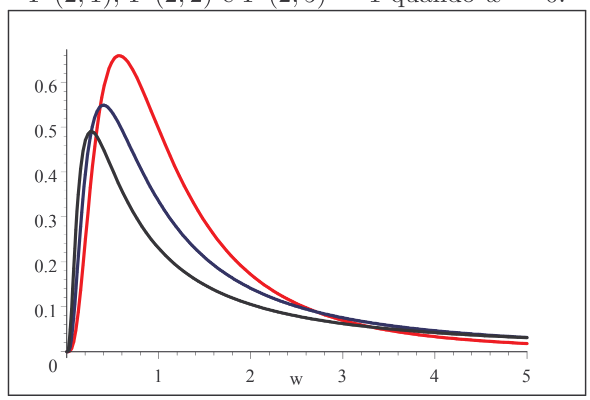
$F(1,1)$, $F(1,2)$ e $F(1,5) \to \infty$ quando $w \to 0$.



$F(2,1)$, $F(2,2)$ e $F(2,5) \to 1$ quando $w \to 0$.



$F(3,1)$, $F(3,2)$ e $F(3,5) \to 0$ quando $w \to 0$.



$F(10,1)$ (pico baixo), $F(10,2)$ e $F(10,5)$ (pico alto)

Os cálculos do valor esperado e variância de W são ainda mais feios e difíceis do que os anteriores – e, portanto, sero deixados para o leitor.

Proposição 30 *Se $W \sim F(m,n)$, então*

$$\boxed{E(W) = \frac{n}{n-2}} \quad \textit{(desde que } n \geq 3\textit{)}$$

$$\boxed{Var(W) = \frac{2n^2(m+n-2)}{m(n-2)^2(n-4)}} \quad \textit{(desde que } n \geq 5\textit{)}$$

Nos outros casos, o valor esperado e variância de W não existem.

Prova. Temos

$$\begin{aligned} E(W) &= \frac{\Gamma\left(\frac{m+n}{2}\right) m^{m/2} n^{n/2}}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)} \int_0^\infty w \frac{w^{(m/2)-1}}{(n+mw)^{(m+n)/2}} dw = \\ &= \frac{\Gamma\left(\frac{m+n}{2}\right) m^{m/2} n^{n/2}}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)} \int_0^\infty w^{m/2} \frac{d\left((n+mw)^{-(m+n)/2+1}\right)}{m\left(-\frac{(m+n)}{2}+1\right)} \end{aligned}$$

Fazendo a integração por partes, o termo "uv" será 0 desde que $\frac{m+n}{2} - 1 > \frac{m}{2}$, isto é, desde que $n > 2$. Neste caso, ficamos apenas com o outro termo $\int vdu$:

$$E(W) = \frac{\Gamma\left(\frac{m+n}{2}\right) m^{m/2} n^{n/2}}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)} \frac{2}{m(m+n-2)} \int_0^\infty \frac{m}{2} \frac{w^{m/2-1}}{(n+mw)^{(m+n)/2-1}} dw$$

Agora façamos $u = \frac{(n-2)w}{n}$. Note que

$$n + mw = n + m\frac{nu}{n-2} = \frac{n}{n-2}\left((n-2) + mu\right)$$

então

$$
\begin{aligned}
E(W) &= \frac{\Gamma\left(\frac{m+n}{2}\right) m^{m/2} n^{n/2}}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)} \frac{1}{m+n-2} \int_0^\infty \left(\frac{n}{n-2}\right)^{m/2-1} \frac{u^{m/2-1}}{((n-2)+mu)^{(m+n-2)/2}} \left(\frac{n-2}{n}\right)^{(m+n)/2-1} \frac{n}{n-2} du = \\
&= \frac{\Gamma\left(\frac{m+n}{2}\right) m^{m/2} n^{n/2}}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)} \frac{1}{m+n-2} \left(\frac{n-2}{n}\right)^{n/2-1} \int_0^\infty \frac{u^{m/2-1}}{((n-2)+mu)^{(m+(n-2))/2}} du
\end{aligned}
$$

