

# Capítulo 1

## Princípios Básicos de Estatística

### 1.1. A Natureza dos Problemas Científicos e da Experimentação

Desde o início da História do Homem, temos sentido a necessidade de entender o funcionamento do mundo que nos cerca. Essa necessidade de compreensão sempre foi motivada por questões muito práticas, como por exemplo, entender o comportamento do tempo, para prever a ocorrência de chuvas ou de secas, que têm implicações diretas sobre a sobrevivência das comunidades que dependem da agricultura para subsistir. (Aliás, essa é uma questão que ainda consome o trabalho diário de milhares de pessoas em todo o mundo, o que mostra como o conhecimento sobre certas questões fundamentais da vida pode se acumular muito lentamente ao longo dos anos.) Como bem demonstra o exemplo, a necessidade de compreender o mundo, embora algumas vezes busque apenas satisfazer a curiosidade de alguns “curiosos” sobre certas questões que os cercam, quase sempre nasce da vontade de se controlar ou prever um conjunto de fenômenos naturais, de forma a tornar possível melhorar, otimizar ou fazer com que a natureza funcione de forma a nos beneficiar de alguma maneira particular. No exemplo, todos esses elementos estão presentes, como vemos abaixo:

**Problema prático:** É necessário plantar para que se produzam alimentos. Secas e enxurradas destroem as plantações, consomem o trabalho e provocam falta de alimento. Seria bom saber onde e quando secas e enxurradas vão ocorrer, pois assim poderíamos escolher o momento certo para plantar e para armazenar os alimentos.

**Questão fundamental:** Como funciona o tempo?

**Finalidade básica da resposta:** Prever o momento adequado para o plantio e armazenamento de alimentos.

Embora o exemplo proposto seja extremamente simples, ele permite identificar os elementos fundamentais do problema científico:

- 1- O problema prático motivador;
- 2- A necessidade de compreensão do fenômeno;
- 3- A necessidade de previsão.

O problema prático motivador pode ser compreendido como a chama que aguça a curiosidade do investigador. Qualquer um que já teve a oportunidade de desenvolver e submeter um projeto a uma agência de financiamento já teve também que preencher um formulário onde se pergunta para quê serve o projeto e quais são os objetivos do projeto.

É difícil acreditar que alguém esteja interessado em um problema sem que haja qualquer objetivo a ser alcançado ou resposta a ser obtida. (Frequentemente as pessoas discordam sobre a relevância dos objetivos a serem alcançados numa investigação, embora eles nunca estejam ausentes.) O problema prático constitui a mola fundamental da era tecnológica e movimentou milhões de pessoas em todo mundo, com uma infinidade de pequenos e grandes problemas que precisam ser resolvidos.

Para que o problema possa ser resolvido de forma adequada, é necessário compreender os fenômenos naturais que geram o problema prático. Quais são as causas do fenômeno? Quais são as conseqüências? Como as causas e conseqüências estão relacionadas? A busca de respostas para essas questões é frequentemente denominada de **modelagem** do fenômeno. As causas e conseqüências são usualmente denominadas de variáveis do problema analisado. A estrutura que relaciona as variáveis do problema é denominada de **modelo**.

Nesse ponto, uma questão fundamental deve ser colocada: a identificação das variáveis de um problema implica necessariamente na observação do fenômeno e na obtenção de dados (atividade empírica), enquanto a construção de uma estrutura que relaciona as variáveis implica necessariamente em um processo abstrato para explicação e justificativa dos resultados observados (atividade teórica). Esse íntimo relacionamento existente entre as atividades empírica e teórica foi compreendido desde o Iluminismo. (Ainda hoje alguns “investigadores” continuam insistindo na discussão sem sentido sobre o que é mais importante - investigação experimental ou teórica. Não entre nessa, pois experimento sem teoria ou teoria sem experimento não faz sentido!!!!) Só podemos dizer que compreendemos um fenômeno se somos capazes de identificar as variáveis relevantes do problema e se somos capazes de dizer como certos grupos de variáveis influenciam os demais; ou seja, se temos um modelo para o fenômeno. Nessa fase, a atividade experimental tem como principais objetivos permitir a identificação adequada das variáveis relevantes do problema e a construção do modelo.

Finalmente, atinge-se a fase em que o conhecimento acumulado deve ser utilizado para resolver o problema proposto. Assim, o modelo deve ser utilizado para prover as respostas do problema. É a etapa de predição. A resposta é então implementada, visando resolver o problema prático que originou a investigação. Caso a resposta predita de fato resolva o problema prático, dizemos que o modelo desenvolvido é válido; caso contrário, a compreensão do fenômeno não foi adequada para resolver o problema e precisa ser reavaliada. Novamente a teoria e a prática estão inter-relacionadas, haja visto que a compreensão teórica só ganha importância se pode ser aplicada para resolver o problema prático original. Se isto não é possível, a teoria construída não tem validade no mundo que nos interessa de fato e tem que ser revista.

Tomando como base a discussão acima, vê-se que é através da experimentação que os problemas práticos são construídos, que as variáveis relevantes do problema são identificadas e que o modelo pode ser montado e validado. A prática teórica permite correlacionar as variáveis e fazer previsões, que fornecem as respostas para os problemas práticos originalmente propostos (e outros que porventura venham a serem propostos).

## 1.2. Metodologia Científica e Experimentação

As discussões apresentadas anteriormente podem ser colocadas num contexto mais geral, definindo-se a Metodologia Científica de tratar um problema. Este contexto mais genérico está apresentado resumidamente na Figura 1.1.

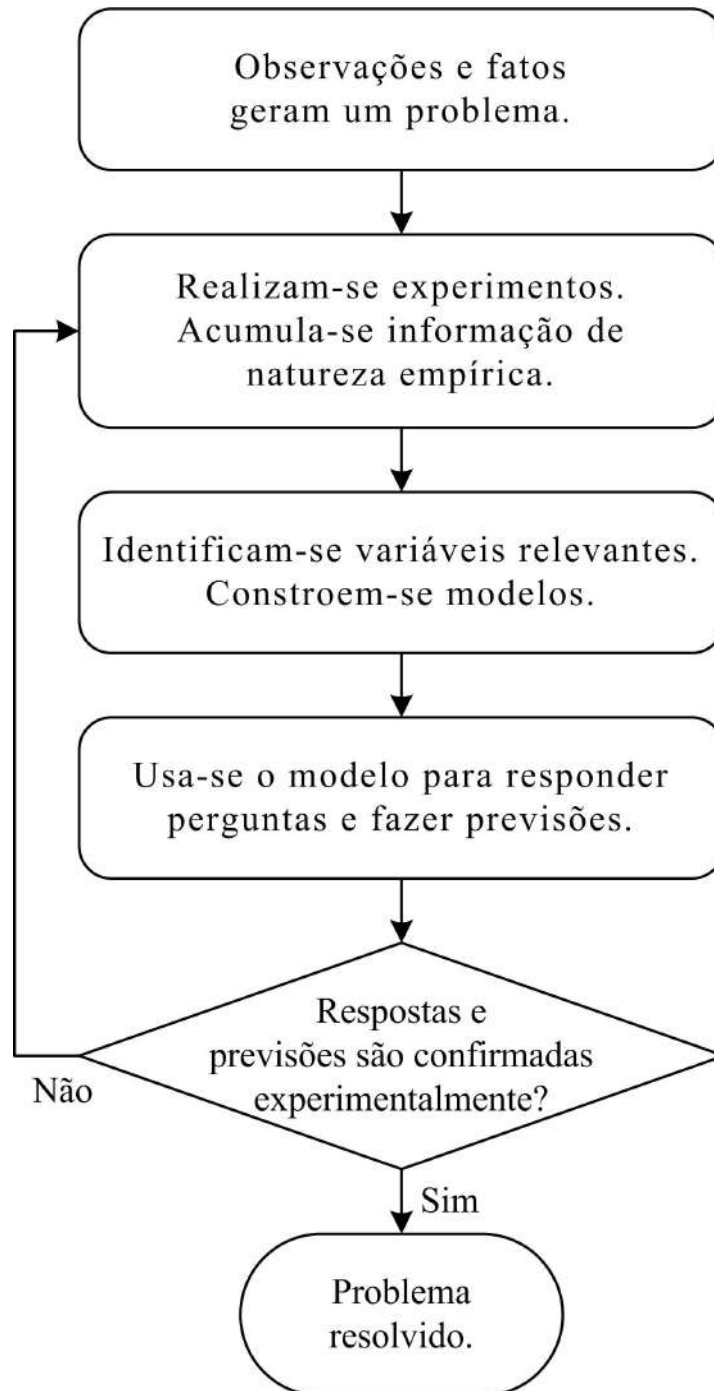


Figura 1.1 - Esquema Geral do Método Científico.

O que a Figura 1.1 não acentua, no entanto, são os seguintes pontos:

### **1.2.1. A natureza cíclica do trabalho científico**

Como o conhecimento acumulado é sempre utilizado para resolver os mais variados problemas, mesmo aqueles que não foram originalmente propostos e utilizados para gerar os modelos, os modelos são continuamente testados. Isso faz com que a abrangência do modelo aumente continuamente (desde que as respostas providas sejam de fato úteis e confirmadas experimentalmente) e que ele seja continuamente revisto e melhorado (o que ocorre sempre que uma resposta obtida seja inadequada e seja negada pela observação empírica).

O exemplo clássico desta “natureza cíclica” é a da Teoria do Movimento de Newton. As Leis de Newton foram utilizadas nos mais diversos campos da Ciência de forma bem sucedida para descrever os mais variados fenômenos. No momento em que os físicos tentaram utilizar as Leis de Newton para descrever o movimento dos sistemas de alta energia, tanto na Astronomia quanto na Física Atômica, as respostas obtidas com o modelo foram negadas pelas observações experimentais. Nesse momento, houve a necessidade de modificar o modelo, para que as novas observações pudessem ser também descritas pela estrutura teórica - e nasceu a Teoria da Relatividade. Note que mais de 100 anos separam as Leis de Newton da Teoria da Relatividade, o que demonstra que a Teoria do Movimento de Newton foi testada durante muito tempo até se demonstrar incompleta. Quanto mais tempo uma estrutura teórica permanece viva e mais ela é testada, mais bem sucedida ela é. Hoje as Leis de Newton podem ser vistas como aproximações excelentes da teoria mais geral, válidas para sistemas de baixas energias.

Estes fatos mostram que o investigador jamais deve acreditar em verdades absolutas e deve estar sempre preparado para contestar o conhecimento estabelecido. Pense que apenas uma fração muito pequena de todas as observações possíveis já foi de fato feita. Tudo ainda está por ser descoberto.

### **1.2.2. A natureza imparcial do trabalho científico**

Se um conhecimento científico é de fato obtido, ele deve poder ser utilizado por todos para resolver problemas semelhantes. Desta forma, observações experimentais devem ser reprodutíveis e os mesmos resultados devem ser obtidos sempre que as mesmas condições forem impostas ao problema. Se condições similares levam a observações distintas, não há como sistematizar o conhecimento, não há como construir modelos e não há como fazer previsões. Não há Ciência, portanto. O conhecimento e a metodologia científicos não são manifestações individualizadas nem profissões de fé (o que de forma nenhuma invalida estas manifestações do espírito humano, como forma de compreender a vida e o universo). Por isso, o bom investigador sempre reproduz suas observações, para garantir que estas são válidas e representam de fato um fenômeno real que pode ser controlado.

### **1.2.3. A natureza limitada do trabalho científico**

Para que as observações sejam feitas, diversas condições devem ser impostas ao sistema experimental investigado, de maneira que as conclusões obtidas só são válidas no contexto limitado em que as observações são feitas. Algumas destas condições são impostas sem mesmo que saibamos disto. Por exemplo, são clássicos os estudos sobre a natureza ondulatória ou particulada das radiações eletromagnéticas, particularmente da luz. A depender de como as condições experimentais são fixadas, conclui-se ou uma coisa ou outra. Hoje, sabe-se que toda partícula em movimento tem a ela associado um movimento ondulatório e vice-versa. O investigador e o ambiente interagem de forma nem sempre bem definida com o experimento que está sendo realizado e podem interferir nos resultados finais obtidos. Como não podemos controlar os efeitos que não conhecemos, é natural que os resultados experimentais obtidos em condições semelhantes não sejam exatamente os mesmos. Por isso, toda a observação experimental está sujeita a flutuações ou a um certo grau de incerteza. Não é possível obter um resultado experimental 100% correto, pois não é possível controlar todo o universo para que realizemos o experimento. O ideal é que as **flutuações** (ou **incertezas** ou **erro experimental**) sejam tão pequenas quanto possível, indicando um controle bastante efetivo sobre as variáveis mais relevantes para a consecução dos dados experimentais obtidos.

#### **1.2.4. A natureza limitada do modelo**

Como toda observação experimental está sujeita a flutuações e deve ter seu escopo limitado ao contexto experimental em que foi executado, não é possível construir modelos perfeitos. Desta forma, nenhum modelo reflete exatamente a realidade e incertezas teóricas devem também ser esperadas. Um modelo bem sucedido é aquele que consegue explicar os resultados experimentais com incertezas compatíveis com aquelas observadas experimentalmente. Não é possível descrever a realidade com precisão maior do que aquela permitida pela observação experimental. Como o modelo é utilizado para fazer previsões e prover respostas a perguntas feitas, toda previsão e resposta obtida através do modelo também apresentam um certo grau de incerteza, que deve ser considerada.

Por tudo o que foi discutido, observa-se que tão ou mais importante que a própria observação experimental é a caracterização apropriada das incertezas a que tais observações estão sujeitas.

### **1.3. As Fontes de Erro e o Ideal Determinístico**

O homem tem procurado através dos tempos as leis que regem o funcionamento do universo. Segundo o ideal positivista, uma vez conhecidas as leis que regem o universo seríamos capazes de entender todo o passado e todo o futuro, já que o desenrolar da vida e da história nada mais seria do que a solução do complexo sistema de equações que representaria estas leis supremas. O destino teria sido ditado quando as condições iniciais foram fixadas e todo o universo foi colocado em movimento.

Diz-se que um sistema ou processo é **determinista** ou determinístico se, fazendo-se sempre a mesma pergunta, obtém-se sempre a mesma resposta. Esse é o resultado típico que se obtém ao se resolver um conjunto de equações matemáticas,

como aquelas que descreveriam o funcionamento do universo. Por exemplo, seja o caso de um tanque de reação continuamente alimentado por uma corrente de processo (Figura 1.2), que flui com vazão (volume/tempo) conhecida e que contém um composto A numa concentração também conhecida (massa/volume). Suponha ainda que é conhecida a vazão da corrente de retirada (volume/tempo), que contém A numa concentração  $C_A$  (massa/volume) desconhecida. Sabe-se que A se transforma em um segundo composto B dentro do tanque, fenômeno esse chamado de reação química. A velocidade com que essa transformação ocorre é conhecida pelos químicos e descrita pela relação

$$R_A = KC_A V \quad (1.1)$$

onde  $R_A$  (massa/tempo) é a velocidade da transformação,  $K$  (1/tempo) é uma constante característica do sistema e  $V$  (volume) é o volume ocupado do tanque. Usando a “lei” desse pequeno universo que diz que “a massa se conserva”, é possível dizer que todo o composto A que entra na alimentação ou sai na corrente de retirada ou vira B. Nesse caso, é possível escrever as seguintes relações matemáticas, que representam essa “lei” do universo:

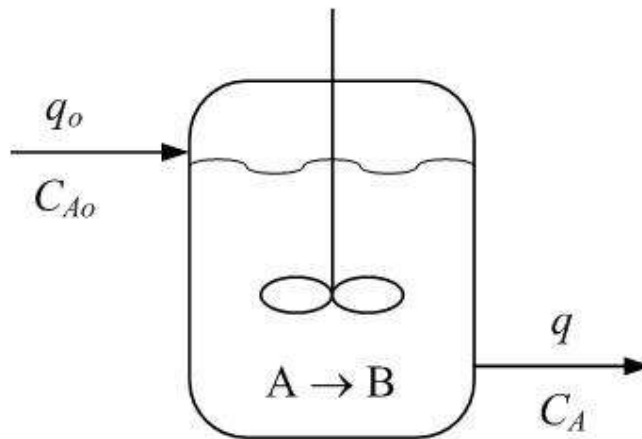


Figura 1.2 - O Tanque de Reação Continuamente Agitado.

$$q_0 C_{A_0} = q C_A + KC_A V \quad (1.2)$$

$$C_A = \frac{q_0 C_{A_0}}{q + KV} \quad (1.3)$$

Dessa forma, repare que sob as mesmas condições de operação ( $q_0$ ,  $q$ ,  $C_{A_0}$ ,  $V$ ), obtém-se sempre o mesmo valor de  $C_A$ . A solução desse problema, na forma proposta, está completamente determinada pelas condições da experimentação.

Sabe-se que isso nem sempre é verdade. Todos já experimentaram a sensação de tentar tirar o número seis no dado, sem sucesso. Vários fatores contribuem para que o resultado de um experimento seja desconhecido, mesmo que a princípio todas as

variáveis pareçam estar bem definidas. É o chamado pesadelo determinista. Vejamos alguns exemplos:

### 1.3.1. O livre arbítrio

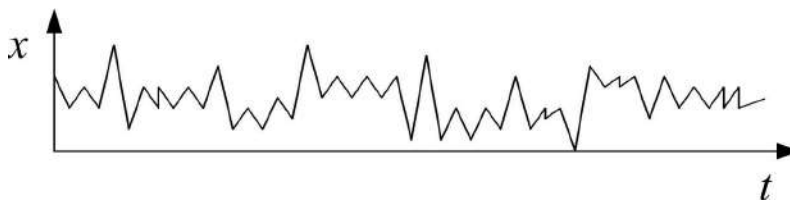
Sob condições idênticas o indivíduo pode optar por soluções diferentes. Embora esta seja uma questão extremamente complexa, com aspectos religiosos, filosóficos e morais que não pretendemos aprofundar aqui, o fato é que a alma humana é bastante complexa e resolve problemas de formas inusitadas e imprevistas. Por isso a dificuldade de se fazer previsões nas áreas de Ciências Humanas e Sociais.

### 1.3.2. A heterogeneidade dos indivíduos

Os indivíduos de um grupo não são idênticos e respondem de forma diferente a diferentes impulsos. Dessa forma, a não ser que todos os elementos do grupo sejam conhecidos com detalhes, previsões sobre comportamentos coletivos são complexos. Isso é verdade tanto nas áreas de Ciências Humanas e Sociais quanto nas áreas de Ciências Exatas. Isso ocorre, por exemplo, sempre que se tentam prever as propriedades da gasolina ou outras frações de petróleo, que são misturas complexas de um número enorme de compostos químicos distintos. Problemas similares ocorrem durante a análise de sistemas biológicos, dado que as células dos organismos que constituem esses sistemas não são necessariamente iguais.

### 1.3.3. A precisão finita dos instrumentos de medidas

Mesmo que fossem conhecidas todas as “leis” do universo, ainda assim teríamos dificuldades de fazer previsões absolutamente corretas, porque os instrumentos de medida têm capacidade finita de aferição. Não conseguimos nunca observar uma grandeza com todas as infinitas casas decimais. As medidas reais se aproximam mais do esquema apresentado na Figura 1.3, onde se observam flutuações (ruídos) por causa da precisão finita do instrumento. Qual o valor real da medida apresentada no registro da Figura 1.3?



**Figura 1.3** - Registro de uma Variável  $x$  com Ruído como uma Função do Tempo.

### 1.3.4. A medição indireta e a necessidade de calibração

Muitas vezes é necessário inferir uma variável a partir da medida de uma outra variável. Por exemplo, quando se mede a temperatura com um termômetro de mercúrio, mede-se de fato o volume do mercúrio em um cilindro graduado. Como o volume do mercúrio muda com o aumento da temperatura (como ocorre com todas as demais substâncias), relaciona-se o volume medido com a temperatura do sistema. Isso gera a

necessidade de construir uma função que relaciona o volume com a temperatura, chamada de **modelo de calibração**. Contudo, como é possível escolher o melhor modelo de calibração? Como é possível garantir que o modelo de calibração permanece válido em todas as condições de experimentação? Esses fatos introduzem incertezas adicionais ao processo de medição e aos valores experimentais medidos.

### **1.3.5. A possível existência de falha no processo de medição**

Instrumentos são constituídos por equipamentos e processos; portanto, estão sujeitos a falhas. Por exemplo, uma régua plástica pode se deformar quando é mal acondicionada em mochilas e pastas escolares, introduzindo erros e imprecisões adicionais no processo de medida. De forma similar, a existência de mau contato em um circuito elétrico pode causar ruído e desvios nas medidas fornecidas por um equipamento. O problema é que esses desvios e deformações nem sempre são percebidos pelo experimentador.

