Parte IV
Noções de sistemas de fila

- Um *sistema de filas* consiste em um ou mais servidores que fornecem algum serviço para clientes que chegam. Quem chega e encontra todos os servidores ocupados entra em alguma fila.
- Grande parte das simulações de evento discreto modelam sistemas de filas do mundo real.

Sistema	**Servidor**	**Cliente**
Banco	Caixas	Clientes
Hospital	Médicos Enfermeiros Leitos	Pacientes
Computacional	CPU Dispositivos	Tarefas
Fábrica	Trabalhadores Máquinas	Produtos
Aeroporto	Pistas Portões Check-in	Aviões Viajantes
Comunicação	Linhas Circuitos Operadores	Chamadas Chamadores Mensagens

8 Componentes e Notação

- 3 componentes: processo de chegada, mecanismo de serviço e disciplina da fila.
- *Processo de chegada*: como os clientes chegam no sistema.
  - Seja A_i a duração entre a chegada do $(i-1)$-ésimo e do i-ésimo clientes.
  - Se $A_1, A_2, \ldots$ são v.a. i.i.d., então denotamos por $E[A]$ o *tempo médio entre chegadas*, e por $\lambda = 1/E[A]$ a *taxa de chegada* de clientes.
- *Mecanismo de serviço*: (i) número de servidores, (ii) fila única ou uma por servidor, e (iii) distribuição de probabilidades do tempo de atendimento.
  - Seja S_i a duração do i-ésimo atendimento de um servidor.
  - Se $S_1, S_2, \ldots$ são v.a. i.i.d., então denotamos por $E[S]$ o *tempo médio de atendimento* do servidor, e por $\omega = 1/E[A]$ a *taxa de serviço* do servidor.
- *Disciplina da fila*: regra para escolher o próximo cliente na fila. Ex.:
  - FIFO: primeiro o cliente que chegou a mais tempo.
  - LIFO: primeiro o cliente que chegou a menos tempo. Ex.: produtos perecíveis no estoque.
  - Prioridade: baseado em ordem de importância (ex.: uso da CPU, idoso no banco), ou característica do serviço requerido (ex.: descarga de caminhões mais leves).
- Notação (fila única, FIFO, chegada e atend. i.i.d.): *dist. chegada / dist. atend. / num. servidores*
  - G: distribuição qualquer.
  - M: distribuição exponencial.
  - E_k: k-Erlang (soma de k exponenciais).

9 Medidas de performance

- Sejam
  - D_i: atraso na fila do i-ésimo cliente.
  - $W_i = D_i + S_i$: tempo que cliente i passou no sistema.
  - $Q(t)$: número de clientes da fila no instante t.
  - $L(t)$: número de clientes no sistema no instante t (ou seja, $Q(t)$ mais o número de clientes sendo atendidos no instante t).
- *Atraso médio (estacionário)*:

$$d = \lim_{n\to\infty} \frac{\sum_{i=1}^{n} D_i}{n}$$

- *Tempo médio (estacionário) no sistema*:

$$w = \lim_{n\to\infty} \frac{\sum_{i=1}^{n} W_i}{n}$$

- *Média no tempo (estacionária) do número de clientes na fila*:

$$Q = \lim_{T\to\infty} \frac{\int_0^T Q(t)dt}{T}$$

- *Média no tempo (estacionária) do número de clientes no sistema*:

$$L = \lim_{T\to\infty} \frac{\int_0^T L(t)dt}{T}$$

- Para filas $G/G/s$, o *fator de utilização* é definido como:

$$\rho = \frac{\lambda}{s\omega}$$

  - $\rho < 1$ é condição necessária para a existência de d, w, Q e L.

10 Equações de conservação

- Para todo sistema de fila onde d e w existem,

$$Q = \lambda d \qquad \text{e} \qquad L = \lambda w.$$

$$w = d + E[S]$$

- Quando a distribuição do tempo entre chegadas, ou a distribuição do tempo de atendimento (ou ambas), é exponencial (ou uma variação da exponencial, como a k-Erlang), podemos obter solução analítica para as medidas de performance.
  - Ex.: para filas $M/G/1$,

$$d = \lambda \left(\frac{\text{Var}[S] + (E[S])^2}{2(1 - \lambda E[S])} \right).$$

Parte V
Revisão de probabilidade e estatística

- Probabilidade e estatística são utilizados para (1) modelar sistemas estocásticos, (2) validar o modelo, (3) escolher as distribuições de entrada, (4) gerar números aleatórios, (5) realizar análise da saída do simulador e (6) projetar experimentos.

11 Variáveis aleatórias e suas propriedades

- O conjunto de todos os resultados possíveis de um experimento é chamada *espaço amostral*, denotado por S.
  - Ex.: lançamento de moeda ($S = \{cara, coroa\}$), lançamento de dado ($S = \{1, 2, 3, 4, 5, 6\}$).
- Uma *variável aleatória* é uma função que associa um número real a cada ponto do espaço amostral.
  - Ex.: lançamento de moeda 10 vezes (v.a. número de caras), lançamento de dado 2 vezes (v.a. soma dos valores).
- Uma *função de distribuição acumulada* $F(x)$ de uma v.a. X é definida como

$$F(x) = \Pr[X \leq x], \quad \text{para todo } x \text{ real.}$$

- A função $F(x)$ tem as propriedades:
  - $0 \leq F(x) \leq 1$ para todo x.
  - $F(x)$ é não decrescente (se $x_1 < x_2$, então $F(x_1) \leq F(x_2)$).
  - $\lim_{x \to +\infty} F(x) = 1$ e $\lim_{x \to -\infty} F(x) = 0$.
- Uma v.a. X é *discreta* se pode assumir no máximo uma quantidade "contável"de valores.
  - "Contável"significa que existe uma bijeção entre este conjunto e os números naturais.
  - Toda v.a. que pode assumir uma quantidade finita de valores é discreta.
- A probabilidade de uma v.a. X assumir valor x_i é denota por $p(x_i) = \Pr[X = x_i]$ (*função de densidade de probabilidade*).
  - Portanto, $\sum_{i=1}^{\infty} p(x_i) = 1$.
  - $F(x) = \sum_{x_i \leq x} p(x_i)$.
  - Ex.: fig. L4.1 e L4.2.

- Uma v.a. X é *contínua* se existe uma função não negativa $f(x)$ (*função de densidade de probabilidade*) tal que para todo conjunto de números reais B,

$$\Pr[X \in B] = \int_B f(x)dx.$$

  - Logo, $\int_{-\infty}^{+\infty} f(x)dx = 1$.
  - $\Pr[X = x] = \Pr[X \in [x, x]] = \int_x^x f(y)dy = 0$.
  - Para todo $\Delta x > 0$,

$$\Pr[X \in [x, x + \Delta x]] = \int_x^{x+\Delta x} f(y)dy,$$

    ou seja, a probabilidade é a área abaixo de $f(x)$ no intervalo $[x, x + \Delta x]$.
    Ex.: fig. L4.3.
  - $F(x) = \Pr[X \in [-\infty, x]] = \int_{-\infty}^x f(y)dy$.
  - Utilizando o teorema fundamental do cálculo,

$$\Pr[X \in [a, b]] = \int_a^b f(x)dx = F(b) - F(a).$$

- Ex.: v.a. uniforme no intervalo $[0, 1]$.

$$f(x) = \begin{cases} 1, & 0 \le x \le 1 \\ 0, & \text{caso contrário.} \end{cases}$$

  - $F(x) = \int_{-\infty}^x f(y)dy = \int_0^1 1dx = x$.
  - $\Pr[X \in [a, b]] = F(b) - F(a) = b - a$.
  - Fig. L4.4 e L4.5.

- Se X e Y são v.a. discretas, então

$$p(x, y) = \Pr[X = x, Y = y]$$

  é chamada *função de densidade de probabilidade conjunta* de X e Y.

