

# Capítulo 2

## Distribuições de Probabilidade

No Capítulo 1 foram introduzidos os conceitos de aleatoriedade e probabilidade. Foi mostrado que um fenômeno aleatório pode ser descrito em termos de distribuições de probabilidade. No entanto, foi também mostrado que o conceito de probabilidade (e, portanto, de distribuições de probabilidade) está intimamente relacionado com a possibilidade de repetir um experimento infinitas vezes. Essa possibilidade, contudo, não é factível, pois nenhum experimento pode ser repetido infinitas vezes por limitações reais de tempo e custo. Sendo assim, as distribuições de probabilidade usadas para descrever problemas reais não podem ser obtidas unicamente por métodos experimentais. Na realidade, as distribuições de probabilidade usadas para descrever problemas reais estão fortemente baseadas em considerações idealizadas sobre o sistema estudado e estão fundamentadas em hipóteses a respeito do comportamento do sistema considerado. São necessários, portanto, modelos idealizados de comportamento ou **modelos probabilísticos**.

Para exemplificar, consideremos o caso do dado ideal analisado no Exemplo 1.2. Para dizermos que o histograma de um dado comporta seis resultados possíveis, e que cada um dos resultados tem a mesma probabilidade de ocorrência, é necessário admitir que o dado é composto por seis faces absolutamente iguais. Repare que há ao menos duas hipóteses fortes embutidas na última afirmação. A primeira admite implicitamente que o dado é um cubo, o que não é um fato óbvio. Por exemplo, por que não poderíamos construir um dado com 4 faces (uma pirâmide) ou 8 faces (um octaedro)? Obviamente que essas duas alternativas de construção são possíveis. Um dado não precisa ter necessariamente seis faces. No entanto, usamos o dado de seis faces com tamanha frequência que associamos o dado ao cubo imediatamente. A segunda hipótese é a de que as faces são absolutamente iguais. Será que as faces de um dado real, construídas com ferramentas e materiais reais (não ideais, não necessariamente homogêneos, sujeitos a falhas, etc.), são de fato iguais? Não há como resolver essa questão facilmente, sem experimentação, mas admitimos para fins práticos quase sempre que as faces são iguais. Portanto, o Exemplo 1.2 ilustra a construção de um modelo probabilístico. Em outras palavras, o histograma do Exemplo 1.2 é de fato um modelo probabilístico em que confiamos e não necessariamente uma realidade experimental.

É muito interessante mostrar como algumas distribuições de probabilidade surgem naturalmente em problemas de Matemática e Engenharia. Por isso, nesta seção são construídos alguns modelos probabilísticos usados comumente para análise de problemas reais, enfatizando-se em particular as hipóteses que permitem a construção desses modelos. São também mostrados alguns exemplos típicos de aplicação dos

modelos de distribuição de probabilidade na prática da análise de dados e interpretação de experimentos aleatórios. Deve ficar claro que o objetivo perseguido aqui não é fazer uma descrição extensiva de modelos probabilísticos descritos na literatura. Uma descrição mais extensa e profunda de modelos probabilísticos, assim como a maior parte das provas de teoremas e resultados, pode ser encontrada nas referências adicionais apresentadas no final deste capítulo. O objetivo perseguido nessa seção é a apresentação e análise de alguns modelos probabilísticos que serão muito úteis para as discussões apresentadas nos próximos capítulos.

## 2.1. A Distribuição Binomial

Admita que em um evento aleatório apenas dois resultados sejam possíveis. Por exemplo, sucesso ou insucesso, cara ou coroa, acima ou abaixo, certo ou errado, sim ou não, etc. Admita ainda que as probabilidades de que os resultados ocorram sejam conhecidas e iguais respectivamente a  $p$  e a  $q = 1 - p$ . A distribuição binomial procura descrever o seguinte problema: após  $m$  repetições do experimento, qual a probabilidade de que se obtenham  $n$  sucessos? Por exemplo, após jogar a moeda para o alto  $m$  vezes, qual a probabilidade de tirar cara ( $p = 0.5$ ;  $q = 0.5$ )  $n$  vezes? Como será mostrado adiante, esse problema aparece com extrema frequência durante a análise de dados reais.

Para que se possa construir a distribuição binomial, é necessário primeiro reconhecer que são possíveis  $NR = m + 1$  resultados distintos, já que o número de sucessos pode variar de zero até  $m$ . Suponha que  $n$ ,  $0 \leq n \leq m$ , é o número particular de sucessos avaliado no momento. Então parece claro que o número de insucessos é igual a  $(m - n)$ . Nesse caso, se os experimentos são realizados de forma independente, a probabilidade de que um certo arranjo de  $n$  sucessos e  $(m - n)$  insucessos seja obtido é igual a

$$P_{A1} = P(S)_1 \dots P(S)_n P(I)_{n+1} \dots P(I)_{n+(m-n)} = p^n q^{m-n} \quad (2.1)$$

onde  $P_{A1}$  é a probabilidade do arranjo particular considerado,  $P(S)_i$  é a probabilidade de sucesso no experimento  $i$  e  $P(I)_i$  é a probabilidade de insucesso no experimento  $i$ . Obviamente, a Equação (2.1) descreve apenas o resultado de um dos possíveis arranjos, em que sucesso é obtido nos  $n$  primeiros experimentos e insucesso é obtido nos últimos  $(m - n)$  experimentos. Suponha que esse arranjo é invertido, de maneira que os primeiros resultados são insucessos, enquanto os últimos resultados são sucessos. Nesse caso,

$$P_{A2} = P(I)_1 \dots P(I)_{(m-n)} P(S)_{(m-n)+1} \dots P(S)_{(m-n)+n} = p^n q^{m-n} \quad (2.2)$$

cujo resultado é idêntico ao anterior. Na realidade, a probabilidade de se obter  $n$  sucessos independe da ordem com que os  $n$  experimentos bem sucedidos sejam distribuídos no arranjo final de resultados. Portanto, a probabilidade de se obter  $n$  sucessos de  $m$  experimentos pode ser escrita na forma

$$P(n; m, p) = N_A p^n q^{m-n} \quad (2.3)$$

onde  $N_A$  é o número total de combinações possíveis de  $n$  sucessos e  $(m - n)$  insucessos em  $m$  experimentos. O número  $N_A$  é uma operação clássica da matemática combinatorial, denominado como a combinação de  $m$ ,  $n$  a  $n$ , dado na forma

$$N_A = \binom{m}{n} = C_n^m = \frac{m!}{n!(m-n)!} \quad (2.4)$$

Dessa maneira, a solução do problema proposto, denominada de distribuição binomial, pode ser representada na forma:

$$\text{Bin}(n; m, p) = \frac{m!}{n!(m-n)!} p^n q^{m-n} \quad (2.5)$$

A distribuição binomial é uma distribuição discreta univariada e bi-paramétrica, pois descreve a variação de probabilidades de uma única variável discreta,  $n$ , e depende de dois parâmetros,  $m$  e  $p$ . Isso significa que apenas dois dos momentos da curva de distribuição podem ser fixados independentemente pelo usuário, ficando os demais automaticamente definidos pela forma da curva da Equação (2.5). Além disso, não é difícil mostrar que

$$\mu_N = mp \quad (2.6)$$

$$\sigma_N^2 = mpq \quad (2.7)$$

Desta forma, os dois parâmetros  $m$  e  $p$  que caracterizam a distribuição binomial podem ser relacionados fácil e diretamente aos parâmetros  $\mu_N$  e  $\sigma_N^2$ , que caracterizam o posicionamento e o espalhamento da curva de distribuição de probabilidades. Isso pode ser muito útil para interpretação de alguns problemas práticos, como o dimensionamento do tamanho das amostras em procedimentos de avaliação e controle de qualidade, analisados a seguir. É também interessante observar que o índice de polidispersão tende a 1 quando  $m$  cresce, independentemente dos valores de  $p$  e  $q$ . Isso indica que o grau de espalhamento em relação à média decresce continuamente com o aumento do conjunto amostral.

**Exemplo 2.1** - Para o caso da moeda ideal, suponha que se deseja saber como a distribuição de probabilidades de se obter  $n$  caras ( $p = 0.5$ ) varia, à medida que  $m$  aumenta. A Figura 2.1 mostra a evolução da curva binomial para diferentes valores de  $m$ .

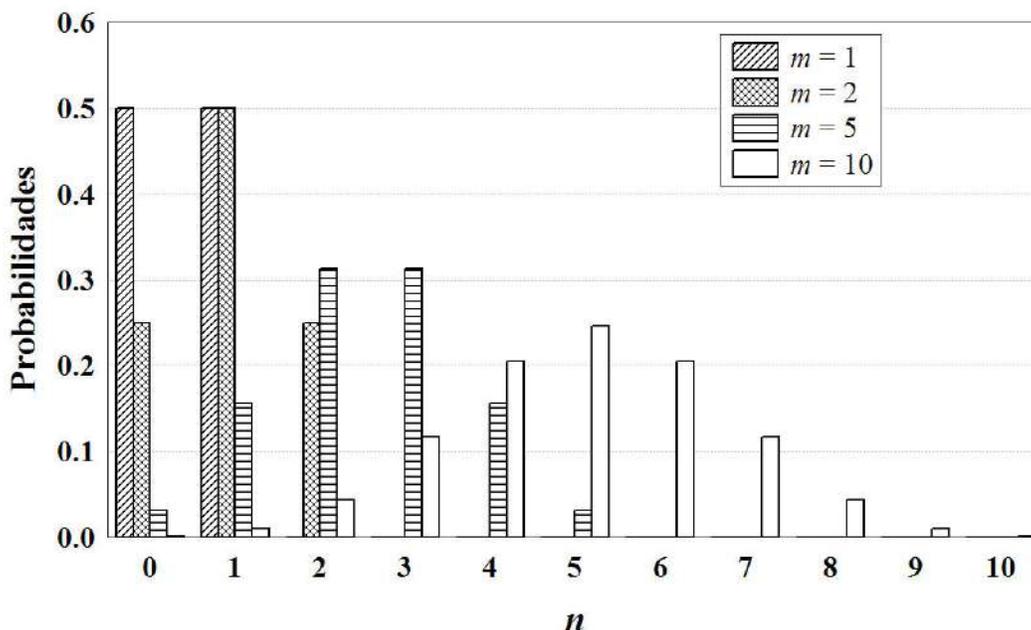


Figura 2.1 - Exemplo da curva binomial:  $Bin(n; m, 0.5)$ .

Observe que a curva binomial vai ganhando a aparência de um sino, à medida que o número de experimentos aumenta. Nesse caso particular, observe ainda que a média  $\mu_N = mp = m/2$ , o que mostra que na média o número de caras deverá ser igual à metade do número de experimentos realizados (embora se saiba que isso não significa que esse é o resultado que necessariamente deverá ocorrer, como mostra a Figura 2.1 e como discutido nas seções anteriores).

**Exemplo 2.2** - Pesquisadores envolvidos com atividades experimentais sabem que nem sempre experimentos podem ser realizados a contento por uma série de fatores: falta de energia, quebra de equipamentos, contaminação de reagentes, etc. Muitas vezes os resultados experimentais têm que ser descartados porque as falhas ocorrem durante a realização dos experimentos ou são detectadas depois de terminada a atividade experimental.

Suponha que a probabilidade de um certo experimento dar certo é de 70%. A experiência acumulada indica que esse é um número bastante representativo da realidade laboratorial, embora obviamente possa variar de caso para caso. Por exemplo, durante a instalação de uma nova unidade experimental e/ou na presença de deficiências de infra-estrutura, a probabilidade de sucesso pode diminuir drasticamente. Suponha ainda que o número de bons experimentos necessários para que se possa desenvolver uma boa análise científica seja de pelo menos 20 experimentos. (Questões relacionadas ao número de experimentos necessários para desenvolver um trabalho de investigação serão analisadas no Volume II dessa série de publicações.) Nesse caso, qual o número de experimentos que devem ser programados para que se garanta com 95% de confiança que ao menos 20 bons experimentos serão obtidos?

Primeiramente deve ser observado que esse é um problema típico de aplicação da curva binomial. Observe que apenas dois resultados são possíveis: certo ou errado.

Observe ainda que as probabilidades individuais de cada resultado são conhecidas:  $p = 0.70$  e  $q = 0.30$ . Assim, a pergunta original pode ser formulada da seguinte forma: dados  $p = 0.70$  e  $q = 0.30$ , qual é o número de experimentos,  $m$ , que faz com que a probabilidade de se obter um valor de  $n$  menor do que 20 é inferior a 5%? Ou ainda, qual o menor valor de  $m$  para o qual

$$\sum_{n=0}^{19} \frac{m!}{n!(m-n)!} p^n q^{m-n} = P_{AC}(19; m, p) \leq 0.05$$

$$\sum_{n=20}^m \frac{m!}{n!(m-n)!} p^n q^{m-n} \geq 0.95$$

O resultado pode ser obtido avaliando-se valores crescentes de  $m$  a partir de  $m = 20$ , uma vez que 20 experimentos bem sucedidos não podem ser obtidos com número inferior de experimentos, como mostrado na Figura 2.2. Para o caso estudado, o valor mínimo de experimentos executados para garantir com 95% de certeza que 20 bons experimentos serão obtidos é igual a 35. Vê-se, portanto, que a programação experimental deve prever a realização de número de experimentos bastante superior ao desejado, para que sejam considerados os problemas eventuais que levam ao descarte dos resultados experimentais. Se a probabilidade de sucesso cai a 60%, o valor mínimo de experimentos necessários para garantir 20 bons resultados com 95% de certeza sobe para 42, caindo respectivamente para 29 e 25 quando a probabilidade de sucesso sobe para 80% e 90%. Se a programação experimental não incluir a possibilidade eventual de falhas, é bastante provável que o cronograma e o orçamento originais não sejam obedecidos, causando frustração.

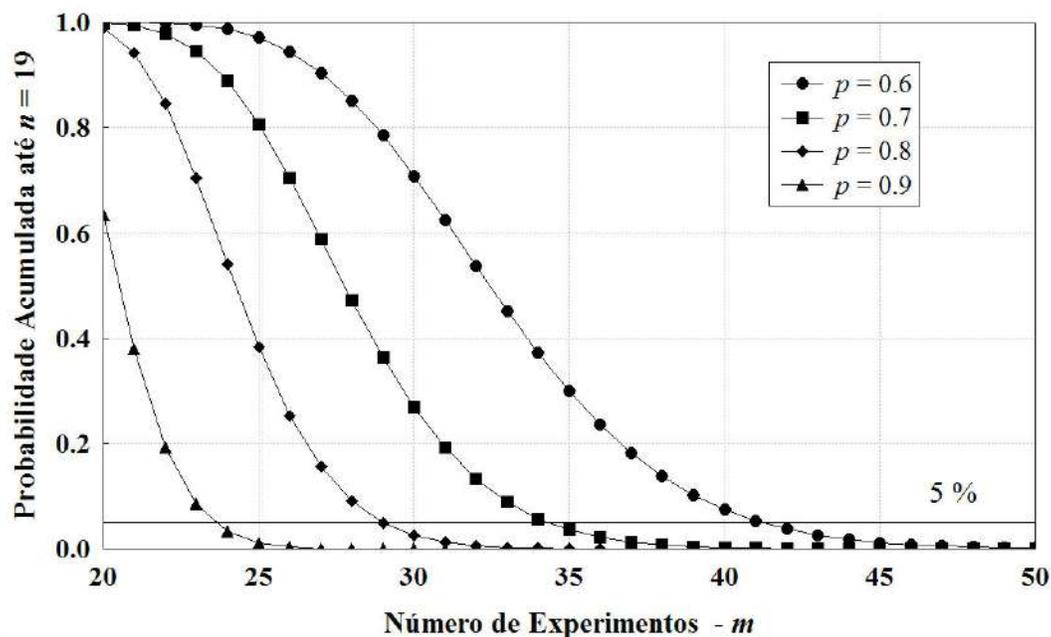
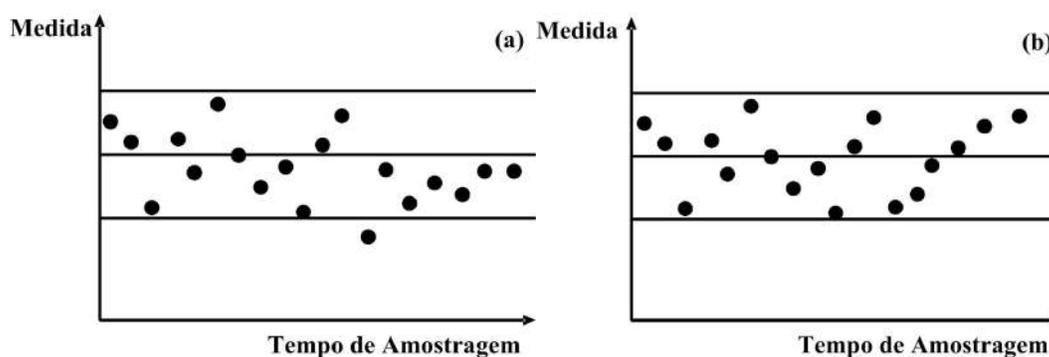


Figura 2.2 - Probabilidade acumulada da curva binomial até  $n = 19$ , para  $m$  e  $p$  variáveis.