### **1.3.6. O controle limitado sobre um número pequeno de variáveis**

E um fato adicional é que não conhecemos todas as variáveis relevantes para um dado problema com toda a precisão. Em geral, apenas as variáveis mais importantes são levadas em consideração durante a análise de um problema real, de forma que flutuações podem ser esperadas por conta das variáveis não controladas do problema. Por exemplo, será que todos os possíveis contaminantes da corrente de alimentação são conhecidos? Será que o isolamento é perfeito e não há nenhuma perda de calor no sistema?

E qual é a consequência desses fatos? A principal delas é que, mesmo quando conhecemos bastante um sistema, há sempre algum grau de incerteza, algum grau de variabilidade, algum grau de imprecisão. Nunca é possível garantir com certeza absoluta qual é o resultado de um determinado experimento. Diferentes equipamentos de medidas e diferentes experimentadores obtêm valores medidos diferentes para uma mesma variável medida. Obviamente, alguns sistemas apresentam maior ou menor grau de imprecisão que outros. Parece óbvio que uma coisa é a precisão obtida quando se prevê o comportamento meteorológico e outra é a precisão obtida quando se prevê o tempo que um objeto que cai do 3º andar de um bloco de apartamentos leva para atingir o chão. E, portanto, já sentimos aqui a necessidade de caracterizar o grau de variabilidade existente num sistema experimental qualquer.

Diz-se que sistemas que apresentam variabilidades ou incertezas quanto ao resultado final têm natureza estatística ou **estocástica**. O exemplo clássico de comportamento estocástico é o experimento dos dados ou da roleta. Estes são casos limites de aleatoriedade, no entanto, haja visto que é sempre possível estabelecer algum grau de determinismo em problemas preponderantemente estocásticos e vice-versa. Por exemplo, sabemos que, ao lançarmos um dado, nunca obteremos valores maiores do que 6 e menores do que 1. De forma similar, correntes químicas sempre têm algum grau de impureza e os instrumentos de medida não são perfeitos, de forma que o valor de  $C_A$  no tanque de reação da Figura 1.2 só pode ser obtido com um certo grau de precisão. Além disso, desde a década de 70 sabe-se que sistemas determinísticos regidos por equações diferenciais não lineares podem apresentar dependência exponencial aos dados iniciais



(o caos). Nesse caso, qualquer pequena incerteza cometida nas condições iniciais crescerá exponencialmente e tornaria qualquer previsão sobre o comportamento do sistema inócua após um certo tempo. Vê-se, assim, que a fronteira entre os mundos determinístico e estocástico pode ser abrangente, mal definida e espessa.

**Exemplo 1.1** – A Figura 1.4 mostra duas seqüências de dados experimentais. Em ambas as seqüências é possível observar o típico padrão aleatório de flutuação dos dados. Os registros sobem e descem sem um padrão definido.

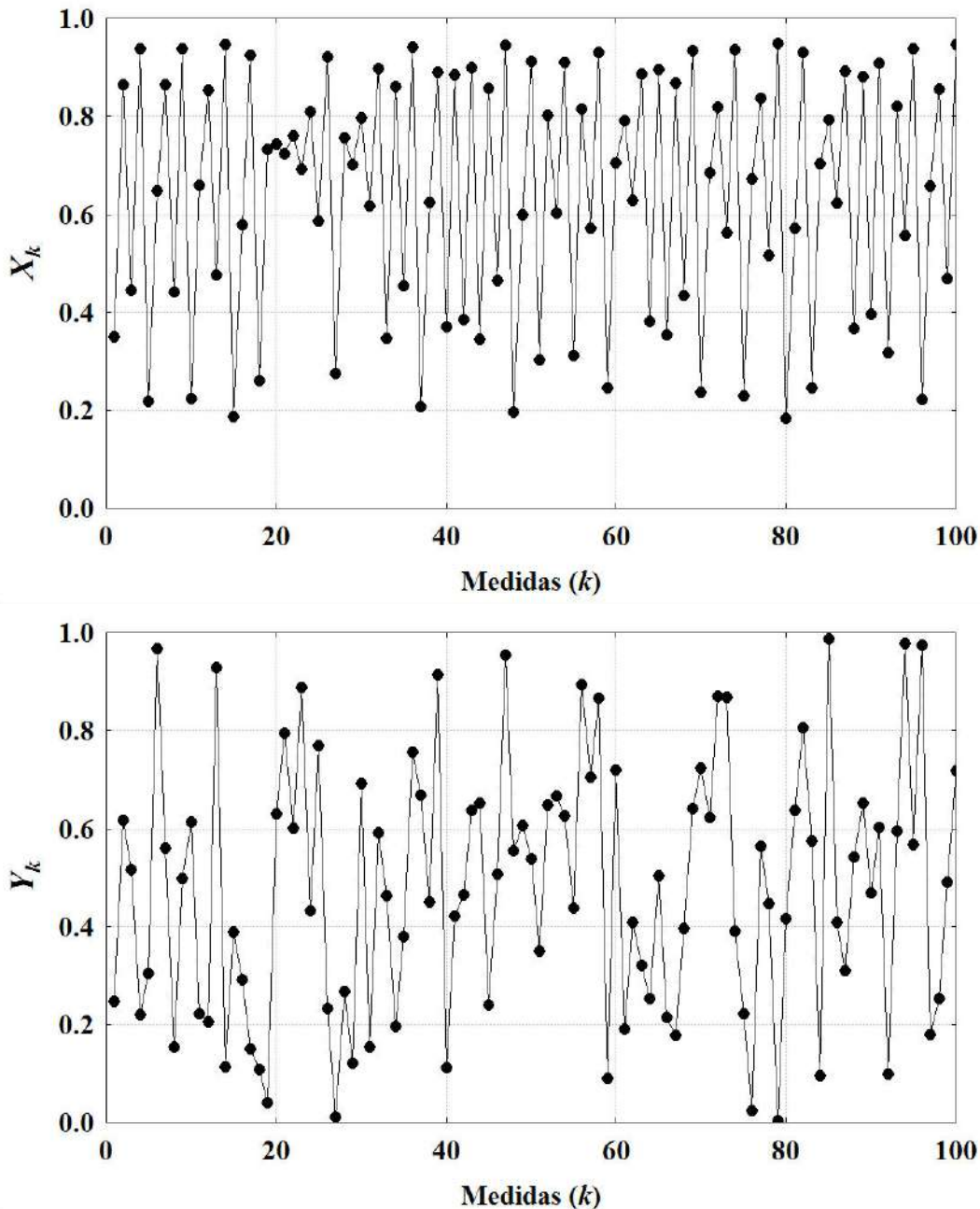


Figura 1.4 - Registro de Duas Seqüências de Medidas  $X$  e  $Y$ .

A variável  $X$  parece flutuar mais que a variável  $Y$  e parece também flutuar de forma um pouco mais regular, embora não seja possível identificar um padrão de comportamento na Figura 1.4. No entanto, uma observação um pouco mais profunda dos dados é apresentada na Figura 1.5.

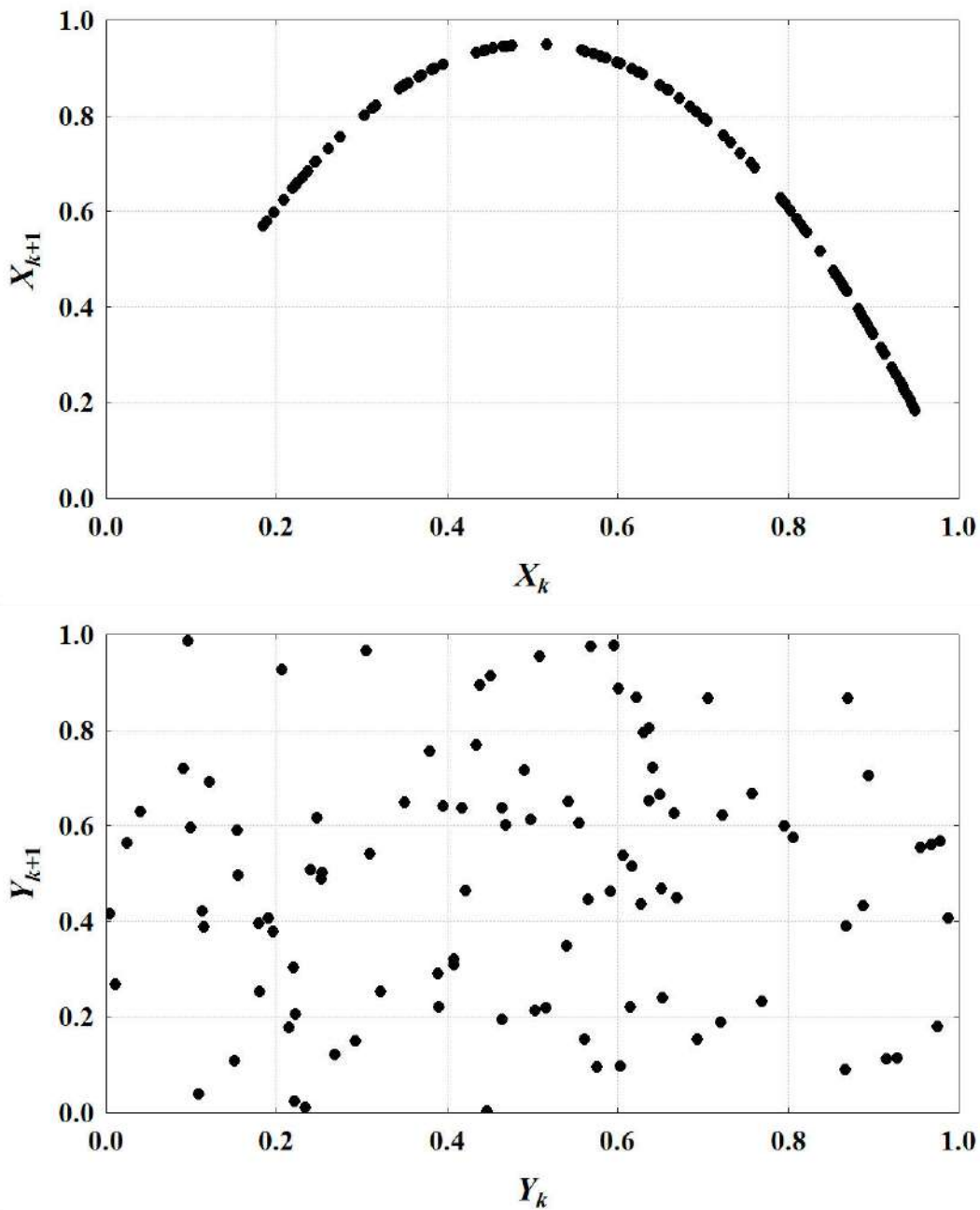


Figura 1.5 - Análise de Duas Sequências de Medidas  $X$  e  $Y$  Deslocadas.

A Figura 1.5 mostra com clareza que a flutuação observada na medida de  $X$  nada tem de aleatória; muito pelo contrário, a medida seguinte ( $X_{k+1}$ ) é uma função determinística da medida anterior ( $X_k$ ). Isso mostra que a identificação do grau de aleatoriedade ou de determinismo de um sinal experimental constitui um problema

relevante para o experimentador per si. O sinal da variável  $Y$  parece ter um grau maior de aleatoriedade que o sinal da variável  $X$ . Contudo, apenas uma investigação mais profunda das propriedades da medida, com o auxílio das ferramentas matemáticas e numéricas apresentadas nos próximos capítulos desse livro, pode permitir que o experimentador defina em bases sólidas se uma medida pode ser considerada aleatória ou não.

### 1.4. Os Conceitos de Probabilidade e de Média

Um conjunto de medidas da variável  $x$  é feito, resultando nos resultados apresentados na Tabela 1.1.

**Tabela 1.1** - Conjunto de Medidas Experimentais Obtidas para a Variável  $x$ .

medida	1	2	3	4	5	6	7	8	9	10
$x$	0.50	0.60	0.50	0.40	0.50	0.50	0.50	0.40	0.50	0.50

Dados os diferentes valores obtidos durante as várias medidas efetuadas, parece lícito perguntar: qual o valor "real" de  $x$ ? A resposta correta para esta pergunta é: NÃO SEI! Supõe-se aqui que todas as medidas foram feitas corretamente e que, portanto, espelham de forma semelhante o valor de  $x$ . Não há como adivinhar a princípio qual é a melhor medida ou qual medida representa melhor o conjunto de medidas. Apesar de tudo isto, ainda assim é necessário definir um valor para  $x$ , pois vários processos de tomada de decisão podem depender disto. Por exemplo, se  $x$  for a medida da quantidade de um contaminante industrial presente num efluente lançado em um rio, a definição do valor de  $x$  pode resultar numa multa emitida pela Secretaria de Meio Ambiente.

Quando as medidas estão sujeitas a flutuações, podemos apenas fornecer um valor que represente o conjunto de medidas de  $x$  de forma conveniente. Por exemplo:

FORMA 1:  $x = 0.50$

0.5 é o valor que aparece mais freqüentemente no conjunto de medidas. Este valor é usualmente chamado de MODA do conjunto de medidas.

FORMA 2:  $x = \frac{0.6 + 7 \cdot 0.5 + 2 \cdot 0.4}{10} = 0.49$

Este é um valor usado comumente para representar um conjunto de números, chamado de MÉDIA ARITMÉTICA. Este valor é uma soma ponderada dos vários números que apareceram no conjunto original de dados. A ponderação utilizada é a freqüência com que o número aparece no conjunto.

FORMA 3:  $x = (0.6 \cdot 0.5^7 \cdot 0.4^2)^{1/10} = 0.48697$

Este é um valor usado também com freqüência para representar um conjunto de números, chamado de MÉDIA GEOMÉTRICA. Este valor é um produto ponderado dos vários números que apareceram no conjunto original de dados. A ponderação utilizada é a freqüência com que o número aparece no conjunto.

Qual destas (ou possivelmente outras) é a melhor forma de representar  $x$ ? Para responder esta pergunta é conveniente introduzir primeiro o conceito de probabilidade. Define-se como probabilidade à **EXPECTATIVA** que se tem de que um certo valor (ou conjunto de valores) possa ocorrer como resultado de um experimento. A probabilidade é expressa como a **FRAÇÃO** das vezes que se espera que o resultado ocorra, quando o experimento é realizado um número muito grande de vezes, tendendo ao **INFINITO**.

Observe que, na definição proposta para probabilidade, alguns pontos merecem ser enfatizados. Primeiramente, a probabilidade é apenas uma **EXPECTATIVA** de que o resultado ocorra e não deve ser confundida com o resultado experimental propriamente dito. Expectativas nem sempre são confirmadas e a vida real está cheia destes exemplos. Azarões surpreendem nos esportes, crises econômicas parecem que às vezes nascem do nada, pessoas dadas como mortas nas UTIs renascem inexplicavelmente, etc. Esta é uma característica que nunca deve ser esquecida: probabilidade é uma coisa e resultado é outra. No fundo, a probabilidade sempre expressa um certo desconhecimento do problema analisado, uma vez que não garante o resultado obtido.

Em segundo lugar, a probabilidade é expressa como a **FRAÇÃO** de vezes que se espera que o resultado analisado seja de fato obtido, se o experimento for realizado várias vezes. Desta forma, a probabilidade é sempre um número positivo, contido no intervalo  $[0,1]$ . Mais ainda: a soma das probabilidades de todas as respostas possíveis é necessariamente igual a 1, pois sempre pelo menos um dos resultados possíveis vai ser obtido experimentalmente. Se a soma das probabilidades não for igual a 1, é porque existem resultados possíveis que não estão sendo analisados.

Finalmente, a probabilidade é definida como uma fração de vezes que se espera que o resultado seja obtido, quando o número de experimentos é **INFINITAMENTE** grande. Portanto, a probabilidade só ganha significado real mais profundo quando infinitos experimentos podem ser realizados, o que nunca é possível na prática. Por maior que seja o número de vezes que se conduz um experimento, esse número é sempre finito. Há, portanto, um enorme esforço de abstração para a definição de probabilidade. Usualmente, experimentos são realizados uma única vez ou um número muito pequeno de vezes, de forma que as decisões tomadas com bases em expectativas, descritas por probabilidades, devem ser tomadas com prudência e conhecimento técnico aprofundado sobre os critérios de tomada de decisão. Por exemplo, ao se dizer que uma usina atômica é 99% segura, diz-se indiretamente que ela é 1% insegura. O problema é que se a expectativa menos provável se confirmar, milhares ou milhões de pessoas podem ser grandemente prejudicadas, a despeito das próximas usinas atômicas instaladas na região para substituírem a usina insegura funcionarem a contento. Na realidade, depois da primeira falha, milhares de pessoas não sobrevivem para confirmar o sucesso das outras 99 tentativas. Isto se ainda forem viáveis novas tentativas.

Com base nestas discussões, é possível introduzir um linguajar matemático mais preciso na forma:

$$p_i = \lim_{\substack{f_j \rightarrow \infty \\ j=1 \dots NR}} \left( \frac{f_i}{\sum_{j=1}^{NR} f_j} \right) = \lim_{N_T \rightarrow \infty} \left( \frac{f_i}{N_T} \right) \quad (1.4)$$

onde  $p_i$  é a probabilidade associada ao evento (resultado)  $i$ ,  $f_i$  é a frequência ou número de vezes que o resultado  $i$  é obtido no conjunto de repetições do experimento,  $NR$  é o número de resultados possíveis para o experimento e  $N_T$  é o número total de observações. Como já discutido:

$$0 \leq p_i \leq 1 \quad (1.5)$$

$$\sum_{i=1}^{NR} p_i = 1 \quad (1.6)$$

**Exemplo 1.2** - Baseado na discussão anterior, qual a probabilidade de tirar 6 num dado? Admitindo-se que as expectativas quanto a qualquer dos possíveis seis resultados são idênticas e que, portanto, os seis resultados possíveis são igualmente prováveis, conclui-se que:

$$p_1 = p_2 = p_3 = p_4 = p_5 = p_6 = p$$

$$\sum_{i=1}^{NR} p_i = 6p = 1 \quad \Rightarrow \quad p = \frac{1}{6}$$

É importante observar que a hipótese de que as seis faces são igualmente prováveis pode não ser verdadeira e que pequenos defeitos de fabricação façam com que certas faces ocorram mais freqüentemente que outras. Por isto, o resultado acima é usualmente utilizado para definir o dado ideal.

Uma vez conhecidos os possíveis resultados de um problema e as expectativas associadas a cada um destes resultados, conhece-se praticamente tudo sobre o destino do experimento. Este acúmulo de conhecimento pode ser representado numa forma gráfica bastante conveniente chamada de Histograma. Um histograma é um gráfico que mostra todos os possíveis resultados experimentais e as respectivas expectativas ou probabilidades de que de fato se realizem. Um histograma ilustra, portanto, uma certa **distribuição de probabilidades**, característica do experimento analisado. Um exemplo é apresentado na Figura 1.6 abaixo.

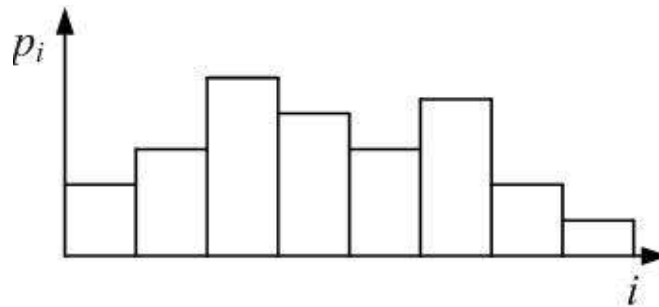


Figura 1.6 - Exemplo de um Histograma.

**Exemplo 1.3** - Para o problema proposto no Exemplo 1.2, apresentam-se abaixo os histogramas de probabilidades para o dado ideal (Figura 1.7) e para um dado real (Figura 1.8). É muito importante que se perceba, no entanto, que a Figura 1.8 pressupõe que o experimento (jogar o dado) tenha sido realizado infinitas vezes. Como isso não é possível, a Figura 1.8 deve ser encarada de fato como uma aproximação de um certo grau do verdadeiro histograma de probabilidades do dado real.