  - As funções de densidade de probabilidade *marginais* de X e Y são

$$p_X(x) = \sum_{\forall y} p(x, y), \text{ e } \quad p_Y(y) = \sum_{\forall x} p(x, y).$$

  - X e Y são *independentes* se $p(x, y) = p_X(x)p_Y(y)$ para todo x, y.
    Conhecer o valor de uma das v.a. não afeta a distribuição da outra.
  - Ex.: $p(x, y) = xy/27$ para $x \in \{1, 2\}$ e $y \in \{2, 3, 4\}$, zero caso contrário. Verificar independência.

$$p_X(x) = \frac{x \times 2}{27} + \frac{x \times 3}{27} + \frac{x \times 4}{27} = \frac{x}{3}$$

$$p_Y(y) = \frac{1 \times y}{27} + \frac{2 \times y}{27} = \frac{y}{9}$$

$$p_X(x)p_Y(y) = \frac{xy}{27} = p(x, y)$$

- As v.a. X e Y são *conjuntamente contínuas* se existe uma função não negativa $f(x, y)$ (chamada *função de densidade de probabilidade conjunta* de X e Y) tal que para todo par de conjuntos reais A e B,

$$\Pr[X \in A, Y \in B] = \int_B \int_A f(x, y)dxdy.$$

  - As funções de densidade de probabilidade *marginais* de X e Y são

$$f_X(x) = \int_{-\infty}^{+\infty} f(x, y)dy, \text{ e } \quad f_Y(y) = \int_{-\infty}^{+\infty} f(x, y)dx.$$

  - X e Y são *independentes* se $f(x, y) = f_X(x)f_Y(y)$ para todo x, y.
  - Ex.: $f(x, y) = 24xy$ para $x \ge 0, y \ge 0, x+y \le 1$, zero caso contrário. Verificar independência.

$$f_X(x) = \int_0^{1-x} 24xydy = 24x \left.\frac{y^2}{2}\right|_0^{1-x} = 12x(1-x)^2$$

$$f_Y(y) = 12y(1-y)^2$$

    $*$ contra-exemplo:

$$f_X\left(\frac{1}{2}\right) f_Y\left(\frac{1}{2}\right) = \left(\frac{3}{2}\right)^2 \neq f\left(\frac{1}{2}, \frac{1}{2}\right) = 6$$

- A *média* ou *valor esperado* (denotado por μ ou $E[X]$) de uma v.a. X é definido como

$$\mu = \begin{cases} \sum_{i=1}^{\infty} x_i p(x_i), & \text{se } X \text{ é discreta,} \\ \int_{-\infty}^{+\infty} x f(x)dx, & \text{se } X \text{ é contínua.} \end{cases}$$

  - "Centro de gravidade"dos dados.
  - Ex.: uniforme entre 0 e 1. $\mu = \int_0^1 x.1.dx = 1/2$.
  - Valor esperado de função de v.a. ($g(X)$):

$$E[g(X)] = \begin{cases} \sum_{i=1}^{\infty} g(x_i)p(x_i), & \text{se } X \text{ é discreta,} \\ \int_{-\infty}^{+\infty} g(x)f(x)dx, & \text{se } X \text{ é contínua.} \end{cases}$$

- Propriedades do valor esperado:

  - $E[aX + b] = aE[X] + b$ (linearidade).
  - $E[\sum_{i=1}^n c_i X_i] = \sum_{i=1}^n c_i E[X_i]$, mesmo quando as v.a. não são independentes.

- A *mediana* é o menor valor de x tal que $F(x) \ge 0.5$.

  - A mediana é menos sensível a valores extremos do que a média.
  - Ex.: 1,2,3,4,5 com prob. 0.2 tem média e mediana 3.
    Ex.: 1,2,3,4,100 com prob. 0.2 tem média 22 e mediana 3.

- A *variância* de uma v.a. X (denotada por σ^2 ou $\mathrm{Var}(X)$) é definida como

$$\sigma^2 = E[(X-\mu)^2] = E[X^2 - 2\mu X + \mu^2] = E[X^2] - \mu^2.$$

  - Medida de dispersão: quanto maior a variância, maior a chance da v.a. se afastar do valor esperado. Fig. L4.9.
  - Por que não usar $E[X-\mu]$? E usar $E[|X-\mu|]$?
  - Ex.: uniforme entre 0 e 1.

$$E[X^2] = \int_0^1 x^2.1.dx = \left.\frac{x^3}{3}\right|_0^1 = \frac{1}{3}.$$

$$\sigma^2 = \frac{1}{3} - \left(\frac{1}{2}\right)^2 = \frac{1}{12}.$$

  - O *desvio padrão* (denotado por σ) é definido como $\sigma = \sqrt{\sigma^2}$.

- Propriedades da variância:
  - $\mathrm{Var}(X) \geq 0$.
  - $\mathrm{Var}(aX + b) = a^2\mathrm{Var}(X)$.
  - $\mathrm{Var}(\sum_{i=1}^n c_i X_i) = \sum_{i=1}^n c_i^2 \mathrm{Var}(X_i)$, quando as v.a. são independentes (na verdade basta que sejam não correlacionadas).

- A *covariância* entre duas v.a. X_i e X_j (denotada por C_{ij} ou $\mathrm{Cov}(X_i, X_j)$) é definida como

$$C_{ij} = E[(X_i - \mu_i)(X_j - \mu_j)] = E[X_i X_j] - \mu_i \mu_j.$$

  - Mede dependência **linear** entre as v.a. X_i e X_j: o quanto o gráfico entre elas se aproxima de uma reta.
  - Se forem independentes, o gráfico vai se afastar de uma reta. Porém, também se afasta se houver relação não linear!
  - Ou seja, independência implica em covariância nula, mas covariância nula não implica em independência.
  - Exceção: quando a dist. é normal multivariada (marginais são normais), covariância nula implica em indepedência.
  - As covariâncias são simétricas: $C_{ij} = C_{ji}$.
  - $C_{ii} = \sigma_i^2$.
  - Se $C_{ij} > 0$ (*positivamente correlacionada*), $X_i > \mu_i$ e $X_j > \mu_j$ tendem a ocorrer juntos, bem como $X_i < \mu_i$ e $X_j < \mu_j$. Portanto, se uma v.a. tem valor alto, a outra provavelmente também tem.
  - Se $C_{ij} < 0$ (*negativamente correlacionada*), $X_i > \mu_i$ e $X_j < \mu_j$ tendem a ocorrer juntos, bem como $X_i < \mu_i$ e $X_j > \mu_j$. Portanto, se uma v.a. tem valor alto, a outra provavelmente tem valor baixo.
  - Unidade: se a v.a. está em minutos, a covariância está em minutos ao quadrado. Para facilitar a interpretação utilizamos a *correlação*.

- A *correlação* (denotada por ρ_{ij}) é definida como

$$\rho_{ij} = \frac{C_{ij}}{\sqrt{\sigma_i^2 \sigma_j^2}}.$$

  - $-1 \leq \rho_{ij} \leq 1$.
  - Se ρ_{ij} é muito próxima a 1, as variáveis são altamente positivamente correlacionadas.
  - Se ρ_{ij} é muito próxima a -1, as variáveis são altamente negativamente correlacionadas.

12 Processos estocásticos

- Um *processo estocástico* é uma coleção de v.a. ordenadas no tempo, todas definidas em um mesmo espaço amostral.
  - A saída do simulador é um processo estocástico.
  - Se a coleção é $X_1, X_2, \ldots$, temos um processo estocástico *de tempo discreto*.
  - Se a coleção é $\{X(t), t \geq 0\}$, temos um processo estocástico *de tempo contínuo*.
  - Ex.: número de clientes na fila (contínuo), custo total em cada ponto de avaliação do sistema de estoque (discreto).

- Um processo estocástico tem *covariância estacionária* se

$$\begin{array}{ll} \mu_i = \mu & \text{para } i = 1, 2, \ldots, \text{ e } -\infty < \mu < +\infty \\ \sigma_i^2 = \sigma^2 & \text{para } i = 1, 2, \ldots, \text{ e } \sigma^2 < +\infty \end{array}$$

  $C_j = \mathrm{Cov}(X_i, X_{i+j})$ independe de i, para $j = 1, 2, \ldots$

  $C_v = \mathrm{Cov}(X(t), X(t+v))$ independe de t, para $v > 0$
  - A covariância depende apenas da distância (*lag*) entre as variáveis.
  - A correlação entre X_i e X_{i+j} é então definida como

$$\rho_{i,i+j} = \frac{C_{i,i+j}}{\sqrt{\sigma_i^2 \sigma_{i+j}^2}} = \frac{C_j}{\sigma^2} = \rho_j.$$

  - Esta suposição permite a análise estatística de processos estocásticos, mas nem sempre é válida na prática (testar).
  - Frequentemente a saída do simulador não tem covariância estacionária no começo (devido à escolha das condições iniciais), mas torna-se estacionária após um tempo de execução (warmup).
    - Ex.: como a fila do banco está inicialmente vazia, $E[Q(0)] = 0$, mas $E[Q(t)] > 0$ para $t > 0$. Porém, existe prova matemática de que sistemas de filas com 1 servidor são estacionários.