**Exemplo 2.3** - A distribuição binomial é uma poderosa ferramenta usada para fins de controle de qualidade e análise de dados, como discutido no Volume III dessa série de publicações. Nesse caso, a distribuição binomial é usada para identificar padrões de formação entre pontos experimentais obtidos diretamente do processo.

A Figura 2.3 ilustra dois tipos clássicos de aplicação da curva binomial em problemas de controle de qualidade. Em ambos os casos, a reta central representa um valor médio histórico do processo, enquanto as duas retas adicionais representam a faixa histórica que concentra 98% dos resultados obtidos no processo. Os pontos se distribuem de forma supostamente aleatória em torno da média, com probabilidades iguais a 50% de estarem acima ou abaixo do valor médio. (No Capítulo 3 deste volume e no Volume III dessa série de publicações serão construídos métodos que permitem a definição formal das faixas consideradas na Figura 2.3.)



**Figura 2.3** - Padrões de medidas obtidas durante o monitoramento do processo.

A intervenção no processo para corrigir eventuais falhas de produção, embora desejada, não deve ser feita de forma exagerada por vários razões. Por exemplo, toda vez que se perturba o processo são introduzidos transientes ou efeitos de longo prazo que podem obrigar o operador a perturbar o processo novamente no futuro muitas vezes. Além disso, a modificação freqüente das condições de operação acelera o envelhecimento de válvulas e demais equipamentos, aumentando os custos de produção. E não deve ser esquecido o fato de que as medidas experimentais contêm erros. Dessa forma, o operador pode ser induzido a corrigir desvios que de fato não existem, introduzindo assim alterações aleatórias e complexas na cadeia de produção. Por todas essas razões parece lícito perguntar quando é de fato o momento de intervir no processo.

Um primeiro procedimento de controle pode ser projetado a partir da Figura 2.3a. Por segurança, toda vez que uma medida do processo cair fora dos limites estabelecidos pela carta de controle de qualidade, medidas de correção devem ser imediatamente tomadas. Esse procedimento está embasado no fato de que esses desvios aleatórios são observados muito pouco freqüentemente (no caso analisado, apenas 2% das vezes, ou 1 em cada 50 pontos), não sendo possível descartar a hipótese de falha do processo.

Um segundo procedimento de controle pode também ser projetado a partir da Figura 2.3a. O operador desconfia de que algo anda errado com o processo porque os últimos sete valores medidos se encontram abaixo da média. Nesse caso, dado um conjunto de  $m$  pontos, qual a probabilidade de que ao menos  $n$  pontos estejam do

mesmo lado da carta de controle, se a probabilidade do ponto estar de um lado ou de outro é a mesma e igual a  $p = 0.5$ ? Esse problema pode ser resolvido no contexto da curva binomial na forma

$$\sum_{i=0}^n \frac{m!}{i!(m-i)!} \left(\frac{1}{2}\right)^i \left(\frac{1}{2}\right)^{m-i} = P_{AC}(n; m, 0.5)$$

já que apenas dois resultados são possíveis (estar acima ou abaixo da linha central) e as probabilidades são conhecidas e iguais a 50%. Supondo que um certo grau de confiança  $P_{X\%}$  é exigido para que se tome a decisão de intervir no processo, toda vez que um padrão for observado e tiver probabilidade inferior a  $P_{X\%}$  de ocorrer, toma-se a decisão de introduzir uma perturbação reguladora de controle. Nesse caso, como o mesmo padrão pode ocorrer de um lado ou de outro da carta de controle (ou seja, a curva de distribuição de probabilidades é simétrica), o problema de controle fica na forma

$$\sum_{i=0}^n \frac{m!}{i!(m-i)!} \left(\frac{1}{2}\right)^i \left(\frac{1}{2}\right)^{m-i} = P_{AC}(n; m, 0.5) \leq \left(\frac{100\% - P_{X\%}}{2}\right)$$

De acordo com a Figura 2.4, admitindo um grau de confiança de 98%, observa-se que padrões pouco prováveis são os seguintes:

- a) seqüências com sete ou mais pontos seguidos do mesmo lado do diagrama;
- b) seqüências de dez ou mais pontos, com apenas um ponto de um lado do diagrama;
- c) seqüências de 14 ou mais pontos, com apenas dois pontos de um mesmo lado do diagrama;
- d) seqüências de 17 ou mais pontos, com apenas três pontos de um mesmo lado do diagrama.

Observe também que o grau de confiança é um importante parâmetro de sintonia do problema de controle. Se o grau de confiança for trocado para 95%, os padrões pouco prováveis são os seguintes:

- a) seqüências com seis ou mais pontos seguidos do mesmo lado do diagrama;
- b) seqüências de nove ou mais pontos, com apenas um ponto de um lado do diagrama;
- c) seqüências de 12 ou mais pontos, com apenas dois pontos de um mesmo lado do diagrama;
- d) seqüências de 15 ou mais pontos, com apenas três pontos de um mesmo lado do diagrama.

Quanto menor o grau de confiança exigido, mais freqüente é a intervenção no processo, aumentando o risco de que se tomem medidas de controle desnecessárias. Por sua vez, quanto maior o grau de confiança exigido, mais se demora para que se faça uma intervenção no processo, aumentando o risco de se corrigir tardiamente um desvio real do processo. Por isso, a definição do grau de confiança depende do processo, dos objetivos perseguidos e do custo que o engenheiro está disposto a pagar por uma decisão eventualmente equivocada. Para a maior parte das aplicações, os níveis de confiança mais amplamente utilizados são os níveis de 95% e 98%. Em ambos os casos, de acordo com a Figura 2.3a, deve-se iniciar um procedimento de correção do processo, já que o valor medido encontra-se improvavelmente baixo (sete valores seguidos abaixo da linha central).

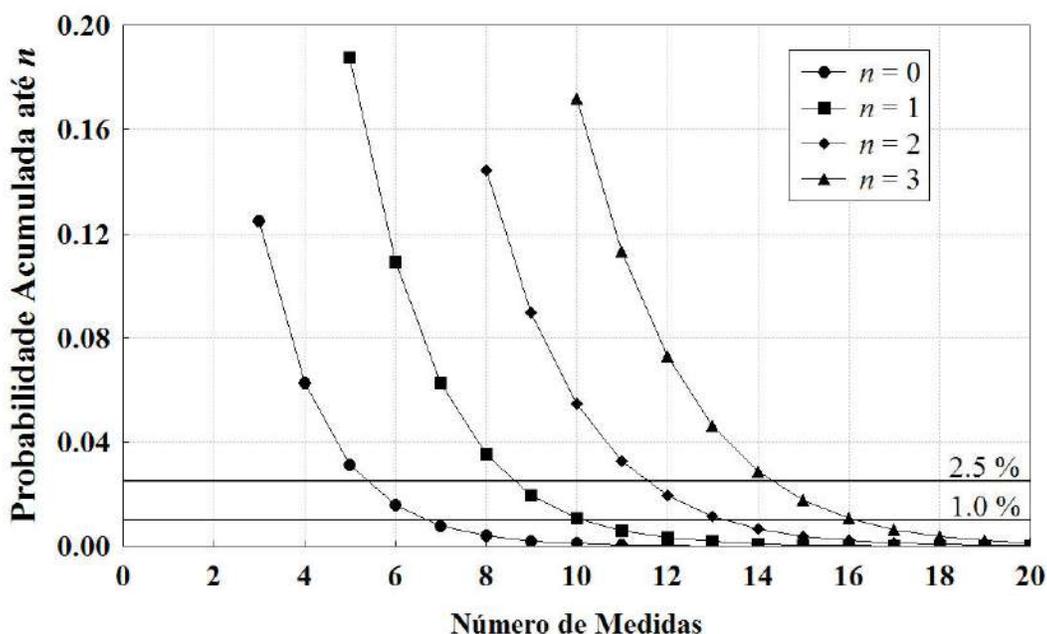


Figura 2.4 - Probabilidade de se obter ao menos  $n$  pontos do mesmo lado do diagrama, à medida que o número de medidas aumenta.

Um terceiro procedimento de controle pode também ser projetado a partir da Figura 2.3b. O operador desconfia de que algo anda errado com o processo porque os últimos seis valores medidos se encontram consistentemente subindo. Nesse caso, dado um conjunto de  $m$  pontos, qual a probabilidade de que eles estejam alinhados de forma consistente na mesma direção da carta de controle, admitindo que a probabilidade do ponto estar alinhado para um lado ou para o outro é a mesma e igual a  $p = 0.5$ ? (Repare que a probabilidade de alinhamento deve ser, na realidade, dependente da posição em que o ponto se encontra no diagrama. Por exemplo, supondo uma distribuição aleatória dos pontos, a probabilidade de que um segundo ponto esteja acima do primeiro é muito maior quando este primeiro se encontra na extremidade inferior da carta. Por isso, a hipótese de probabilidade constante e igual a  $p = 0.5$  deve ser vista como uma aproximação do problema real e tomada com cautela.) Esse problema pode ser resolvido novamente no contexto da curva binomial, de forma semelhante à realizada anteriormente. Deve-se atentar apenas para o fato de que quando  $m$  pontos estão alinhados em uma certa direção,  $m-1$  pontos caracterizam de fato o experimento, já que o primeiro ponto da série é a referência que não está sendo comparada a ninguém. Assim, dizer que **dois** pontos estão alinhados em forma crescente é o mesmo que dizer que **um** experimento resultou em um valor maior que o valor anterior.

Assim, admitindo um grau de confiança de 98% e em conformidade com a discussão anterior, observa-se que padrões pouco prováveis são os seguintes:

- a) seqüências com oito ou mais pontos seguidos alinhados na mesma direção do diagrama;
- b) seqüências com onze ou mais pontos alinhados na mesma direção do diagrama, com apenas uma falha na direção contrária;

- c) seqüências com quinze ou mais pontos alinhados na mesma direção do diagrama, com apenas duas falhas na direção contrária;
- d) seqüências com dezoito ou mais pontos alinhados na mesma direção do diagrama, com apenas três falhas na direção contrária.

Como no caso anterior, se o grau de confiança for trocado para 95%, os padrões pouco prováveis são os seguintes:

- a) seqüências com sete ou mais pontos seguidos alinhados na mesma direção do diagrama;
- b) seqüências com dez ou mais pontos alinhados na mesma direção do diagrama, com apenas uma falha na direção contrária;
- c) seqüências com treze ou mais pontos alinhados na mesma direção do diagrama, com apenas duas falhas na direção contrária;
- d) seqüências com dezesseis ou mais pontos alinhados na mesma direção do diagrama, com apenas três falhas na direção contrária.

No caso da Figura 2.3b, o operador deveria esperar ao menos mais um resultado, antes de intervir no processo, pois o padrão observado (seis pontos alinhados de forma crescente) ainda não é um padrão improvável, nos limites de confiança de 95% e 98%.

**Exemplo 2.4** - Um problema clássico de uso da distribuição binomial é a avaliação da qualidade de conjuntos de resultados (lotes). Suponha que a cada um dos elementos de um conjunto possa ser atribuído um dos seguintes atributos: “adequado” ou “inadequado”. Suponha ainda que a qualidade do conjunto (produção) possa ser avaliada a partir da fração observada de elementos com o atributo “adequado” em uma amostra de tamanho  $m$ . Admita que a fração mínima de elementos “adequados” do conjunto todo deva ser igual a  $p$ . Qual deve ser o tamanho  $m$  do conjunto amostral para que a observação de  $d$  peças defeituosas não seja suficiente para condenar a produção com grau de confiança especificado e igual a  $P_{X\%}$ ?

Esse problema é semelhante ao problema analisado no Exemplo 2.2. Fundamentalmente, a solução do problema é dada pela expressão

$$\sum_{n=0}^d \frac{m!}{n!(m-n)!} (1-p)^n p^{m-n} = P_{AC}(d; m, 1-p) \leq 1 - P_{X\%}$$

Nesse problema, o grau de confiança  $P_{X\%}$  é especificado pelo projetista e reflete o grau de conservadorismo ou rigor do analista.  $d$  é um número máximo admissível de observações “inadequadas” no lote e usualmente reflete a dificuldade da análise. Quando a análise de cada elemento é fácil,  $d$  pode ser grande, para permitir uma tomada de decisão mais robusta. Não chega a ser absurdo, no entanto, que  $d$  seja considerado igual a 0 ou 1, como nos casos em que a análise do atributo é muito custosa em termos econômicos e/ou de tempo.  $p$  reflete uma expectativa de qualidade e, num ambiente de produção, costuma ser especificado pelo cliente.

**Tabela 2.1** - Tamanho dos lotes como função da especificação das análises para 95% de confiança.

$p$	$d$	$m$
0.10	0	29
	1	46
	2	61
	3	75
	4	89
	5	103
0.05	0	59
	1	93
	2	124
	3	153
	4	181
	5	209

Repare na Tabela 2.1 que os lotes têm que ser muito maiores que os valores médios sugeridos pelo número de defeitos observados. Para aprovar um lote produzido, com 95% de confiança, ao menos 29 peças têm que ser analisadas e nenhum defeito pode ser detectado, para que se possa garantir que a fração de defeitos é inferior a 10%. O tamanho do lote sobe para 59 peças, sem quaisquer defeitos observados, para que seja possível garantir que a fração de defeitos é inferior a 5%. Isso mostra como uma boa precisão pode requerer a análise de número bastante grande de experimentos.

## 2.2. A Distribuição de Poisson

Admita que num certo problema a probabilidade de sucesso (ou insucesso) seja muito pequena. Esse é o caso típico, por exemplo, numa linha de produção em que a maioria esmagadora dos produtos fabricados é muito boa e as falhas ocorrem apenas eventualmente. Nesse caso, o uso da distribuição binomial pode ser muito inconveniente, porque nos obriga a trabalhar com longas somas ( $m$  tem que ser muito grande para resultar em valores de  $n$  mensuráveis) de valores muito pequenos (por causa dos baixos valores de  $p$  ou de  $q$ ). Portanto, é conveniente avaliar o que acontece com a distribuição binomial no limite em que  $p$  vai a zero,  $m$  vai a infinito, mas  $\mu_N = mp$  permanece constante. Para tanto, vale a pena reescrever a curva binomial da Equação (2.5) na forma:

$$\text{Bin}(n; m, p) = \frac{m(m-1)\dots(m-n+1)}{n!} p^n q^{m-n} \quad (2.8)$$

Multiplicando o numerador e o denominador por  $m^n$  lembrando que  $\mu_N = mp$ , chega-se a:

$$\begin{aligned}
 \text{Bin}(n; m, p) &= \frac{m(m-1)\dots(m-n+1)}{m^n n!} (mp)^n q^{m-n} = \\
 &= \frac{m(m-1)\dots(m-n+1)}{m^n} \frac{\mu_N^n}{n!} (1-p)^{m-n} = \\
 &= \left(1 - \frac{1}{m}\right) \left(1 - \frac{2}{m}\right) \dots \left(1 - \frac{n-1}{m}\right) \frac{\mu_N^n}{n!} (1-p)^{m-n} = \\
 &= \frac{\left(1 - \frac{1}{m}\right) \left(1 - \frac{2}{m}\right) \dots \left(1 - \frac{n-1}{m}\right)}{(1-p)^n} \frac{\mu_N^n}{n!} (1-p)^m
 \end{aligned} \tag{2.9}$$

Usando agora as seguintes relações matemáticas

$$(1-p)^m = \left[ (1-p)^{-1/p} \right]^{-mp} = \left[ (1-p)^{-1/p} \right]^{-\mu_N} \tag{2.10}$$

$$\lim_{z \rightarrow 0} \left[ (1+z)^{1/z} \right] = e \tag{2.11}$$

então

$$\lim_{p \rightarrow 0} \left[ (1-p)^{-1/p} \right]^{-\mu_N} = e^{-\mu_N} \tag{2.12}$$

Como

$$\lim_{\substack{m \rightarrow \infty \\ p \rightarrow 0}} \left[ \frac{\left(1 - \frac{1}{m}\right) \left(1 - \frac{2}{m}\right) \dots \left(1 - \frac{n-1}{m}\right)}{(1-p)^n} \right] = 1 \tag{2.13}$$

a Equação (2.9) fica na forma

$$\lim_{m \rightarrow \infty} \left[ \text{Bin} \left( n; m, \frac{\mu_N}{m} \right) \right] = \frac{\mu_N^n}{n!} e^{-\mu_N} = \text{Poisson}(n; \mu_N) \tag{2.14}$$

Portanto, a Equação (2.14) mostra que a distribuição de Poisson é o limite da distribuição binomial para valores muito pequenos de  $p$  (ou  $q$ ) e número muito grande de experimentos. A distribuição de Poisson é uma distribuição discreta univariada e uniparamétrica, pois descreve a variação de probabilidades de uma única variável discreta,  $n$ , e depende de um único parâmetro,  $\mu_N$ . Isso significa que apenas um dos momentos da curva de distribuição pode ser fixado independentemente pelo usuário, ficando os demais automaticamente definidos pela forma da curva da Equação (2.14). Nesse caso muito particular, o parâmetro pode ser confundido com o próprio valor médio da distribuição, o que é muito conveniente do ponto de vista prático. Além disso, não é difícil mostrar que

$$\sigma_N^2 = \mu_N \tag{2.15}$$

mostrando que não é possível fixar de forma independente os valores da média e da variância da distribuição de Poisson. Além disso, como no caso da curva binomial, o índice de polidispersão diminui continuamente com o aumento de  $\mu$ , indicando redução continuada do grau de espalhamento relativo à medida que o valor médio se desloca para valores mais altos.