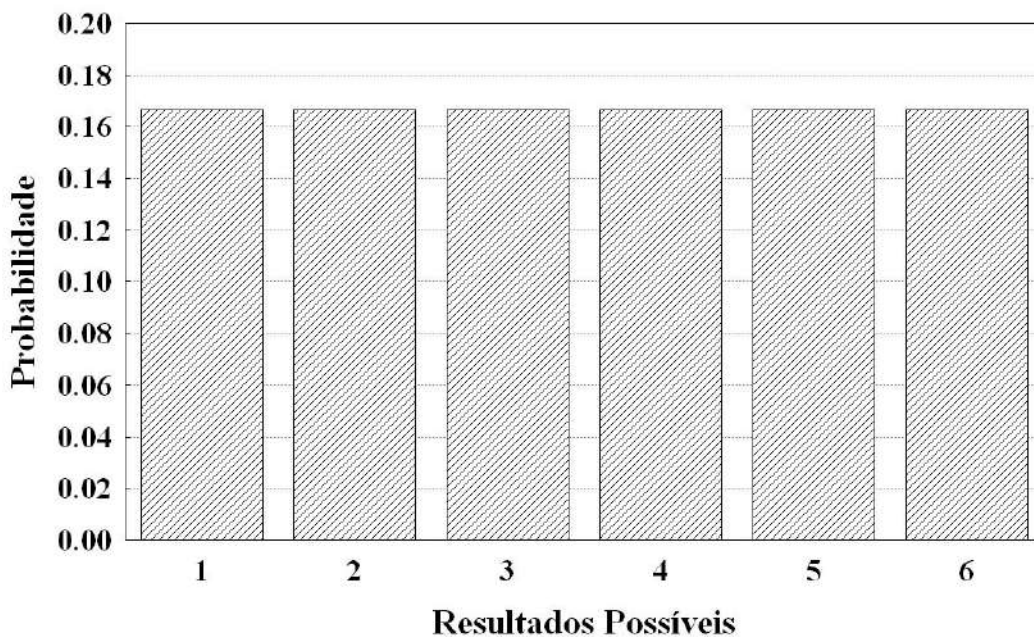
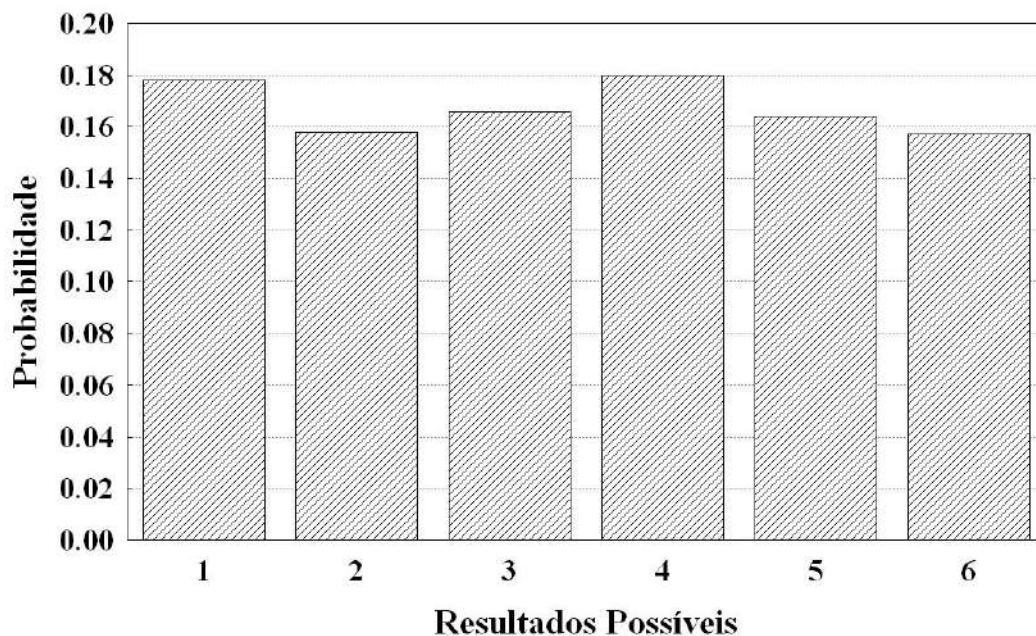


Figura 1.7 - Histograma de Probabilidades para o Dado Ideal.



**Figura 1.8** - Histograma de Probabilidades para um certo Dado Real (obtido a partir de 1000 experimentos).

Voltemos agora à questão de como representar um conjunto de medidas sujeitas a flutuações. Observe que, colocadas na forma de um histograma, a questão que se coloca é como escolher um número que represente a totalidade do histograma de probabilidades. De outra forma, a questão que se coloca é escolher um número que identifique de alguma maneira um valor em torno do qual as probabilidades se distribuem. Podemos dizer que buscamos um número que caracterize o histograma quanto ao movimento de translação, capaz de servir como base para tomadas de decisão e comparações. Por motivações práticas, algumas propriedades devem ser satisfeitas por esse número:

1- Deve ter uma posição central, no sentido de que as probabilidades devem se distribuir em torno deste número (ou seja, o número deve representar de alguma forma os possíveis resultados do experimento);

2- Deve ser unicamente determinado, no sentido de que deve resultar de uma transformação injetora, de forma que cada histograma resulte num único valor de referência (ou seja, a aplicação da operação sobre o histograma deve resultar em um único valor, para que se eliminem ambigüidades de definição).

É fácil mostrar com contra-exemplos que a **moda** (valor que aparece mais frequentemente) e a **mediana** (valor que divide o histograma em dois subconjuntos de iguais probabilidades) não satisfazem a segunda condição descrita acima; ou seja, são medidas ambíguas do histograma. Por exemplo, na Figura 1.7 todos os números são igualmente prováveis, donde não é possível definir a moda. Nesta mesma figura, qualquer número real no intervalo (3,4) divide o histograma em dois subconjuntos de probabilidade igual a 50%, donde se conclui que a mediana também é ambígua. Assim, embora a moda e a mediana possam ser definidas e utilizadas em muitos problemas,

elas não servem de forma inequívoca para fins de caracterização e comparação de histogramas (e distribuições de probabilidade).

Os conceitos de média aritmética e média geométrica podem ser estendidos para o histograma de probabilidades na forma:

$$\mu_X = \sum_{i=1}^{NR} p_i x_i \quad (1.7)$$

$$\mu_X^G = \prod_{i=1}^{NR} x_i^{p_i} \quad (1.8)$$

É fácil mostrar que ambas as definições satisfazem as condições 1 e 2 impostas anteriormente. A comprovação da propriedade 2 é trivial para ambos os casos, pois para cada conjunto de valores  $x_1, \dots, x_{NR}$  e  $p_1, \dots, p_{NR}$  as operações representadas pelas Equações (1.7) e (1.8) resultam em um único número. Pode-se dizer, portanto, que a definição das médias aritmética e geométrica não resulta em qualquer tipo de ambigüidade. Isso não deve ser confundido com a afirmação inversa; ou seja, a média **NÃO** caracteriza inequivocamente a distribuição de probabilidades que a gerou. Portanto, diferentes distribuições de probabilidade podem gerar os mesmos valores de média. Essa afirmação pode ser provada com um contra-exemplo simples, como mostrado na Figura 1.9. Portanto, a média não substitui de forma alguma a informação contida no histograma de probabilidades; apenas fornece um valor em torno do qual os resultados flutuam.

Para provar a validade da primeira condição imposta, suponha que os valores  $x_1, \dots, x_{NR}$  estão organizados em ordem crescente. Então:

$$\sum_{i=1}^{NR} p_i x_1 = x_1 \leq \sum_{i=1}^{NR} p_i x_i = \mu_X \leq \sum_{i=1}^{NR} p_i x_{NR} = x_{NR} \quad (1.9)$$

$$\prod_{i=1}^{NR} x_1^{p_i} = x_1 \leq \prod_{i=1}^{NR} x_i^{p_i} = \mu_X^G \leq \prod_{i=1}^{NR} x_{NR}^{p_i} = x_{NR} \quad (1.10)$$

Logo, as médias aritmética e geométrica são sempre centrais, no sentido de que assumem valores contidos no intervalo formado pelos valores admissíveis máximo e mínimo do experimento. Isso **NÃO** significa dizer, como usualmente admitido, que a média expresse o valor mais provável ou que tenha algum significado físico especial. Por exemplo, no Histograma 1 da Figura 1.9 observa-se que, apesar da média aritmética ser igual a 2, esse valor não é de fato admissível, por ocorrer com probabilidade zero. Os valores mais prováveis nesse caso são os resultados  $x = 1$  e  $x = 3$ , cada um com frequência relativa de 50%. A média deve ser encarada, portanto, como uma entidade numérica que apenas eventualmente pode admitir algum tipo de interpretação física ou de fato refletir um resultado que apresente máxima probabilidade de ocorrer. A Figura 1.10 procura ilustrar os diferentes conceitos de média.



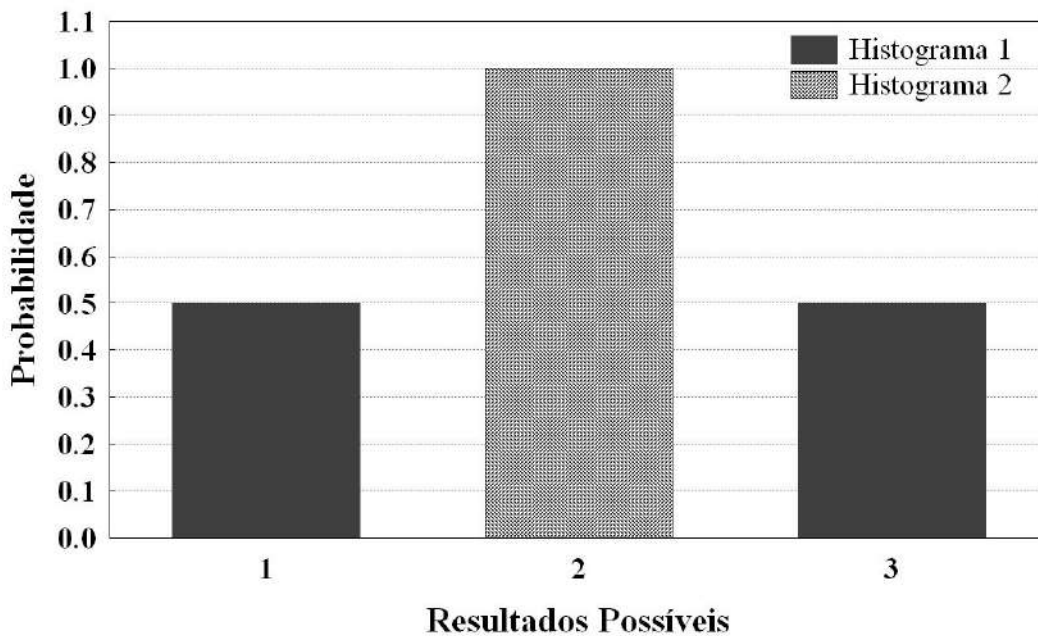


Figura 1.9 - Exemplos de Histogramas de Probabilidade com  $\mu_X = 2$ .

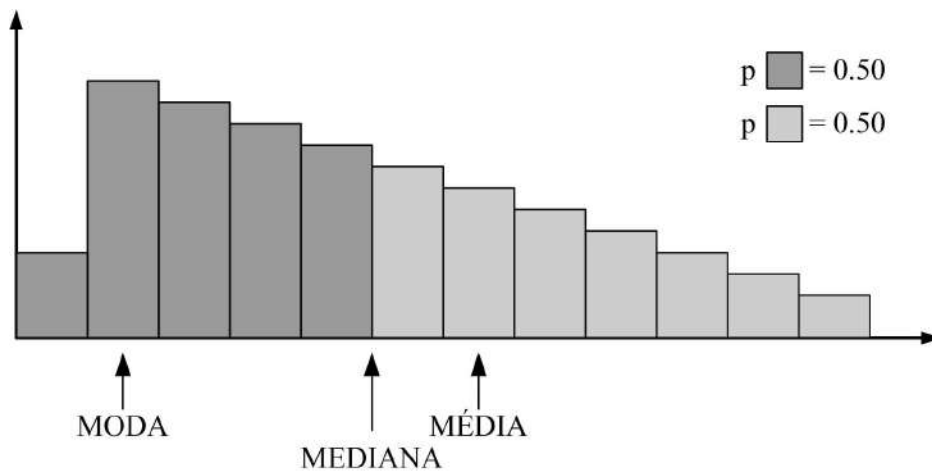


Figura 1.10 - Ilustração dos Diferentes Conceitos de Média.

**Exemplo 1.4** - Para o dado ideal apresentado nos Exemplos 1.2 e 1.3, a média aritmética pode ser calculada como

$$\mu_X = 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 5 \cdot \frac{1}{6} + 6 \cdot \frac{1}{6} = 3.5$$

O valor 3.5 certamente nunca pode ser obtido do lançamento de um dado, ilustrando que a média não é necessariamente o valor mais provável do experimento nem precisa ser um resultado físico real.

## 1.5. O Conceito de Variáveis Independentes e as Propriedades da Média

É importante observar que podem ocorrer problemas com valores negativos no caso da média geométrica, o que pode tornar esse número inconveniente para aplicações em certos problemas. Portanto, há motivações matemáticas adicionais para se escolher uma ou outra operação de média, a depender do problema estudado. Pode-se dizer que a média aritmética é uma definição muito conveniente de média, pois pode ser calculada facilmente a partir do histograma de probabilidades e apresenta uma série de propriedades que facilitam a sua aplicação em problemas de análise matemática. Deve ser aqui salientado que três propriedades de enorme importância para o uso de médias são:

**Propriedade 1.1** - Sejam o conjunto  $(x_i, p_i)$  um histograma de probabilidades e  $\alpha$  um escalar. Então,  $\mu_{\alpha X} = E\{\alpha x\} = \alpha E\{x\} = \alpha \mu_X$ .

$$\mu_{\alpha X} = \sum_{i=1}^{NR} p_i (\alpha x_i) = \alpha \sum_{i=1}^{NR} p_i x_i = \alpha \mu_X \quad (1.11)$$

Portanto, ao multiplicar os resultados possíveis por um escalar  $\alpha$  qualquer, a média aritmética fica multiplicada pelo mesmo escalar  $\alpha$ .

**Propriedade 1.2** - Sejam o conjunto  $(x_i, p_i)$  um histograma de probabilidades e  $\alpha$  um escalar. Então,  $\mu_{\alpha X}^G = \alpha \mu_X^G$ .

$$\mu_{\alpha X}^G = \prod_{i=1}^{NR} (\alpha x_i)^{p_i} = \alpha \prod_{i=1}^{NR} (x_i)^{p_i} = \alpha \mu_X^G \quad (1.12)$$

Portanto, ao multiplicar os resultados possíveis por um escalar  $\alpha$  qualquer, a média geométrica fica multiplicada pelo mesmo escalar  $\alpha$ .

**Propriedade 1.3** - Sejam os dois histogramas de probabilidades  $(x_i, p_{xi})$  e  $(y_i, p_{yi})$ . Então,  $\mu_{X+Y} = E\{x+y\} = E\{x\} + E\{y\} = \mu_X + \mu_Y$ .

Para provarmos a Propriedade 1.3, é bastante conveniente introduzirmos alguns conceitos relativos à probabilidade conjunta de resultados. Diz-se que dois experimentos aleatórios são independentes quando os respectivos histogramas de probabilidade  $(x_i, p_{xi})$  e  $(y_i, p_{yi})$  não dependem dos resultados obtidos. Por exemplo, para o caso do dado ideal, espera-se que a probabilidade de se tirar o número 1 na segunda vez que se rola o dado independa do valor obtido da primeira vez que se rolou o dado. Ou seja, ao se repetir o experimento, o histograma de probabilidades independe do primeiro resultado encontrado. Quando experimentos são independentes, a probabilidade de obter uma certa seqüência de resultados pode ser dada por:

$$P(x_1, \dots, x_N) = \prod_{i=1}^N p_{xi} \quad (1.13)$$

Para entendermos a Expressão (1.13), basta verificarmos que, à medida que se estende o número de experimentos a infinito, uma fração  $p_{x_1}$  destes experimentos terá  $x_1$  como primeiro resultado. Desta fração, uma fração  $p_{x_2}$  terá  $x_2$  como segundo resultado; ou seja, uma fração  $p_{x_1} \cdot p_{x_2}$  destes experimentos terá  $x_1$  e  $x_2$  como primeiros resultados, nesta ordem. Por indução, chega-se à Equação (1.13). Desta forma, se os experimentos são independentes, o histograma que descreve a probabilidade de se obter uma certa  $N$ -tupla ordenada de resultados é:

$$\left( [x_1, \dots, x_N], \prod_{i=1}^N p_{x_i} \right)$$

Podemos agora voltar à Propriedade 1.3. Admitimos, por comodidade da apresentação, que os histogramas  $(x_i, p_{x_i})$  e  $(y_i, p_{y_i})$  estendem-se ao domínio de todos os números inteiros contidos em  $(-\infty, +\infty)$ . Isto em nada restringe o problema, já que podemos associar probabilidades iguais a zero àqueles valores que não fazem parte de fato do histograma particular estudado e já que podemos multiplicar cada número natural por um número real  $\Delta\alpha$  arbitrariamente pequeno, se quisermos trabalhar com intervalos de números reais.

Sejam  $x$  e  $y$  dois experimentos aleatórios obtidos dos histogramas  $(x_i, p_{x_i})$  e  $(y_i, p_{y_i})$ . Neste caso

$$\mu_X = E\{x\} = \sum_{x=-\infty}^{\infty} xp_x(x) \quad (1.14)$$

$$\mu_Y = E\{y\} = \sum_{y=-\infty}^{\infty} yp_y(y) \quad (1.15)$$

O valor médio do histograma da soma de  $x$  e  $y$  deve ser representado como:

$$\mu_{X+Y} = E\{x+y\} = \sum_{x+y=-\infty}^{\infty} (x+y)p_{(x+y)}(x+y) \quad (1.16)$$

onde  $p_{(x+y)}(x+y)$  é a probabilidade de, dados dois experimentos  $x$  e  $y$ , obtermos a soma  $x+y$ . Para facilitar a notação, chamemos  $m = x + y$ .

$$\mu_M = E\{m\} = \sum_{m=-\infty}^{\infty} mp_m(m) \quad (1.17)$$

A questão então é calcular a distribuição de probabilidades de  $m$ . Se  $x$  e  $y$  são eventos independentes, considerando-se que  $m$  pode ser obtido de várias maneiras diferentes (por exemplo,  $m = 4$  pode ser obtido como  $1+3$ ,  $2+2$ ,  $3+1$ ,  $4+0$ , etc...), a Equação (1.13) pode ser usada para calcularmos a probabilidade de cada uma das possíveis combinações, de forma que:

$$p_m(m) = \sum_{x=-\infty}^{\infty} p_x(x) p_y(m-x) \quad (1.18)$$

Logo,

$$\mu_M = E\{m\} = \sum_{m=-\infty}^{\infty} m \sum_{x=-\infty}^{\infty} p_x(x) p_y(m-x) \quad (1.19)$$

Agora, vejamos que o somatório da Equação (1.19) pode ser visualizado na forma:

$$\begin{aligned} & \vdots \\ & + (-3) \{ \cdots + p_x(-3) p_y(0) + p_x(-2) p_y(-1) + p_x(-1) p_y(-2) + p_x(0) p_y(-3) + \cdots \} \\ & + (-2) \{ \cdots + p_x(-3) p_y(1) + p_x(-2) p_y(0) + p_x(-1) p_y(-1) + p_x(0) p_y(-2) + \cdots \} \\ & + (-1) \{ \cdots + p_x(-3) p_y(2) + p_x(-2) p_y(1) + p_x(-1) p_y(0) + p_x(0) p_y(-1) + \cdots \} \\ & + (0) \{ \cdots + p_x(-3) p_y(3) + p_x(-2) p_y(2) + p_x(-1) p_y(1) + p_x(0) p_y(0) + \cdots \} \\ & + (1) \{ \cdots + p_x(-3) p_y(4) + p_x(-2) p_y(3) + p_x(-1) p_y(2) + p_x(0) p_y(1) + \cdots \} \\ & \vdots \end{aligned}$$

Lendo o somatório de cima para baixo

$$\begin{aligned} & \dots + (y-3) p_x(-3) \sum_{y=-\infty}^{\infty} p_y(y) + (y-2) p_x(-2) \sum_{y=-\infty}^{\infty} p_y(y) + \\ & + (y-1) p_x(-1) \sum_{y=-\infty}^{\infty} p_y(y) + (y-0) p_x(-0) \sum_{y=-\infty}^{\infty} p_y(y) + \dots \end{aligned} \quad (1.20)$$

ou seja

$$\sum_{m=-\infty}^{\infty} m \sum_{x=-\infty}^{\infty} p_x(x) p_y(m-x) = \sum_{x=-\infty}^{\infty} \sum_{y=-\infty}^{\infty} (y+x) p_x(x) p_y(y) \quad (1.21)$$

Portanto:

$$\mu_M = \sum_{x=-\infty}^{\infty} \sum_{y=-\infty}^{\infty} y p_x(x) p_y(y) + \sum_{x=-\infty}^{\infty} \sum_{y=-\infty}^{\infty} x p_x(x) p_y(y) = \mu_Y + \mu_X \quad (1.22)$$

Mas e se as distribuições de probabilidade das variáveis  $x$  e  $y$  não fossem independentes? Nesse caso, admitindo que  $x$  é o evento determinante, a distribuição de probabilidades de  $y$  dependeria do valor particular de  $x$  encontrado. Parece complicado, mas estamos acostumados a lidar com esse conceito no dia-a-dia. Por exemplo, qual é a probabilidade de encontrarmos um amigo na praia? Se o dia estiver nublado ou chuvoso, a probabilidade deve ser muito baixa, pois poucas pessoas costumam ir à praia nessas condições. Se o dia estiver ensolarado, as praias enchem e aumentam as chances de encontrarmos pessoas conhecidas tomando seu banho de mar. Nesse caso, o evento

principal ou condicionante é o estado do tempo ( $x$ ), enquanto encontrarmos uma pessoa conhecida na praia ( $y$ ) é o evento secundário ou condicionado. Como a distribuição de probabilidades de  $y$  muda com  $x$ , diz-se que a probabilidade de  $y$  é condicionada por  $x$ , representada usualmente por  $p_y(y/x)$  (lida quase sempre como "probabilidade de  $y$  dado  $x$ "), e que  $y$  e  $x$  são variáveis dependentes.