13 Estimativas de médias, variâncias e correlações

- Sejam $X_1, X_2, \ldots, X_n$ v.a. i.i.d. (observações), onde a população tem média finita μ e variância finita σ^2.
- Desejamos estimar a média μ e variância σ^2 populacionais utilizando esta amostra.
- A *média amostral*

$$\overline{X} = \frac{\sum_{i=1}^n X_i}{n}$$

é um estimador não viesado ($E[\overline{X}] = \mu$) de μ.

- A *variância amostral*

$$S^2 = \frac{\sum_{i=1}^n (X_i - \overline{X})^2}{n-1}$$

é um estimador não viesado de σ^2.

$$(X_i - \overline{X})^2 = X_i^2 - 2X_i\overline{X} + \overline{X}^2$$

$$X_i\overline{X} = \frac{X_i^2 + \sum_{j\neq i} X_iX_j}{n}$$

$$E[X_i^2] = \sigma^2 + \mu^2, \quad E[X_iX_j] = \mu^2 \text{ (pois } C_{ij} = 0)$$

$$E[X_i\overline{X}] = \frac{\sigma^2 + \mu^2 + (n-1)\mu^2}{n} = \frac{\sigma^2}{n} + \mu^2$$

$$\overline{X}^2 = \frac{\sum_{i=1}^n \sum_{j=1}^n X_iX_j}{n^2} = \frac{\sum_{i=1}^n (X_i^2 + \sum_{j\neq i} X_iX_j)}{n^2}$$

$$E[X_i^2 + \sum_{j\neq i} X_iX_j] = \sigma^2 + \mu^2 + (n-1)\mu^2 = \sigma^2 + n\mu^2$$

$$E[\overline{X}^2] = \frac{\sigma^2}{n} + \mu^2 = E[X_i\overline{X}]$$

$$E[(X_i - \overline{X})^2] = \sigma^2 + \mu^2 - \left(\frac{\sigma^2}{n} + \mu^2\right) = \frac{n-1}{n}\sigma^2$$

- O quão afastado $\overline{X}$ está de μ?
  Como as v.a. X_i são independentes,

$$\text{Var}[\overline{X}] = \text{Var}\left[\frac{1}{n}\sum_{i=1}^n X_i\right] = \frac{1}{n^2}\sum_{i=1}^n \text{Var}[X_i] = \frac{\sigma^2}{n}$$

  - Ou seja, quanto maior o tamanho n da amostra, mais próximo $\overline{X}$ deve estar de μ.
  - Podemos usar o estimador de σ^2 para estimar $\text{Var}[\overline{X}]$:

$$\hat{\text{Var}}[\overline{X}] = \frac{S^2}{n} = \frac{\sum_{i=1}^n (X_i - \overline{X})^2}{n(n-1)}.$$

- Como as v.a. são independentes, temos que $\rho_{i,j} = 0$ para todo par (i, j) (não precisamos estimar).
- O que acontece com estes estimadores quando as observações não são independentes?
  - "If there are simulations with independent output data, we have never seen one."[Law]
  - Ex.: o tempo que o cliente esperou tem correlação com o tempo esperado pelo cliente anterior?
  - Vamos assumir que o conjunto de v.a. tem pelo menos covariância estacionária.
    * Neste caso, $\overline{X}$ é um estimador não viesado para μ, mas S^2 passa a ser um estimador viesado para σ^2.

$$E[S^2] = \sigma^2\left[1 - 2\frac{\sum_{j=1}^{n-1}(1 - j/n)\rho_j}{n-1}\right]$$

$$E[\text{Var}[\overline{X}]] = \frac{\sigma^2}{n}\left[1 + 2\sum_{j=1}^{n-1}(1 - j/n)\rho_j\right]$$

$$E\left[\frac{S^2}{n}\right] = \text{Var}[\overline{X}]\frac{n/a - 1}{n-1},$$

$$\text{onde } a = 1 + 2\sum_{j=1}^{n-1}(1 - j/n)\rho_j$$

    * Ou seja, se $\rho_j > 0$ (caso comum) então $E[S^2] < \sigma^2$, $a > 1$ e $E[S^2/n] < \text{Var}[\overline{X}]$. Estamos subestimando a variância!
    * As correlações podem ser estimadas com

$$\hat{\rho_j} = \frac{\hat{C_j}}{S^2}, \text{ onde}$$

$$\hat{C_j} = \frac{\sum_{i=1}^{n-j}(X_i - \overline{X})(X_{i+j} - \overline{X})}{n-j}$$

    * Com estes estimadores para as correlações podemos melhorar o estimador da variância.
    * Porém, ρ_j é bom estimador apenas quando n é grande e j é bem menor que n (viesado, grande variância, e correlacionado com outros $\rho_k, k \neq j$).
    * Note que $\hat{\rho_j}$ pode ser diferente de zero mesmo quando $\rho_j = 0$, pois $\hat{\rho_j}$ é uma v.a..

14 Intervalo de confiança e teste de hipótese para a média

- Sejam $X_1, X_2, \ldots, X_n$ v.a. i.i.d. com média μ e variância σ^2 finitas.
- *Teorema do Limite Central*: para n "suficientemente grande", a v.a.

$$Z = \frac{\overline{X} - \mu}{\sqrt{\sigma^2/n}}$$

tem distribuição aproximadamente normal com média 0 e variância 1 (*normal padrão*), independente da distribuição dos X_i's.

- Geralmente a variância não é conhecida, mas para n grande podemos substituir σ^2 por S^2.
- O *ponto crítico* z_β de uma v.a. normal padrão Z é tal que $\Pr[Z \leq z_\beta] = \beta$.
  - Ex.: se $\beta = 0.975$, então $z_\beta = 1.96$ (tabela).
- Então, (Fig. L4.15)

$$\Pr[-z_{1-\alpha/2} \leq Z \leq z_{1-\alpha/2}] = 1 - \alpha$$

  - Ex.: $\alpha = 0.05$, $\Pr[-1.96 \leq Z \leq 1.96] = 0.95$.
- *Intervalo de confiança* $100(1-\alpha)\%$ (para n grande):

$$\Pr\left[-z_{1-\alpha/2} \leq \frac{\overline{X} - \mu}{\sqrt{S^2/n}} \leq z_{1-\alpha/2}\right] = 1 - \alpha$$

$$\Pr\left[\overline{X} - z_{1-\alpha/2}\sqrt{\frac{S^2}{n}} \leq \mu \leq \overline{X} + z_{1-\alpha/2}\sqrt{\frac{S^2}{n}}\right] = 1-\alpha$$

  - Interpretação: se construirmos vários intervalos de confiança com base em amostras independentes de tamanho n (grande), $100(1-\alpha)\%$ destes intervalos conterão a média populacional μ.
- Problema: o tamanho n ("suficientemente grande") depende do quanto a distribuição dos X_i's se afasta de uma normal.
- Quando os X_i's têm distribuição normal, Z tem distribuição t com $n-1$ graus de liberdade.
  - Logo, o intervalo de confiança $100(1-\alpha)\%$ vale

$$\overline{X} \pm t_{n-1,1-\alpha/2}\sqrt{\frac{S^2}{n}},$$

    onde $t_{n-1,1-\alpha/2}$ é o ponto crítico da dist. t com $n-1$ graus de liberdade tal que $\Pr[Z \leq t_{n-1,1-\alpha/2}] = 1-\alpha$.
  - A distribuição t tem caudas maiores que a normal, fornecendo intervalos de confiança mais largos (mais conservadores).
- Na prática os X_i's raramente tem distribuição normal, logo os intervalos de confiança utilizando a distribuição t também são aproximados.
  - Mas como estes intervalos são mais conservadores, são mais recomendados que aqueles utilizando a distribuição normal.
  - Mesmo porque, a distribuição t converge para a distribuição normal com o aumento de n.
- Ex.: (matlab)

```
x = [1.2 1.5 1.68 1.89 0.95 1.49 1.58 1.55
0.5 1.09];
mean(x), tinv(0.95,9)*sqrt(var(x)/10)
```

  - $\mu = 1.34 \pm 0.24$ (90% de confiança da média ocorrer neste intervalo).
- Cobertura da média por intervalo de confiança 90% em 500 experimentos:

Distribuição	Skew	$n=5$	$n=10$	$n=20$	$n=40$
Normal	0.00	0.910	0.902	0.898	0.900
Exponencial	2.00	0.854	0.878	0.870	0.890
Chi quadrada	2.83	0.810	0.830	0.848	0.890
Lognormal	6.18	0.758	0.768	0.842	0.852
Hiper-exp	6.43	0.584	0.586	0.682	0.774

- Se a amostra $X_1, X_2, \ldots, X_n$ é i.i.d. com distribuição aproximadamente normal (ou com n grande), podemos utilizar a estatística Z para testar a hipótese $H_0 : \mu = \mu_0$.
  - Se a hipótese for verdadeira, então Z tem distribuição t com $n-1$ graus de liberdade.
  - Portanto, com nível de confiança α:
    Rejeitamos H_0 se $|Z| > t_{n-1,1-\alpha/2}$.
    "Aceitamos" H_0 caso contrário.
  - Podemos cometer dois tipos de erros: tipo 1 (rejeitar H_0 quando for verdadeira) e tipo 2 (aceitar H_0 quando for falsa).
  - Quando H_0 é verdadeira, a chance de rejeitar H_0 vale α.
    - Ou seja, temos controle sobre a chance de cometer este erro.
    - Geralmente escolheremos confiança $\alpha = 0.05$ ou $\alpha = 0.10$, antes de realizar o teste.
  - Quando H_0 é falsa, não conhecemos a distribuição de Z.
    - Ou seja, não temos controle da confiança do teste se a hipótese for falsa.
    - Neste sentido, na verdade nunca aceitamos H_0, mas apenas deixamos de rejeitar.
    - Sabemos apenas que a probabilidade de cometer o erro tipo 2 diminui com o aumento de n.
  - Ex.: com os dados do exemplo anterior (matlab), rejeitamos a hipótese de que $\mu = 1$ com nível $\alpha = 0.1$?

15 Problemas ao substituir uma distribuição pela sua média

- Para simplificar o modelo, o analista pode ficar tentado a substituir uma v.a. de entrada pelo valor esperado.
- Ex.: fila $M/M/1$ com $\lambda = 1$ e $\omega = 0.99$ (em minutos).
  - Substituindo o tempo entre chegadas pela média, teríamos intervalos fixos de 1 min entre chegadas.

– Substituindo o tempo de atendo pela média, teríamos sempre o atendimento em 0.99 min.
– Portanto, o atraso do cliente seria sempre zero.
– Entretanto,

$$d = 1 \times (0.99^2 + 0.99^2)/(1 - 0.99) = 98.01 \text{ min.}$$

Parte VI
Distribuições clássicas contínuas e discretas

- Para cada distribuição apresentaremos sua definição, uso, média ($\mu = \mathrm{E}[X]$), variância ($\sigma^2 = \mathrm{E}[(X-\mu)^2] = \mathrm{E}[X^2] - \mu^2$) e como obter números aleatórios nesta distribuição partindo da distribuição uniforme entre 0 e 1 ($U(0,1)$). Apresentaremos a expressão da acumulada apenas quando ela não for simplesmente o somatório de densidades de probabilidades.

- A "função geratriz de momentos" $M_X(t)$ de uma VA X é o valor esperado de e^{tX}.

$$M_X(t) = \mathrm{E}[e^{tX}] = \begin{cases} \sum_x e^{tx} f(x), & \text{se } X \text{ é discreta} \\ \int_{-\infty}^{+\infty} e^{tx} f(x) dx, & \text{se } X \text{ é contínua} \end{cases}$$

$$\mathrm{E}[X^r] = \left. \frac{d^r M_X(t)}{dt^r} \right|_{t=0}$$

  – Se X é uma VA e a uma constante, então

$$M_{X+a}(t) = e^{at} M_X(t)$$

$$M_{aX}(t) = M_X(at)$$

  – Se $X_1, \ldots, X_n$ são VA independentes, e $Y = \sum_{i=1}^n X_i$, então

$$M_Y(t) = M_{X_1}(t) \times \ldots \times M_{X_n}(t).$$

- Alguns somatórios utilizados:

  – Como $\sum_{k=0}^n k^2 + (n+1)^2 = \sum_{k=0}^n (k+1)^2 = \sum_{k=0}^n k^2 + 2\sum_{k=0}^n k + \sum_{k=0}^n 1$,

$$2\sum_{k=0}^{n} k = (n+1)^2 - (n+1) = n(n+1).$$

  – Como $\sum_{k=0}^n k^3 + (n+1)^3 = \sum_{k=0}^n (k+1)^3 = \sum_{k=0}^n k^3 + 3\sum_{k=0}^n k^2 + 3\sum_{k=0}^n k + \sum_{k=0}^n 1$,

$$6\sum_{k=0}^{n} k^2 = n(n+1)(2n+1).$$

  – Como $a\sum_{k=0}^n a^k = \sum_{k=0}^n a^{k+1} = \sum_{k=0}^n a^k + a^{n+1} - 1$,

$$(1-a)\sum_{k=0}^{n} a^k = 1 - a^{n+1}.$$

  – Como $\sum_{k=0}^n (k+1)a^{k+1} = \sum_{k=0}^n ka^k + (n+1)a^{n+1} = a\sum_{k=0}^n ka^k + a\sum_{k=0}^n a^k$,

$$(1-a)^2 \sum_{k=0}^{n} ka^k = a(1-a^{n+1}) - (1-a)(n+1)a^{n+1}.$$

$$\lim_{n\to\infty}\sum_{k=0}^{n} ka^k = \frac{a}{(1-a)^2}, \quad 0<a<1.$$

$$\lim_{n\to\infty}\sum_{k=0}^{n} k^2a^k = \frac{a(a+1)}{(1-a)^3}, \quad 0<a<1.$$

- Para gerar números aleatórios em algumas distribuições utilizaremos o fato de que, se uma VA X tem função de densidade acumulada $F(x)$, então $u = F(x)$ tem distribuição uniforme entre 0 e 1.
  - Assim, x pode ser gerado utilizando algum número aleatório uniforme u, pois $x = F^{-1}(u)$.
  - Prova: $F_U(u) = \Pr[U \le u] = \Pr[X \le F^{-1}(u)] = F(F^{-1}(u)) = u$ (ver figura 28.1, Jain). Além disso, $f_U(x) = dF_U/du = 1$. Logo, $u \sim U(0,1)$.
  - Para distribuições discretas, nos casos onde a expressão de $F^{-1}(u)$ é complexa, colocamos os valores de $F(x)$ em um array e procuramos o valor de x tal que $F(x) \le u < F(x+1)$.

16 Distribuições Discretas

- Se X uma v.a. discreta, então $\mathrm{E}[X] = \sum_x xp(x)$, onde $p(x)$ é a probabilidade de ocorrência de x.

16.1 Uniforme Discreta

- Uma v.a. X tem distribuição "uniforme discreta" se os n valores que pode assumir $x_1, \ldots, x_n$ tem probabilidades iguais. Portanto,

$$p(x_i) = \frac{1}{n}$$

$$F(x) = \frac{i}{n}, \quad x_i \le x < x_{i+1},\ x_0 = -\infty,\ x_{n+1} = +\infty$$

- Quando os valores de X são inteiros consecutivos $a, a+1, \ldots, b$, para $a \le b$, temos

$$F(x) = \begin{cases} 0, & x < a \\ \frac{x-a+1}{b-a+1}, & a \le x < b \\ 1, & x \ge b \end{cases}$$

$$\mu = \frac{1}{b-a+1}\sum_{k=a}^{b} k = \frac{(b+a)(b-a+1)}{2(b-a+1)} = \frac{a+b}{2},$$

$$\sigma^2 = \frac{1}{b-a+1}\sum_{k=a}^{b}\left(k - \frac{a+b}{2}\right)^2 = \frac{(b-a+1)^2-1}{12}.$$

- É utilizada quando acreditamos que o valor é igualmente provável dentro de um intervalo.
- Ex.: resultado do lançamento de um dado tem $\mu = (1+6)/2 = 3.5$ e $\sigma = \sqrt{((6-1+1)^2-1)/12} \approx 1.7$.
- Para gerar um valor de X, geramos $u \sim U(0,1)$, e retornamos $\lfloor a + (b-a+1) \times u \rfloor$.