**Exemplo 2.5** - Para o caso em que  $\mu_N = 10$ , a Figura 2.5 mostra como a curva binomial converge em direção à curva de Poisson, à medida que  $m$  cresce e  $p$  decresce. Vê-se que a convergência é muito rápida.

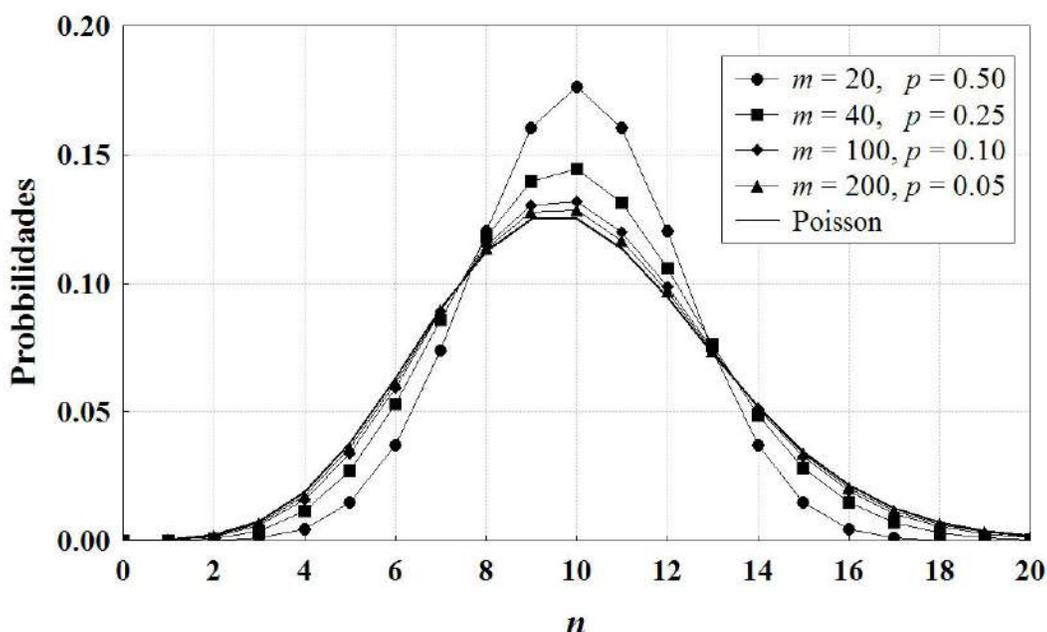


Figura 2.5 - Comparação entre as distribuições Bin( $n; m, p$ ) e Poisson( $n; 10$ ).

**Exemplo 2.6** - A curva de Poisson é muito utilizada como ferramenta de controle de processos, quando a probabilidade de se detectar falhas é pequena. Por exemplo, suponha que em um processo de manufatura qualquer é aceitável encontrar 2% de peças defeituosas. (Esse limite, obviamente, caracteriza a economia do processo e o grau de exigência imposto pelo cliente, não devendo ser tomado como valor absoluto de referência.) Nesse caso, se caixas contendo 200 peças são encaminhadas ao cliente, qual a probabilidade de se encontrar  $n$  ou menos peças defeituosas na caixa?

Para resolver esse problema, é conveniente expressar o problema de forma matemática precisa. Deseja-se saber o valor de

$$P_{AC}(n) = \sum_{i=0}^n \text{Poisson}(i; 4)$$

O uso da distribuição de Poisson é justificado por causa do problema tipicamente binomial (apenas dois resultados são possíveis: peça perfeita ou peça defeituosa) e do baixo valor da probabilidade e respectivo alto valor de dados analisados ( $m = 200, p = 0.02, \mu_N = 4$ ). A soma inclui o fato de que um número de defeitos em até  $n$  peças está sendo considerado.

A Figura 2.6 mostra como  $P_{AC}(n)$  varia com  $n$ , para diferentes valores de  $m$  (lembrando que no problema analisado  $\mu_N = 0.02 m$ ). Observe que os limites de 95% e 98% impostos definem as quantidades prováveis de defeitos encontradas em lotes de tamanhos variáveis. Assim, para lotes de 25 peças, encontrar duas peças defeituosas já condena a produção. De forma similar, encontrar três ou cinco peças defeituosas em lotes de tamanho 50 ou 100 respectivamente também condena a produção. Para lotes de tamanho 200, 8 peças defeituosas condenam o processo no limite de 95%, enquanto 9 peças condenam o processo no limite de 98%. Se for tomado como procedimento heurístico a recomendação de nunca condenar o processo sem a detecção de pelo menos 5 peças defeituosas, vê-se que o tamanho do lote amostral nunca deve ser inferior a 100 peças.

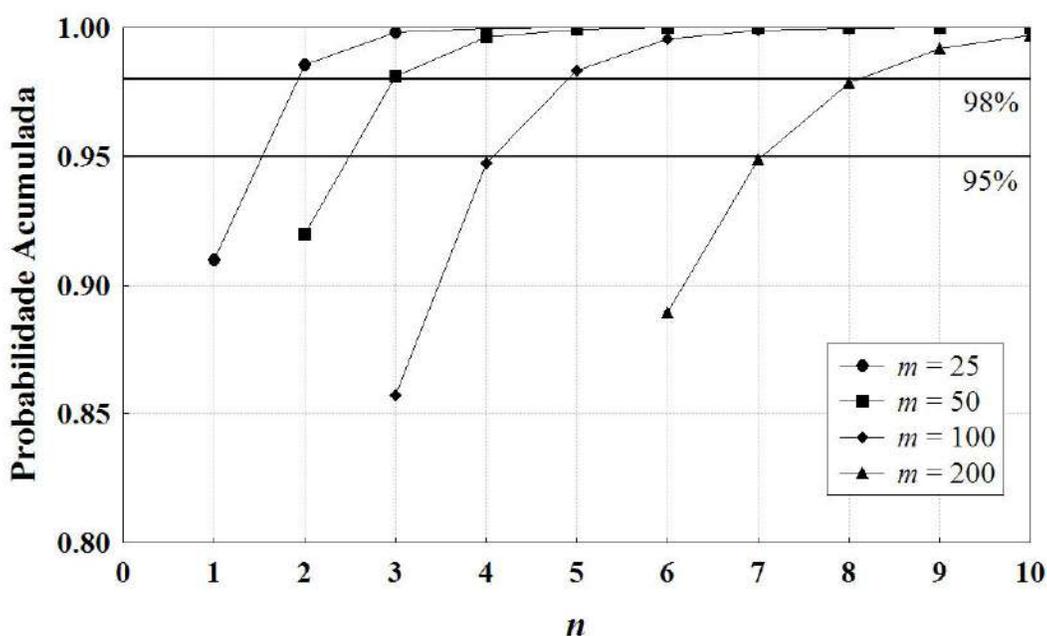
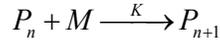


Figura 2.6 - Probabilidade acumulada da distribuição de Poisson( $n; 0.02 m$ ).

**Exemplo 2.7** - A distribuição de Poisson também aparece naturalmente em vários problemas de interesse da engenharia. Um dos problemas clássicos que levam ao aparecimento da distribuição de Poisson é o problema do crescimento de aglomerados, como na formação de cadeias moleculares longas (polimerização), de agregados bacterianos ou de estruturas coloidais.

Suponha que em um meio diluído, com composições aproximadamente constantes, ocorre uma transformação na forma:



onde  $P_n$  é a espécie que cresce,  $M$  é a unidade fundamental de formação do aglomerado e  $K$  é uma constante de velocidade, que diz quão rapidamente a transformação pode ocorrer. Nesse caso, a espécie  $P_n$  é formada de acordo com a seguinte equação:

$$\frac{dP_n}{dt} = KMP_{n-1} - KMP_n$$

onde o termo diferencial representa o acúmulo da espécie de tamanho  $n$ , o primeiro termo do lado direito representa a velocidade com que a espécie de tamanho  $n$  é formada a partir da espécie de tamanho  $(n - 1)$  e o segundo termo do lado direito representa a velocidade com que a espécie de tamanho  $n$  é consumida para formar a espécie de tamanho  $(n + 1)$ .

Para resolver a equação de balanço formada, é preciso reconhecer primeiro que a espécie de tamanho 1 não pode ser formada a partir de nenhuma outra espécie. Nesse caso,

$$\frac{dP_1}{dt} = -KMP_1$$

Além disso, é preciso fornecer as condições de contorno do problema (nesse caso, condições iniciais). Normalmente, em problemas práticos fornece-se uma quantidade conhecida da espécie fundamental (número de bactérias, espécies químicas que promovem a reação, etc.) e observa-se a evolução das quantidades das demais espécies. Assim

$$P_1(0) = P_{10}, P_n(0) = 0, n \geq 2$$

Para resolver as equações de balanço é conveniente dividi-las pelo produto  $(KMP_{10})$ , de maneira que as equações fiquem na forma mais simples

$$\frac{d\left(\frac{P_n}{P_{10}}\right)}{d(KMt)} = \left(\frac{P_{n-1}}{P_{10}}\right) - \left(\frac{P_n}{P_{10}}\right) \Rightarrow \frac{dp_n}{d\tau} = p_{n-1} - p_n, \quad p_n(0) = 0$$

$$\frac{d\left(\frac{P_1}{P_{10}}\right)}{d(KMt)} = -\frac{P_1}{P_{10}} \Rightarrow \frac{dp_1}{d\tau} = -p_1, \quad p_1(0) = 1$$

onde  $p_n$  e  $\tau$  são chamados respectivamente de concentração adimensional da espécie  $n$  e tempo adimensional do processo. A quantidade  $p_n$  pode também ser interpretada como uma probabilidade, já que ela representa a fração de aglomerado que tem comprimento  $n$ , dentre todos os aglomerados possíveis formados no sistema.

As equações podem ser resolvidas recursivamente a partir de  $n = 1$ . Para a primeira equação,

$$p_1(t) = \exp(-\tau)$$

Substituindo o valor de  $p_1(t)$  no balanço de  $p_2$ , chega-se a

$$\frac{dp_2}{dt} + p_2 = \exp(-\tau)$$

cuja solução é

$$p_2(t) = \tau \exp(-\tau)$$

Repetindo-se o procedimento para  $n = 3, 4, \dots$

$$p_3(t) = \frac{\tau^2}{2} \exp(-\tau)$$

$$p_4(t) = \frac{\tau^3}{3 \cdot 2} \exp(-\tau)$$

$$p_n(t) = \frac{\tau^{n-1}}{(n-1)!} \exp(-\tau)$$

Comparando-se a equação anterior com a Equação (2.14), observa-se que a solução do problema é a distribuição de Poisson deslocada uma unidade para frente; ou seja,

$$p_n(t) = \text{Poisson}(n-1; \tau)$$

O deslocamento é por causa do início da contagem dos tamanhos ( $n = 1$ ), maior que o valor inicial válido para a distribuição de Poisson ( $n = 0$ ). É muito curioso observar que a curva de distribuição de tamanhos dos aglomerados se desloca com valor médio igual ao valor do tempo adimensional (portanto, cresce sempre), que é o parâmetro fundamental do processo de crescimento. A Figura 2.7 ilustra a evolução dos tamanhos dos aglomerados, à medida que o tempo passa.

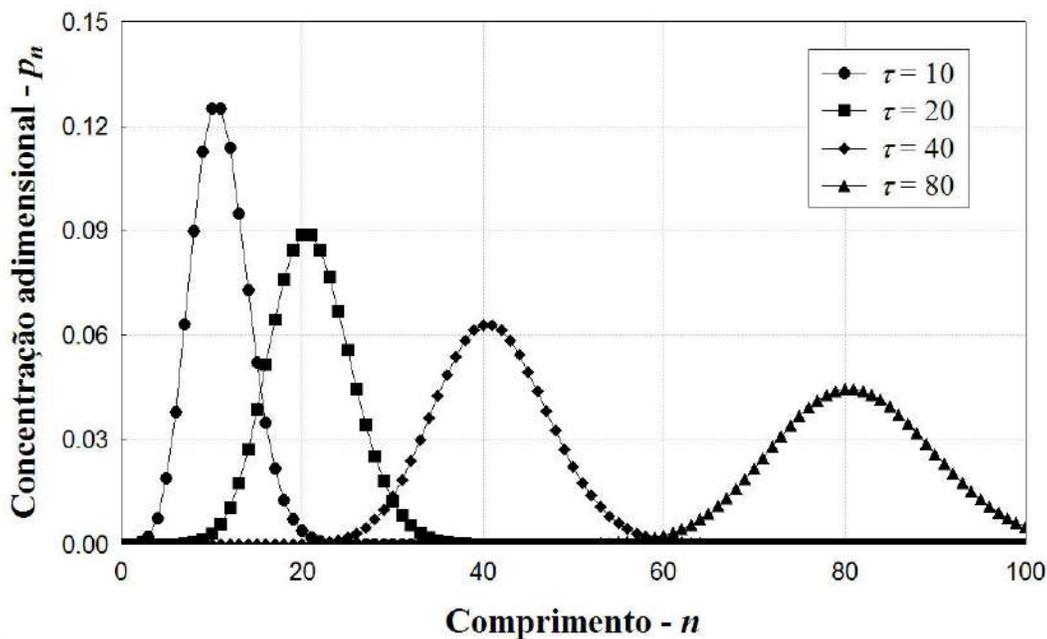


Figura 2.7 - Evolução temporal da concentração de aglomerados de tamanho  $n$ .

### 2.3. A Distribuição Hipergeométrica

Admita que num certo problema a probabilidade de sucesso (ou insucesso) se modifique, à medida que os experimentos se sucedem. Esse é o caso típico que ocorre quando a realização do experimento interfere nos tamanhos das populações investigadas. Por exemplo, imagine um saco onde são acondicionadas 10 bolinhas, sendo cinco delas pretas e as cinco restantes, vermelhas. Qual a probabilidade de se retirar do saco duas bolas pretas, se as bolas **NÃO** são devolvidas ao saco? Nesse caso, embora a probabilidade de se retirar a primeira bola preta seja de 50% (5 possibilidades dentre 10), a probabilidade de se retirar a segunda bola preta cai para 44.44% (4 possibilidades dentre 9). Portanto, a probabilidade de serem retiradas duas bolas pretas em seguida é

$$P_{pp} = \frac{5}{10} \cdot \frac{4}{9} = \frac{2}{9} = 22.22\% \neq \frac{5}{10} \cdot \frac{5}{10} = 25\%$$

Vê-se, portanto, que toda vez que o procedimento de amostragem ou realização do experimento modifica a natureza das populações avaliadas, esse efeito deve ser levado em consideração. (É por esse motivo que as amostras devem ser tão pequenas e representativas da população investigada quanto possível, quando o material amostrado modifica a população investigada. É também por esse motivo que medidas experimentais não intrusivas e não destrutivas são preferíveis no ambiente de laboratório.)

No caso mais geral, seja  $N$  o tamanho da população,  $m$  o tamanho da amostra ou número de experimentos realizados,  $n$  o número de sucessos observado e  $p$  a

probabilidade inicial de sucesso. Como no caso da distribuição binomial, a probabilidade de se obter  $n$  sucessos seguidos pode ser dado na forma

$$P_{.A1} = P(S)_1 \dots P(S)_n P(I)_{n+1} \dots P(I)_{n+(m-n)}$$

$$P_{.A1} = \frac{pN}{N} \frac{pN-1}{N-1} \dots \frac{pN-(n-1)}{N-(n-1)} \frac{(1-p)N}{N-n} \frac{(1-p)N-1}{N-(n+1)} \dots \frac{(1-p)N-(m-n)}{N-(n+(m-n))} \quad (2.16)$$

Repare que se a posição dos sucessos e insucessos for modificada na Equação (2.16), a equação não muda, indicando que qualquer arranjo que contenha o mesmo número de sucessos tem a mesma probabilidade de ocorrer. Como o número de diferentes arranjos pode ser dado pela Equação (2.4), a probabilidade de  $n$  sucessos ocorrerem pode ser dada pelo produto da Equação (2.16) - probabilidade de um arranjo qualquer de tamanho  $m$  que contém  $n$  sucessos - pela Equação (2.4) - número de arranjos de tamanho  $m$  que contém  $n$  sucessos. Fazendo-se essa operação, chega-se a

$$\text{Hiper}(n; N, m, p) = \frac{\binom{Np}{n} \binom{N-Np}{m-n}}{\binom{N}{m}} \quad (2.17)$$

que é a curva de distribuição hipergeométrica.