Note que, a menos das constantes, a integral que nos resta calcular é a densidade de uma variável $X \sim F(m, n-2)$. De fato:

$$
\begin{aligned}
\int_0^\infty f(x)\, dx &= 1 \Rightarrow \frac{\Gamma\left(\frac{m+n-2}{2}\right) m^{m/2} (n-2)^{(n-2)/2}}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n-2}{2}\right)} \int_0^\infty \frac{x^{m/2-1}}{((n-2)+mx)^{(m+n-2)/2}} dy = 1 \Rightarrow \\
&\Rightarrow \int_0^\infty \frac{x^{m/2-1}}{((n-2)+mx)^{(m+n-2)/2}} dy = \frac{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n-2}{2}\right)}{\Gamma\left(\frac{m+n-2}{2}\right) m^{m/2} (n-2)^{(n-2)/2}}
\end{aligned}
$$

Juntando tudo:

$$
\begin{aligned}
E(W) &= \frac{\Gamma\left(\frac{m+n}{2}\right) m^{m/2} n^{n/2}}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)} \frac{1}{m+n-2} \left(\frac{n-2}{n}\right)^{n/2-1} \frac{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n-2}{2}\right)}{\Gamma\left(\frac{m+n-2}{2}\right) m^{m/2} (n-2)^{(n-2)/2}} = \\
&= \frac{m+n-2}{2} \frac{2n}{n-2} \frac{1}{m+n-2} = \frac{n}{n-2}
\end{aligned}
$$

O cálculo de $E\left(W^2\right)$ pode ser feito de forma análoga, usando integrais por partes duas vezes e então ajustando a variável de integração para ficar com uma integral conhecida. Ao final, encontramos (desde que $n > 4$):

$$
E\left(W^2\right) = \frac{n^2(m+2)}{(n-4)(n-2)m}
$$

e, portanto

$$
Var(W) = E\left(W^2\right) - (E(W))^2 = \frac{n^2(m+2)}{(n-4)(n-2)m} - \left(\frac{n}{n-2}\right)^2 = \frac{2n^2(m+n-2)}{m(n-2)^2(n-4)}
$$

■

8.6.1 Para que serve?

Teorema 31 *Seja X_1, X_2, ..., X_n uma AAS da variável aleatória $X \sim N\left(\mu_1, \sigma_1^2\right)$ e seja $Y_1, Y_2, ..., Ym$ uma AAS da variável aleatória $Y \sim N\left(\mu_2, \sigma_2^2\right)$ (onde X e Y são independentes). Então:*

$$
W = \frac{S_Y^2}{S_X^2}\frac{\sigma_1^2}{\sigma_2^2} \sim F(m-1, n-1)
$$

Prova. Sabemos que

$$
\frac{(n-1)S_X^2}{\sigma_1^2} \sim \chi^2(n-1) \text{ e } \frac{(m-1)S_Y^2}{\sigma_2^2} \sim \chi^2(m-1)
$$

e, como X e Y são independentes, S_X e S_Y também o serão. Portanto, pela definição da distribuição F de Snedecor:

$$
\frac{\frac{(m-1)S_Y^2}{\sigma_2^2}\frac{1}{m-1}}{\frac{(n-1)S_X^2}{\sigma_1^2}\frac{1}{n-1}} \sim F(m-1, n-1)
$$

ou seja, simplificando o lado esquerdo

$$
\frac{S_Y^2}{S_X^2}\frac{\sigma_1^2}{\sigma_2^2} \sim F(m-1, n-1)
$$

■

Exemplo 32 *Sejam* $X \sim N(\mu_1, \sigma^2)$ *e* $Y \sim N(\mu_2, \sigma^2)$ *(com a mesma variância). Tome uma AAS de* X *com* 9 *amostras e uma AAS de* Y *com* 13 *amostras. Calcule* $\Pr(S_X > S_Y)$ *e encontre* a *tal que* $\Pr(S_Y > aS_X) = 0.95$. *Seja* $W = \frac{S_Y}{S_X}$. *Pelo teorema anterior, sabemos que* $W \sim F(12, 8)$. *Então:*

$$\begin{aligned} \Pr(S_X > S_Y) &= \Pr(W < 1) = \text{FDist}(1; 12, 8) = 48.261\% \\ \Pr(S_Y > aS_X) &= \Pr(W > a) = 95\% \Rightarrow a = \text{FInv}(0.05; 12, 8) = 0.351 \end{aligned}$$