16.2 Bernoulli

- Uma v.a. de Bernoulli pode assumir apenas dois valores 0 (falha) ou 1 (sucesso). Portanto, se p é a probabilidade de sucesso, então $1-p$ será a probabilidade de falha. Assim

$$p(x) = \begin{cases} 1-p, & \text{se } x = 0 \\ p, & \text{se } x = 1 \\ 0, & \text{caso contrário} \end{cases}$$

$$\mu = 1 \times p + 0 \times (1-p) = p$$

$$\sigma^2 = (1-p)^2 \times p + (0-p)^2 \times (1-p) = p(1-p)$$

- Ex.: no lançamento de uma moeda, podemos considerar cara como sucesso e coroa como falha. Se a moeda for justa, $p = 0.5$, $\mu = 0.5$ e $\sigma = \sqrt{0.5 \times 0.5} = 0.5$.
- Utilizando $u \sim U(0,1)$, geramos um valor para esta v.a. retornando 1 se $u \le p$, e 0 caso contrário.

16.3 Binomial

- Considere um experimento onde são coletadas n observações independentes de uma VA de Bernoulli com probabilidade de sucesso $0 < p < 1$. A VA X que é igual ao número de sucessos neste experimento é chamada de "binomial" com parâmetros p e n.
  - Como os n eventos são independentes, a probabilidade de ocorrer uma dada combinação com x sucessos é $p^x(1-p)^{n-x}$. Como existem $\binom{n}{x}$ destas combinações,

$$p(x) = \binom{n}{x} p^x (1-p)^{n-x}.$$

  - Se Y é uma VA de Bernoulli, $M_Y(t) = 1-p+e^t p$. Então,

$$M_X(t) = (1-p+e^t p)^n.$$

$$\mu = n(1-p+e^t p)^{n-1}(e^t p)|_{t=0} = np$$

$$\begin{aligned} \mathrm{E}[X^2] &= n(1-p+e^t p)^{n-1}(e^t p) \\ &\quad + n(n-1)(1-p+e^t p)^{n-2}(e^t p)^2|_{t=0} \\ &= np + n(n-1)p^2 \end{aligned}$$

$$\sigma^2 = np + n(n-1)p^2 - (np)^2 = np(1-p)$$

- Ex.: (matlab)

```
p=.9;n=10;x=1:n;bar(x, binopdf(x,n,p));grid on
p=.5;n=10;x=1:n;bar(x, binocdf(x,n,p));grid on
p=.9;n=100;r=binornd(n,p),[p,pci]=binofit(r,n)
```

- Ex.: Se a probabilidade de um computador falhar em 2 anos é de 10%, qual a probabilidade de x computadores terem falhados neste período em um laboratório com 10 computadores? Figura 3-8(b), Montgomery.
- Podemos gerar um valor para X gerando n valores independentes para uma variável de Bernoulli com probabilidade de sucesso p, e somando estes n valores.

16.4 Geométrica

- Considere um experimento onde são coletadas uma sequência de observações independentes de uma VA de Bernoulli com probabilidade de sucesso $0 < p < 1$. A VA X que é igual ao número de observações até o primeiro sucesso é chamada de "geométrica" com parâmetro p.

$$p(x) = (1-p)^{x-1}p,\ x = 1, 2, \ldots$$

$$F(x) = \sum_{k=1}^{x}(1-p)^{k-1}p = 1-(1-p)^x$$

$$M_X(t) = \sum_{x=1}^{\infty} e^{tx}(1-p)^{x-1}p = \frac{pe^t}{1-(1-p)e^t}$$

$$\mu = M'_X(t)|_{t=0} = \frac{1}{p}$$

$$\sigma^2 = M''_X(t)|_{t=0} - \mu^2 = \frac{1-p}{p^2}$$

- Ex.: (matlab)

```
p=.5;n=10;x=1:n;bar(x, geopdf(x,p));grid on
p=.5;n=10;x=1:n;bar(x, geocdf(x,p));grid on
Estimativa de p usa a média.
```

- Ex.: Se a chance de falha na transmissão de uma mensagem é de 10%, qual a probabilidade de precisarmos transmitir a mensagem x vezes? Figura 3-9.

- Esta distribuição é "sem memória", no sentido de que as observações passadas não afetam as observações futuras.
  - Ex.: Se enviamos 100 mensagens e todas elas falharam, a probabilidade de que o primeiro sucesso ocorra na mensagem 105 é $(1-p)^4p$, ou seja, igual a probabilidade de sucesso na quinta mensagem de uma nova sequência.

- Geração utilizando $F^{-1}(x)$: $x = \lceil \ln(1-u)/\ln(1-p) \rceil$. Como $1-u \sim U(0,1)$, podemos usar

$$x = \left\lceil \frac{\ln(u)}{\ln(1-p)} \right\rceil.$$

16.5 Pascal

- A distribuição de "Pascal" é uma generalização da geométrica, onde a VA X é a quantidade necessária de observações independentes (de uma VA de Bernoulli) até atingirmos r sucessos.
  - A probabilidade de uma dada combinação com r sucessos e $x-r$ falhas ($x \geq r$) é $(1-p)^{x-r}p^r$. Como a última observação é de sucesso, temos $\binom{x-1}{r-1}$ destas combinações. Portanto,

$$f(x) = \binom{x-1}{r-1}(1-p)^{x-r}p^r,\quad x = r, r+1, \ldots$$

- Seja a VA Y_1 o número de observações até o primeiro sucesso, Y_2 o número de observações entre o primeiro sucesso e o segundo sucesso, assim sucessivamente.
  - Portanto, podemos interpretar uma VA binomial negativa X como a soma $Y_1 + \cdots + Y_r$.
  - Note que as VAs Y_i são geométricas. Além disso, como a distribuição geométrica não tem memória, as VAs Y_i são independentes.

$$\mu = \sum_{i=1}^{r} \mathrm{E}[Y_i] = \frac{r}{p}$$

$$\sigma^2 = \sum_{i=1}^{r} \mathrm{E}[(Y_i - \mathrm{E}[Y_i])^2] = \frac{r(1-p)}{p^2}$$

- Ex.: Se a chance de falha na transmissão de uma mensagem é de 10%, qual a probabilidade de precisarmos fazer x transmissões na tentativa de enviar r mensagens?

- Geração: produza uma sequência de $u_i \sim U(0,1)$ até que r observações sejam menores que p. Retorne o número de observações.

16.6 Binomial Negativa

- Uma VA X tem distribuição "binomial negativa" se ela representa o número de falhas em experimentos de Bernoulli até atingirmos r sucessos.
  - A probabilidade de uma combinação com x falhas e r sucessos é $(1-p)^x p^r$. Como o último evento é sucesso, temos $\binom{x+r-1}{r-1}$ destas combinações. Assim,

$$f(x) = \binom{x+r-1}{r-1}(1-p)^x p^r.$$

  - Podemos interpretar uma VA X com distribuição "binomial negativa" como uma VA Y com distribuição de "Pascal" subtraida de r (ou seja, contabiliza apenas as falhas). Portanto,

$$\mu = \mathrm{E}[Y] - r = \frac{r(1-p)}{p}$$

$$\sigma^2 = \sigma_Y^2 = \frac{r(1-p)}{p^2}$$

- Ex.: (matlab)

```
p=.5;r=1;x=0:20;bar(x, nbinpdf(x,r,p));grid on
p=.5;r=3;x=0:20;bar(x, nbinpdf(x,r,p));grid on
p=.5;r=5;x=0:20;bar(x, nbinpdf(x,r,p));grid on

Acidentes por dia:
a=[2  3  4  2  3  1  12  8  14  31  23  1  10
7  0];
mean(a),var(a),[p,pci]=nbinfit(a)
Var. > média => não Poisson/Binomial
x=0:20;bar(x, nbinpdf(x,p(1),p(2)));grid on
```

- Geração: produzimos um valor para uma variável com distribuição de Pascal e subtraímos de r.