A distribuição hipergeométrica é uma distribuição discreta univariada e triparmétrica, pois descreve a variação de probabilidades de uma única variável discreta,  $n$ , e depende de três parâmetros:  $N$ , que caracteriza o tamanho do sistema investigado;  $m$ , que caracteriza o tamanho da amostra; e  $p$ , que caracteriza o estado inicial da população. Isso significa que três momentos da curva de distribuição podem ser fixados independentemente pelo usuário, ficando os demais automaticamente definidos pela forma da curva da Equação (2.17). Além disso, não é difícil mostrar que

$$\mu_N = mp \quad (2.18)$$

e

$$\sigma_N^2 = mp(1-p) \left( \frac{N-m}{N-1} \right) \quad (2.19)$$

Parece também intuitivo e, portanto, desnecessário provar rigorosamente, que a distribuição hipergeométrica converge para a distribuição binomial, à medida que o tamanho  $N$  da população aumenta. Nesse caso, o procedimento de amostragem não consegue mudar de forma significativa o estado da população, independentemente do resultado obtido.

**Exemplo 2.8** - Para o caso proposto originalmente, em que há cinco bolas pretas e cinco bolas vermelhas inicialmente no saco,  $N = 10$  e  $p = 0.5$ . Se o tamanho da amostra é  $m = 2$ , as probabilidades ficam:

$$P_{PP} = \frac{5}{10} \frac{4}{9} = \frac{2}{9}, P_{PV} = \frac{5}{10} \frac{5}{9} = \frac{5}{18}, P_{VP} = \frac{5}{10} \frac{5}{9} = \frac{5}{18}, P_{VV} = \frac{5}{10} \frac{4}{9} = \frac{2}{9}$$

Repare que, de fato, as probabilidades dos arranjos  $VP$  e  $PV$  ocorrerem são as mesmas, como previsto pela Equação (2.16). Se associamos a condição de sucesso às bolas pretas, as probabilidades acima podem ser escritas como

$$P_2 = P_{PP} = \frac{2}{9}, P_1 = P_{PV} + P_{VP} = \frac{5}{9}, P_0 = P_{VV} = \frac{2}{9}$$

Utilizando a Equação (2.17) para calcular as probabilidades de sucesso, chega-se a

$$P_0 = \text{Hiper}(0;10,2,0.5) = \frac{\binom{5}{0} \binom{5}{2}}{\binom{10}{2}} = \frac{\left(\frac{5!}{0!5!}\right) \left(\frac{5!}{2!3!}\right)}{\left(\frac{10!}{2!8!}\right)} = \frac{1 \cdot 10}{45} = \frac{2}{9}$$

$$P_1 = \text{Hiper}(1;10,2,0.5) = \frac{\binom{5}{1} \binom{5}{1}}{\binom{10}{2}} = \frac{\left(\frac{5!}{1!4!}\right) \left(\frac{5!}{1!4!}\right)}{\left(\frac{10!}{2!8!}\right)} = \frac{5 \cdot 5}{45} = \frac{5}{9}$$

$$P_2 = \text{Hiper}(2;10,2,0.5) = \frac{\binom{5}{2} \binom{5}{0}}{\binom{10}{2}} = \frac{\left(\frac{5!}{2!3!}\right) \left(\frac{5!}{0!5!}\right)}{\left(\frac{10!}{2!8!}\right)} = \frac{10 \cdot 1}{45} = \frac{2}{9}$$

o que mostra a consistência da análise.

**Exemplo 2.9** - Para o caso proposto originalmente, em que 50% das bolas são pretas no início do processo, associa-se a essa cor o resultado bem sucedido. Então, para um tamanho amostral constante e igual a  $m = 10$ , calcula-se a distribuição de probabilidades de  $n$  sucessos para valores crescentes de  $N$ . Os resultados são mostrados na Figura 2.8 e comparados à curva binomial. Repare a rápida convergência para a curva binomial, à medida que  $N$  aumenta.

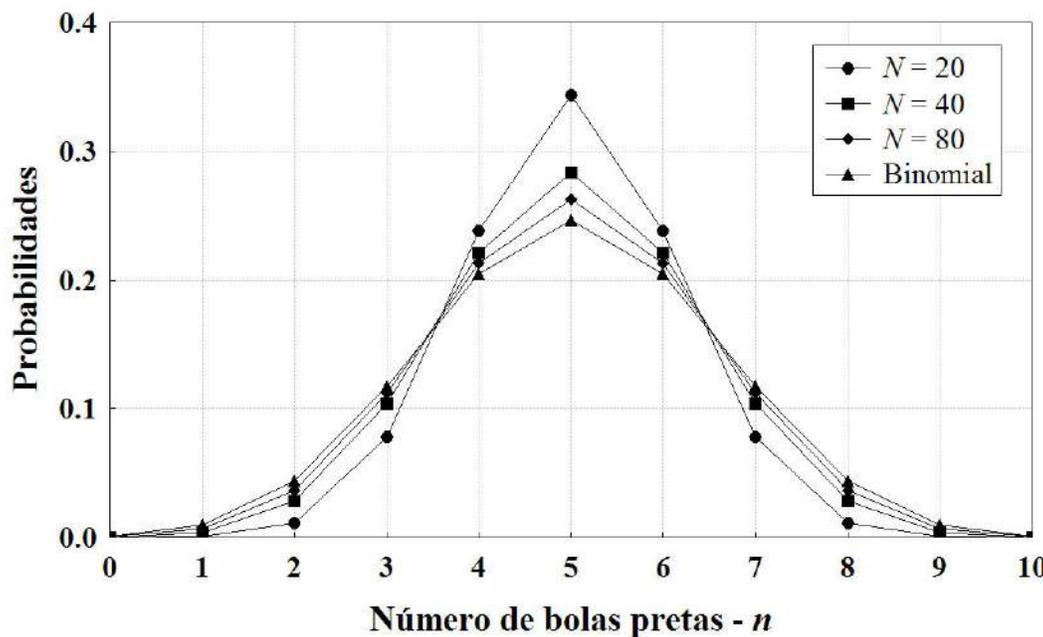


Figura 2.8 - Comparação das distribuições hipergeométricas e binomial para diferentes tamanhos de população.

### 2.4. A Distribuição Uniforme ou Retangular

Admita que num certo problema números reais são gerados aleatoriamente e cobrem de forma uniforme o segmento de reta  $[a, b]$ , de tal maneira que qualquer subintervalo de comprimento inferior a  $(b - a)$  contenha o mesmo número de pontos e, portanto, seja igualmente provável. Nesse caso, obtém-se a distribuição uniforme ou retangular, cuja definição é:

$$\text{Unif}(x; a, b) = \begin{cases} 0, & x < a \\ \frac{1}{b-a}, & a \leq x \leq b \\ 0, & x > b \end{cases} \quad (2.20)$$

A distribuição uniforme está ilustrada na Figura 2.9. A distribuição uniforme é uma distribuição contínua univariada e biparamétrica, pois descreve a variação de probabilidades de uma única variável contínua,  $x$ , e depende de dois parâmetros,  $a$  e  $b$ , que são os extremos do intervalo real em que está definida. Isso significa que dois momentos da curva de distribuição podem ser fixados independentemente pelo usuário, ficando os demais automaticamente definidos pela forma da curva da Equação (2.20). Além disso, não é difícil mostrar que

$$\mu_x = \frac{a+b}{2} \quad (2.21)$$

e

$$\sigma_x^2 = \frac{(b-a)^2}{12} \tag{2.22}$$

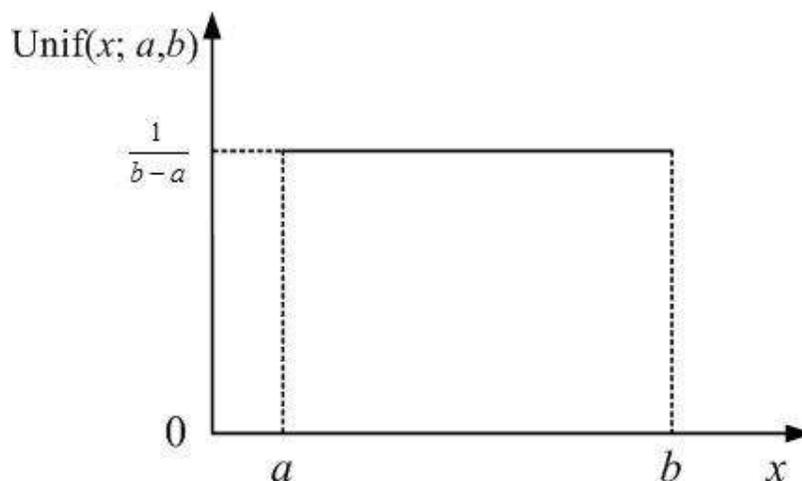


Figura 2.9 - A distribuição uniforme.

Embora a distribuição uniforme seja bastante simples, ela aparece em uma variedade de problemas práticos, como por exemplo no arredondamento de erros. Seja a  $i$ -ésima casa decimal de um número real, que se deseja arredondar. Se a  $(i + 1)$ -ésima casa decimal é inferior a cinco, a  $i$ -ésima casa é mantida constante e as casas decimais menos significativas são descartadas. Se a  $(i + 1)$ -ésima casa decimal é igual ou superior a cinco, a  $i$ -ésima casa é incrementada de uma unidade, enquanto as casas decimais menos significativas são descartadas. Por exemplo,

$$1.\underline{5}3453876 \rightarrow 1.\underline{5}, 1.53\underline{4}53876 \rightarrow 1.53, 1.534\underline{5}3876 \rightarrow 1.535$$

A operação de arredondamento é muito importante porque não faz sentido usar mais algarismos significativos que aqueles consistentes com a precisão do equipamento usado para fazer a medida. Ao se ler um número arredondado, não é possível dizer como as casas decimais menos significativas estavam preenchidas, de maneira que é bastante razoável supor que o erro de arredondamento, desconhecido, segue a distribuição uniforme no intervalo da casa decimal imediatamente menos significativa. Por exemplo, nas operações definidas acima seriam,

$$1.\underline{5} \rightarrow [1.45, 1.55], 1.53\underline{4} \rightarrow [1.525, 1.535], 1.534\underline{5} \rightarrow [1.5345, 1.5355]$$

Essa argumentação torna possível implementar uma técnica computacional de geração de números aleatórios extremamente eficiente, que é amplamente utilizada pela maior parte dos simuladores e ambientes de programação. A técnica consiste em provocar o truncamento e o conseqüente arredondamento de casas decimais no computador. Por exemplo, se um número inteiro é dividido por um outro número inteiro, o resultado pode ter um número infinito de casas decimais não nulas. Se o computador tem precisão finita, parte do resultado é perdida para sempre. Portanto, ao se multiplicar o resultado truncado pelo divisor, obtém-se um número que não é

necessariamente igual ao dividendo original. Por exemplo, suponha que a precisão da nossa operação é de duas casas decimais apenas. Nesse caso,

$$\frac{1}{7} = 0.14 \rightarrow 0.14 \cdot 7 = 0.98 \rightarrow 1 - 0.98 = 0.02$$

Vê-se que a operação  $1 - 7\left(\frac{1}{7}\right)$  resultou num resíduo de 0.02. Se os números manipulados são suficientemente grandes, por exemplo na forma  $N_1 - N_2\left(\frac{N_1}{N_2}\right)$ , e se  $N_1$  e/ou  $N_2$  é gerado através de algum procedimento aleatório, como por exemplo um número obtido da transformação do horário em que o programa está rodando com precisão de décimo de segundo na forma,

$$HH : MM : SS.D \rightarrow N_2 = \left( \frac{MM \cdot 600 + SS \cdot 10 + D}{36000} \right) N_3 = X_0 N_3$$

então é possível gerar números distribuídos no intervalo (0,1) de maneira praticamente uniforme. (Na expressão acima, *HH*, *MM*, *SS* e *D* representam respectivamente a hora, os minutos, os segundos e os décimos de segundo. A transformação acima pode ser considerada aleatória na suposição de que a operação pode ser realizada a qualquer momento do dia, sem horário marcado. Não é conveniente introduzir a hora *HH* na operação porque em geral o trabalho é realizado no horário comercial, o que acabaria por introduzir significativo grau de determinismo na operação.  $N_3$  é um número de referência, do qual o número  $N_2 < N_3$  pode ser considerado uma fração  $X_0$ . No caso considerado,  $N_3$  deveria ser o número 36000, que é o número de décimos de segundo contidos em uma hora.) A operação entre parênteses gera um número inicial no intervalo (0,1) chamado de **semente**. A operação de truncamento pode ser então repetida de maneira iterativa, usando o resultado da iteração prévia como semente da próxima iteração, na forma

$$X_{k+1} = X_k \frac{N_3}{N_2} - \text{Trunc} \left( X_k \frac{N_3}{N_2} \right) = (X_k N_3) \text{mod} (N_2) \quad (2.23)$$

O significado da Equação (2.23) é semelhante à operação de divisão executada anteriormente com o número 7. O primeiro termo consiste em gerar um número maior do que 1 com um certo número de casas decimais, enquanto o segundo termo consiste em gerar o mesmo número sem casas decimais. Dessa forma, o número resultante da operação é um número entre 0 e 1, com parte inteira nula e número arbitrário de casas decimais. Na Equação (2.23) **Trunc** representa a operação de abandonar a parte não inteira do número resultante, enquanto **mod** representa manter apenas a parte decimal da divisão entre dois números. A seqüência de números gerada, então, não é verdadeiramente aleatória, pois a repetição da semente inicial resultará sempre na mesma seqüência de números. É a geração aleatória da semente que garante de fato um certo grau de aleatoriedade da seqüência de números. Por isso, a seqüência obtida é dita **pseudo-aleatória** e é, para todos os fins práticos, uma seqüência de números aleatórios excelente para fins de simulação. Os exemplos abaixo ilustram esses conceitos.

**Exemplo 2.10** - As Figuras 2.10 e 2.11 mostram os primeiros 3600 números obtidos do procedimento recursivo implementado computacionalmente, com  $X_0 = 0.35312356$ .

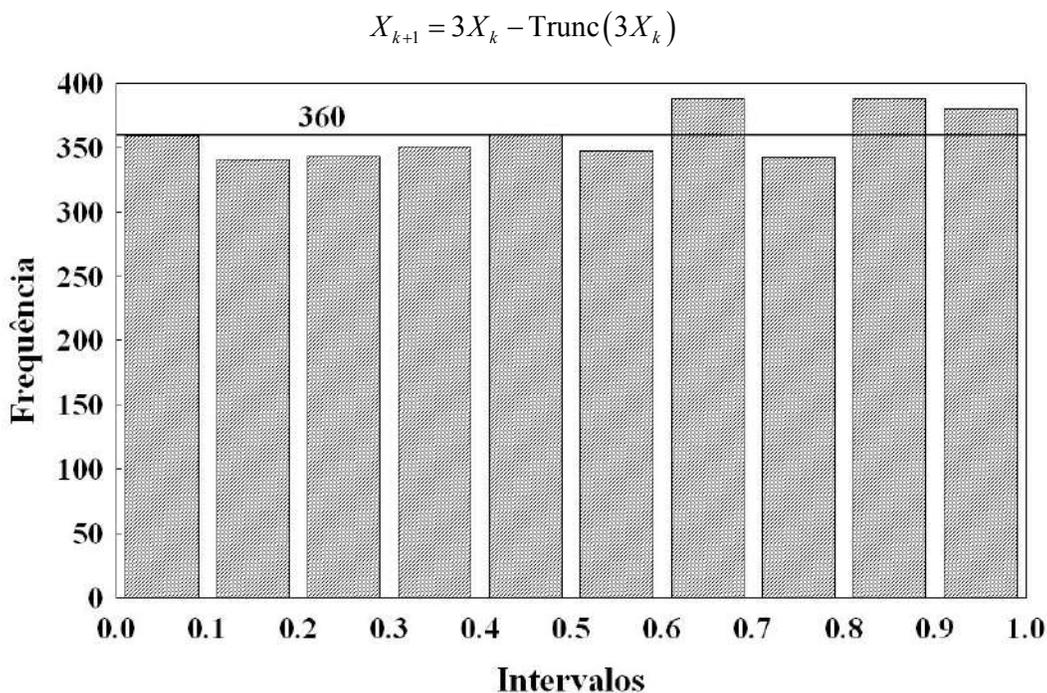


Figura 2.10 - Distribuição dos pontos obtidos da recursão.