Nota 33 *Ao consultar uma tabela, lembre-se que*

$$\Pr(W > a) + \Pr\left(\frac{1}{W} > \frac{1}{a}\right) = 1$$

ou seja,

$$\Pr(W > a) = p \Leftrightarrow \Pr\left(\frac{1}{W} > \frac{1}{a}\right) = 1 - p$$

Mas, como $W \sim F(m, n) \Leftrightarrow \frac{1}{W} \sim F(n, m)$, *você pode encontrar o valor de* a *na tabela* $F(m, n)$ *procurando a probabilidade* p *ou encontrar o valor de* $\frac{1}{a}$ *na tabela* $F(n, m)$ *procurando* $1 - p$. *Em outras palavras:*

$$\boxed{\text{FInv}(p; m, n) = \frac{1}{\text{FInv}(1-p; n, m)}}$$

Graças a esta propriedade, basta termos tabelas com $0 \leq p \leq 0.5$ *– outros valores serão obtidos pela inversão acima.*

Em suma, a distribuição F de Snedecor é utilizada quando queremos avaliar probabilidades e hipóteses que lidam com as variâncias de duas populações distintas (**análise de variância**).

8.6.2 Exercícios

Ex. 18 *Seja* $X \sim t(n)$. *Determine a distribuição de* $Y = X^2$.

Ex. 19 *Seja* $X \sim F(5, 2)$. *Com o auxílio de uma tabela ou computador, calcule* $\Pr(X < 4)$.

Ex. 20 *Seja* $X \sim t(5)$. *Com o auxílio de uma tabela ou computador, calcule* $\Pr(X > 2)$.

Ex. 21 *Encontre os percentis* $P_{0.05}$ *e* $P_{0.95}$ *para* $F(3; 5)$ *e compare-os com os mesmos percentis de* $F(5; 3)$.

Ex. 22 *Encontre os percentis* $P_{0.05}$ *e* $P_{0.95}$ *para* $t(5)$. *Qual a relação entre eles?*

Ex. 23 *Mostre que a moda de* $F(m; n)$ *é* $\frac{m-2}{m}\frac{n}{n+2}$ *para* $m > 2$. *Para* $m < 2$, *mostre que a densidade é decrescente e, portanto, a moda é* 0.

Ex. 24 *Seja* $X \sim N(\mu, \sigma^2)$. *Tomando* 15 *amostras desta distribuição, calculam-se* $\bar{X}$ *e* S^2.
a) Calcule $\Pr(\bar{X} > \mu + 2\sigma \mid S^2 = 25)$.
b) Calcule $\Pr(\bar{X} > \mu + 2S)$.
c) Encontre o valor de a *tal que* $\Pr(\bar{X} - \mu > aS) = 95\%$
d) Encontre o valor de b *tal que* $\Pr(-bS < \bar{X} - \mu < bS) = 95\%$. *A partir daqui, conclua que* $\mu \in [\bar{X} - bS, \bar{X} + bS]$ *com* 95% *de confiança.*

Ex. 25 *Suponha que* $X \sim N(\mu_1, \sigma^2)$ *e* $Y \sim N(\mu_2, 4\sigma^2)$ *(isto é, o desvio-padrão de* Y *é o dobro do desvio-padrão de* X*). Você coleta* 17 *amostras de* X *e* 7 *amostras de* Y *e encontra* $S_X = S_Y$. *Para verificar se isto é verossímil, calcule*

$$\Pr(S_X \geq S_Y)$$

para 17 *amostras independentes de* X *e* 7 *de* Y.