No caso em que a probabilidade de um evento é condicionada por um outro evento, a Equação (1.13) tem que ser modificada para

$$P(x_1, \dots, x_N) = \prod_{i=1}^N p(x_i) p(x_2/x_1) \dots p(x_N/x_1, x_2, \dots, x_{N-1}) \quad (1.23)$$

sendo que

$$\sum_{x_N=-\infty}^{\infty} p(x_N/x_1, x_2, \dots, x_{N-1}) = 1, \quad \forall x_1, x_2, \dots, x_{N-1} \quad (1.24)$$

isto para que seja satisfeita a Equação (1.6), um dos requisitos básicos da probabilidade.

Dessa forma, se o evento  $y$  é condicionado pelo evento  $x$ , as Equações (1.18) e (1.19) ganham a forma:

$$p_m(m) = \sum_{x=-\infty}^{\infty} p_x(x) p_y(m-x/x) \quad (1.25)$$

$$\mu_M = E\{m\} = \sum_{m=-\infty}^{\infty} m \sum_{x=-\infty}^{\infty} p_x(x) p_y(m-x/x) \quad (1.26)$$

de maneira que as Equações (1.21) e (1.22) ficam

$$\mu_M = E\{m\} = \sum_{m=-\infty}^{\infty} m \sum_{x=-\infty}^{\infty} p_x(x) p_y(m-x/x) = \sum_{x=-\infty}^{\infty} \sum_{y=-\infty}^{\infty} (y+x) p_x(x) p_y(y/x) \quad (1.27)$$

$$\begin{aligned} \mu_M &= \sum_{x=-\infty}^{\infty} \sum_{y=-\infty}^{\infty} y p_x(x) p_y(y/x) + \sum_{x=-\infty}^{\infty} \sum_{y=-\infty}^{\infty} x p_x(x) p_y(y/x) = \\ \mu_M &= \sum_{x=-\infty}^{\infty} p_x(x) \sum_{y=-\infty}^{\infty} y p_y(y/x) + \sum_{x=-\infty}^{\infty} x p_x(x) \sum_{y=-\infty}^{\infty} p_y(y/x) = \\ \mu_M &= \sum_{x=-\infty}^{\infty} p_x(x) \mu_y(x) + \sum_{x=-\infty}^{\infty} x p_x(x) = \mu_y + \mu_x \end{aligned} \quad (1.28)$$

Portanto, as Propriedades 1.1 e 1.3 são sempre satisfeitas, independentemente das variáveis serem dependentes ou independentes. Conclui-se que a **Operação Média Aritmética é LINEAR**. Isto torna a operação média aritmética, definida pela Equação (1.7), extremamente conveniente do ponto de vista matemático, sendo por isso usualmente escolhida como melhor maneira de representar o ponto em torno do qual se

distribuem as probabilidades num histograma de probabilidades. A linearidade da operação média aritmética garante que a média da soma é a soma das médias e que ao multiplicar a variável por um escalar, a média fica multiplicada pelo mesmo escalar. Mas o que ocorre se outros operadores forem aplicados sobre as variáveis  $x$  e  $y$ ?

$$E\{f(x)\} = \mu_f = \sum_{i=1}^{NR} f(x_i) p(x_i) = \sum_{i=1}^{NR} f_i p_i \neq f(\mu_x) \quad (1.29)$$

Para mostrar a Equação (1.29), podemos usar o Histograma 1 da Figura 1.9. Por exemplo, admitamos que a operação  $f(x) = x^2$  é aplicada sobre o histograma. Neste caso, o valor médio obtido é

$$E\{x^2\} = 1^2 \cdot \frac{1}{2} + 3^2 \cdot \frac{1}{2} = 5 \neq \mu_f = \mu_{x^2} = 4$$

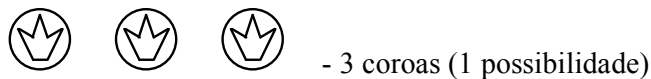
Portanto, a linearidade da média não permite afirmar que o valor médio de uma função aplicada sobre o histograma é o valor da função calculada no ponto médio do histograma. Isso só é verdadeiro se a função for linear. Por exemplo,

$$E\{f(x) = \alpha x + \beta\} = \alpha E\{x\} + \beta = \alpha \mu_x + \beta = f(\mu_x)$$

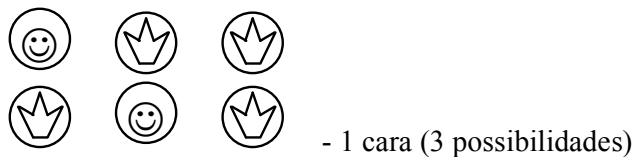
**Exemplo 1.5** - Suponhamos que um cidadão jogue uma moeda para o alto três vezes e que receba 1 real por cada cara que tirar. Se o experimento for repetido  $N$  vezes, quanto o cidadão ganhará na média?

Primeiramente é interessante perceber que o experimento "jogar a moeda" resulta em resultados independentes, de forma que a Equação (1.13) pode ser aplicada. Portanto, pode-se imaginar que cada configuração particular de três resultados tem probabilidade  $p_i = \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{8}$ , já que a probabilidade de cada resultado (cara ou coroa) é sempre igual a  $1/2$ . Vejamos:

- Nenhuma cara

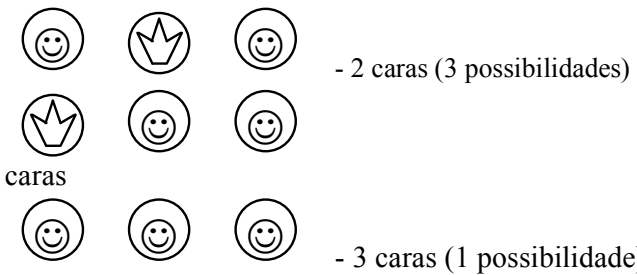


- Apenas uma cara



- Apenas duas caras





- Três caras

Portanto, o histograma de probabilidades tem a forma:

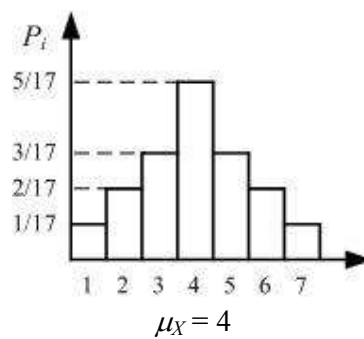
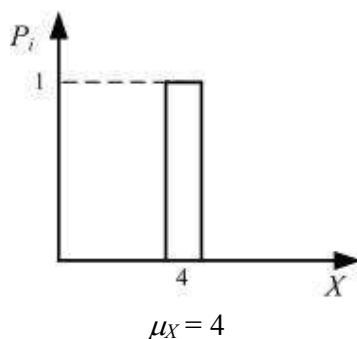
- 0 cara - 1/8 das vezes
- 1 cara - 3/8 das vezes
- 2 caras - 3/8 das vezes
- 3 caras - 1/8 das vezes

cuja média é  $\mu_X = 0 \cdot \frac{1}{8} + 1 \cdot \frac{3}{8} + 2 \cdot \frac{3}{8} + 3 \cdot \frac{1}{8} = \frac{12}{8} = 1.5$

O dinheiro total arrecadado é uma transformação linear do histograma e pode ser dado na forma  $f(x) = Nx$ . Portanto,  $\mu_f = N\mu_X = 1.5N$ .

### 1.6. Os Conceitos de Espalhamento, Variância e Covariância

Considere os histogramas da Figura 1.9 e da Figura 1.11 mostrada abaixo. Em ambos os casos, as médias dos histogramas apresentados são idênticas. No entanto é óbvio que as distribuições são muito diferentes. No segundo histograma da Figura 1.9 e no primeiro histograma da Figura 1.11, apenas um valor é possível. Logo, não há qualquer dúvida sobre a observação que será feita após o experimento. É como colocar uma única pedra de bingo no interior de um saco e perguntar que número será sorteado. No segundo caso, há um espalhamento de possíveis valores em torno do valor médio e não é possível mais garantir o resultado do experimento. No primeiro histograma da Figura 1.9, dois resultados são possíveis, enquanto 7 diferentes resultados são possíveis no histograma 2 da Figura 1.11. Portanto, pode ser dito de forma pouco precisa que o resultado do experimento descrito pelo segundo histograma da Figura 1.11 é o mais incerto dentre todos os histogramas analisados.



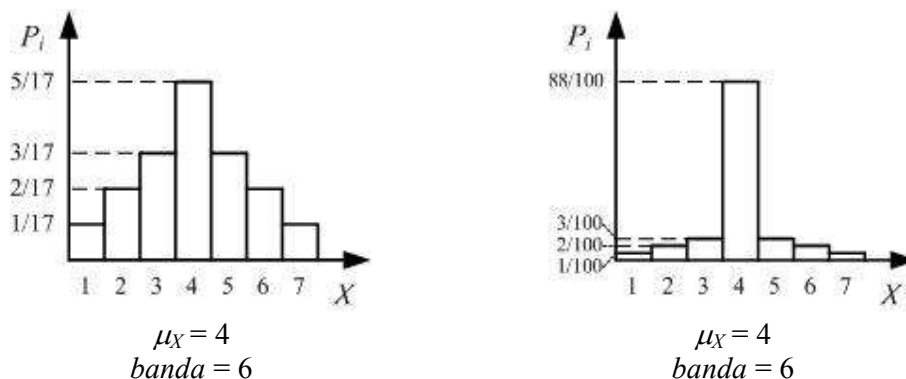
**Figura 1.11** - Exemplos de histogramas bem diferentes, mas com médias iguais.

Quando um único resultado é possível em um histograma, diz-se que a distribuição de probabilidades é singular ou monodispersa e que a população de resultados possíveis é uniforme ou homogênea. Caso contrário, o histograma é dito polidisperso, enquanto a população de resultados é dita heterogênea. Portanto, a definição de uma grandeza que possa caracterizar de forma precisa a heterogeneidade da população a partir da distribuição de probabilidades parece ser bastante útil, já que as Figuras 1.9 e 1.11 ilustram que diferentes histogramas podem apresentar diferentes graus de espalhamento, apesar de terem a mesma média aritmética.

Várias maneiras distintas podem ser usadas para caracterizar o espalhamento. Uma das formas mais simples e intuitivas de caracterização do espalhamento é usar o conceito de **banda**. A banda pode ser definida como a diferença entre o maior e o menor resultados admissíveis da distribuição. Assim,

$$banda = x_{\max} - x_{\min} \tag{1.30}$$

Embora muito usada em problemas práticos, essa definição de espalhamento não é muito adequada para estudo da maior parte dos problemas. Primeiramente, ela não resulta em propriedades matemáticas convenientes, como a linearidade da operação de média. Em segundo lugar, essa definição não permite caracterizar de forma adequada o espalhamento de histogramas que admitem infinitos resultados, como analisado nas próximas seções. E até quando a banda pode ser definida de forma precisa, como na Figura 1.12 abaixo, ela não reflete o fato de que o grau de homogeneidade dos resultados pode ser muito diferente mesmo quando o número de resultados possíveis é idêntico. Por exemplo, é muito mais provável obter como resultado do experimento um valor próximo do valor médio no segundo histograma da Figura 1.12 do que no primeiro histograma dessa figura. Por isso, parece razoável dizer que a população de resultados do segundo histograma é mais homogênea, a despeito da banda resultar no mesmo valor em ambos os casos.



**Figura 1.12** - Exemplos de histogramas com diferentes graus de homogeneidade, mas com bandas iguais.



Uma outra forma muito comum de definir o espalhamento é usar o conceito de **percentil**. Diz-se que os percentis, representados aqui como  $x_{i\%}$ , são os valores que separam regiões de resultados admissíveis com probabilidades iguais a um certo valor estabelecido, como por exemplo 1%. Assim,

$$P(x_i \leq x_{i\%}) = i\% \quad (1.31)$$

e

$$P(x_i \leq x_{(i+1)\%}) - P(x_i \leq x_{i\%}) = 1\% \quad (1.32)$$

Baseado nos percentis, é possível redefinir a banda de forma mais adequada, como por exemplo:

$$banda_{p\%} = x_{(100-p)\%} - x_{p\%} \quad (1.33)$$

Dessa forma, a banda definida pela Equação (1.30) seria equivalente à  $banda_{0\%}$  definida pela Equação (1.33). Para definição do espalhamento, é muito utilizado o conceito de quartil, que nada mais é do que o conjunto constituído por  $x_{0\%} = x_{\min}$ ,  $x_{25\%}$ ,  $x_{50\%}$ ,  $x_{75\%}$ ,  $x_{100\%} = x_{\max}$ , que divide o histograma em quatro regiões de probabilidades iguais a 25%. Nesse caso:

$$banda_{25\%} = x_{75\%} - x_{25\%} \quad (1.34)$$

As Equações (1.33-34) permitem eliminar dois defeitos embutidos na definição original de banda: tornam possível a caracterização de espalhamento em problemas com infinitos resultados admissíveis e são sensíveis a mudanças de espalhamento como os ilustrados na Figura 1.12. No entanto, a manipulação matemática de expressões envolvendo percentis não é simples. Além disso, da mesma forma que no caso da definição da moda e da mediana, a definição dos percentis pode não ser precisa. Por exemplo, no segundo histograma da Figura 1.12 é fácil definir os percentis  $x_{1\%}=1$ ,  $x_{3\%}=2$ ,  $x_{6\%}=3$ ,  $x_{94\%}=4$ ,  $x_{97\%}=5$ ,  $x_{99\%}=6$  e  $x_{100\%}=7$ . Contudo, e os demais 94 percentis? Como defini-los de forma inequívoca a partir do histograma? Dessa maneira, a definição da  $banda_{25\%}$  baseada nos quartis não seria possível nesse caso.

Uma forma precisa e conveniente de se caracterizar o espalhamento é utilizar o conceito de média desenvolvido anteriormente. Por exemplo, o **espalhamento médio** poderia ser definido como a média das diferenças observadas entre os vários resultados possíveis e o valor médio desses resultados, na forma

$$E\{|x_i - \mu_X|\} = \sum_{i=1}^{NR} p_i |x_i - \mu_X| \quad (1.35)$$

Para os histogramas 1 e 2 da Figura 1.12, os resultados seriam respectivamente iguais a:

$$E\{|x_i - \mu_x|\} = \frac{1}{17}|1-4| + \frac{2}{17}|2-4| + \frac{3}{17}|3-4| + \frac{5}{17}|4-4| + \\ + \frac{3}{17}|5-4| + \frac{2}{17}|6-4| + \frac{1}{17}|7-4| = \frac{20}{17}$$

$$E\{|x_i - \mu_x|\} = \frac{1}{100}|1-4| + \frac{2}{100}|2-4| + \frac{3}{100}|3-4| + \frac{88}{100}|4-4| + \\ + \frac{3}{100}|5-4| + \frac{2}{100}|6-4| + \frac{1}{100}|7-4| = \frac{20}{100}$$

Os resultados obtidos refletem exatamente o sentimento de que o grau de espalhamento no segundo caso é menor que no primeiro. Além disso, a obtenção das medidas de espalhamento pode ser feita diretamente a partir do histograma de probabilidades, sem que haja qualquer ambigüidade. No entanto, a Equação (1.35) tem o inconveniente de usar o módulo da diferença como medida de distância. Como o módulo é uma função descontínua, isso causa certos inconvenientes de manipulação matemática e induz a definição do conceito de **variância**.

Define-se como **variância de x** (representada por  $\text{Var}\{x\}$ ,  $E\{(x-\mu_x)^2\}$ ,  $\sigma_{xx}^2$ ,  $\sigma_x^2$  ou simplesmente  $\sigma^2$ ) a média do quadrado das diferenças observadas entre os vários resultados possíveis e o valor médio desses resultados, na forma

$$\sigma_{xx}^2 = \text{Var}\{x\} = E\{(x_i - \mu_x)^2\} = \sum_{i=1}^{NR} p_i (x_i - \mu_x)^2 \quad (1.36)$$

Para os histogramas 1 e 2 da Figura 1.12, os resultados seriam respectivamente iguais a:

$$\sigma_{xx}^2 = \frac{1}{17}(1-4)^2 + \frac{2}{17}(2-4)^2 + \frac{3}{17}(3-4)^2 + \frac{5}{17}(4-4)^2 + \\ + \frac{3}{17}(5-4)^2 + \frac{2}{17}(6-4)^2 + \frac{1}{17}(7-4)^2 = \frac{40}{17}$$

$$\sigma_{xx}^2 = \frac{1}{100}(1-4)^2 + \frac{2}{100}(2-4)^2 + \frac{3}{100}(3-4)^2 + \frac{88}{100}(4-4)^2 + \\ + \frac{3}{100}(5-4)^2 + \frac{2}{100}(6-4)^2 + \frac{1}{100}(7-4)^2 = \frac{40}{100}$$

Comparada às diferentes medidas de espalhamento apresentadas anteriormente, a definição de variância apresenta muitas vantagens. Primeiramente, a variância pode ser obtida diretamente do histograma de probabilidades sem qualquer ambigüidade. Em segundo lugar, a utilização das operações de média e do quadrado da distância em relação à média permite a manipulação relativamente simples de expressões matemáticas, como será mostrado a seguir. No entanto, da mesma forma que no caso da definição da média, o usuário deve resistir à tentação de explicar em bases físicas e concretas o significado da variância. A variância deve ser encarada apenas como uma

medida matemática conveniente de espalhamento, e que por isso pode ser utilizada para caracterizar e comparar histogramas também de forma matemática conveniente. Algumas propriedades relevantes da operação de cálculo da variância são apresentadas a seguir.

**Propriedade 1.4** - A variância é um número positivo, sendo igual a zero se e somente se a distribuição de probabilidades é monodispersa.

A comprovação dessa propriedade a partir da Equação (1.36) é trivial. Como cada uma das parcelas da soma representada pela Equação (1.36) é positiva ou nula, então a variância tem que ser necessariamente um número positivo. Se a distribuição é monodispersa, como no primeiro histograma da Figura 1.11, apenas um termo tem probabilidade diferente de zero. Nesse caso, como para esse termo o resultado admissível coincide com a média, a variância fica identicamente nula. Por outro lado, se a variância é nula, todos os termos da soma têm que ser iguais a zero. Nesse caso, ou as probabilidades são iguais a zero ou os resultados para os quais as probabilidades não são iguais a zero são iguais ao valor médio. Portanto, a distribuição tem que ser necessariamente monodispersa.

**Propriedade 1.5** - Sejam o conjunto  $(x_i, p_i)$  um histograma de probabilidades e  $\alpha$  um escalar. Então,  $\text{Var}\{\alpha x\} = \alpha^2 \text{Var}\{x\}$ .

$$\begin{aligned} \text{Var}\{\alpha x\} &= \sum_{i=1}^{NR} p_i (\alpha x_i - \mu_{\alpha x})^2 = \sum_{i=1}^{NR} p_i (\alpha x_i - \alpha \mu_x)^2 \\ \text{Var}\{\alpha x\} &= \alpha^2 \sum_{i=1}^{NR} p_i (x_i - \mu_x)^2 = \alpha^2 \text{Var}\{x\} \end{aligned} \quad (1.37)$$

Portanto, ao multiplicar os resultados possíveis por um escalar  $\alpha$  qualquer, a variância fica multiplicada pelo quadrado do escalar  $\alpha$ .