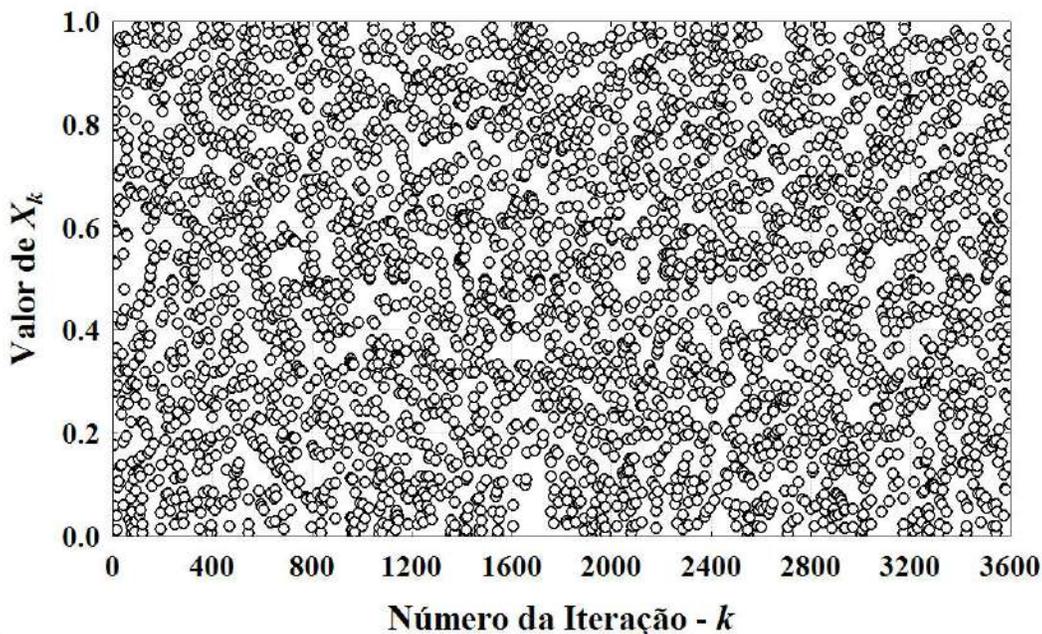


Figura 2.11 - Seqüência dos pontos obtidos da recursão.

As Figuras 2.10 e 2.11 mostram que, a despeito da regra determinística que deu origem aos pontos, não parece haver qualquer padrão de formação entre os pontos, que se aproximam de fato de uma distribuição uniforme. O fator de correlação calculado entre pontos deslocados de uma iteração é inferior a 0.3, indicando que há realmente

muito pouca ordem no sistema. A Figura 2.12 ilustra claramente o procedimento determinístico utilizado para gerar os pontos da distribuição.

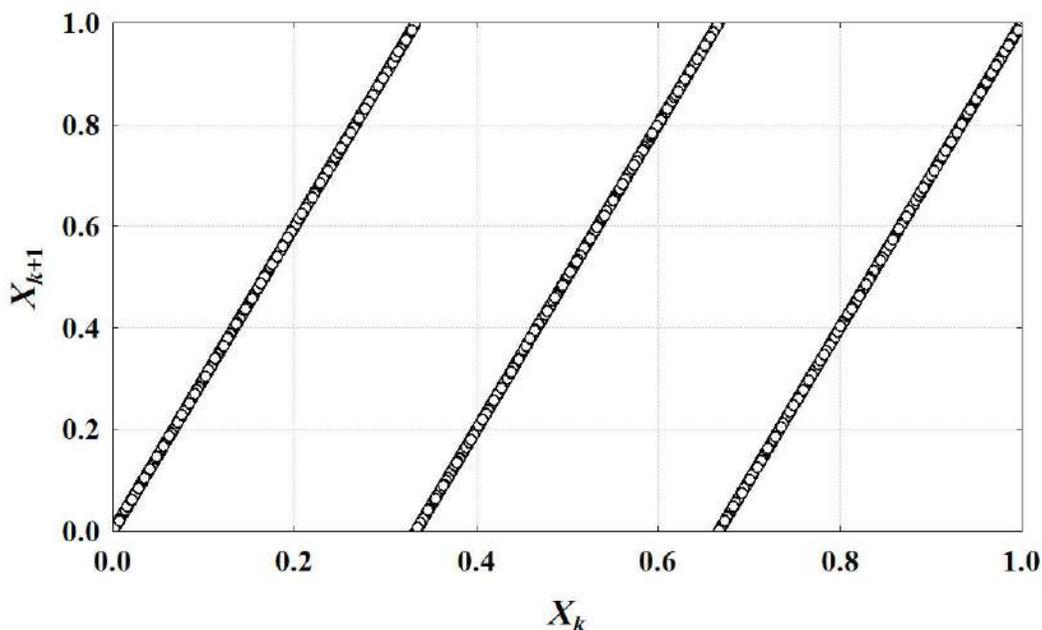


Figura 2.12 - Regra de recursão utilizada.

As Figuras 2.10 a 2.12 são impressionantes porque indicam que um excelente grau de aparente aleatoriedade já pode ser obtido com valores muito pequenos de  $N_1$ ,  $N_2$  e  $N_3$ . Modificando esses valores de maneira apropriada é possível obter seqüências pseudo-aleatórias onde a componente determinística é virtualmente inexistente.

A forma extremamente simples com que é possível gerar números aleatórios no computador faz com que a distribuição uniforme seja utilizada como padrão de comparação e geração de números aleatórios para outras funções de densidade de probabilidade de variáveis contínuas. Para que isso seja possível, dois conjuntos de dados aleatórios  $x_i$  e  $y_i$  podem ser chamados de **seqüências aleatórias equivalentes** se

$$\int_{x_{min}}^{x_i} \phi_1(x) dx = \int_{y_{min}}^{y_i} \phi_2(y) dy \quad (2.24)$$

A Equação (2.24) define seqüências aleatórias equivalentes como aquelas que geram probabilidades acumuladas iguais para diferentes distribuições de probabilidade. Como o procedimento definido pela Equação (2.23) resulta na distribuição uniforme no intervalo (0,1), a Equação (2.24) pode ser escrita como

$$x_i = \int_{y_{min}}^{y_i} \phi_2(y) dy = P_{AC,2}(y_i) \quad (2.25)$$

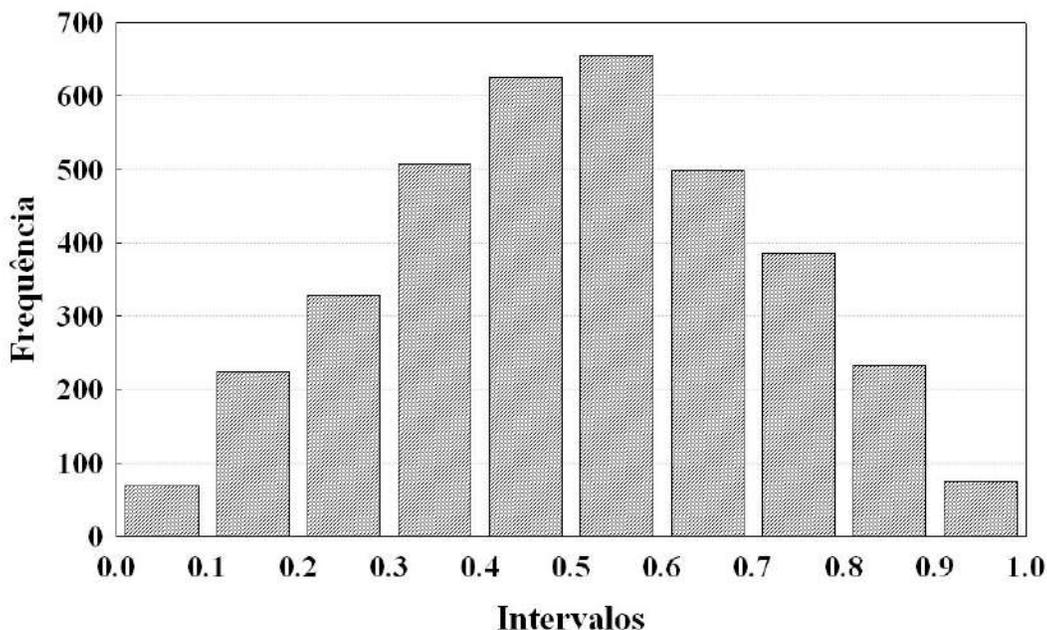
de maneira que o número  $x_i$  gerado pelo computador, que segue a distribuição uniforme no intervalo  $(0,1)$ , pode ser transformado em um número aleatório que segue qualquer distribuição  $\varphi_2(y)$ , bastando para isso encontrar o valor  $y_i$  que faz com que a probabilidade acumulada da segunda distribuição seja igual ao valor  $x_i$  gerado pelo computador. Este procedimento permite manipular praticamente qualquer distribuição estatística por via computacional. Por isso, esses procedimentos são muito explorados durante o desenvolvimento e implementação de técnicas de Monte Carlo, como descrito no Capítulo 5 desse Volume da série.

**Exemplo 2.11** - No Exemplo 1.13 mostrou-se que a probabilidade acumulada da distribuição triangular tem a forma

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

Usando a regra de transformação definida pela Equação (2.25) conclui-se que

$$y_i = \begin{cases} 0, & x_i \leq 0 \\ \sqrt{\frac{x_i}{2}}, & 0 \leq x_i \leq 0.5 \\ 1 - \sqrt{1 - \frac{(x_i + 1)}{2}}, & 0.5 \leq x_i \leq 1 \\ 1, & x_i \geq 1 \end{cases}$$



**Figura 2.13** - Distribuição triangular obtida a partir dos pontos gerados computacionalmente no Exemplo 2.10, com distribuição próxima da uniforme.

As Figuras 2.13 e 2.14 ilustram a qualidade da transformação obtida quando os pontos gerados no Exemplo 2.10 são usados para gerar a distribuição triangular. Os resultados podem ser considerados excelentes. Repare como os pontos estão de fato mais concentrados ao redor do valor 0.5, como esperado. No entanto, do mesmo jeito que foi observado no exemplo anterior, os números não constituem de fato uma seqüência aleatória, dado que a repetição da semente resultaria na repetição da seqüência de números obtidos. A Figura 2.15 ilustra a regra de transformação utilizada para geração da seqüência pseudo-aleatória. Repare que, a despeito da existência de claríssima regra determinística de obtenção dos números, é virtualmente impossível observar isso a partir da seqüência de valores obtidos.

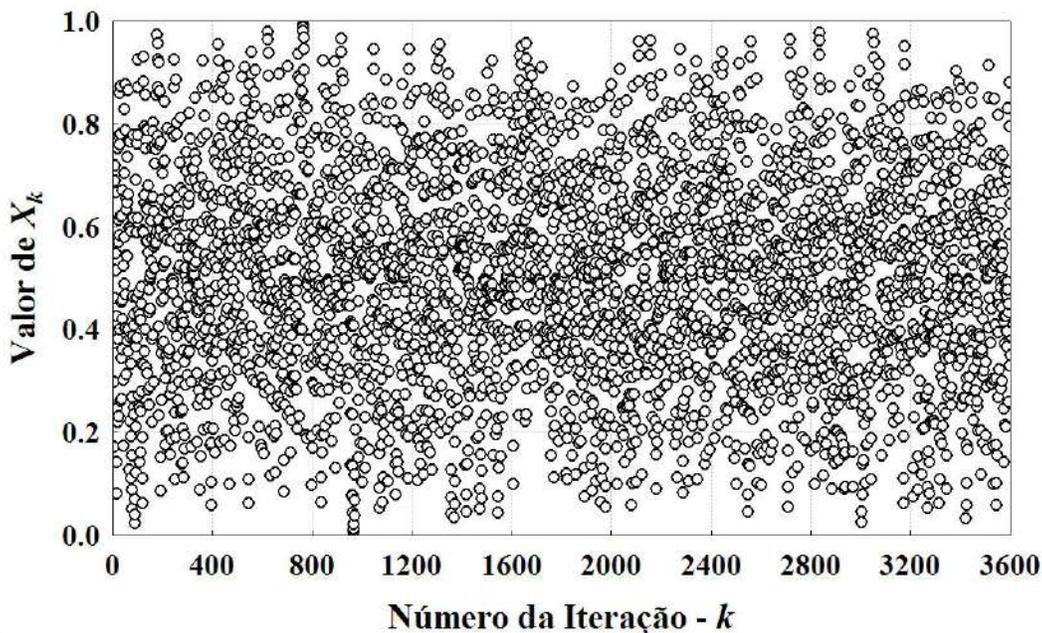


Figura 2.14 - Seqüência dos pontos obtidos da recursão para a distribuição triangular.

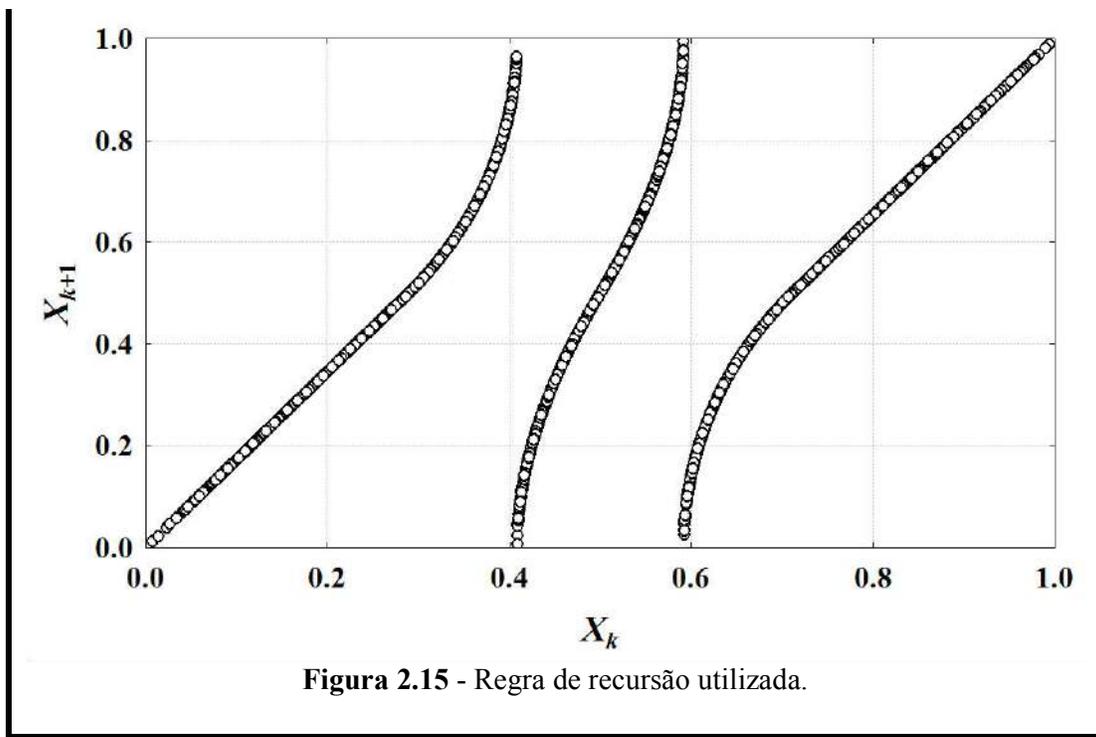


Figura 2.15 - Regra de recursão utilizada.

### 2.5. A Distribuição Exponencial

Admita que num certo problema é possível associar uma velocidade média de perda de alguma propriedade específica do sistema com a quantidade dessa propriedade. Por exemplo, suponha que a velocidade de desaparecimento de indivíduos em uma população seja proporcional ao número total de indivíduos que constituem a população. Esse é o caso típico que ocorre quando não há qualquer termo que contribua com o aumento do número de indivíduos da população, como no caso de um lote de partículas de catalisador que é colocado no interior de um reator químico ou num lote de partículas radioativas que é utilizado como fonte de energia em um determinado processo radiativo. Nesse caso, um problema bastante importante para o projeto do processo é saber qual a probabilidade de se encontrar um certo número de partículas "vivas" ou ativas após um determinado período de tempo. Para que o problema possa ser respondido, é conveniente associar ao problema a seguinte equação de balanço:

$$\frac{dN}{dt} = -KN, N(0) = N_0 \tag{2.26}$$

onde  $K$  é a constante de velocidade do problema e  $N$  é o número de indivíduos ativos ou "vivos" da população. Como no Exemplo 2.7, é conveniente dividir a Equação (2.26) por  $K$  e  $N_0$ , resultando em

$$\frac{d\left(\frac{N}{N_0}\right)}{d(Kt)} = -\left(\frac{N}{N_0}\right) \Rightarrow \frac{dp}{d\tau} = -p, p(0) = 1 \tag{2.27}$$

cuja solução é

$$p(\tau) = \exp(-\tau) \quad (2.28)$$

que é a chamada forma absoluta da distribuição exponencial. Para fins de generalização da forma da distribuição exponencial, já que nem sempre é possível determinar *a priori* o valor da constante de velocidade do problema, é mais comum encontrá-la na forma:

$$\wp(t) = \text{Expon}(t; \mu_T) = \frac{\exp\left(-\frac{t}{\mu_T}\right)}{\mu_T}, \quad \mu_T = \frac{1}{K} \quad (2.29)$$

onde  $\mu_T$  é o valor médio da distribuição ou o tempo médio de vida dos indivíduos.