16.7 Poisson

- Considere um intervalo T e uma VA U uniforme em T. Se dividirmos T em k intervalos $T_1, \ldots, T_k$ de mesmo tamanho L, e gerarmos n valores para U, qual a probabilidade de X valores ocorrerem em T_i?
  - Note que o resultado será o mesmo para qualquer T_i escolhido, $i = 1, \ldots, k$.
  - Temos que X é uma VA binomial com parâmetros n e $p = 1/k$. E portanto, $\mu = n/k$.
- Se o número de intervalos de tamanho L aumentar para ak ($a > 1$), então a probabilidade de uma observação ocorrer em um determinado intervalo cai para $p = 1/(ak)$.
  - Então, para manter a média de observações nos intervalos constante ($\mu = np = \lambda$), devemos aumentar o número de observações para an.
  - Se aumentarmos indefinidamente o número de intervalos, a distribuição da VA X converge para uma distribuição de "Poisson". Ou seja,

$$f(x) = \lim_{n\to\infty} \binom{n}{x} \left(1 - \frac{\lambda}{n}\right)^{n-x} \left(\frac{\lambda}{n}\right)^{x} = \frac{\lambda^x e^{-\lambda}}{x!}$$

$$M(t) = e^{-\lambda} \sum_{x=0}^{\infty} \frac{\lambda^x e^{tx}}{x!} = e^{(e^t-1)\lambda}$$

(pois $e^x = \sum_{k=0}^{\infty} x^k/k!$ – série de Taylor)

$$\mu = M'(t)|_{t=0} = \lambda e^{(e^t-1)\lambda+t}|_{t=0} = \lambda$$

$$\sigma^2 = M''(t)|_{t=0} - \lambda^2 = \lambda e^{(e^t-1)\lambda+t}(\lambda e^t+1)|_{t=0} - \lambda^2 = \lambda$$

- Ex.: (matlab)

```
l=5;n=15;x=0:n;bar(x, poisspdf(x,l));grid on
```

- Esta distribuição é apropriada para modelar casos onde os eventos são produzidos por muitas fontes independentes, e a taxa de ocorrência de eventos é mantida.
  - Ex.: A chegada de clientes em uma fila de banco, em um determinado período do dia. Se os clientes chegam em uma taxa de $\lambda = 2$ cliente por hora, qual a probabilidade de que cheguem 4 cliente na próxima hora (figura 3-14(b))?
- Geração: produze uma sequência $u_i \sim U(0, 1)$ até que $u_0 \cdots u_{n-1} > e^{-\lambda} \geq u_0 \cdots u_n$, e retorne n. Em média, $\lambda + 1$ números aleatórios serão necessários.

16.8 Relação entre distribuições

- Figura 29.1, Jain.
- A variância da binomial é sempre menor que a média, a variância da binomial negativa é sempre maior que a média, e a variância da Poisson é sempre igual a média. Esta observação ajuda na escolha da distribuição mais apropriada para o modelo.

16.8.1 Discreta para Contínua: aproximação normal da binomial

17 Distribuições Contínuas

17.1 Uniforme Contínua

- Uma VA X com densidade de probabilidades

$$f(x) = \frac{1}{b-a}, \quad a < x < b,$$

é "uniforme contínua" com parâmetros a e b.

  - Utilizamos quando a VA é limitada e nenhuma outra informação está disponível. Ex.: tempo de acesso ao HD.

$$F(x) = \int_{u=a}^{x} \frac{d_u}{b-a} = \frac{x-a}{b-a}, \quad a < x < b$$

$$\mu = \int_{x=a}^{b} \frac{x}{b-a} dx = \frac{b^2-a^2}{2(b-a)} = \frac{a+b}{2}$$

$$\sigma^2 = \int_{x=a}^{b} \frac{(x-(a+b)/2)^2}{b-a} dx = \frac{(b-a)^2}{12}$$

- Ex.: figuras 4-8 e 4-9, Mont.
- Geração: $x = a + (b - a)u$.

17.2 Normal

- Uma VA X com densidade de probabilidade

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad -\infty < x < +\infty$$

é uma VA "normal" com média μ e desvio padrão $\sigma > 0$.

  - Esta é a distribuição mais utilizada, pois sempre que um experimento aleatório é repetido, a distribuição da VA que representa o resultado médio deste experimento tende para uma normal (com o aumento do número de repetições) ("Teorema Central do Limite").
  - É utilizada sempre que a aleatoriedade é causada por várias fontes independentes atuando aditivamente. Ex.: Erros em medidas. Média amostral de um grande número de observações independentes de uma dada distribuição.

- Ex.: Figuras 4-10, 4-11 e 4-12, Mont.
- Se X tem distribuição normal com parâmetros μ e σ, então
$$Z = \frac{X - \mu}{\sigma}$$
tem distribuição normal com parâmetros $\mu = 0$ e $\sigma = 1$ (chamada de "normal padrão").
  - Portanto, uma VA normal X com parâmetros μ e σ pode ser obtida partindo de uma normal padrão Z:
$$X = \mu + \sigma Z.$$
- Geração: a média Y de várias observações independentes $u_i \sim U(0,1)$ tende a uma distribuição normal com $\mu = \mu_{u_i} = 1/2$ e $\sigma^2 = \sigma^2_{u_i}/n = 1/(12n)$, onde n é o número de observações. Assim,
$$X = \mu + \sigma\frac{Y - 1/2}{1/\sqrt{12n}} = \mu + \sigma\frac{\sum u_i - n/2}{\sqrt{n/12}}.$$

Exponencial

- Quando os eventos ocorrem de acordo com um processo de Poisson, os tempos entre eventos tem distribuição exponencial.
- Utilizada para modelar o tempo entre chegadas de consumidores, tempo de atendimento/reparo.
- $F(x,a) = 1 - e^{-x/a}, a > 0, x \geq 0$
$$f(x,a) = e^{-x/a}/a$$
$$M(t) = \int_0^\infty e^{tx}\frac{e^{-x/a}}{a}dx = \left.\frac{e^{-x/a}e^{tx}}{at-1}\right|_{x=0}^{\infty} = \frac{1}{1-at}$$
$$\mu = -(1-at)^{-2}(-a)\big|_{t=0} = a(1-at)^{-2}\big|_{t=0} = a$$
$$\sigma^2 + a^2 = 2a^2(1-at)^{-3}\big|_{t=0} = 2a^2$$
- Tem a propriedade **sem memória**:
$$\begin{aligned}\Pr[X \geq b + c \mid X \geq b] &= \frac{\Pr[X \geq b + c]}{\Pr[X \geq b]} = \frac{e^{-(b+c)/a}}{e^{-b/a}} \\ &= e^{-c/a} = \Pr[X \geq c]\end{aligned}$$

Weibull

- Se $Y \sim$ Exponencial$(\lambda^{-1/b})$, então $X = Y^b$ é uma Weibull(a,b), para $a > 0, b > 0, x \geq 0$.
- $F(x,a,b) = 1 - e^{-(x/a)^b}$.
- **Flexível:** os parâmetros podem tornar a distribuição mais próxima de uma exponencial ou mais próxima de uma normal.
- Se $b = 1$, temos uma distribuição exponencial.
- Se $b < 1$, podemos interpretar como uma exponencial onde a taxa $(1/a)$ cai com o tempo.
- Se $b > 1$, a taxa da exponencial cresce com o tempo.

Erlang

- Quando os eventos são independentes (processo de Poisson) com taxa λ, o tempo para a ocorrência de k eventos tem distribuição Erlang com parâmetros λ e k.
- Ou seja, é a soma de k VAs exponenciais IID.
- Ex.: a soma do tempo de atendimento de k servidores idênticos consecutivos.

Gama

- A distribuição Gama é uma generalização da Erlang, onde o parâmetro k não precisa ser inteiro.

Beta

- Esta distribuição é utilizada para VAs com valores restritos a um intervalo $[a,b]$, e tem forma mais flexível que a uniforme. Ex.: proporções.
- Tem dois parâmetros positivos que determinam a forma da distribuição. Figura Wikipedia.

Cauchy

- Se $Y_1 \sim N(0,1)$ e $Y_2 \sim N(0,1)$ são independentes, então $X = Y_1/Y_2$ tem distribuição de Cauchy.

Chi-quadrada (χ^2)

- Soma dos quadrados de k VAs $N(0,1)$ IID tem distribuição χ^2 com k graus de liberdade.
- Ex.: distribuição da variância amostral de uma VA $N(0,1)$.
- Utilizada também para testar o ajuste dos dados a uma determinada distribuição.

F

- Se $Y_1 \sim \chi^2(k_i)$ e $Y_2 \sim \chi^2(k_2)$ são independentes, então $X = (Y_1/k_1)/(Y_2/k_2)$ tem distribuição F com k_1 e k_2 graus de liberdade.
- Utilizada em análise de variância e regressão.

t-Student

- Se $Y_1 \sim N(0,1)$ e $Y_2 \sim \chi^2(n)$ são independentes, então $X = Y_1/\sqrt{Y_2/n}$ tem distribuição t com k graus de liberdade.
- Utilizada em testes de hipótese sobre a média, onde a variância também é estimada através da amostra.

17.3 Relação entre distribuições

- Figura 29.2, Jain.