8.7 Exercícios de Provas

Ex. 26 (A2 2005.2) *Seja X uma variável aleatória com distribuição uniforme no intervalo $[0,a]$. Considere uma AAS X_1, X_2, ..., X_n de tamanho n de X. Seja $Y = \max(X_1, X_2, ..., X_n)$. É fácil ver que a função de distribuição* ***acumulada*** *de Y é*

$$F(y) = \begin{cases} 0, \text{ para } y < 0 \\ \frac{y^n}{a^n}, \text{ para } 0 \leq y \leq a \\ 1, \text{ para } y > a \end{cases}$$

a) Mostre que

$$E(Y) = \frac{n}{n+1}a \quad e \quad Var(Y) = \frac{n}{(n+1)^2(n+2)}a^2$$

b) Seja $Z = \frac{n+1}{n}Y$. Considerando Z como um estimador para a, calcule $Viés(Z)$ e $EQM(Z;a)$.
c) Os estimadores Y e Z são consistentes para estimar a? Explique.

Ex. 27 (A2 2005.2) *Encontre a moda da distribuição $\chi^2(n)$ para $n \geq 2$.*

Ex. 28 (A2 2005.2) *Uma população tem distribuição $X \sim N(50, 10^2)$. Desta população, retira-se uma amostra com 10 elementos e calculam-se $\bar{X}$ e S^2. Encontre:*
a) $\Pr(S^2 < 163,16)$.
b) Os valores de a e b tais que $\Pr(S^2 < a) = \Pr(S^2 > b) = 5\%$.
c) O valor de c tal que $\Pr(|\bar{X} - 50| < cS) = 90\%$.

Ex. 29 (AS 2005.2) *Os pacotes de açúcar da Onion Sugar Inc. tem pesos com distribuição normal de média $\mu = 1005g$ e desvio-padrão $\sigma = 10g$. O jornal Tabajara está prestes a avaliar a denúncia de que a Onion rouba nos pesos dos pacotes (ou tem um péssimo controle de qualidade) pesando 9 pacotes de açúcar e calculando sua média $\bar{X}$ e seu $S^2 = \frac{\sum(X_i - \bar{X})^2}{8}$.*
a) Qual a probabilidade do primeiro destes 9 pacotes ter menos de $1kg$?
b) Qual a probabilidade dos 9 pacotes terem, somados, $9kg$ ou menos?
c) Qual a probabilidade dos 9 pacotes terem um S^2 maior do que $100g^2$?
d) ***Sem usar os valores exatos de μ e σ****, calcule* $\Pr(|\bar{X} - \mu| > 0.5\sigma)$ *e compare-a com* $\Pr(|\bar{X} - \mu| > 0.5S)$.

Ex. 30 (AS 2005.2) *A distribuição de probabilidade da variável X é uniforme em $[0, 2a]$. Desta distribuição, retira-se uma amostragem aleatória simples com apenas duas amostras independentes X_1 e X_2. Seja $Z = \frac{\sqrt{X_1X_2}}{2a}$. Pode-se mostrar que $E(Z) = \frac{4}{9}$, $Var(Z) = \frac{17}{324}$ e a f.d.p. e a f.d.a de Z são respectivamente*

$$f(z) = \begin{cases} -4z \ln z \text{ para } 0 < z < 1 \\ 0, \text{ caso contrário} \end{cases} \quad e \; F(z) = \begin{cases} 0, \text{ para } z \leq 0 \\ z^2(1 - \ln z^2), \text{ para } 0 < z < 1 \\ 1, \text{ para } z \geq 1 \end{cases}$$

cujos principais quantis estão na tabela abaixo

$F(z)$	0.05	0.1	0.15	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.85	0.9	0.95
z	0.093	0.143	0.185	0.224	0.295	0.364	0.432	0.502	0.578	0.662	0.711	0.767	0.837

a) Considere o uso do estimador $G = \sqrt{X_1X_2}$ para estimar a. Este estimador é viesado? Calcule o erro quadrático médio de G (como estimador para a).
b) A média geométrica de duas amostras foi $G = \sqrt{X_1X_2} = 2.47$. Encontre um intervalo de confiança para a com 80% de confiança.