Na forma da Equação (2.29), a distribuição exponencial é uma distribuição contínua univariada e uniparamétrica, pois descreve a variação de probabilidades de uma única variável contínua,  $t$ , e depende de um único parâmetro,  $K$  ou  $\mu_T$ . Isso significa que apenas um dos momentos da curva de distribuição pode ser fixado independentemente pelo usuário, ficando os demais automaticamente definidos pela forma da curva da Equação (2.29). Além disso, não é difícil mostrar que

$$\sigma_T^2 = \mu_T^2 = \left(\frac{1}{K}\right)^2 \quad (2.30)$$

o que também mostra que o índice de polidispersão é constante e igual a 2 para a distribuição exponencial, indicando um grau de espalhamento relativo em relação à média constante; ou seja, à medida que o valor médio cresce, o grau de espalhamento cresce na mesma proporção.

A distribuição exponencial é muito usada para descrever problemas práticos de engenharia, como aqueles que envolvem o cálculo do tempo de vida útil de peças e equipamentos. Nesse sentido, a distribuição exponencial é um membro particular de uma família muito mais ampla de distribuições de probabilidade que podem ser chamadas genericamente de **distribuições de tempo de vida**. Essas distribuições poderiam ser descritas genericamente a partir da Equação (2.27) como

$$\frac{d[\ln(\wp)]}{dt} = -K(t), \quad \int_0^{\infty} \wp(t) dt = 1 \quad (2.31)$$

onde  $K(t)$  é uma constante de velocidade que pode variar com o tempo, ao invés de admitir sempre o mesmo valor, como considerado anteriormente. Para que se compreenda como essa variação pode ser importante em certos problemas, as estatísticas médicas comprovam que o índice de mortalidade infantil diminui consideravelmente, à medida que a idade da criança aumenta. Logo, a constante de velocidade  $K(t)$  é alta para idades pequenas e cai continuamente, à medida que a criança vai ficando mais velha. O mesmo fenômeno ocorre com máquinas e equipamentos. À

medida que o tempo passa, para valores de tempo pequenos, os riscos de falha diminuem progressivamente, até que um valor mínimo é atingido. A partir de certa idade, o risco de falha dos equipamentos começa a aumentar progressivamente, em função do envelhecimento de peças e demais componentes. Há, portanto, incentivos para se analisar o comportamento de distribuições de tempo de vida obtidas a partir da Equação (2.31), para diferentes formas da constante de velocidade  $K(t)$ .

Antes de analisar essa família de funções, é conveniente introduzir um linguajar característico da área de análise de riscos. Chamemos inicialmente de  $S(t)$  a probabilidade de que o indivíduo sobreviva ao tempo  $t_1$ . Nesse caso

$$S(t > t_1) = 1 - P_{AC}(t_1) = P(t > t_1) = \int_{t_1}^{\infty} p(t) dt \quad (2.32)$$

Nesse caso, a curva de densidade de probabilidades pode ser descrita como

$$\wp(t) = \frac{dP_{AC}(t)}{dt} = -\frac{dS(t)}{dt} \quad (2.33)$$

Definamos como risco à função

$$h(t) = \frac{\wp(t)}{S(t)} \quad (2.34)$$

que é uma medida relativa da velocidade de decaimento da população. Substituindo a Equação (2.34) na Equação (2.33), chega-se a:

$$\frac{d[\ln(S(t))]}{dt} = -h(t) \quad (2.35)$$

Como  $S(0) = 1$ , a Equação (2.35) pode ser integrada na forma:

$$S(t_1) = 1 - P_{AC}(t_1) = \exp\left[-\int_0^{t_1} h(t) dt\right] \quad (2.36)$$

Logo, combinando as Equações (2.31), (2.34) e (2.35), chega-se a

$$\begin{aligned} \frac{d[\ln(\wp(t))]}{dt} &= -K(t) \\ \frac{d[\ln(S(t))]}{dt} + \frac{d[\ln(h(t))]}{dt} &= -K(t) \\ \frac{d[\ln(h(t))]}{dt} - h(t) &= -K(t) \end{aligned} \quad (2.37)$$

A distribuição exponencial da Equação (2.29) pode ser obtida fazendo-se  $h(t)=K(t)$ . Outras formas convenientes de distribuições de tempo de vida são apresentadas a seguir.

### 2.5.1. A Distribuição de Gompertz

$$h(t) = \exp[\alpha + \beta t], \quad -\infty < \alpha < \infty, \quad \beta \geq 0 \quad (2.38)$$

$$S(t) = \exp \left\{ -\exp(\alpha) \frac{[\exp(\beta t) - 1]}{\beta} \right\} \quad (2.39)$$

$$\wp(t) = \text{Gomp}(t; \alpha, \beta) = \exp(\alpha + \beta t) \exp \left\{ \exp(\alpha) \frac{[\exp(\beta t) - 1]}{\beta} \right\} \quad (2.40)$$

A **distribuição de Gompertz** é um modelo bi-paramétrico muito usado para descrever a taxa de mortalidade da população adulta. Observe que o risco aumenta exponencialmente com o aumento da idade do indivíduo. A restrição imposta sobre  $\beta$  é necessária para garantir a existência da integral da função de densidade de probabilidades.

### 2.5.2. A Distribuição de Weibull

$$h(t) = \alpha t^\beta, \quad \alpha > 0, \quad \beta > -1 \quad (2.41)$$

$$S(t) = \exp \left\{ -\alpha \frac{t^{\beta+1}}{\beta+1} \right\} \quad (2.42)$$

$$\wp(t) = \text{Weibull}(t; \alpha, \beta) = \alpha t^\beta \exp \left\{ -\alpha \frac{t^{\beta+1}}{\beta+1} \right\} \quad (2.43)$$

A **distribuição de Weibull** é um modelo bi-paramétrico muito usado para descrever o tempo de vida útil de equipamentos, peças e componentes eletrônicos. Observe que o risco aumenta como uma potência do tempo de vida do indivíduo. As restrições impostas sobre  $\alpha$  e  $\beta$  são necessárias para garantir a existência da integral da função de densidade de probabilidades. Para a distribuição de Weibull é possível mostrar que

$$\mu_T = \left( \frac{\alpha}{\beta+1} \right)^{\left( \frac{1}{\beta+1} \right)} \Gamma \left( \frac{1}{\beta+1} + 1 \right) \quad (2.44)$$

$$\sigma_T^2 = \left( \frac{\alpha}{\beta+1} \right)^{\left( \frac{2}{\beta+1} \right)} \left[ \Gamma \left( \frac{2}{\beta+1} + 1 \right) - \Gamma^2 \left( \frac{1}{\beta+1} + 1 \right) \right] \quad (2.45)$$

onde  $\Gamma(x)$  é a **função gama**, definida como

$$\Gamma(x) = \int_0^{\infty} z^{x-1} e^{-z} dz \quad (2.46)$$

Uma curiosidade sobre a função gama é que

$$\Gamma(x+1) = x\Gamma(x) \quad (2.47)$$

que resulta no valor  $\Gamma(x+1) = x!$  quando  $x$  é inteiro. Por isso, a função gama é chamada usualmente de **função fatorial generalizada**.

### 2.5.3. A Distribuição Gama

Quando a constante de velocidade da Equação (2.31) se aproxima assintoticamente de um valor limite na forma

$$\frac{d\wp}{dt} = -C \left[ \frac{\beta t - \alpha}{t} \right] \wp, \quad \alpha > 0, \beta > 0, C > 0 \quad (2.48)$$

então

$$\wp(t) = \text{Gama}(t; \alpha, \beta) = \frac{\beta^{\alpha+1}}{\Gamma(\alpha+1)} t^{\alpha} e^{-\beta t} \quad (2.49)$$

que é a conhecida **distribuição gama**. Repare que a constante de velocidade indica que a ocorrência de falhas para tempos muito curtos é virtualmente inexistente, havendo convergência para uma constante de velocidade aproximadamente estável, à medida que o tempo passa. A distribuição gama é um modelo bi-paramétrico (a constante  $C$  é calculada de forma a garantir que a integral de  $\wp(t)$  seja igual a 1) também muito usado para descrever o tempo de vida útil de equipamentos, peças e componentes eletrônicos. As restrições impostas sobre  $\alpha$  e  $\beta$  são necessárias para garantir a existência da integral da função de densidade de probabilidades. Para a distribuição gama é possível mostrar que

$$\mu_T = \left( \frac{\alpha+1}{\beta} \right) \quad (2.50)$$

$$\sigma_T^2 = \left( \frac{\alpha+1}{\beta^2} \right) \quad (2.51)$$

**Exemplo 2.12** - Uma curva de distribuição discreta, análoga à curva de distribuição exponencial contínua, pode ser desenvolvida para sistemas em crescimento, como no Exemplo 2.7. Nesse caso, suponha a seguinte equação de balanço

$$\frac{dP_n}{dt} = K_p P_{n-1} - (K_p + K_t) P_n, P_n(0) = 0, n > 1$$

$$\frac{dP_1}{dt} = F - (K_p + K_t) P_1, P_1(0) = 0$$

onde  $K_p$  e  $K_t$  são respectivamente as constantes de velocidade para o crescimento e desaparecimento da espécie em crescimento  $P_n$ .  $F$  é uma fonte de espécies de tamanho mínimo. Como nos casos anteriores, é conveniente dividir as equações de balanço pela constante global de velocidade  $(K_p + K_t)$  na forma:

$$\frac{dP_n}{d\tau} = q P_{n-1} - P_n, P_n(0) = 0, n > 1$$

$$\frac{dP_1}{d\tau} = f - P_1, P_1(0) = 0$$

onde  $\tau = (K_p + K_t)t$  é o tempo adimensional,  $q = K_p / (K_p + K_t)$  é a probabilidade de crescimento e  $f = F / (K_p + K_t)$  é a fonte adimensional. A solução do problema fica então na forma:

$$P_n = q^{n-1} f \left[ 1 - \sum_{i=0}^{n-1} \frac{\tau^i e^{-\tau}}{i!} \right]$$

Quando o tempo é muito grande, o termo exponencial faz com que a parte dinâmica da solução vá a zero, resultando na seguinte solução estacionária:

$$P_n = q^{n-1} f$$

Para que essa solução possa ser interpretada como uma probabilidade, é necessário que a soma de todos os valores possíveis seja igual a 1. Nesse caso,

$$\sum_{n=1}^{\infty} P_n = f \sum_{n=1}^{\infty} q^{n-1} = f \frac{1}{(1-q)}$$

Portanto, definindo

$$p_n = \frac{P_n}{\frac{f}{(1-q)}} = (1-q) q^{n-1}$$

que é a chamada **distribuição de Flory**, análoga discreta da curva exponencial, pois:

$$p_n = (1-q) \exp\{(n-1) \ln(q)\} = \left(\frac{1-q}{q}\right) \exp\left\{-\left[\ln\left(\frac{1}{q}\right)\right] n\right\}$$

que pode ser comparada à Equação (2.29). Portanto, a curva exponencial discreta aparece como solução estacionária dos problemas de crescimento de espécies, quando a espécie mínima é gerada continuamente e quando há desaparecimento simultâneo das espécies em crescimento no meio.

## 2.6. A Distribuição Normal

Admita que na Equação (2.31),  $K(t)$  é uma função linear de  $t$ . Nesse caso,

$$\frac{d[\ln(\wp)]}{dt} = -\frac{t}{\alpha} \quad (2.52)$$

$$\wp(t) = C(\alpha) \exp\left(-\frac{t^2}{2\alpha}\right)$$

onde  $C(\alpha)$  é uma constante de integração que garante que a integração da curva de densidade sobre o domínio de  $t$  é igual a 1. A Equação (2.52) é a chamada **distribuição Gaussiana** ou **distribuição normal**. Na forma apresentada na Equação (2.52), a curva de distribuição normal pode ser interpretada como uma distribuição de tempos de vida em que a probabilidade de falhas aumenta linearmente com o aumento da vida útil do material. O parâmetro  $\alpha$  controla a intensidade dessa variação.

Na realidade, a curva de distribuição normal ou Gaussiana apresenta utilidade muito maior que a sugerida somente pela interpretação da distribuição de tempos de vida. Em primeiro lugar, a curva normal pode ser estendida e utilizada para todo o domínio real da variável contínua  $t$ , já que ela é simétrica em relação ao eixo  $\wp(t)$ . Logo, ela pode ser associada a um número muito maior de problemas físicos de interesse prático, onde a variável aleatória pode assumir valores positivos ou negativos. Em segundo lugar, a variável  $t$  pode ser escalada convenientemente na forma

$$\wp(t) = \text{Normal}(t; \mu_T, \sigma_T) = \frac{1}{\sigma_T \sqrt{2\pi}} \exp\left(-\frac{1}{2} \left(\frac{t - \mu_T}{\sigma_T}\right)^2\right), \quad -\infty < t < \infty \quad (2.53)$$

ganhando a forma de uma distribuição contínua univariada bi-paramétrica. (O escalonamento realizado consiste em fazer com que o ponto de máximo da curva normal coincida com o valor médio.) A forma bi-paramétrica é extremamente prática porque os parâmetros da curva normal coincidem com os valores da média e do desvio padrão (ou variância) usados anteriormente para caracterizar o posicionamento e o grau de espalhamento da distribuição de probabilidades. Dessa forma, apenas a caracterização da média e da variância da distribuição é suficiente para a utilização direta da curva de distribuição normal. Como nos demais casos, fixadas a média e a variância (ou desvio padrão), os demais momentos da curva normal estão automaticamente definidos pela forma da Equação (1.26). Em terceiro lugar, por ser simétrica e monomodal (apresenta apenas um máximo), a média, a moda e a mediana da curva normal são coincidentes.

Além disso, a distribuição normal é matematicamente tratável; isto é, muitos resultados analíticos só podem ser explicitados para esta distribuição. Por esse motivo, as distribuições t de Student,  $\chi^2$  (chi-quadrado) e F de Fisher puderam ser desenvolvidas para a análise dos resultados obtidos a partir de variáveis com distribuição normal. Essas três distribuições e os testes estatísticos relacionados a elas serão discutidos no Capítulo 3 desse Volume da série.

Por todas essas características, a curva normal foi adotada como modelo adequado de distribuição de variáveis aleatórias em um infindável número de aplicações. Isso é particularmente verdadeiro em problemas que lidam com flutuações aleatórias causadas por imprecisões e erros de medidas. Em problemas de medida, quando o instrumento de medição está bem calibrado e funcionando adequadamente (o que não é necessariamente verdade!), parece razoável admitir que:

- a) Os erros podem ser positivos ou negativos, distribuindo-se aleatoriamente em torno do valor zero de forma simétrica. Dessa forma, parece razoável admitir que as medidas se distribuam aleatoriamente e de forma simétrica em torno do valor médio;
- b) Quanto maiores os erros em valor absoluto, menos provável é a sua ocorrência.

Ambas as condições são naturalmente satisfeitas pela curva normal, que ainda exige apenas a caracterização de dois parâmetros naturalmente utilizados para a caracterização de densidades de probabilidade - a média e a variância. Isso faz da curva normal um bom modelo probabilístico para boa parte das medidas experimentais. Atente para o fato de que isso **NÃO** significa dizer que os erros de medição seguem necessariamente o modelo normal.

No entanto, um resultado extremamente importante, conhecido como o **Teorema do Limite Central**, faz da curva normal a *rainha* de todas as curvas de distribuição de probabilidades.

**Teorema do Limite Central** - Sejam  $x_1, x_2, x_3, \dots, x_N$ , números gerados por distribuições de probabilidades quaisquer, com médias  $\mu_{x_i}$  e variância  $\sigma_{x_i}^2$ . Seja ainda a soma  $S_N$  definida como

$$S_N = \sum_{i=1}^N x_i$$

Então  $f(S_N)$  se aproxima de uma curva normal quando  $N$  é suficientemente grande.