Parte VII
Seleção das distribuições de entrada

18 Verificando independência nos dados

- Algumas técnicas estatísticas assumem independência nos dados.
  - Antes de utilizar esta hipótese, necessário testar.
- Gráfico dos estimadores $\hat{\rho}_j$ das autocorrelações de lag $j = 1, \ldots, k$.
  - Se o estimador se afasta muito de zero, então descartamos a independência.
- Gráfico dos pares (x_i, x_{i+1}).
  - Se as observações são independentes, esperamos que os pontos estejam dispersos aleatoriamente.
  - Quando existe correlação, esperamos encontrar algo próximo a uma reta.
- Ex.: (matlab)

```
x=normrnd(zeros(100,1),1); autocorr(x)
load carsmall; x = Acceleration; autocorr(x)
scatter(x(2:end), x(1:end-1))
```

19 Distribuição empírica

- Distribuição obtida diretamente dos dados, ao invés de ajustar os dados a uma dada distribuição teórica.
- Observações: $x_1, x_2, \ldots, x_n$ (v.a. contínua).
- Vamos assumir que a distribuição acumulada é linear por partes. Fig. L6.17.
- Neste caso, a acumulada no ponto x_i vale $(i-1)/(n-1)$.

$$y = ax + b \;\Rightarrow\; \frac{i-1}{n-1} = ax_i + b, \quad \frac{i}{n-1} = ax_{i+1} + b$$

$$F(x) = \begin{cases} 0 & x < x_1 \\ \frac{i-1}{n-1} + \frac{x - x_i}{(n-1)(x_{i+1} - x_i)} & x_i \le x < x_{i+1} \\ 1 & x \ge x_n \end{cases}$$

- Desvantagens:
  - Números aleatórios gerados a partir desta distribuição não podem assumir valores que não estejam entre o menor e o maior valor observado na amostra.
    - Valores altos podem ter impacto grande na simulação. Ex.: tempo de atendimento muito alto pode gerar grande atraso na fila.
    - Alguns autores sugerem colocar uma dist. exponencial no final.
  - A média obtida através de $F(x)$ não necessariamente é igual à média da amostra.
- Em alguns casos os dados estão agrupados em intervalos (histograma).
  - Intervalos: $[a_0, a_1), [a_1, a_2), \ldots, [a_{k-1}, a_k)$.
  - n_j é o número de observ. no j-ésimo intervalo.
  - $n = n_1 + n_2 + \cdots + n_k$.
  - Acumulada em a_j vale $(n_1 + n_2 + \cdots + n_j)/n$, (vale zero em a_0).

$$F(x) = \begin{cases} 0 & x < a_0 \\ F(a_{j-1}) + \frac{(x - a_{j-1})(F(a_j) - F(a_{j-1}))}{a_j - a_{j-1}} & a_{j-1} \le x < a_j \\ 1 & x \ge a_k \end{cases}$$

- Para distribuições discretas, basta observar a proporção das observações que possuem cada valor.
- Ex.: (matlab)

```
n=100; close all; rand('seed',12345);
x=normrnd(zeros(n,1),1); ecdf(x); hold on;
x=sort(x); plot(x,normcdf(x,0,1))
```

20 Seleção de distribuição

- Quando coletamos dados sobre uma v.a. de entrada, podemos especificar a distribuição em uma das formas abaixo (em ordem de preferência):
  1. Os dados são usados diretamente na simulação (trace).
  2. Geramos uma dist. empírica, e sorteamos valores de acordo com esta distribuição.
  3. Utilizamos métodos estatísticos para determinar a distribuição teórica que melhor se ajusta aos dados. Podemos então gerar números aleatórios de acordo com esta distribuição teórica.
- Desvantagens do 1o método: (i) reproduzimos exatamente o que ocorreu no passado, e (ii) podemos não ter dados suficientes.
- Uma dist. teórica é melhor que a empírica, pois
  - A dist. teórica é mais "suave"que a empírica, principalmente para poucos dados.
  - A empírica não permite valores fora da região observada na amostra.
  - A representação da distribuição teórica é mais compacta: armazenamos apenas os parâmetros da distribuição.

    - A empírica exige o dobro dos pontos observados na amostra.
    - Por esta razão, a geração de números aleatórios é menos eficiente.
- Quando nenhuma teórica se ajusta bem aos dados, temos que utilizar a empírica.

21 Identificando o tipo da distribuição

- Inicialmente tentamos identificar o tipo da distribuição (normal, uniforme, exponencial..), para depois determinar os parâmetros da distribuição.
- Em algumas situações temos algum *conhecimento prévio* sobre a distribuição.
  - Ex.: quando os clientes chegam de forma independente com taxa constante, podemos esperar uma dist. exponencial para o tempo entre chegadas.
  - Ex.: dist. que podem assumir valores negativos (ex.: normal) não servem como tempo de atendimento.
  - Ex.: o percentual de eleitores de um candidato não pode ter dist. exponencial, pois pode assumir valores maiores que 1.
  - Na prática temos pouca informação prévia sobre a distribuição.

21.1 Histogramas

- Quando temos uma dist. contínua, o *histograma* das observações $x_1, x_2, \ldots, x_n$ fornece uma idéia da forma da distribuição.
  - Dividimos o range de valores observados (menor até o maior) em k intervalos disjuntos $[b_0, b_1), [b_1, b_2), \ldots, [b_{k-1}, b_k)$ de mesmo tamanho Δb.
  - Pode ser necessário descartar alguns outliers (valores muito baixo ou muito alto), para poder visualizar melhor a região de interesse.
  - Se h_j é a proporção das observações no j-ésimo intervalo, então plotamos
$$h(x) = \begin{cases} 0 & x < b_0 \\ h_j & b_{j-1} \leq x < b_j \\ 0 & x \geq b_k \end{cases}$$
  - Note que $h_j \approx \Pr[b_{j-1} \leq X < b_j]$, e para algum $y \in [b_{j-1}, b_j)$,
$$\int_{b_{j-1}}^{b_j} f(x)dx = \Delta b f(y).$$
    Ou seja, se Δb é pequeno e temos muitas observações, então o histograma converge para a densidade de probabilidade.
  - Comparamos então a gráfico gerado com o *shape* das dist. teóricas, ignorando os parâmetros de escala e localização.
- Uma dificuldade na construção do histograma é decidir o tamanho Δb.
  - Um Δb pequeno produz um histograma irregular, devido ao aumento da variância de h_j (pois temos menos amostras por intervalo).
  - Um Δb grande compromete o entendimento da forma, devido ao excesso de agrupamento de pontos. Ex.: podemos perder picos próximos a zero.
  - Solução heurística: teste vários tamanhos de intervalo. Não existe um guia para todo caso.
```
x=csvread('service.csv');
hist(x,5), hist(x,20), hist(x,40)
```
- Quando a dist. é discreta, colocamos uma barra para cada valor possível da v.a., ao invés de construir intervalos.
  - Temos então histogramas mais corretos para v.a. discretas, pois nenhum agrupamento é feito.
```
x=csvread('demand.csv');
a=min(x); b=max(x);
for n=a:b, y(n-a+1)=sum(x==n); end;
stem(a:b, y/length(x));
xlim([min(x)-1, max(x)+1])
```
- Quando o histograma apresenta dois ou mais picos, nenhuma dist. teórica poderá ser ajustada. F. L6.24.
  - Em alguns casos podemos dividir as observações entre as distribuições.
  - Ex.: o tempo para consertar um equipamento depende da necessidade de adquirir uma peça.
  - Sejam $f_1(x)$ e $f_2(x)$ as dist. do tempo com e sem aquisição de peça, respectivamente. Se p_1 é a proporção das peças com aquisição, então
$$f(x) = p_1 f_1(x) + (1 - p_1) f_2(x).$$

21.2 Box Plots

- Permite observar a assimetria da distribuição.
```
x=csvread('service.csv');
boxplot(x,'orientation','horizontal')
```
- A caixa indica a região entre o 1o e o 3o quartil (25% e 75% dos dados).
- A linha no centro da caixa indica a mediana (2o quartil, 50% dos dados).
  - Se a mediana não está no centro da caixa, então temos indicação de assimetria.

- São considerados outliers os pontos que estão afastados da caixa mais de 1,5 vezes a largura da caixa. Indicados com uma cruz.
- As linhas que saem das caixas indicam o restante dos pontos.