**Propriedade 1.6** - Sejam o conjunto  $(x_i, p_i)$  e  $(y_i, p_{yi})$  dois histogramas de probabilidades de eventos independentes. Então,  $\text{Var}\{x+y\} = \text{Var}\{x\} + \text{Var}\{y\}$ . Para que seja possível demonstrar essa propriedade, é preciso lembrar que

$$\text{Var}\{x+y\} = E\left\{\left[(x+y) - \mu_{x+y}\right]^2\right\} \quad (1.38)$$

Inserindo a Equação (1.22) na Equação (1.38), chega-se a

$$\text{Var}\{x+y\} = E\left\{\left[(x+y) - (\mu_x + \mu_y)\right]^2\right\} \quad (1.39)$$

O termo quadrático da Equação (1.39) pode ser aberto, resultando em

$$\begin{aligned}
 \text{Var}\{x+y\} &= E\left\{\left[(x-\mu_x)+(y-\mu_y)\right]^2\right\} = \\
 &E\left\{(x-\mu_x)^2 + 2(x-\mu_x)(y-\mu_y) + (y-\mu_y)^2\right\} = \\
 &E\left\{(x-\mu_x)^2\right\} + 2E\left\{(x-\mu_x)(y-\mu_y)\right\} + E\left\{(y-\mu_y)^2\right\} = \\
 &\text{Var}\{x\} + 2\text{Covar}\{x,y\} + \text{Var}\{y\}
 \end{aligned} \tag{1.40}$$

Na Equação (1.40), define-se como **covariância** entre as variáveis  $x$  e  $y$ , representada por  $\text{Covar}\{x,y\}$  ou simplesmente  $\sigma_{XY}^2$ , à seguinte operação de média:

$$\sigma_{XY}^2 = \text{Covar}\{x,y\} = E\left\{(x-\mu_x)(y-\mu_y)\right\} \tag{1.41}$$

Para que a operação de covariância seja compreendida, é conveniente escrever a Equação (1.41) na forma

$$\sigma_{XY}^2 = \sum_{x=-\infty}^{\infty} p_x(x) \sum_{y=-\infty}^{\infty} p_y(y/x)(x-\mu_x)(y-\mu_y) \tag{1.42}$$

onde a soma dupla identifica todas as possíveis combinações de resultados que podem ser obtidas a partir dos dois histogramas de probabilidades. Se os eventos  $x$  e  $y$  são independentes, então

$$\begin{aligned}
 \sigma_{XY}^2 &= \sum_{x=-\infty}^{\infty} p_x(x) \sum_{y=-\infty}^{\infty} p_y(y)(x-\mu_x)(y-\mu_y) = \\
 &\sum_{x=-\infty}^{\infty} p_x(x)(x-\mu_x) \sum_{y=-\infty}^{\infty} p_y(y)(y-\mu_y) = \\
 &\sum_{x=-\infty}^{\infty} p_x(x)(x-\mu_x)(\mu_y - \mu_y) = 0
 \end{aligned} \tag{1.43}$$

Portanto, quando os eventos são independentes, a covariância entre os resultados obtidos a partir dos dois experimentos é igual a zero. Por isso, a covariância é usada com frequência como uma medida de independência entre resultados obtidos a partir de diferentes experimentos. (Essa técnica de inferência do grau de dependência entre variáveis, no entanto, deve ser usada com cautela. Como será mostrado posteriormente, a afirmação inversa não é necessariamente verdadeira; ou seja, resultados de experimentos distintos podem ser fortemente dependentes uns dos outros, resultando contudo em covariância igual ou próxima de zero.) Assim, se os resultados dos experimentos  $x$  e  $y$  são independentes, e portanto resultam em covariância nula, a Equação (1.40) fica finalmente na forma:

$$\text{Var}\{x+y\} = \text{Var}\{x\} + \text{Var}\{y\} \tag{1.44}$$

Se os resultados obtidos para  $x$  e  $y$  não são independentes, então a Equação (1.42) tem que ser escrita na forma

$$\begin{aligned} \sigma_{XY}^2 &= \sum_{x=-\infty}^{\infty} p_x(x)(x - \mu_X) \sum_{y=-\infty}^{\infty} p_y(y/x)(y - \mu_Y) = \\ &= \sum_{x=-\infty}^{\infty} p_x(x)(x - \mu_X)(\mu_{Y/X} - \mu_Y) = \sum_{x=-\infty}^{\infty} p_x(x)\mu_{Y/X}(x - \mu_X) = \quad (1.45) \\ &= \sum_{x=-\infty}^{\infty} p_x(x)x\mu_{Y/X} - \mu_X\mu_Y = \sum_{x=-\infty}^{\infty} p_x(x)x\mu_Y(x) - \mu_X\mu_Y \end{aligned}$$

que é uma operação de média conjunta dos valores de  $x$  e de como a média de  $y$  depende de  $x$ . A Equação (1.45) mostra também de uma outra forma que a covariância entre eventos independentes é igual a zero. Para tanto, basta fazer  $\mu_{Y/X} = \mu_Y$ .

**Exemplo 1.6** - Para o dado ideal dos Exemplos 1.2 e 1.3, a variância pode ser calculada como

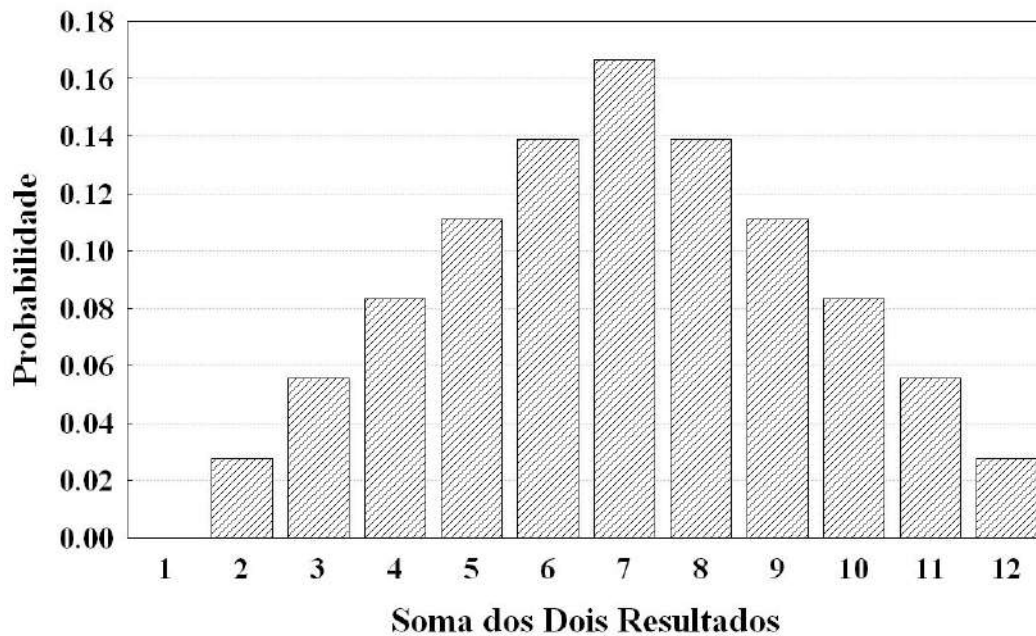
$$\begin{aligned} \sigma^2 &= \frac{1}{6}(1-3.5)^2 + \frac{1}{6}(2-3.5)^2 + \frac{1}{6}(3-3.5)^2 + \\ &= \frac{1}{6}(4-3.5)^2 + \frac{1}{6}(5-3.5)^2 + \frac{1}{6}(6-3.5)^2 = \frac{17.5}{6} \end{aligned}$$

**Exemplo 1.7** - Para o dado ideal dos Exemplos 1.2 e 1.3, suponha que dois dados são lançados simultaneamente em um jogo e que a soma dos valores obtidos é usada para movimentar as pedras do tabuleiro. Nesse caso, a distribuição de probabilidades dos valores obtidos pode ser obtida da seguinte forma:

**Tabela 1.2** - Distribuição de probabilidades da soma dos valores obtidos a partir do lançamento de dois dados ideais.

Valores Admissíveis	1	2	3	4	5	6	7	8	9	10	11	12
Combinações	-	1:1	1:2 2:1	1:3 2:2 3:1	1:4 2:3 3:2 4:1	1:5 2:4 3:3 4:2 5:1	1:6 2:5 3:4 4:3 5:2 6:1	2:6 3:5 4:4 5:3 6:2	3:6 4:5 5:4 6:3	4:6 5:5 6:4	5:6 6:5	6:6
Probabilidade	0	1/36	2/36	3/36	4/36	5/36	6/36	5/36	4/36	3/36	2/36	1/36

A Figura 1.13 abaixo mostra o histograma com a distribuição de probabilidades do problema considerado.



**Figura 1.13** - Distribuição de probabilidades da soma dos valores obtidos a partir do lançamento de dois dados ideais.

A partir da distribuição de probabilidades da Tabela 1.2 e da Figura 1.13, é possível obter os seguintes valores para a média e para a variância:

$$\begin{aligned} \mu_{X+Y} &= 1\frac{0}{36} + 2\frac{1}{36} + 3\frac{2}{36} + 4\frac{3}{36} + 5\frac{4}{36} + 6\frac{5}{36} + 7\frac{6}{36} + \\ & 8\frac{5}{36} + 9\frac{4}{36} + 10\frac{3}{36} + 11\frac{2}{36} + 12\frac{1}{36} = \frac{252}{36} = 7 \end{aligned}$$

$$\begin{aligned} \sigma_{X+Y}^2 &= \frac{0}{36}(1-7)^2 + \frac{1}{36}(2-7)^2 + \frac{2}{36}(3-7)^2 + \frac{3}{36}(4-7)^2 + \frac{4}{36}(5-7)^2 + \\ & \frac{5}{36}(6-7)^2 + \frac{6}{36}(7-7)^2 + \frac{5}{36}(8-7)^2 + \frac{4}{36}(9-7)^2 + \frac{3}{36}(10-7)^2 + \\ & \frac{2}{36}(11-7)^2 + \frac{1}{36}(12-7)^2 = \frac{210}{36} = \frac{35}{6} \end{aligned}$$

Como os experimentos dos lançamentos dos dados são independentes, as Equações (1.22) e (1.44) dizem que

$$\mu_{X+Y} = \mu_X + \mu_Y = 3.5 + 3.5 = 7$$

$$\sigma_{X+Y}^2 = \sigma_X^2 + \sigma_Y^2 = \frac{17.5}{6} + \frac{17.5}{6} = \frac{35}{6}$$

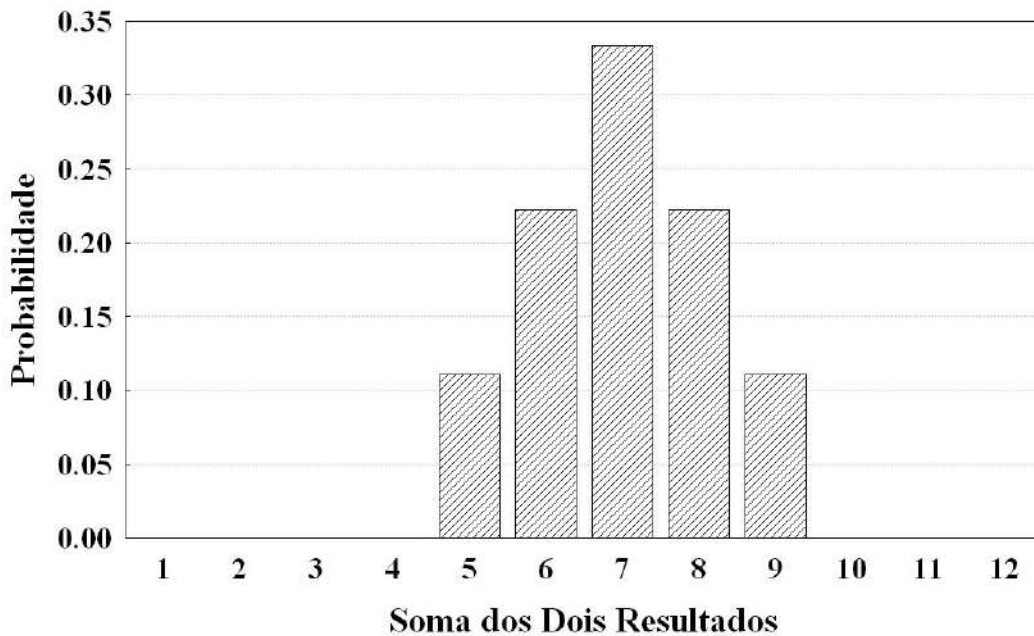
que confirmam os resultados obtidos anteriormente.

**Exemplo 1.8** - Para o dado ideal dos Exemplos 1.2 e 1.3, suponha que dois dados são lançados em seqüência em um jogo e que a soma dos valores obtidos é usada para movimentar as pedras do tabuleiro. No entanto, uma regra do jogo impõe que se o valor obtido no primeiro conjunto de dados for 1, 2 ou 3, o segundo valor só é aceito se for igual a 4, 5 ou 6, e vice-versa. Nesse caso, a distribuição de probabilidades dos valores obtidos pode ser obtida da seguinte forma:

**Tabela 1.3** - Distribuição de probabilidades da soma dos valores obtidos a partir do lançamento de dois dados ideais, com regra definida no Exemplo 1.7.

Valores Admissíveis	1	2	3	4	5	6	7	8	9	10	11	12
Combinações	-	-	-	-	1:4 4:1	1:5 2:4 4:2 5:1	1:6 2:5 3:4 4:3 5:2 6:1	2:6 3:5 5:3 6:2	3:6 6:3	-	-	-
Probabilidade	0	0	0	0	1/9	2/9	3/9	2/9	1/9	0	0	0

A Figura 1.14 abaixo mostra o histograma com a distribuição de probabilidades do problema considerado. A Figura 1.15 mostra as distribuições de probabilidade do dado ideal no primeiro lançamento e no segundo lançamento, segundo as regras estabelecidas.



**Figura 1.14** - Distribuição de probabilidades da soma dos valores obtidos a partir do lançamento de dois dados ideais.

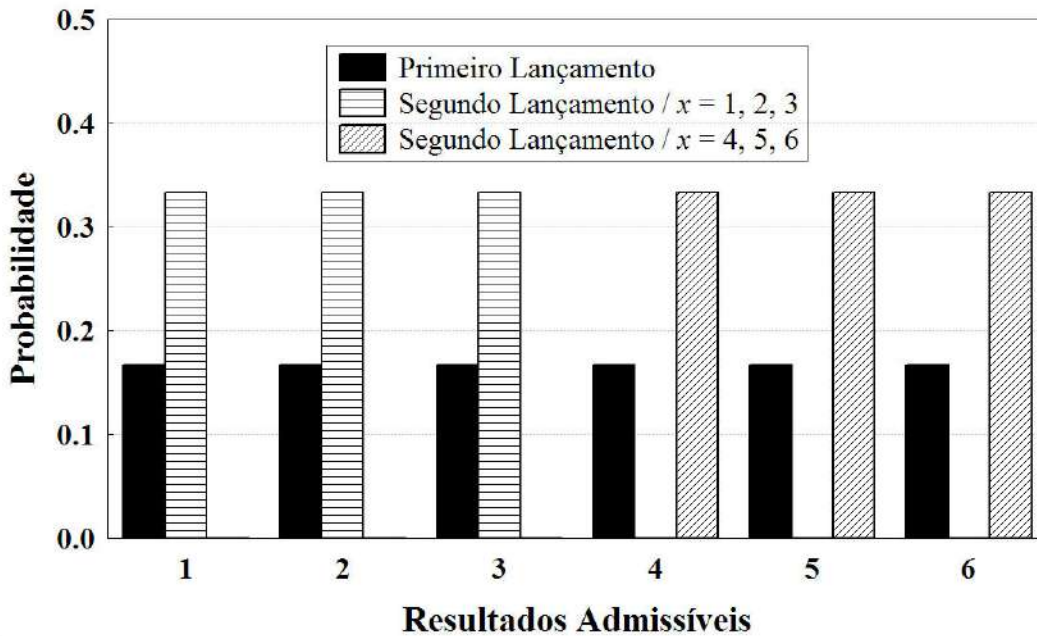


Figura 1.15 - Distribuições de probabilidades dos resultados durante o primeiro lançamento e durante o segundo lançamento.

Para o primeiro lançamento, os Exemplos 1.2, 1.3 e 1.6 mostram que  $\mu_x = 3.5$  e  $\sigma_x^2 = 17.5/6$ . Para o segundo lançamento esses valores têm que ser recalculados, pois os resultados do primeiro lançamento interferem nos resultados obtidos no segundo lançamento. Assim, para o cálculo da média dos valores obtidos no segundo lançamento,

$$\begin{aligned} \mu_y &= \sum_{i=1}^6 y_i \left[ \sum_{j=1}^6 p_x(x_j) p_y(y_i/x_j) \right] = 1 \left( \frac{10}{63} + \frac{10}{63} + \frac{10}{63} + \frac{11}{63} + \frac{11}{63} + \frac{11}{63} \right) + \\ &+ 2 \left( \frac{10}{63} + \frac{10}{63} + \frac{10}{63} + \frac{11}{63} + \frac{11}{63} + \frac{11}{63} \right) + 3 \left( \frac{10}{63} + \frac{10}{63} + \frac{10}{63} + \frac{11}{63} + \frac{11}{63} + \frac{11}{63} \right) + \\ &+ 4 \left( \frac{11}{63} + \frac{11}{63} + \frac{11}{63} + \frac{10}{63} + \frac{10}{63} + \frac{10}{63} \right) + 5 \left( \frac{11}{63} + \frac{11}{63} + \frac{11}{63} + \frac{10}{63} + \frac{10}{63} + \frac{10}{63} \right) + \\ &+ 6 \left( \frac{11}{63} + \frac{11}{63} + \frac{11}{63} + \frac{10}{63} + \frac{10}{63} + \frac{10}{63} \right) = \frac{63}{18} = 3.5 \end{aligned}$$

onde os termos entre parênteses representam a probabilidade do resultado ser obtido no segundo lançamento, dados os resultados obtidos no primeiro. Para o cálculo da variância dos resultados obtidos no segundo lançamento,

$$\begin{aligned} \sigma_y^2 &= \sum_{i=1}^6 (y_i - \mu_y)^2 \left[ \sum_{j=1}^6 p_x(x_j) p_y(y_i/x_j) \right] = (1-3.5)^2 \left( \frac{1}{6} \right) + (2-3.5)^2 \left( \frac{1}{6} \right) + \\ &+ (3-3.5)^2 \left( \frac{1}{6} \right) + (4-3.5)^2 \left( \frac{1}{6} \right) + (5-3.5)^2 \left( \frac{1}{6} \right) + (6-3.5)^2 \left( \frac{1}{6} \right) = \frac{17.5}{6} \end{aligned}$$



Assim, apesar das distribuições de probabilidade serem bastante diferentes no primeiro e no segundo lançamento e dos resultados obtidos não serem independentes, as médias e variâncias em ambos os casos são idênticas. A partir da distribuição de probabilidades da Tabela 1.3 e da Figura 1.14, é possível obter os seguintes valores para a média e para a variância da soma dos resultados:

$$\mu_{X+Y} = 1\frac{0}{9} + 2\frac{0}{9} + 3\frac{0}{9} + 4\frac{0}{9} + 5\frac{1}{9} + 6\frac{2}{9} + 7\frac{3}{9} + 8\frac{2}{9} + 9\frac{1}{9} + 10\frac{0}{9} + 11\frac{0}{9} + 12\frac{0}{9} = \frac{63}{9} = 7$$

$$\begin{aligned} \sigma_{X+Y}^2 &= \frac{0}{9}(1-7)^2 + \frac{0}{9}(2-7)^2 + \frac{0}{9}(3-7)^2 + \frac{0}{9}(4-7)^2 + \frac{1}{9}(5-7)^2 + \frac{2}{9}(6-7)^2 + \\ &+ \frac{3}{9}(7-7)^2 + \frac{2}{9}(8-7)^2 + \frac{1}{9}(9-7)^2 + \frac{0}{9}(10-7)^2 + \frac{0}{9}(11-7)^2 + \frac{0}{9}(12-7)^2 = \frac{12}{9} = \frac{8}{6} \end{aligned}$$

Como os experimentos dos lançamentos dos dados nesse caso não são independentes, é necessário calcular a covariância entre os resultados obtidos do primeiro e do segundo lançamento dos dados através, por exemplo, da Equação (1.45). Nesse caso, o valor médio obtido do segundo lançamento  $\mu_Y(x)$  é igual a 5, se  $i = 1, 2$  ou 3, e é igual a 2, se  $x = 4, 5$  ou 6. Portanto,

$$\sigma_{XY}^2 = 1\frac{1}{6}5 + 2\frac{1}{6}5 + 3\frac{1}{6}5 + 4\frac{1}{6}2 + 5\frac{1}{6}2 + 6\frac{1}{6}2 - (3.5)^2 = 10 - 12.25 = -2.25$$

O valor negativo da covariância indica que o valor obtido do segundo lançamento tende a diminuir se o valor obtido do primeiro lançamento aumenta.