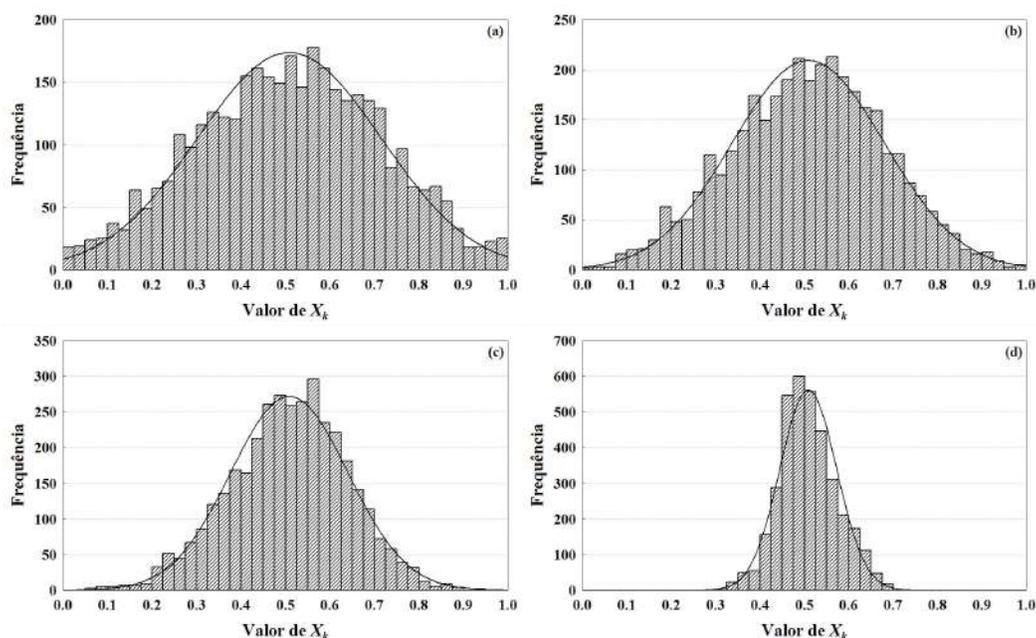
O Teorema do Limite Central, que não será provado aqui por falta de espaço, diz que, independentemente das distribuições de probabilidades que deram origem às flutuações aleatórias fundamentais, resultados obtidos da soma de muitos eventos aleatórios apresentam distribuição aproximadamente normal. Isso significa que eventos complexos, gerados a partir da soma de pequenas flutuações aleatórias, apresentam distribuição de probabilidades próxima da distribuição normal. Talvez seja essa a razão principal que faz com que a distribuição normal encontre uso generalizado como modelo probabilístico da distribuição de erros de medida.

A despeito da força do Teorema do Limite Central, deve-se evitar a falsa impressão de que toda distribuição de erros ou de que toda distribuição de probabilidades contínua é normal. Tal associação é absolutamente equivocada e uns poucos exemplos bastam para mostrar que a curva normal não é uma panacéia para todos os problemas e aplicações. Por exemplo, a distribuição normal apresenta um grande defeito para sua utilização em grande número de problemas, que é o fato das flutuações aleatórias ocorrerem no intervalo  $(-\infty, +\infty)$ . Obviamente algumas variáveis não podem ser infinitamente grandes e outras não podem jamais assumir valores negativos. Por exemplo, se a variável estudada for a altura de pessoas numa população, não parece razoável acreditar que seja possível encontrar pessoas com mais de 3 metros de altura, por menor que essa probabilidade seja. Da mesma forma, não parece razoável acreditar que seja possível encontrar pessoas com altura negativa, por menor que seja essa probabilidade. Por isso, o modelo de distribuição normal de probabilidades deve ser encarado como um modelo conveniente para uso, por todas as razões descritas anteriormente. Isso não significa que a distribuição real de probabilidades de qualquer problema físico possa ou deva ser descrita necessariamente pela curva normal, como algumas pessoas teimam em pensar e afirmar.

**Exemplo 2.13** - Para a seqüência pseudo-aleatória do Exemplo 2.10, foram calculadas médias de seqüências de 3, 5, 10 e 50 pontos respectivamente. 3600 valores médios foram tomados em cada caso para construir os histogramas de distribuição. As curvas normais foram traçadas usando-se como valor médio a média aritmética dos 3600 pontos e como variância os desvios médios quadráticos dos pontos em relação à média, na forma:

$$\mu_X = \frac{\sum_{i=1}^{3600} x_i}{3600}, \quad \sigma_X^2 = \frac{\sum_{i=1}^{3600} (x_i - \mu_X)^2}{3600}$$

Os resultados estão apresentados na Figura 2.16 abaixo.



**Figura 2.16** - Distribuição da média aritmética de seqüências de (a) 3, (b) 5, (c) 10 e (d) 50 pontos obtidos da distribuição pseudo-aleatória do Exemplo 2.10.

Repare como a distribuição uniforme converge rapidamente para a distribuição normal. Repare ainda que a distribuição normal vai se estreitando em torno do valor médio, à medida que o número de pontos usados para cálculo da média vai aumentando.

## 2.7. A Distribuição Log-Normal

Como discutido anteriormente, o uso da curva normal pode apresentar sérios inconvenientes práticos quando a variável analisada não pode ser negativa. Esse é o caso típico da descrição de distribuições de tamanho, massa, etc., comuns em problemas de engenharia. Nesse caso, é comum admitir que a distribuição do logaritmo da variável considerada, que pode assumir valores positivos ou negativos, segue uma distribuição normal. Esse modelo dá origem à curva de distribuição de probabilidade log-normal, na forma

$$f(x) = \text{LogNorm}(x; \alpha, \beta) = \frac{1}{\beta \sqrt{2\pi}} \exp\left(-\frac{1}{2} \left(\frac{\ln(x) - \alpha}{\beta}\right)^2\right) \quad (2.54)$$

onde

$$\mu_x = \exp\left(\alpha + \frac{\beta^2}{2}\right) \quad (2.55)$$

$$\sigma_x^2 = \exp\left(2\alpha + \beta^2(\exp(\beta^2) - 1)\right) \quad (2.56)$$

A distribuição log-normal é uma distribuição contínua, univariada e bi-paramétrica. Dessa forma, é possível definir de forma independente os dois primeiros momentos da curva de distribuição, ficando os demais determinados automaticamente pela forma da Equação (2.54).

## 2.8. Extensão de Conceitos para Sistemas Multidimensionais

Apesar de todo o estudo realizado até aqui ter sido baseado em distribuições de probabilidade de uma única variável aleatória, todos os conceitos podem ser estendidos convenientemente para distribuições de múltiplas variáveis. Modelos de histogramas de probabilidade de múltiplas variáveis discretas e funções densidade de probabilidade de múltiplas variáveis contínuas serão usados muitas vezes nas próximas seções. Em particular, esses modelos são muito úteis porque permitem considerar de uma vez só e de forma integrada as múltiplas influências que as diversas variáveis aleatórias exercem umas sobre as outras.

Deve ser observado que a extensão dos conceitos de probabilidades para sistemas multidimensionais nada tem de arbitrário. Por exemplo, é comum que jogos de azar façam usos de múltiplos dados simultaneamente, de forma que cada evento (jogar os dados) resulta simultaneamente em vários valores aleatórios (os resultados em cada um dos dados). Contudo, esse exemplo pode ser convenientemente estudado por técnicas de uma única variável, admitindo-se que cada resultado aleatório é obtido independentemente dos demais (ou seja, que cada dado resulta em um valor que em nada está relacionado aos valores obtidos nos demais dados). Em outras palavras, seja uma densidade de probabilidades de múltiplas variáveis aleatórias simultâneas na forma

$$\wp(x_1, x_2, \dots, x_{NX}) = \wp(x) \quad (2.57)$$

Então, no caso considerado,

$$\wp(x_1, x_2, \dots, x_{NX}) = \wp(x_1) \wp(x_2) \dots \wp(x_{NX}) \quad (2.58)$$

Isso nem sempre é verdade. Um exemplo típico diz respeito à distribuição de pesos e tamanhos em uma população de pessoas. O evento, nesse caso, pode ser a mensuração do peso e da altura (variáveis aleatórias) de pessoas escolhidas de forma aleatória em uma certa região geográfica. Obviamente as pessoas mais altas tendem a apresentar maior peso. Portanto, a medida de uma das variáveis (altura) certamente está correlacionada à medida da outra (peso); ou seja, a distribuição de probabilidades da variável peso depende do valor especificado para a variável altura. Portanto, a Equação (2.58) não pode ser utilizada para representar o problema e o sistema bi-dimensional não pode ser compreendido a partir das respectivas distribuições unidimensionais. A consideração de outras características pessoais pode justificar o aumento da dimensão do problema (cor da pele, tipo de cabelo, cor dos olhos, etc.) e ajudam a compreender como as distribuições de probabilidade multidimensionais podem ser complexas e mal caracterizadas pelas respectivas distribuições unidimensionais. (Pense, por exemplo, na proporção de indivíduos de cabelo louro numa população indígena.)

É importante observar que, se a Equação (2.57) representa uma distribuição de probabilidades, então

$$\int_{x_1} \int_{x_2} \dots \int_{x_{NX}} \wp(x_1, x_2, \dots, x_{NX}) dx_{NX} \dots dx_2 dx_1 = \int_x \wp(x) dx = 1 \quad (2.59)$$

De forma similar, as médias e covariâncias podem ser obtidas na forma

$$\mu_i = \int x_i \wp(x) dx \quad (2.60)$$

$$\sigma_{ij}^2 = \int (x_i - \mu_i)(x_j - \mu_j) \wp(x) dx \quad (2.61)$$

onde a Equação (2.60) define os valores médios para cada uma das variáveis do problema e a Equação (2.61) define as diversas covariâncias existentes entre os diversos pares de variáveis do problema. É então conveniente definir em notação vetorial

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \dots \\ x_{NX} \end{bmatrix}, \boldsymbol{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \dots \\ \mu_{NX} \end{bmatrix}, \mathbf{V}_X = \begin{bmatrix} \sigma_{11}^2 & \sigma_{12}^2 & \dots & \sigma_{1NX}^2 \\ \sigma_{21}^2 & \sigma_{22}^2 & \dots & \sigma_{2NX}^2 \\ \dots & \dots & \dots & \dots \\ \sigma_{NX1}^2 & \sigma_{NX2}^2 & \dots & \sigma_{NXNX}^2 \end{bmatrix} \quad (2.62)$$

onde  $\mathbf{x}$  é o vetor de variáveis aleatórias,  $\boldsymbol{\mu}$  é o vetor de médias e  $\mathbf{V}_X$  é a matriz de covariâncias. Se a matriz de covariâncias é diagonal na forma

$$\mathbf{V}_X = \begin{bmatrix} \sigma_{11}^2 & 0 & \dots & 0 \\ 0 & \sigma_{22}^2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \sigma_{NXNX}^2 \end{bmatrix} \quad (2.63)$$

as variáveis flutuam de forma independente umas das outras e o sistema é formado por variáveis independentes. Caso contrário, as flutuações experimentadas por algumas variáveis influenciam as flutuações das demais. A extensão das Equações (2.58-2.63) para sistemas discretos é imediata, bastando para isso substituir os termos integrais por somas sobre o domínio discreto.

**Propriedade 2.1** - A matriz de covariâncias  $\mathbf{V}_X$  é simétrica.

A Propriedade 2.1, muito importante para aplicações práticas, como discutido ao longo dos próximos capítulos, decorre da seguinte igualdade

$$\sigma_{ij}^2 = \int_x (x_i - \mu_i)(x_j - \mu_j) \wp(x) dx = \int_x (x_j - \mu_j)(x_i - \mu_i) \wp(x) dx = \sigma_{ji}^2 \quad (2.64)$$

**Propriedade 2.2** - A matriz de covariâncias  $\mathbf{V}_X$  é positiva definida.

A Propriedade 2.2 é também muito importante para aplicações práticas, como discutido ao longo dos próximos capítulos. Para que se compreenda essa propriedade, é interessante observar o comportamento do sistema bi-dimensional na forma

$$\mathbf{x}^T \mathbf{V}_X \mathbf{x} = [x_1 \quad x_2] \begin{bmatrix} \sigma_1^2 & \sigma_{12}^2 \\ \sigma_{21}^2 & \sigma_2^2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \sigma_1^2 x_1^2 + \sigma_{12}^2 x_1 x_2 + \sigma_{21}^2 x_1 x_2 + \sigma_2^2 x_2^2 \quad (2.65)$$

A Propriedade 2.1 permite então que se escreva a Equação (2.65) na forma

$$\mathbf{x}^T \mathbf{V}_X \mathbf{x} = \sigma_1^2 x_1^2 + 2 \sigma_{12}^2 x_1 x_2 + \sigma_2^2 x_2^2 \quad (2.66)$$

Contudo, a Equação (1.51) permite escrever que

$$-\sigma_1 \sigma_2 \leq \sigma_{12}^2 \leq \sigma_1 \sigma_2 \quad (2.67)$$

de maneira que a Equação (2.66) pode ser também escrita como

$$\sigma_1^2 x_1^2 - 2\sigma_1\sigma_2 |x_1||x_2| + \sigma_2^2 x_2^2 \leq \mathbf{x}^T \mathbf{V}_X \mathbf{x} \leq \sigma_1^2 x_1^2 + 2\sigma_1\sigma_2 |x_1||x_2| + \sigma_2^2 x_2^2 \quad (2.68)$$

ou ainda

$$(\sigma_1 |x_1| - \sigma_2 |x_2|)^2 \leq \mathbf{x}^T \mathbf{V}_X \mathbf{x} \leq (\sigma_1 |x_1| + \sigma_2 |x_2|)^2 \quad (2.69)$$

o que mostra que o produto vetorial definido na Equação (2.65) resulta sempre em um número positivo, sendo identicamente nulo no caso em que  $\mathbf{x}$  é o vetor nulo. Esse resultado pode ser estendido de forma absolutamente análoga para sistemas de dimensões maiores do que 2. Por isso, a matriz de covariâncias  $\mathbf{V}_X$  é positiva definida e

$$\mathbf{x}^T \mathbf{V}_X \mathbf{x} \geq 0, \quad \forall \mathbf{x} \neq \mathbf{0} \quad (2.70)$$

**Exemplo 2.14** - Considere a distribuição exponencial bidimensional apresentada abaixo

$$\varphi(x_1, x_2) = A e^{(-x_1 - 2x_2)}$$

definida nos intervalos  $0 \leq x_1 \leq \infty$  e  $0 \leq x_2 \leq \infty$ , onde  $A$  é uma constante. Para definir o valor de  $A$  de forma apropriada, lembremos que

$$\int_0^\infty \int_0^\infty \varphi(x_1, x_2) dx_2 dx_1 = 1$$

Logo,

$$\begin{aligned} \int_0^\infty \int_0^\infty A e^{(-x_1 - 2x_2)} dx_2 dx_1 &= A \int_0^\infty e^{(-x_1)} \int_0^\infty e^{(-2x_2)} dx_2 dx_1 = \\ A \int_0^\infty e^{(-x_1)} \left[ \frac{e^{(-2x_2)}}{-2} \right]_0^\infty dx_1 &= \frac{A}{2} \int_0^\infty e^{(-x_1)} dx_1 = \frac{A}{2} \left[ \frac{e^{(-x_1)}}{-1} \right]_0^\infty = \frac{A}{2} = 1 \end{aligned}$$

Portanto,  $A = 2$ .

Para calcular os valores médios de  $x_1$  e  $x_2$ , faz-se:

$$\begin{aligned} \mu_1 &= \int_0^\infty \int_0^\infty 2x_1 e^{(-x_1 - 2x_2)} dx_2 dx_1 = 2 \int_0^\infty x_1 e^{(-x_1)} \int_0^\infty e^{(-2x_2)} dx_2 dx_1 = \\ 2 \int_0^\infty x_1 e^{(-x_1)} \left[ \frac{e^{(-2x_2)}}{-2} \right]_0^\infty dx_1 &= \int_0^\infty x_1 e^{(-x_1)} dx_1 = \left[ \frac{(x_1 + 1) e^{(-x_1)}}{-1} \right]_0^\infty = 1 \\ \mu_2 &= \int_0^\infty \int_0^\infty 2x_2 e^{(-x_1 - 2x_2)} dx_2 dx_1 = 2 \int_0^\infty e^{(-x_1)} \int_0^\infty x_2 e^{(-2x_2)} dx_2 dx_1 = \end{aligned}$$

$$2 \int_0^{\infty} e^{(-x_1)} \left[ \frac{\left(x_2 + \frac{1}{2}\right) e^{(-2x_2)}}{-2} \right]_0^{\infty} dx_1 = \frac{1}{2} \int_0^{\infty} e^{(-x_1)} dx_1 = \frac{1}{2} \left[ \frac{e^{(-x_1)}}{-1} \right]_0^{\infty} = \frac{1}{2}$$

Para calcular as variâncias, faz-se

$$\sigma_1^2 = \int_0^{\infty} \int_0^{\infty} 2(x_1 - 1)^2 e^{(-x_1 - 2x_2)} dx_2 dx_1 = 2 \int_0^{\infty} (x_1 - 1)^2 e^{(-x_1)} \int_0^{\infty} e^{(-2x_2)} dx_2 dx_1 =$$

$$2 \int_0^{\infty} (x_1 - 1)^2 e^{(-x_1)} \left[ \frac{e^{(-2x_2)}}{-2} \right]_0^{\infty} dx_1 = \int_0^{\infty} (x_1 - 1)^2 e^{(-x_1)} dx_1 = \left[ \frac{(x_1^2 + 1) e^{(-x_1)}}{-1} \right]_0^{\infty} = 1$$

$$\sigma_2^2 = \int_0^{\infty} \int_0^{\infty} 2 \left(x_2 - \frac{1}{2}\right)^2 e^{(-x_1 - 2x_2)} dx_2 dx_1 = 2 \int_0^{\infty} e^{(-x_1)} \int_0^{\infty} \left(x_2 - \frac{1}{2}\right)^2 e^{(-2x_2)} dx_2 dx_1 =$$

$$2 \int_0^{\infty} e^{(-x_1)} \left[ \frac{\left(x_2^2 + \frac{1}{4}\right) e^{(-2x_2)}}{-2} \right]_0^{\infty} dx_1 = \frac{1}{4} \int_0^{\infty} e^{(-x_1)} dx_1 = \frac{1}{4} \left[ \frac{e^{(-x_1)}}{-1} \right]_0^{\infty} = \frac{1}{4}$$

$$\sigma_{12}^2 = \int_0^{\infty} \int_0^{\infty} 2(x_1 - 1) \left(x_2 - \frac{1}{2}\right) e^{(-x_1 - 2x_2)} dx_2 dx_1 =$$

$$2 \int_0^{\infty} (x_1 - 1) e^{(-x_1)} \int_0^{\infty} \left(x_2 - \frac{1}{2}\right) e^{(-2x_2)} dx_2 dx_1 =$$

$$2 \int_0^{\infty} (x_1 - 1) e^{(-x_1)} \left[ \frac{x_2 e^{(-2x_2)}}{-2} \right]_0^{\infty} dx_1 = 0$$

de maneira que as variáveis  $x_1$  e  $x_2$  fluam de forma independente. O vetor de média e a matriz de covariâncias tomam portanto a forma

$$\boldsymbol{\mu} = \begin{bmatrix} 1 \\ 0.5 \end{bmatrix}, \mathbf{V}_x = \begin{bmatrix} 1 & 0 \\ 0 & 0.25 \end{bmatrix}$$

A curva de distribuição de probabilidades estudada é representada na Figura 2.17.

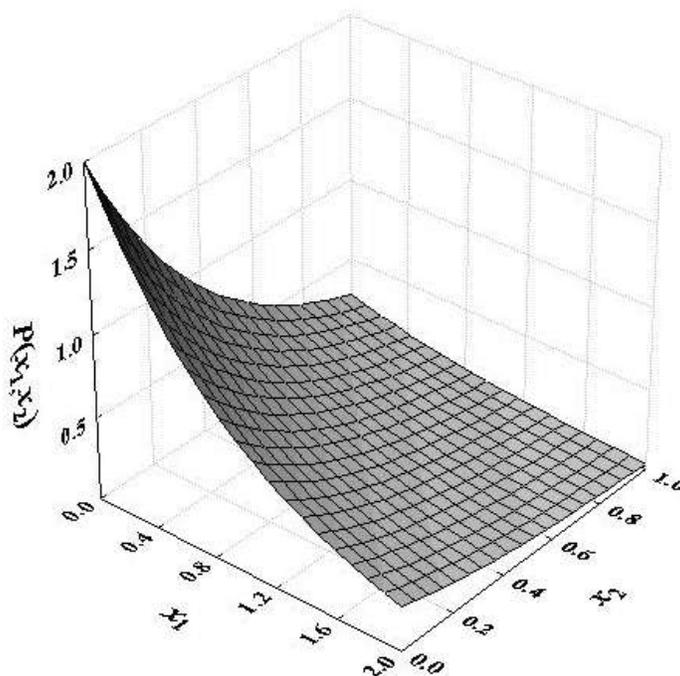


Figura 2.17 - Densidade de probabilidades do Exemplo 2.14.

## 2.9. A Distribuição Multinomial

Um exemplo muito importante de distribuição multidimensional discreta é a conhecida distribuição multinomial, que consiste na generalização da distribuição binomial, apresentada na Seção 2.1, quando três ou mais resultados ou atributos podem ser obtidos de um experimento discreto. Admite-se que um experimento é repetido  $m$  vezes e que se deseja conhecer a probabilidade de se obterem  $n_i$  resultados do tipo  $i$ ,  $i = 1 \dots NR$ , onde  $NR$  é o número de resultados possíveis. Admite-se que as probabilidades  $p_i$  de se obter o resultado  $i$  são conhecidas. Repetindo-se o procedimento apresentado na Seção 2.1, obtém-se

$$\text{Multinom}(n_1, n_2, \dots, n_{NR}; m, p_1, p_2, \dots, p_{NR}) = \left( C_{n_1}^m C_{n_2}^{m-n_1} \dots C_{n_{NR}}^{m-n_1-\dots-n_{NR-1}} \right) (p_1^{n_1} p_2^{n_2} \dots p_{NR}^{n_{NR}}) \quad (2.71)$$

onde  $C_j^i$  é a combinação de  $i$  fatores  $j$  a  $j$ , como definido na Equação (2.4).

**Exemplo 2.15** - Considere a distribuição multinomial para duas variáveis. Nesse caso

$$\text{Multinom}(n_1, n_2; m, p_1, p_2) = \left( \frac{m!}{n_1!(m-n_1)!} \frac{(m-n_1)!}{n_2!(m-n_1-n_2)!} \right) (p_1^{n_1} p_2^{n_2})$$

Como apenas dois resultados são possíveis,

$$m = n_1 + n_2$$

$$p_1 + p_2 = 1$$

Portanto

$$\text{Multinom}(n_1, n_2; m, p_1, p_2) = \frac{m!}{n_1!(m-n_1)!} p_1^{n_1} (1-p_1)^{m-n_1}$$

que é a própria distribuição binomial. Esse resultado já poderia ser esperado, dado que a distribuição binomial é exatamente aquela em que apenas dois resultados são possíveis.

## 2.10. A Distribuição Normal Multidimensional

Dentre os muitos modelos multidimensionais plausíveis que podem ser formulados para descrever flutuações de dados experimentais, não há qualquer dúvida que o mais importante deles é a generalização da curva normal para um conjunto simultâneo de muitas variáveis. As propriedades da curva normal de várias variáveis aleatórias são as mesmas já apresentadas anteriormente para a curva normal de uma única variável aleatória na Seção 2.6. No formato multivariável, a função normal ganha a forma

$$f(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^{NX}}} \frac{1}{\sqrt{\det(\mathbf{V}_x)}} \exp\left[-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T \mathbf{V}_x^{-1}(\mathbf{x}-\boldsymbol{\mu})\right] \quad (2.72)$$

**Propriedade 2.3** - A curva normal multidimensional tem o comportamento normal ao longo de qualquer direção do espaço.

Para provar a Propriedade 2.3, é conveniente admitir que

$$\mathbf{x} = t\mathbf{z} - \mathbf{z}_0 \quad (2.73)$$

onde  $t$  é uma variável e  $\mathbf{z}$  e  $\mathbf{z}_0$  são vetores especificados e constantes. A Equação (2.73) define uma reta no espaço de dimensão  $NX$ . Então, o termo entre colchetes na Equação (2.72) pode ser escrito como

$$\begin{aligned} & (t\mathbf{z} - \mathbf{z}_0 - \boldsymbol{\mu})^T \mathbf{V}_x^{-1} (t\mathbf{z} - \mathbf{z}_0 - \boldsymbol{\mu}) = \\ & [\mathbf{z}^T \mathbf{V}_x^{-1} \mathbf{z}] t^2 - 2 \left[ (\mathbf{z}_0 + \boldsymbol{\mu})^T \mathbf{V}_x^{-1} \mathbf{z} \right] t + \left[ (\mathbf{z}_0 + \boldsymbol{\mu})^T \mathbf{V}_x^{-1} (\mathbf{z}_0 + \boldsymbol{\mu}) \right] \end{aligned} \quad (2.74)$$

A Equação (2.74) pode ser rescrita como  $\alpha^2 (t - \beta)^2 + \gamma$ , onde

$$\alpha^2 = [\mathbf{z}^T \mathbf{V}_x^{-1} \mathbf{z}] \quad (2.75)$$

$$\beta = \frac{[(\mathbf{z}_0 + \boldsymbol{\mu})^T \mathbf{V}_x^{-1} \mathbf{z}]}{[\mathbf{z}^T \mathbf{V}_x^{-1} \mathbf{z}]} \quad (2.76)$$

$$\gamma = [(\mathbf{z}_0 + \boldsymbol{\mu})^T \mathbf{V}_x^{-1} (\mathbf{z}_0 + \boldsymbol{\mu})] - \alpha^2 \beta^2 \quad (2.77)$$

Portanto, a menos de uma constante de normalização que depende do valor de  $\gamma$ , definida para que a integral da Equação (2.72) satisfaça a condição da Equação (2.59), a variável  $t$  tem distribuição normal, com média igual a  $\mu_t = \beta$  e variância igual a  $\sigma_t^2 = \alpha^{-2}$ . A Propriedade 2.3 mostra, portanto, que a Equação (2.72) é uma generalização fantástica da curva normal, que mantém o comportamento normal da distribuição qualquer que seja a combinação de variáveis considerada (direções do espaço). No caso particular em que  $\mathbf{z}_0 = -\boldsymbol{\mu}$ , a variável  $t$  flutua em torno do valor zero.

**Exemplo 2.16** - Considere a distribuição normal para duas variáveis, com os seguintes vetor de médias e matriz de covariâncias

$$\boldsymbol{\mu} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \quad \mathbf{V}_x = \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix}$$

A Figura 2.18 ilustra a forma da curva de densidade normal para as duas variáveis.

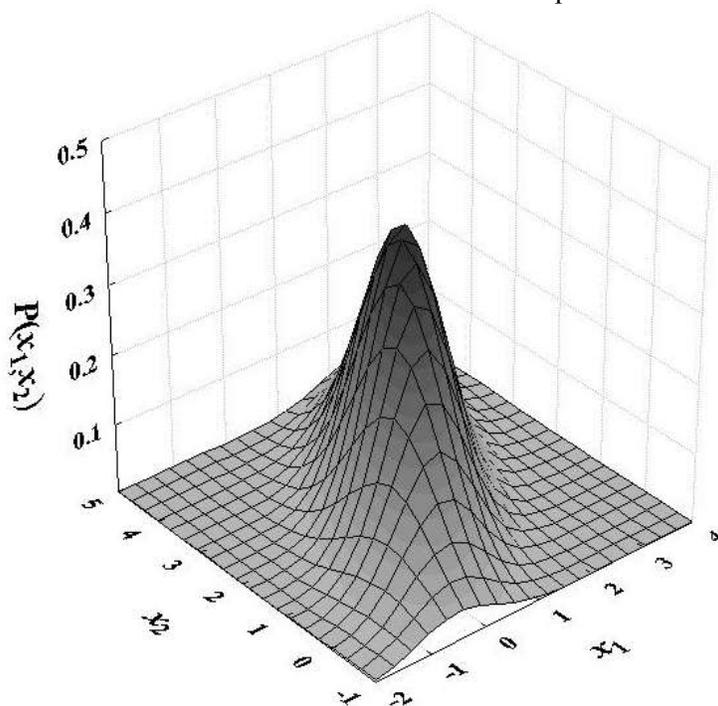


Figura 2.18 - Distribuição normal bivariada do Exemplo 2.16.

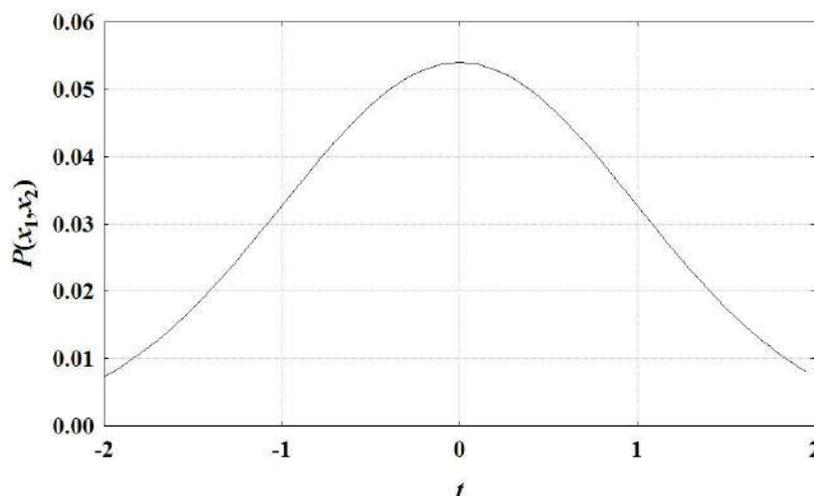


Figura 2.19 - Corte normal da Figura 2.18 ao longo da direção  $x_2 = x_1 - 1$ .

Observe a forma típica da curva normal, como um chapéu pousado sobre uma superfície plana. Observe também que a curva pode estar deslocada do centro de coordenadas e se distribuir ao longo de eixos que não são necessariamente os eixos de coordenadas usuais. A Figura 2.19 ilustra um corte da curva normal ao longo da direção  $\mathbf{x} = t \begin{bmatrix} 1 \\ 1 \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ ; ou seja,  $x_2 = x_1 - 1$ . Observe o comportamento tipicamente normal observado ao longo da direção analisada.

## 2.11. Conclusões

Foram apresentados no Capítulo 2 vários modelos probabilísticos distintos, que serão utilizados nos capítulos seguintes desse volume e nos volumes seguintes dessa série de publicações para resolver problemas práticos de análise. Cada um dos modelos apresentados admite certas hipóteses idealizadas sobre o sistema considerado. Essas hipóteses fundamentais não devem ser desprezadas durante a análise dos dados. Finalmente, os conceitos associados a distribuições de uma única variável foram estendidos para várias variáveis aleatórias, sujeitas a flutuações conjuntas.

## 2.12. Leitura Adicional

Como já discutido ao final do Capítulo 1, a literatura dedicada à apresentação de modelos probabilísticos e distribuições 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 enorme gama de diferentes distribuições de probabilidade, a análise matemática das propriedades dessas distribuições e uma prova formal da validade do Teorema do Limite Central são apresentadas 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.

### 2.13. Exercícios Sugeridos

- 1- Você acha que a curva normal pode descrever satisfatoriamente uma curva de distribuição de tamanhos de partículas muito finas? E de partículas grandes? Justifique.
- 2- Discuta se uma curva normal pode ser usada como modelo de probabilidades para descrever as flutuações de altura numa população de indivíduos. Que modificações poderiam ser introduzidas no modelo para torná-lo mais crível e representativo da realidade.
- 3- Utilizando uma planilha eletrônica ou um programa de computador:
  - a) Gere uma seqüência de 2000 pontos pseudo-aleatórios (começando com a semente 0.50) na forma:

$$X_{k+1} = 11 X_k - \text{Trunc}(11 X_k)$$

$$Y_k = X_{k+1}$$

Repare que  $X$  e  $Y$  identificam seqüências distintas de pontos deslocados no tempo

- b) Gere uma seqüência de 2000 pontos pseudo-aleatórios (começando com a semente 0.51) na forma:

$$Z_{k+1} = 11 Z_k - \text{Trunc}(11 Z_k)$$

$$W_k = Z_{k+1}$$

Repare que  $Z$  e  $W$  identificam seqüências distintas de pontos deslocados no tempo, diferentes das duas seqüências  $X$  e  $Y$  anteriores.

- c) Usando os dados anteriores, construa histogramas de probabilidade como a freqüência com que os pontos aparecem nos dez intervalos definidos pelos valores  $[0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0]$ . Comente os resultados obtidos.
  - d) Calcule os fatores de correlação entre as variáveis  $X$ ,  $Y$ ,  $Z$  e  $W$ . Comente os resultados obtidos.
  - e) Finalmente responda – os dados obtidos são aleatórios ou determinísticos ?
- 4- Utilizando a seqüência de valores anteriormente obtidos para a variável  $Z$ , e admitindo que essa seqüência de valores segue a distribuição uniforme, transforme essa seqüência de dados em outra que segue:
    - a) A distribuição exponencial;
    - b) A distribuição normal.
  - 5- Admita que o modelo normal multidimensional da Equação (2.72) é válido. Nesse caso, comente os resultados obtidos quando:

- a) A matriz de covariâncias é diagonal, como na Equação (2.63). Qual a forma particular da distribuição nesse caso? Como o resultado pode ser interpretado?
- b) O vetor de médias definido na Equação (2.62) e a matriz de covariâncias diagonal e definida na Equação (2.63) são compostos por  $NX$  valores iguais. Qual a forma particular da distribuição nesse caso? Como o resultado pode ser interpretado?