Utilizando-se a Equação (1.22) para cálculo do valor médio da soma dos resultados, obtém-se

$$\mu_{X+Y} = \mu_X + \mu_Y = 3.5 + 3.5 = 7$$

que confirma os resultados anteriores. Utilizando-se a Equação (1.40) para cálculo da variância da soma dos resultados, obtém-se

$$\sigma_{X+Y}^2 = \sigma_X^2 + 2\sigma_{XY}^2 + \sigma_Y^2 = \frac{17.5}{6} - 4.5 + \frac{17.5}{6} = \frac{8}{6}$$

que também confirma os resultados obtidos anteriormente.

**Exemplo 1.9** - Para o dado ideal dos Exemplos 1.2 e 1.3, suponha que um único dado é lançado para gerar simultaneamente dois números. O primeiro número é o valor obtido do experimento. O segundo resultado é escolhido de acordo com uma regra bem simples: para  $x = 1$  ou 2,  $y = 6$ ; para  $x = 3$  ou 4,  $y = 1$ ; para  $x = 5$  ou 6,  $y = 6$ . Portanto, o grau de dependência entre os dois resultados é total e determinística.

A Equação (1.45) é utilizada para calcular a covariância entre as medidas  $x$  e  $y$ . Para tanto, a média  $\mu_y$  pode ser calculada como

$$\mu_y = \frac{1}{6}6 + \frac{1}{6}6 + \frac{1}{6}1 + \frac{1}{6}1 + \frac{1}{6}6 + \frac{1}{6}6 = \frac{26}{6}$$

enquanto a covariância pode ser calculada como

$$\sigma_{XY}^2 = 1\frac{1}{6}6 + 2\frac{1}{6}6 + 3\frac{1}{6}1 + 4\frac{1}{6}1 + 5\frac{1}{6}6 + 6\frac{1}{6}6 - 3.5\frac{26}{6} = \frac{91}{6} - \frac{91}{6} = 0$$

Portanto, apesar das variáveis  $x$  e  $y$  estarem completamente correlacionadas, a covariância entre as duas variáveis no problema proposto é igual a zero. Isso mostra que o fato da covariância ser igual a zero não implica necessariamente que as medidas sejam de fato independentes.

É importante observar na Equação (1.41) que a covariância representa uma expectativa de variação conjunta dos resultados obtidos a partir de diferentes experimentos. Se a covariância entre duas variáveis  $x$  e  $y$  é um número positivo, a Equação (1.41) indica que flutuações do resultado do experimento  $x$  acima da média são também normalmente acompanhadas de flutuações do resultado do experimento  $y$  acima da média, e vice-versa. As variáveis apresentam, portanto, algum grau de **dependência direta**. Se a covariância entre duas variáveis  $x$  e  $y$  é um número negativo, a Equação (1.41) indica que flutuações do resultado do experimento  $x$  acima da média são também normalmente acompanhadas de flutuações do resultado do experimento  $y$  abaixo da média, e vice-versa. As variáveis apresentam, portanto, algum grau de **dependência inversa**. Portanto, a covariância pode ser um importante elemento para análise do grau de dependência funcional existente entre variáveis distintas, a despeito dos resultados apresentados no Exemplo 1.9. A covariância não é uma medida absoluta de dependência funcional porque ela não leva em consideração que a variável  $y$  pode ora aumentar com a variável  $x$  em alguns intervalos, ora diminuir com a variável  $x$  em outros intervalos, como no caso do Exemplo 1.9. Em outras palavras, a operação de covariância não consegue detectar de forma adequada a existência de dependência não linear entre  $x$  e  $y$ .

Para fins de manipulação de expressões matemáticas é importante observar que a operação de covariância satisfaz as seguintes propriedades:

**Propriedade 1.7** - Sejam os conjuntos  $(x_i, p_{xi})$  e  $(y_i, p_{yi})$  dois histogramas de probabilidades e  $\alpha$  e  $\beta$  dois escalares. Então,  $\text{Covar}\{\alpha x, \beta y\} = \text{Covar}\{\beta y, \alpha x\} = \alpha\beta\text{Covar}\{x, y\}$ .

$$\begin{aligned} \text{Covar}\{\alpha x, \beta y\} &= E\{(\alpha x - \mu_{\alpha x})(\beta y - \mu_{\beta y})\} = E\{(\alpha x - \alpha\mu_x)(\beta y - \beta\mu_y)\} = \\ &= E\{\alpha\beta(x - \mu_x)(y - \mu_y)\} = \alpha\beta E\{(x - \mu_x)(y - \mu_y)\} = \alpha\beta\text{Covar}\{x, y\} \end{aligned} \quad (1.46)$$

Portanto, ao multiplicar os resultados possíveis por escalares  $\alpha$  e  $\beta$  quaisquer, a covariância fica multiplicada pelos mesmos escalares.

**Propriedade 1.8** - Sejam os conjuntos  $(x_i, p_{xi})$ ,  $(y_i, p_{yi})$  e  $(z_i, p_{zi})$  três histogramas de probabilidades. Então,  $\text{Covar}\{x, y+z\} = \text{Covar}\{x, y\} + \text{Covar}\{x, z\}$  e  $\text{Covar}\{x+z, y\} = \text{Covar}\{x, y\} + \text{Covar}\{z, y\}$ .

$$\begin{aligned} \text{Covar}\{x, y+z\} &= E\{(x - \mu_x)((y+z) - \mu_{y+z})\} = \\ &E\{(x - \mu_x)((y+z) - (\mu_y + \mu_z))\} = E\{(x - \mu_x)((y - \mu_y) + (z - \mu_z))\} = \quad (1.47) \\ &E\{(x - \mu_x)(y - \mu_y)\} + E\{(x - \mu_x)(z - \mu_z)\} = \text{Covar}\{x, y\} + \text{Covar}\{x, z\} \end{aligned}$$

Portanto, ao somar os resultados possíveis de distribuições de probabilidade distintas, a covariância fica somada de forma análoga.

Como a variância  $\sigma_x^2$  tem dimensão do quadrado da variável  $x$  (de  $x^2$ , portanto) é útil definir o **desvio padrão** da variável  $x$ , representado como  $\sigma_x$ , como

$$\sigma_x = \sqrt{\sigma_x^2} \quad (1.48)$$

O desvio padrão é uma medida adequada de espalhamento na escala métrica da variável  $x$ , obtida a partir da operação de cálculo da variância. Como veremos nas próximas seções, o desvio padrão pode ser somado à média para fornecer regiões onde estão concentrados os resultados mais prováveis, dentro de um certo limite de confiança.

Uma outra normalização frequentemente utilizada para definir a variância é o chamado **índice de polidispersão**,  $IP$ . O índice de polidispersão, **polidispersividade** ou simplesmente **polidispersão** é uma medida relativa da variância da distribuição, na forma

$$IP = 1 + \frac{\sigma_x^2}{\mu_x^2} \quad (1.49)$$

O índice de polidispersão é, portanto, uma medida do grau relativo de espalhamento em relação à média, encontrando várias aplicações práticas para interpretação de problemas físicos reais.

Como a covariância  $\sigma_{xy}^2$  tem dimensão das variáveis  $x$  e  $y$  simultaneamente e como a magnitude dessas variáveis pode mudar de problema para problema, é conveniente definir uma forma normalizada para o grau de **dependência funcional linear** entre as variáveis  $x$  e  $y$ . A forma normalizada mais usada é o chamado **coeficiente ou fator de correlação linear**, ou simplesmente coeficiente ou fator de correlação, representado como  $\rho_{xy}$  e definido como

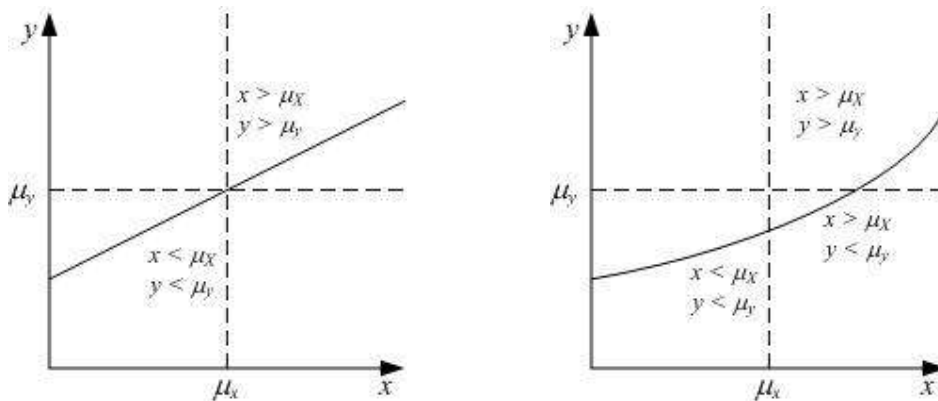
$$\rho_{xy} = \frac{\sigma_{xy}}{\sigma_x \sigma_y} \quad (1.50)$$

Quando as variáveis  $x$  e  $y$  são independentes,  $\sigma_{XY}^2 = 0$  e  $\rho_{XY} = 0$ . Quando  $y = \alpha x + \beta$ , então  $\mu_Y = \alpha\mu_X + \beta$ ,  $\sigma_Y^2 = \alpha^2\sigma_X^2$ ,  $\sigma_{XY}^2 = \alpha\sigma_X^2$  e  $\rho_{XY} = \pm 1$ , dependendo se  $\alpha$  é positivo ou negativo respectivamente. É interessante observar que isso implica na validade da seguinte relação

$$-1 \leq \rho_{XY} \leq 1 \tag{1.51}$$

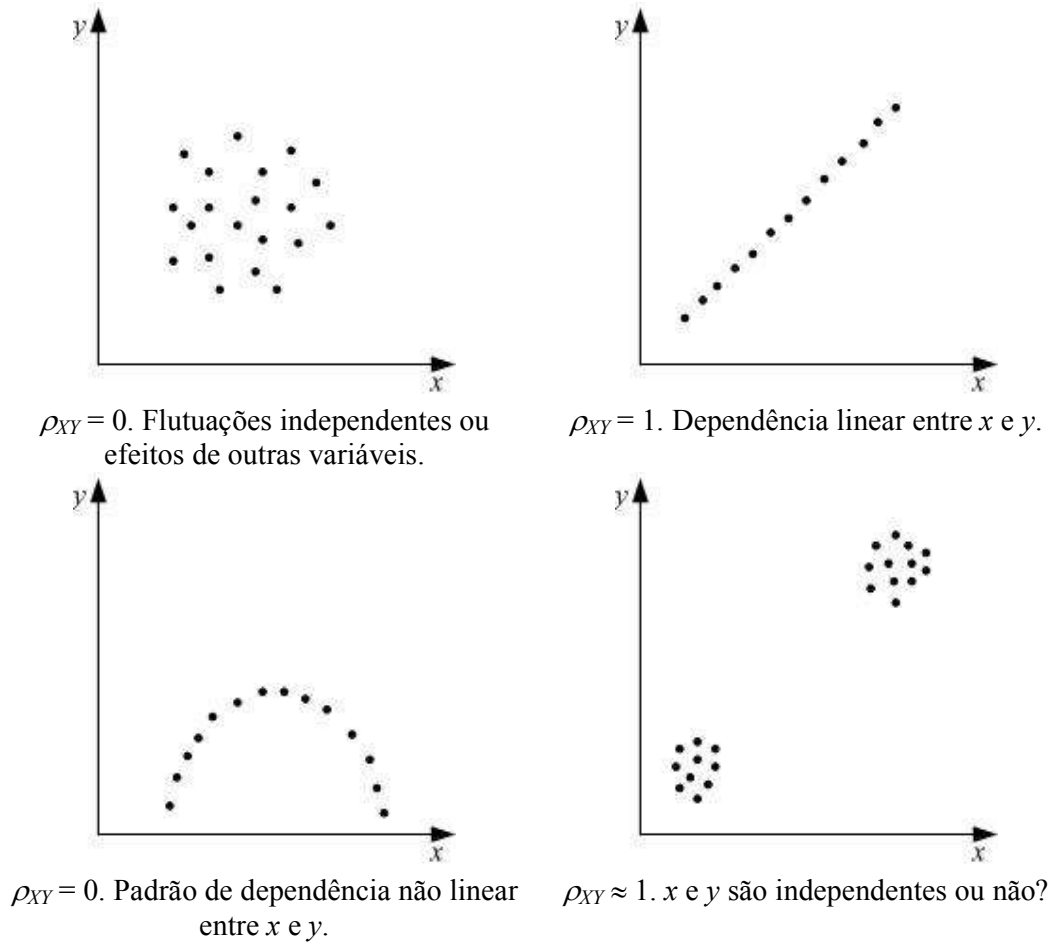
o que mostra que o processo de normalização definido pela Equação (1.50) é bastante eficiente. Se as variáveis  $x$  e  $y$  estão relacionadas linearmente na forma direta,  $\rho_{XY} = 1$ ; se as variáveis  $x$  e  $y$  estão relacionadas linearmente na forma inversa,  $\rho_{XY} = -1$ ; se as variáveis  $x$  e  $y$  são independentes,  $\rho_{XY} = 0$ . Quanto mais próximo de 1 o valor absoluto de  $\rho_{XY}$ , mais perfeito o grau de correlação linear entre as variáveis  $x$  e  $y$ . Quanto mais próximo de 0 o valor absoluto de  $\rho_{XY} = 0$ , maior o grau de flutuação independente das variáveis (o que pode indicar independência verdadeira entre os experimentos, mas também pode indicar a existência de erros pronunciados de medição ou influência de outras variáveis sobre o experimento) ou maior o grau de não linearidade da dependência funcional entre  $x$  e  $y$ .

A Figura 1.16 procura ilustrar como a presença de dependência não linear entre as variáveis  $x$  e  $y$  provoca redução do fator de correlação. Observe que no primeiro gráfico, em que a relação é linear, sempre que  $x$  se eleva em relação ao valor médio, o mesmo ocorre com a variável  $y$ . No entanto, quando a relação é não linear, as médias estão deslocadas no plano do segundo gráfico. Isso faz com que existam regiões onde a variável  $x$  está acima da média e a variável  $y$  está abaixo da média, contribuindo para a redução da covariância entre  $x$  e  $y$ .



**Figura 1.16** - Ilustração do efeito da não linearidade sobre o cálculo da covariância.

Como já discutido anteriormente, o que o coeficiente de correlação linear mede de fato é se existe alguma tendência de variação linear entre  $x$  e  $y$ ; ou seja, se um aumento de  $x$  provoca um aumento proporcional em  $y$ . Portanto, coeficientes de correlação devem ser usados com cautela para a interpretação de resultados, como ilustrado na Figura 1.17.



**Figura 1.17** - Padrões típicos de dependência entre  $x$  e  $y$  e respectivos coeficientes de correlação.

É muito interessante notar que na definição de média introduzida pela Equação (1.7) tem-se

$$\mu_X = \sum_{i=1}^{NR} p_i x_i \tag{1.7}$$

Por sua vez, na definição de variância proposta pela Equação (1.36) tem-se

$$\begin{aligned} \sigma_{XX}^2 &= \sum_{i=1}^{NR} p_i (x_i - \mu_X)^2 = \sum_{i=1}^{NR} p_i (x_i^2 - 2 x_i \mu_X + \mu_X^2) = \\ &= \sum_{i=1}^{NR} p_i x_i^2 - 2 \mu_X \sum_{i=1}^{NR} p_i x_i + \mu_X^2 \sum_{i=1}^{NR} p_i = \sum_{i=1}^{NR} p_i x_i^2 - \mu_X^2 \end{aligned} \tag{1.52}$$

Com frequência, expressões na forma

$$\mu_X^{(k)} = \sum_{i=1}^{NR} p_i x_i^k \quad (1.53)$$

aparecem na análise estatística. Estas expressões são chamadas de **momentos estatísticos** ou momentos da curva de distribuição de probabilidades. Dessa forma, a média e a variância poderiam ser definidas como

$$\mu_X = \sum_{i=1}^{NR} p_i x_i = \mu_X^{(1)} \quad (1.54)$$

$$\sigma_{XX}^2 = \sum_{i=1}^{NR} p_i x_i^2 - \mu_X^2 = \mu_X^{(2)} - [\mu_X^{(1)}]^2 \quad (1.55)$$

de maneira que é possível afirmar que a operação de média registra o primeiro momento da curva de distribuição, enquanto a operação de cômputo da variância registra o segundo momento da curva de distribuição. Momentos estatísticos adicionais podem ser também calculados, já que infinitas distribuições de probabilidade distintas podem apresentar a mesma média e variância. Por exemplo, o momento de ordem 3, normalmente registrado na forma

$$\lambda_X = \frac{\left[ \sum_{i=1}^{NR} p_i (x_i - \mu_X)^3 \right]^{1/3}}{\sigma_X} \quad (1.56)$$

fornece informações sobre a assimetria da distribuição de probabilidades, sendo por isso chamado de **fator de assimetria**. Baseado na dependência cúbica utilizada para definição do fator de assimetria da distribuição, não é difícil compreender que valores positivos do fator de assimetria são resultantes de distribuições muito alongadas no sentido positivo do eixo dos resultados, e vice-versa. O momento de ordem 4, por sua vez, normalmente registrado na forma

$$k_X = \frac{\left[ \sum_{i=1}^{NR} p_i (x_i - \mu_X)^4 \right]^{1/4}}{\sigma_X} \quad (1.57)$$

e chamado de **kurtose**, pode ser associado ao formato achatado ou alongado da distribuição de probabilidades, mostrar multimodalidades, e assim por diante.

É importante observar que uma distribuição de probabilidades possui infinitos momentos e que, de forma inversa, os infinitos momentos da curva de distribuição precisam ser especificados para que a curva de distribuição possa ser especificada também de forma inequívoca. Portanto, a informação completa contida no histograma de probabilidades não pode jamais ser substituída por uma coleção finita de momentos estatísticos, como a média e a variância.

**Exemplo 1.10** - Suponha que uma distribuição de probabilidades definida no intervalo discreto  $(0, \infty)$  possa ser definida como

$$P_i = q^{i-1} (\alpha_0 + \alpha_1 i + \alpha_2 i^2)$$

onde  $q$  é um número real positivo  $0 < q < 1$  e  $\alpha_0$ ,  $\alpha_1$  e  $\alpha_2$  são escalares reais. Nesse caso, para especificar a média e a variância é necessário fazer

$$\begin{aligned} \mu_i^{(0)} = 1 &= \alpha_0 \sum_{i=1}^{\infty} q^{i-1} + \alpha_1 \sum_{i=1}^{\infty} i q^{i-1} + \alpha_2 \sum_{i=1}^{\infty} i^2 q^{i-1} \\ \mu_i^{(1)} = \mu_i &= \alpha_0 \sum_{i=1}^{\infty} i q^{i-1} + \alpha_1 \sum_{i=1}^{\infty} i^2 q^{i-1} + \alpha_2 \sum_{i=1}^{\infty} i^3 q^{i-1} \\ \mu_i^{(2)} = \sigma_i^2 + \mu_i^2 &= \alpha_0 \sum_{i=1}^{\infty} i^2 q^{i-1} + \alpha_1 \sum_{i=1}^{\infty} i^3 q^{i-1} + \alpha_2 \sum_{i=1}^{\infty} i^4 q^{i-1} \end{aligned}$$

São, portanto, três equações e quatro incógnitas, havendo infinitas distribuições de probabilidade distintas com a forma geral proposta, com a mesma média e a mesma variância. Não é difícil compreender que a forma geral proposta pode ser estendida para um grau arbitrariamente grande da expansão polinomial, resultando na necessidade de se especificar um número arbitrariamente grande de momentos estatísticos para fixar de fato a distribuição.

Admitindo-se que  $\mu_i = 100$  e  $\sigma_i = 100$ , pode-se obter, por exemplo, as distribuições de probabilidade apresentadas na Figura 1.18a, para distintos valores de  $q$  e de  $\alpha_0$ ,  $\alpha_1$  e  $\alpha_2$  apresentados na Tabela 1.4. Fica assim patente que a especificação da média e da variância (ou de um conjunto finito de momentos) não é suficiente para definir inequivocamente a curva de distribuição de probabilidades que as originou. Portanto, a média e a variância não substituem a informação contida no histograma de probabilidades.

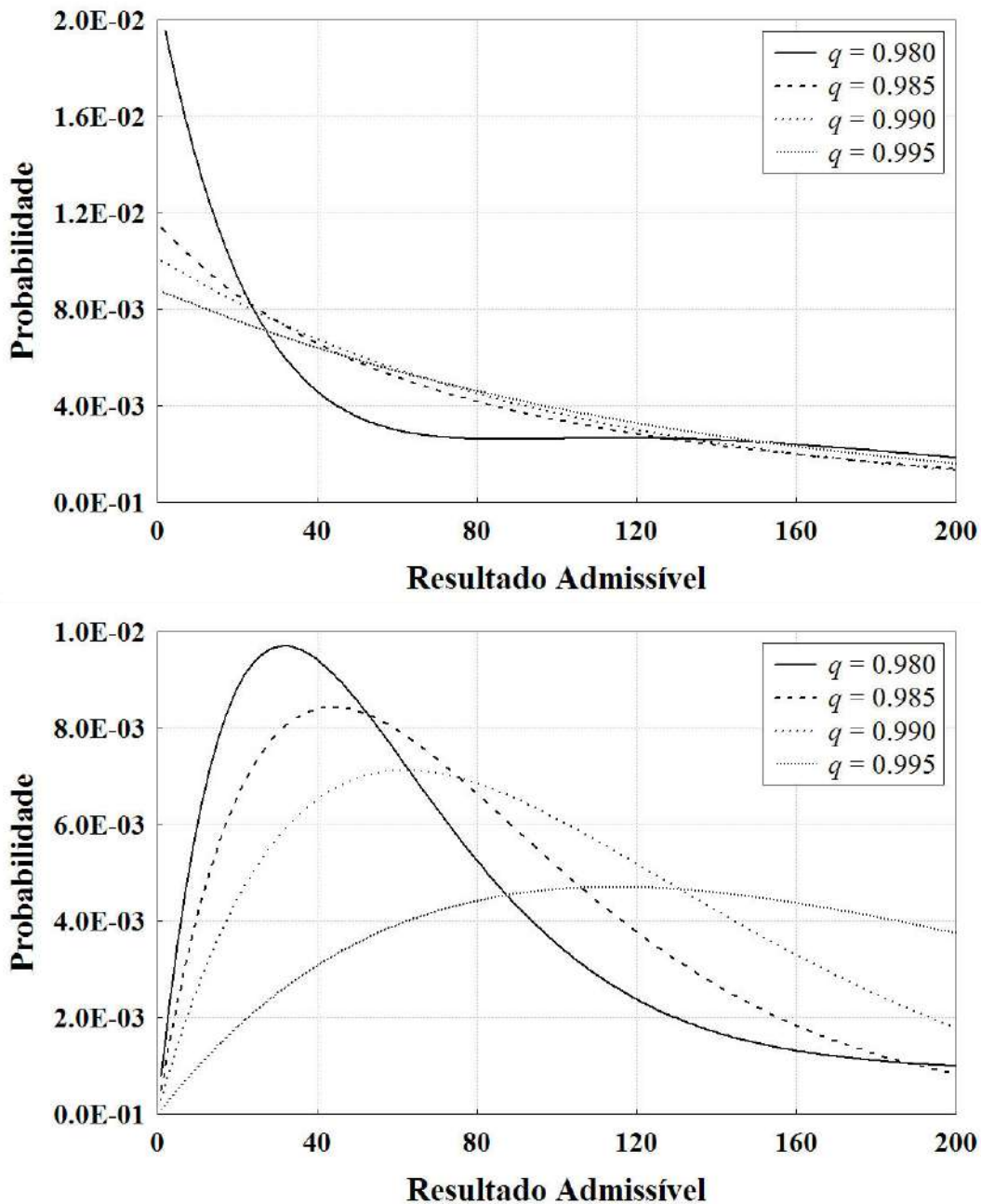


Figura 1.18 - Diferentes distribuições de probabilidade com  $\mu_I = 100$  e  $\sigma_I = 100$ .  
 (a) Formato 1; (b) Formato 2.

É importante observar que a forma da distribuição não precisa necessariamente estar no mesmo formato fixado anteriormente. Por exemplo, histogramas bastante distintos que apresentam  $\mu_I = 100$  e  $\sigma_I = 100$  são também apresentados na Figura 1.18b. Nesse caso, admitiu-se que

$$P_i = iq^{i-1}(\alpha_0 + \alpha_1 i + \alpha_2 i^2)$$

de maneira que



$$\begin{aligned} \mu_I^{(0)} = 1 &= \alpha_0 \sum_{i=1}^{\infty} i q^{i-1} + \alpha_1 \sum_{i=1}^{\infty} i^2 q^{i-1} + \alpha_2 \sum_{i=1}^{\infty} i^3 q^{i-1} \\ \mu_I^{(1)} = \mu_I &= \alpha_0 \sum_{i=1}^{\infty} i^2 q^{i-1} + \alpha_1 \sum_{i=1}^{\infty} i^3 q^{i-1} + \alpha_2 \sum_{i=1}^{\infty} i^4 q^{i-1} \\ \mu_I^{(2)} = \sigma_I^2 + \mu_I^2 &= \alpha_0 \sum_{i=1}^{\infty} i^3 q^{i-1} + \alpha_1 \sum_{i=1}^{\infty} i^4 q^{i-1} + \alpha_2 \sum_{i=1}^{\infty} i^5 q^{i-1} \end{aligned}$$

Tabela 1.4 - Valores usados para construir os histogramas da Figura 1.18.

Formato 1			
$q$	$\alpha_0$	$\alpha_1$	$\alpha_2$
0.980	$2.08329 \times 10^{-2}$	$-4.37231 \times 10^{-4}$	$4.24820 \times 10^{-6}$
0.985	$1.14248 \times 10^{-2}$	$-6.50240 \times 10^{-6}$	$4.54385 \times 10^{-7}$
0.990	$1.00510 \times 10^{-2}$	$-1.01785 \times 10^{-6}$	$2.55075 \times 10^{-9}$
0.995	$8.78147 \times 10^{-3}$	$-2.52360 \times 10^{-5}$	$1.58613 \times 10^{-8}$
Formato 2			
$q$	$\alpha_0$	$\alpha_1$	$\alpha_2$
0.980	$8.00299 \times 10^{-4}$	$-8.16820 \times 10^{-6}$	$2.77772 \times 10^{-8}$
0.985	$5.03770 \times 10^{-4}$	$-3.38219 \times 10^{-6}$	$6.42642 \times 10^{-9}$
0.990	$2.99008 \times 10^{-4}$	$-1.50007 \times 10^{-6}$	$1.67516 \times 10^{-9}$
0.995	$1.06157 \times 10^{-4}$	$-3.12896 \times 10^{-7}$	$1.82948 \times 10^{-10}$

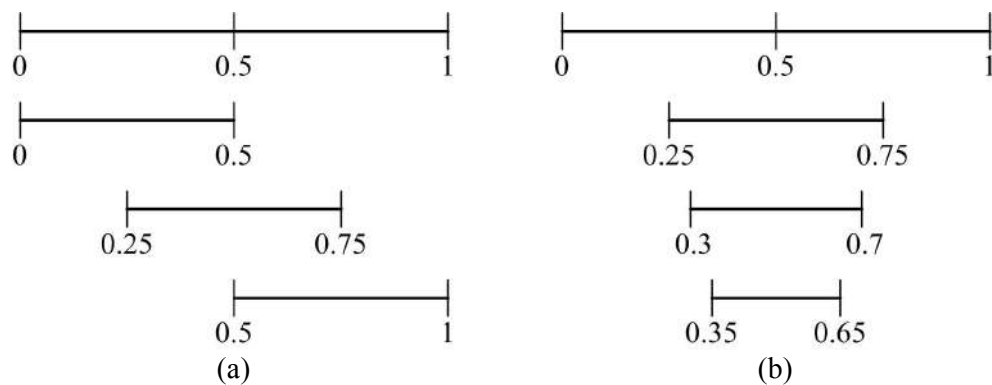
### 1.7. Extensão dos Conceitos de Distribuição, Média e Variância Para Variáveis Contínuas

Tudo o que foi visto até aqui é perfeito para **variáveis discretas**. Por exemplo, o valor que sai do dado ou é 1, 2, 3, 4, 5 ou 6; o resultado esperado ao se jogar uma moeda para o alto ou é cara ou é coroa. Mas uma grande parte dos problemas de interesse prático envolve **variáveis contínuas e não enumeráveis**. Por exemplo, qual é o tempo de vida de uma lâmpada incandescente? Qual é o grau de conversão de um regente químico que deve ser esperado na corrente de saída de um reator? Para que percebamos a diferença que existe entre os dois tipos de variáveis, suponha que todos os números reais contidos no intervalo [0,1] são acondicionados em um saco. O experimento então consiste em sortear um desses números, supondo que todos os números estão perfeitamente embaralhados. Qual deve ser então a probabilidade de se retirar o número 0.5 do saco?

Em um bingo de verdade, a probabilidade de se retirar um número inteiro qualquer entre 0 e 99 (incluindo esses extremos) escrito em uma das 100 pedras do jogo é exatamente igual a 1% (1/100), se as pedras estão perfeitamente misturadas. Isso ocorre porque a probabilidade de que qualquer uma das 100 pedras seja sorteada é a mesma. No caso do problema proposto no parágrafo anterior, há infinitos números naturais contidos no intervalo [0,1]. Logo, a probabilidade de que qualquer desses

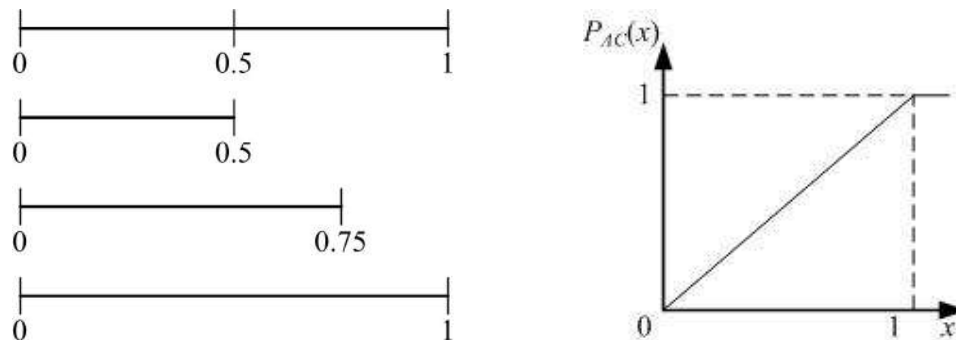
números seja sorteado é sempre igual a zero, de maneira que o histograma que representaria o problema seria trivial. Esse tipo de dificuldade de interpretação ocorre sempre que uma variável contínua é analisada. Por exemplo, qual a probabilidade de se encontrar uma pessoa com altura igual a 1.733333.... m na cidade do Rio de Janeiro, respeitadas todas as infinitas casas decimais da dízima periódica (se isso fosse possível com o nosso sistema de medida, o que não é absolutamente verdade)? Portanto, parece clara a necessidade de modificar a sistemática de análise para os problemas que envolvem variáveis contínuas.

Uma das formas mais simples de se interpretar variáveis contínuas é pensar em intervalos de valores, ao invés de valores absolutos. Por exemplo, suponha que o problema originalmente proposto seja modificado para a seguinte pergunta: Qual a probabilidade de se retirar do saco um número real menor do que 0.5? Nesse caso, supondo que a reta real é igualmente densa em números em qualquer que seja o intervalo numérico analisado, a Figura 1.19 mostra que metade dos números reais existentes no intervalo  $[0,1]$  está contida no intervalo  $[0,0.5]$ . Portanto, a probabilidade de que um número inferior a 0.5 seja sorteado no experimento é de 50%. Mais importante ainda: a modificação da pergunta, focada nos intervalos e não nos valores absolutos, permitiu associar de forma clara uma probabilidade a um certo conjunto de resultados. A Figura 1.19 mostra no entanto que isso não é suficiente para remover a ambigüidade da representação. Em primeiro lugar, ao se movimentar um segmento de tamanho especificado ao longo do intervalo  $[0,1]$ , como na Figura 1.19a, mudam-se os valores que delimitam o subintervalo considerado, mas não a probabilidade dos números contidos nele serem sorteados. Assim, se o tamanho do intervalo for igual a 0.5, a probabilidade 0.5 pode ser associada a qualquer número do intervalo, bastando para isso que o subintervalo esteja colocado sobre o número considerado. Em segundo lugar, ao se estreitar o subintervalo que contém um número qualquer, como na Figura 1.19b, a probabilidade daquele conjunto particular de pontos também muda. Assim, qualquer valor de probabilidade pode ser associado a um número qualquer do intervalo, dependendo da largura do segmento de reta considerado.



**Figura 1.19** - (a) Qualquer segmento de reta de comprimento 0.5 contido no intervalo real  $[0,1]$  contém metade dos números do intervalo. (b) Segmentos de reta de tamanhos distintos ao redor do ponto 0.5 contêm frações distintas dos pontos contidos no intervalo  $[0,1]$ .

Com a finalidade de remover a ambigüidade da representação de probabilidade em problemas envolvendo variáveis contínuas, faz-se necessário fixar ao menos um dos limites do segmento utilizado para definir o intervalo de valores considerados. Por uma questão de conveniência, faz-se aqui o "ancoramento" do segmento no valor mínimo admissível como solução para o problema considerado. Assim, define-se como a **probabilidade acumulada** de um valor  $x$ , representada por  $P_{AC}(x) = P(x' \leq x)$ , como a probabilidade de se encontrar em um determinado problema uma solução igual ou inferior ao valor de  $x$ . A Figura 1.20 ilustra esse conceito para o problema proposto.



**Figura 1.20** - Ilustração do conceito de probabilidade acumulada,  $P_{AC}(x) = P(x' \leq x)$ .

O conceito de probabilidade acumulada permite associar, portanto, a cada valor específico um número sem ambigüidade, que é a probabilidade de se encontrar como resposta do problema um valor menor que ele. Repare que esse conceito pode ser aplicado tanto a problemas discretos quanto a problemas contínuos indistintamente, embora tenha utilidade muito maior nos problemas de natureza contínua. A partir do conceito de probabilidade acumulada, chega-se facilmente ao seguinte conjunto de propriedades:

**Propriedade 1.9** - Seja  $P_{AC}(x) = P(x' \leq x)$  uma função de probabilidade acumulada para um problema particular. Então  $P_{AC}(x)$  é uma função monotônica não decrescente de  $x$ , contida no intervalo  $[0,1]$ .

Seja  $x_2 > x_1$ . Então

$$P_{AC}(x_2) = P(x' \leq x_2) = P(x' \leq x_1 + (x_2 - x_1)) \geq P(x' \leq x_1) = P_{AC}(x_1) \quad (1.58)$$

já que os valores contidos no intervalo  $x_1 < x < x_2$  foram excluídos do conjunto analisado, resultando num conjunto menor de valores possíveis. Logo, a função  $P_{AC}(x)$  é monotônica e não decrescente. Se  $x_{min}$  e  $x_{max}$  são os valores admissíveis mínimo e máximo do problema, então

$$P_{AC}(x_1 \leq x_{min}) = P(x' \leq x_{min}) = 0 \quad (1.59)$$

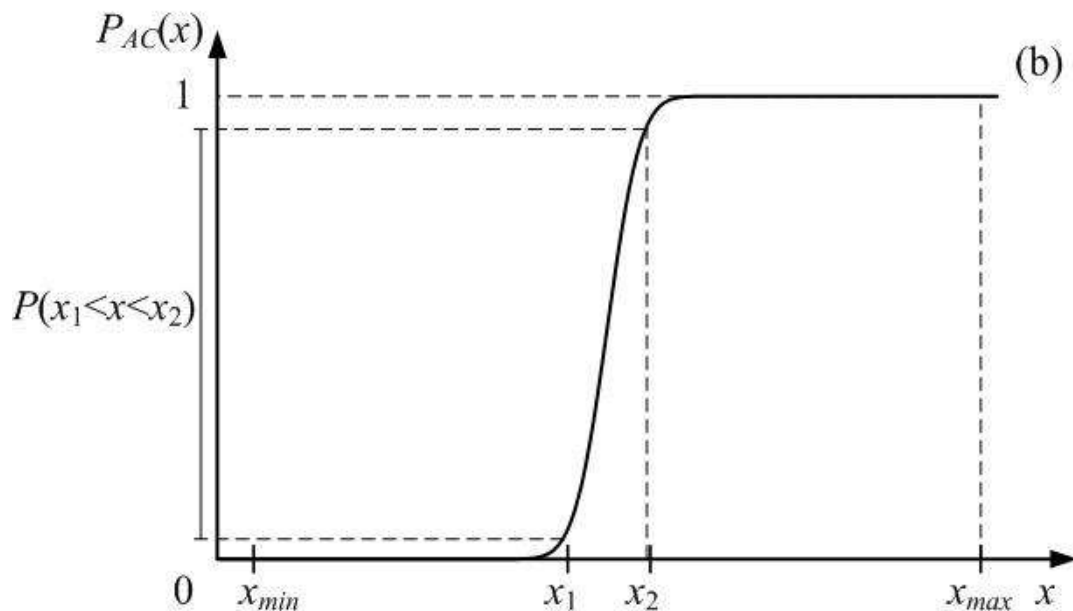
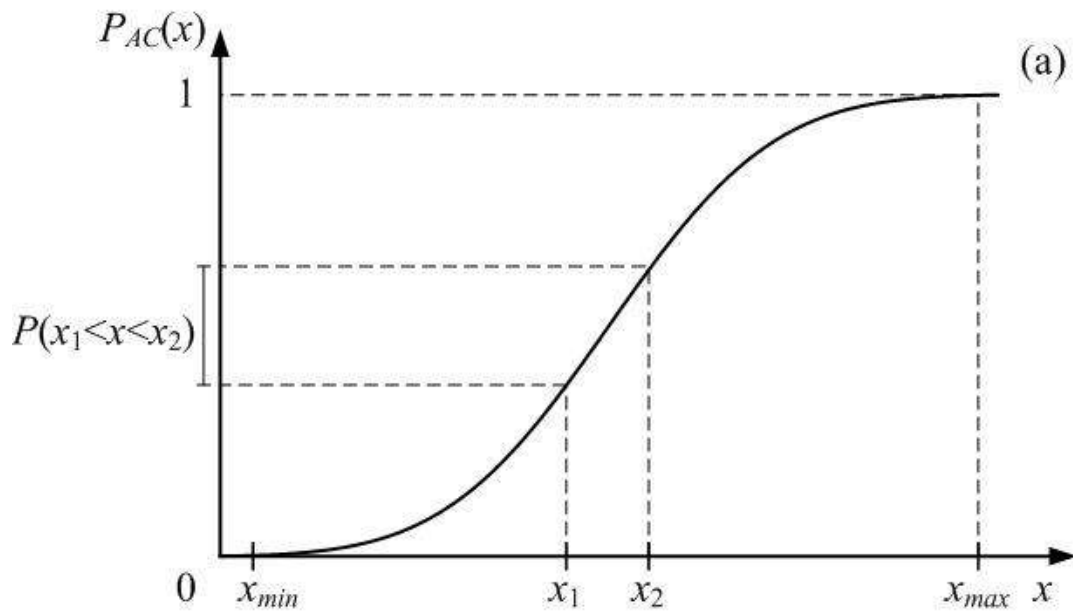
e

$$P_{AC}(x_2 \geq x_{\max}) = P(x' \leq x_{\max}) = 1 \quad (1.60)$$

Obviamente, se os valores mínimo e máximo admissíveis são infinitos, então a função  $P_{AC}(x)$  se aproxima assintoticamente dos valores 0 e 1 respectivamente, estando contida no intervalo aberto  $(0, 1)$ .

**Propriedade 1.10** - Seja  $P_{AC}(x) = P(x' \leq x)$  uma função de probabilidade acumulada para um problema particular. Então  $P(x_1 < x \leq x_2) = P_{AC}(x_2) - P_{AC}(x_1)$ .

$$P(x_1 < x \leq x_2) = P(x \leq x_2) - P(x \leq x_1) = P_{AC}(x_2) - P_{AC}(x_1) \quad (1.61)$$

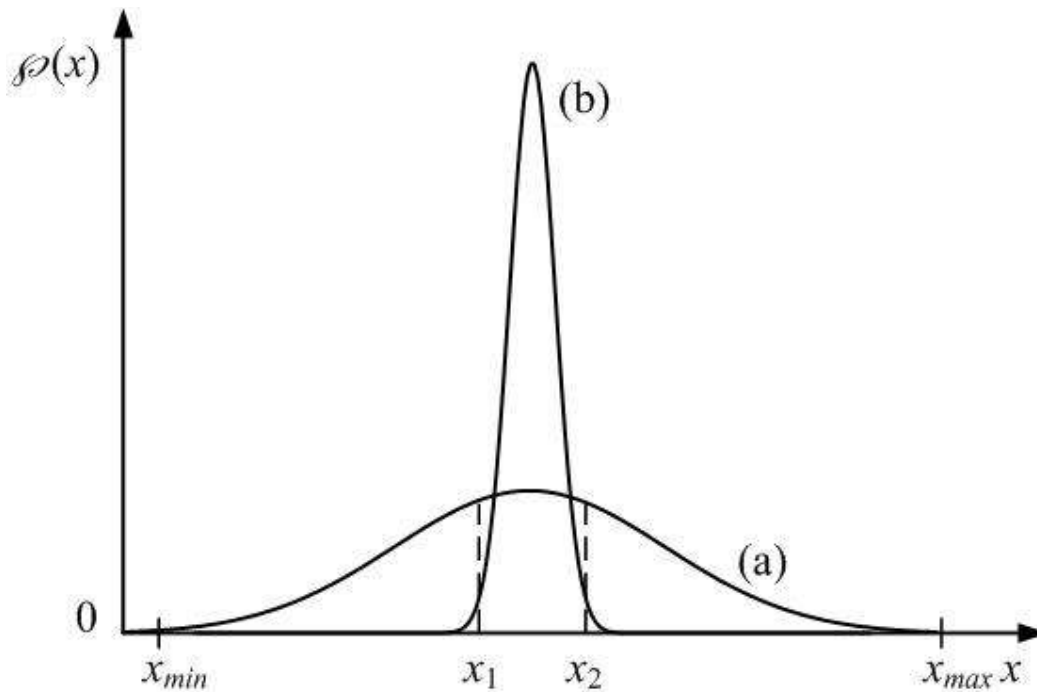


**Figura 1.21** - Interpretação quantitativa de curvas de probabilidade acumulada.

A Equação (1.61) mostra que é possível calcular a probabilidade de que um valor contido em um intervalo contínuo qualquer considerado seja obtido como resultado de um experimento, desde que a curva de probabilidade acumulada característica do experimento seja conhecida. O procedimento gráfico está ilustrado na Figura 1.21. Portanto, a curva de probabilidade acumulada tem o mesmo papel em problemas contínuos que o histograma de probabilidades tem em problemas discretos. A comparação das curvas das Figuras 1.21a e 1.21b sugere ainda que a probabilidade de encontrar um conjunto de valores em um intervalo de tamanho definido pode aumentar muito, quando a curva de probabilidade acumulada varia rapidamente. Por exemplo, a Figura 1.21b mostra que é quase certo que o resultado obtido do experimento esteja contido no intervalo  $x_1 \leq x \leq x_2$ . Comparada à Figura 1.21a, pode-se dizer que é muito mais seguro prever os resultados obtidos no experimento descrito pela Figura 1.21b. Além disso, é fácil compreender que os valores mais prováveis são aqueles ao redor dos quais a curva de probabilidade acumulada varia mais rapidamente. Portanto, há razões suficientes para introduzir a definição de **densidade de probabilidade**,  $\wp(x)$ , na forma:

$$\wp(x) = \frac{dP_{AC}(x)}{dx} \quad (1.62)$$

A densidade de probabilidade é uma medida de quão rapidamente varia a curva de probabilidade acumulada, à medida de aumenta a variável  $x$ . Ela dá, portanto, uma medida relativa de quão mais provável é a obtenção de resultados num pequeno intervalo considerado ao redor do valor  $x$ . Logo, do ponto de vista qualitativo, ela dá uma informação muito semelhante à informação fornecida pelos histogramas de probabilidade apresentados nas seções anteriores. Pode-se dizer sem excesso de rigor que a curva de densidade de probabilidades é o histograma de probabilidades de um problema em que a variável analisada é contínua. A Figura 1.22 ilustra o comportamento das curvas de densidade de probabilidade para as curvas de probabilidade acumulada da Figura 1.21. Vê-se que a curva de densidades de probabilidade da Figura 1.21b é mais estreita e intensa no intervalo  $x_1 \leq x \leq x_2$ , indicando o menor espalhamento de valores possíveis em torno de um valor médio provavelmente contido nesse mesmo intervalo.



**Figura 1.22** - Curvas de densidade de probabilidade obtidas a partir das curvas de probabilidade acumulada da Figura 1.21.

Como a curva de densidade de probabilidade está diretamente relacionada à curva de probabilidade acumulada, é possível mostrar que as seguintes propriedades são satisfeitas:

**Propriedade 1.11** - Seja  $f(x)$  uma função densidade de probabilidade para um problema particular. Então  $f(x)$  é uma **função não negativa** de  $x$ .

Como a curva de probabilidade acumulada é monotônica não decrescente, as derivadas da curva de probabilidade acumulada são nulas ou positivas, fazendo com que a Propriedade 1.11 decorra diretamente da Propriedade 1.9.

**Propriedade 1.12** - Seja  $f(x)$  uma função densidade de probabilidade para um problema particular. Então  $P(x_1 \leq x \leq x_2) = \int_{x_1}^{x_2} f(x) dx$ .

Da Equação (1.62), que define a função densidade de probabilidade, é possível escrever de forma inversa:

$$P_{AC}(x_2) - P_{AC}(x_1) = \int_{x_1}^{x_2} f(x) dx \quad (1.63)$$

que combinada com a Equação (1.61) resulta em

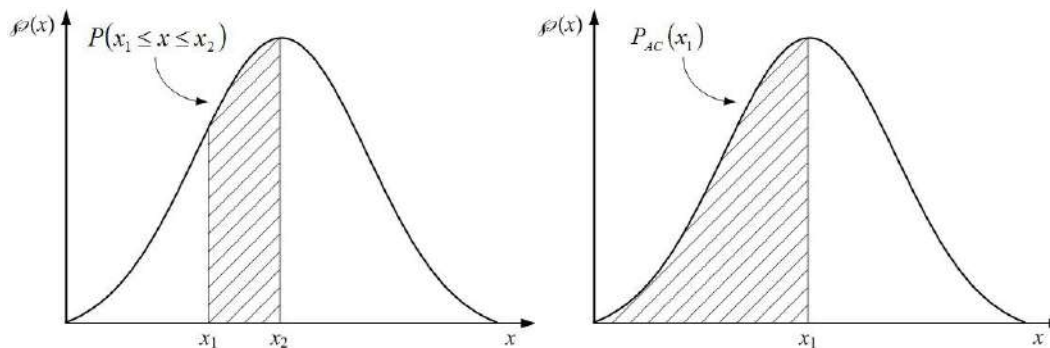
$$P(x_1 \leq x \leq x_2) = \int_{x_1}^{x_2} f(x) dx \quad (1.64)$$

**Propriedade 1.13** - Seja  $f(x)$  uma função densidade de probabilidade para um problema particular. Então  $\int_{x_{min}}^{x_{max}} f(x) dx = 1$ .

Aplicando a Equação (1.64) entre os limites admissíveis mínimo e máximo,

$$P(x_{min} \leq x \leq x_{max}) = 1 = \int_{x_{min}}^{x_{max}} f(x) dx \quad (1.65)$$

As Propriedades 1.11-1.13 são ilustradas graficamente na Figura 1.23. Dada uma curva de densidade de probabilidade, vê-se que a probabilidade de um resultado ocorrer em um certo intervalo  $x_1 \leq x \leq x_2$  é dada pela área (integral) sob a curva de densidade limitada pelo intervalo. Mais ainda, a curva de probabilidade acumulada pode ser vista como a área (integral) sob a curva limitada pelo valor mínimo admissível para o problema e o ponto particular considerado. Como todos os cálculos de probabilidade podem ser obtidos diretamente a partir da curva de densidade de probabilidade e como  $f(x)$  reflete o espalhamento e os valores em torno dos quais os resultados mais prováveis se concentram, faz-se normalmente a apresentação das distribuições de probabilidade de variáveis contínuas na forma de densidades de probabilidades.



**Figura 1.23** - Ilustração Gráfica das Propriedades 1.11-1.13.

Para mostrar como pode ser feita uma analogia direta entre os histogramas de probabilidades, definidos para as variáveis discretas, e as curvas de densidade de probabilidade, definidas para variáveis contínuas, poderíamos representar a curva de densidades de probabilidade como ilustrado na Figura 1.24. De acordo com essa representação, dada uma certa resolução  $\Delta x$  definida pelo usuário, o histograma de probabilidades  $(\bar{x}_i, p_i)$  poderia ser construído na forma:

$$\bar{x}_i = \frac{x_{i+1} + x_i}{2}, \quad x_i = x_{min} + (i-1)\Delta x \quad (1.66)$$

$$p_i = \int_{x_i}^{x_{i+1}} \wp(x) dx \approx \wp(\bar{x}_i) \Delta x \tag{1.67}$$

Portanto, os momentos da curva de distribuição de probabilidades, definidos pela Equação (1.53), poderiam ser redefinidos na forma

$$\mu_X^{(k)} = \sum_{i=1}^{NR} \bar{x}_i^k p_i \approx \sum_{i=1}^{NR} \bar{x}_i^k \wp(\bar{x}_i) \Delta x \tag{1.68}$$

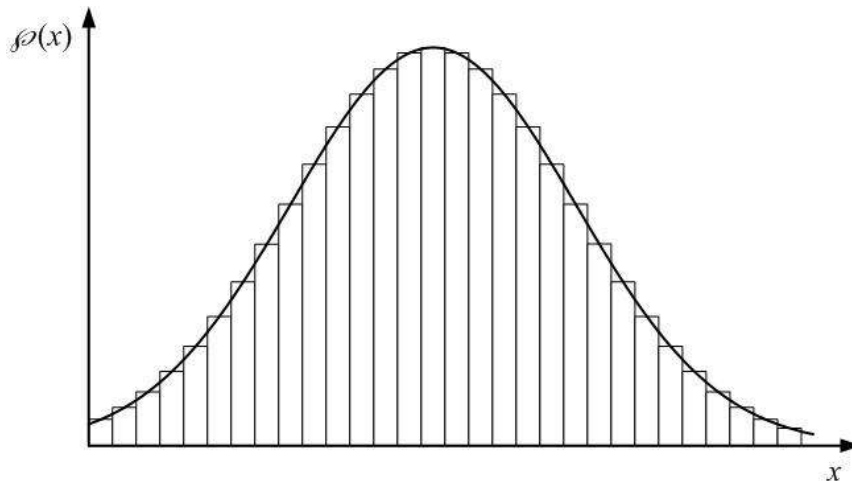
onde o número de valores admissíveis  $NR$  do histograma seria dado por

$$NR = \frac{x_{min} - x_{max}}{\Delta x} \tag{1.69}$$

No limite em que  $\Delta x$  se aproxima de zero, a Equação (1.68) fica na forma

$$\mu_X^{(k)} = \int_{x_{min}}^{x_{max}} x^k \wp(x) dx \tag{1.70}$$

que permite estender o conceito de momento estatístico da curva de distribuição para as variáveis contínuas, apenas trocando o operador soma da Equação (1.53) para o operador integral (que pode ser interpretado como uma soma infinita de fatias muito pequenas).



**Figura 1.24** - Transformação da Curva de Densidade de Probabilidades, para Variáveis Contínuas, em um Histograma de Probabilidades, para Variáveis Discretas.

A partir da Equação (1.70), desenvolvida para momentos de qualquer ordem, fica fácil perceber que os conceitos de média, variância e covariância podem ser estendidos para variáveis contínuas na forma:



$$\mu_X = \int_{x_{min}}^{x_{max}} x \wp(x) dx \tag{1.71}$$

$$\sigma_{XX}^2 = \int_{x_{min}}^{x_{max}} (x - \mu_X)^2 \wp(x) dx \tag{1.72}$$

$$\sigma_{XY}^2 = \int_{x_{min}}^{x_{max}} (x - \mu_X) \left[ \int_{y_{min}}^{y_{max}} (y - \mu_Y) \wp(y/x) dy \right] \wp(x) dx \tag{1.73}$$

Como a analogia entre os histogramas de probabilidade e as curvas de densidade de probabilidade é direta, todas as propriedades mostradas anteriormente para a média, a variância e a covariância, assim como as interpretações e significados apresentados, podem ser estendidos diretamente para as variáveis contínuas. Apesar disso, o leitor interessado pode refazer as provas das propriedades apresentadas nas seções anteriores sem maiores dificuldades, apenas substituindo o operador somatório pelo operador integral onde for cabível.

**Exemplo 1.11** - Suponha que uma distribuição de probabilidades no intervalo contínuo  $[0,1]$ , chamada de distribuição triangular e ilustrada na Figura 1.25, possa ser definida como

$$\wp(x) = \begin{cases} 4x, & 0 \leq x \leq 0.5 \\ 4 - 4x, & 0.5 \leq x \leq 1 \end{cases}$$

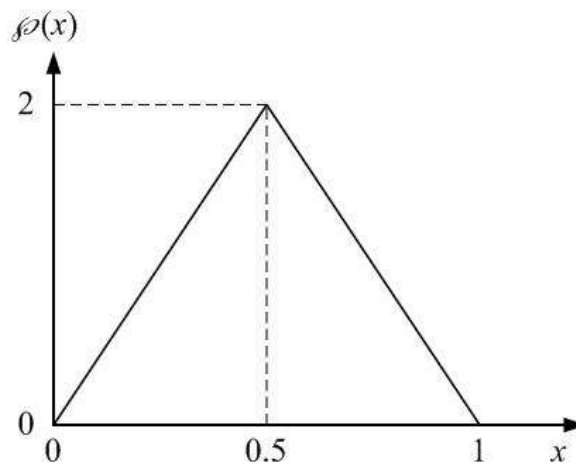


Figura 1.25 - A distribuição Triangular.

Para mostrar a Propriedade 1.13,

$$\int_0^1 \wp(x) dx = \int_0^{0.5} 4x dx + \int_{0.5}^1 (4 - 4x) dx = \frac{4x^2}{2} \Big|_0^{0.5} + 4x \Big|_{0.5}^1 - \frac{4x^2}{2} \Big|_{0.5}^1 = \frac{1}{2} + 2 - \left( \frac{4-1}{2} \right) = 1$$

confirmando que a função proposta é de fato uma densidade de probabilidade. Nesse caso, a curva de probabilidade acumulada ganha a forma

$$P_{AC}(x) = \begin{cases} 0, & x \leq 0 \\ 2x^2, & 0 \leq x \leq 0.5 \\ 0.5 + (4x - 2) - (2x^2 - 0.5), & 0.5 \leq x \leq 1 \\ 1, & x \geq 1 \end{cases}$$

Para obter a média,

$$\begin{aligned} \int_0^1 x\wp(x) dx &= \int_0^{0.5} 4x^2 dx + \int_{0.5}^1 (4x - 4x^2) dx = \frac{4x^3}{3} \Big|_0^{0.5} + \frac{4x^2}{2} \Big|_{0.5}^1 - \frac{4x^3}{3} \Big|_{0.5}^1 = \\ &= \frac{1}{6} + \left( \frac{4-1}{2} \right) - \left( \frac{4-0.5}{3} \right) = \frac{1}{6} + \frac{9}{6} - \frac{7}{6} = \frac{1}{2} \end{aligned}$$

Para obter o momento estatístico de ordem 2,

$$\begin{aligned} \mu_x^{(2)} = \int_0^1 x^2\wp(x) dx &= \int_0^{0.5} 4x^3 dx + \int_{0.5}^1 (4x^2 - 4x^3) dx = \frac{4x^4}{4} \Big|_0^{0.5} + \frac{4x^3}{3} \Big|_{0.5}^1 - \frac{4x^4}{4} \Big|_{0.5}^1 = \\ &= \frac{1}{16} + \left( \frac{4-0.5}{3} \right) - \left( \frac{4-0.25}{4} \right) = \frac{3}{48} + \frac{56}{48} - \frac{45}{48} = \frac{7}{24} \end{aligned}$$

Portanto, a variância é igual a

$$\sigma_x^2 = \mu_x^{(2)} - (\mu_x)^2 = \frac{7}{24} - \frac{1}{4} = \frac{1}{24}$$

## 1.8. Conclusões

No Capítulo 1 foram introduzidos os conceitos de aleatoriedade e determinismo, fundamentais para a compreensão de problemas de medição. Para caracterizar a componente aleatória das medidas, foi introduzido o conceito de probabilidade e de distribuição de probabilidades. Foi acentuado o fato de que uma probabilidade é uma expectativa de que um certo resultado ocorre, não garantindo de fato a consecução do resultado. Para tornar possível a comparação e o processo de tomada de decisão em diferentes problemas, caracterizados por diferentes distribuições de probabilidade, foram definidas a média e a variância. A primeira caracteriza um valor em torno do qual os resultados possíveis flutuam. A segunda caracteriza o quanto os resultados flutuam

em torno do valor médio. Finalmente, foi introduzido o conceito de independência entre medidas e variáveis e foi definida a covariância, que caracteriza o grau de dependência linear entre as variáveis analisadas.

Um problema fundamental que se põe é o de como caracterizar a distribuição de probabilidades que caracteriza um determinado problema estocástico. Outro problema fundamental é o de como utilizar essa informação para julgar e analisar medidas experimentais. Esses serão os tópicos principais abordados nos próximos capítulos.

### **1.9. Leitura Adicional**

A literatura dedicada à apresentação de pontos fundamentais relacionados aos conceitos de aleatoriedade, de independência de medidas e de probabilidades é muito vasta. Não cabe aqui, portanto, uma revisão dessa área. O leitor interessado encontrará centenas de livros que abordam esses assuntos em qualquer biblioteca dedicada à Matemática e à Engenharia.

Uma discussão muito interessante sobre a caracterização do grau de aleatoriedade em problemas físicos e matemáticos é apresentada em

*“What is Random? Chance and Order in Mathematics and Life.”*, E. Beltrami, Springer-Verlag, New York, 1999.

Um texto clássico relacionado ao uso e aplicação dos conceitos discutidos no Capítulo 1 em problemas de Engenharia é apresentado em

*“Process Analysis by Statistical Methods”*, D.M. Himmelblau, John Wiley & Sons, New York, 1970.

Uma discussão mais formal sobre as propriedades matemáticas associadas a distribuições de probabilidades é apresentada em

*“Probability and Statistical Inference. Volume 1: Probability”*, J.G. Kalbfleisch, Springer-Verlag, New York, 1985.

*“Probability and Statistics. Theory and Applications.”*, G. Blom, Springer-Verlag, New York, 1989.

### **1.10. Exercícios Sugeridos**

1- Defina os seguintes eventos como determinísticos ou estocásticos e justifique:

- a) Tempo de cozimento de um tijolo na olaria;
- b) Tempo de espera por um ônibus depois da chegada no ponto;
- c) Tempo da viagem do Rio a Salvador por via terrestre e por via aérea;
- d) Número de telhas necessárias para cobrir um telhado;
- e) Número de equipamentos que falham por ano em uma escola de informática;
- f) Condição do tempo daqui a exatamente dois meses.

- 2- Pegue uma folha de papel e rasgue uma tira com as mãos. Meça a largura dessa tira em diferentes pontos com uma régua milimetrada. Repita o experimento. As medidas obtidas são iguais? Você é capaz de identificar as fontes de erro desse experimento?
- 3- Uma função discreta muito utilizada para descrever a probabilidade de encontrar uma espécie de tamanho  $i$  em sistemas que crescem de forma não contínua (ou seja, em que há um mecanismo que interrompe o crescimento) é a chamada curva de Flory. A curva de Flory pode ser escrita na forma:

$$P_i = (1 - q)q^{i-1}$$

onde  $i$  ( $i = 1, 2, \dots, N, \dots$ ) é o comprimento,  $P_i$  é a probabilidade de se encontrar uma espécie de tamanho  $i$  e  $q$  é uma constante  $0 < q < 1$  que caracteriza o processo.

- a) Prove que  $P_i$  é de fato uma distribuição de probabilidades, provando que as Equações (1.5) e (1.6) são satisfeitas;
- b) Calcule o comprimento médio da população  $\mu_i$ ;
- c) Calcule a variância da população  $\sigma_i^2$ .
- 4- Para a distribuição exponencial,  $f(x) = \alpha \exp(-\alpha x)$ , definida no intervalo contínuo  $[0, \infty)$ :
- a) Calcule o valor de  $\alpha$ , para que  $f(x)$  seja de fato uma densidade de probabilidades;
- b) Calcule a probabilidade acumulada  $P_{AC}(x)$  no intervalo de definição do problema;
- c) Calcule o valor médio de  $x$ ;
- d) Calcule a variância de  $x$ ;
- e) Pense em quantos momentos estatísticos independentes podem ser definidos.
- 5- No laboratório é feita uma medida cromatográfica (separação dos vários componentes químicos de uma mistura) usando uma coluna de separação (um tubo oco) recheada com um composto plástico poroso. Toda vez que um composto ácido é usado na coluna, parte do recheio plástico é corroído e, dessa forma, extraído da coluna. Sabendo que o composto plástico poroso é que de fato promove a separação dos componentes da mistura, as medidas de composição feitas na coluna poderiam ser consideradas independentes? Por quê?
- 6- Suponha que duas variáveis  $x$  e  $y$  estão relacionadas na forma  $y = 4x(1 - x)$ , definida no intervalo contínuo  $[0, 1]$ . Suponha ainda que  $f(x) = 1$  no intervalo de definição do problema.
- a) Mostre que  $f(x)$  define de fato uma distribuição de probabilidades;
- b) Calcule  $f(y)$ ;
- c) Calcule  $f(y/x)$ ;
- d) Calcule  $\text{Covar}(x, y)$  e  $\rho_{xy}$ ;
- e) Comente o significado dos resultados obtidos no item anterior.