21.3 Estatísticas

- Quando temos observações $x_1, x_2, \ldots, x_n$ i.i.d., podemos utilizar estatísticas para ajudar a identificar o tipo da distribuição.
- Os valores mínimo e máximo podem sugerir a região de valores possíveis.
- Quando os estimadores para a média e a mediana são aproximadamente iguais, temos uma indicação que a dist. pode ser simétrica.
- O *skewness* ν mede a simetria de uma distribuição.

$$\hat{\nu} = \frac{\sum_{i=1}^{n}[x_i - \overline{X}]^3/n}{[S^2]^{3/2}}$$

  - Quando $\nu = 0$, a dist. é simétrica.
  - Quando $\nu < 0$, a cauda da esquerda é mais longa, mais observações no lado direito, poucos valores pequenos. Ex.: 1, 100, 101, 102, 103.
  - Quando $\nu > 0$, a cauda da direita é mais longa, mais observações no lado esquerdo, poucos valores grandes. Ex.: 1, 2, 3, 4, 100.
  - Ex.: $\nu = 2$ para a dist. exponencial.
  - Em projetos de simulação encontramos com mais frequência $\nu > 0$.

```
x=csvread('service.csv'); skewness(x)
x=csvread('demand.csv'); skewness(x)
```

- Para dist. contínuas, podemos usar o estimador do *coeficiente de variação*:

$$\hat{cv} = \frac{\sqrt{S^2}}{\overline{X}}.$$

  - A dist. exponencial tem $cv = 1$.
  - As dist. gamma e weibull tem $cv < 1$ para $\alpha > 1$, $cv = 1$ para $\alpha = 1$, e $cv > 1$ para $\alpha < 1$.
  - Note que a dist. exponencial é um caso particular das dist. gamma e weibull.
  - A dist. lognormal sempre tem a forma de uma gamma/weibull com $\alpha > 1$, mas pode assumir qualquer valor de cv.
    - Então, se a dist. tem esta forma (histograma) com $\hat{cv} > 1$, a lognormal modela melhor os dados que a gamma/weibull.
  - O cv não é útil para outras distribuições.
  - Não é bem definida para quando $\mu = 0$. Ex.: $N(0, \sigma^2)$, $U(-c, c)$.

```
x=csvread('service.csv');
min(x), max(x), mean(x), median(x),
mode(x), skewness(x), std(x)/mean(x)
```

- Podemos usar a *razão lexis* τ para diferenciar as dist. discretas poisson, binomial e binomial negativa (a geométrica é um caso particular da binomial neg.).

$$\hat{\tau} = \frac{S^2}{\overline{X}}.$$

  - $\tau = 1 \Rightarrow$ poisson.
  - $\tau < 1 \Rightarrow$ binomial.
  - $\tau > 1 \Rightarrow$ binomial negativa.

```
x=csvread('demand.csv'); autocorr(x);
scatter(x(1:end-1),x(2:end));
boxplot(x,'orientation','horizontal');
a=min(x), b=max(x),
for n=a:b, y(n-a+1)=sum(x==n); end;
stem(a:b, y/length(x));
xlim([min(x)-1, max(x)+1]);
mean(x), median(x), mode(x), skewness(x),
var(x)/mean(x)
```

22 Estimativa de parâmetros da distribuição

- As mesmas observações **i.i.d.** $x_1, x_2, \ldots, x_n$ utilizadas para deteminar a forma da dist. podem ser utilizadas para determinar os parâmetros da dist..
- Existem várias formas de estimar parâmetros. Os estimadores de *máxima verossimilhança* (MLE) são os que reunem as melhores propriedades.
- Para dist. discreta, a função de verossimilhança $L(\theta)$ é a probabilidade de observar a amostra dado que θ é o parâmetro da dist.. Como a amostra é i.i.d.:

$$L(\theta) = p_\theta(x_1)p_\theta(x_2)\cdots p_\theta(x_n).$$

- O estimador de máxima verossimilhança $\hat{\theta}$ é o valor de θ que maximiza $L(\theta)$.
  - Ou seja, é o parâmetro com maior chance de produzir a amostra observada!
- Embora a probabilidade de observar x_i seja zero para dist. contínuas, a função de verossimilhança é definida de forma análoga:

$$L(\theta) = f_\theta(x_1)f_\theta(x_2)\cdots f_\theta(x_n).$$

- Para encontrar o θ que maximiza $L(\theta)$ geralmente tiramos o logarítmo de $L(\theta)$, derivamos e igualamos a zero.

- Nem sempre este procedimento fornece uma expressão fechada, sendo necessário utilizar métodos numéricos. Ex.: gamma, weibull, beta.
- Neste caso, temos funções prontas no matlab, R, octave, maple...

- Propriedades dos estimadores MLE:
  1. Para a maioria das dist mais comuns, $\hat{\theta}$ é o único valor que maximiza $L(\theta)$.
  2. Embora $\hat{\theta}$ seja viesado para algumas dist, $\mathrm{E}[\hat{\theta}]$ converge para θ quando $n \to \infty$.
  3. O MLE de $\phi = h(\theta)$ vale $h(\hat{\theta})$.
     Ex.: como $\mathrm{Var}[\mathrm{expo}(\beta)] = \beta^2$, o MLE desta variância vale $(\overline{X})^2$.
  4. Quando n é grande, temos o seguinte intervalo de confiança para $\hat{\theta}$:

$$\hat{\theta} \pm z_{1-\alpha/2}\sqrt{\frac{\delta(\hat{\theta})}{n}},$$

     onde $\delta(\theta) = -n/\mathrm{E}[d^2(\log L(\theta))/d\theta^2]$.
     - Ex.: geométrica.

$$L(p) = \prod_{i=1}^{n} p(1-p)^{x_i}$$

$$l(p) = \log(L(p)) = n\log(p) + \log(1-p)\sum_{i=1}^{n} x_i$$

$$\frac{d^2 l(p)}{dp^2} = -\frac{n}{p^2} - \frac{\sum_{i=1}^{n} x_i}{(1-p)^2}$$

$$\mathrm{E}\left[\frac{d^2 l(p)}{dp^2}\right] = -\frac{n}{p^2} - \frac{\sum_{i=1}^{n} \mathrm{E}[x_i]}{(1-p)^2} = -\frac{n}{p^2(1-p)}$$

```
x=csvread('service.csv'); [p i]=gamfit(x)
```

- Podemos utilizar este intervalo de confiança para verificar a sensibilidade deste parâmetro na saída do simulador.
  - Simulamos com valores extremos dos intervalos e observamos o impacto nas medidas de performance.
  - Se não for muito sensível, estamos confortáveis em usar o estimador.
  - Caso contrário, deveríamos coletar mais dados para diminuir o intervalo de confiança.

23 Teste de ajuste da distribuição

- Nesta etapa verificamos quão bem a dist teórica se ajusta aos dados.
- Quando várias dist são candidatas, devemos escolher a que melhor se ajusta. Nenhuma terá um ajuste perfeito.

23.1 Métodos gráficos

23.1.1 Comparação de frequências

- Colocamos no mesmo gráfico o histograma $h(x)$ e a proporção $\int_{b_{j-1}}^{b_j} \hat{f}(x)dx$ das observações que devem ocorrer em cada intervalo $[b_{j-1}, b_j)$ de acordo com a dist teórica $\hat{f}(x)$.
  - Para dist contínuas, uma alternativa é plotar o histograma e o gráfico de $\Delta b \times \hat{f}(x)$.

```
x=csvread('service.csv'); k=20;

[h b]=hist(x,k); h=h/length(x); bar(b,h);
hold on; p=gamfit(x); db=b(2)-b(1);
y=b(1)-db/2:0.01:b(end)+db/2;
plot(y,db*gampdf(y,p(1),p(2)))

figure; bar(b,h); hold on; p=wblfit(x);
plot(y,db*wblpdf(y,p(1),p(2)))

figure; bar(b,h); hold on; p=lognfit(x);
plot(y,db*lognpdf(y,p(1),p(2)))

----

x=csvread('demand.csv'); a=min(x); b=max(x);
for n=a:b, y(n-a+1)=sum(x==n); end;
stem(a:b, y/length(x)); hold on;
xlim([min(x)-1, max(x)+1]); p=nbinfit(x);
scatter(a:b,nbinpdf(a:b,p(1),p(2)))
```

23.2 Gráficos de probabilidades

- Podemos comparar o gráfico da acumulada empírica $\hat{F}(x)$ com a acumulada da dist. ajustada $F(x)$. Figura L6.30.

```
x=csvread('service.csv'); ecdf(x); hold on;
p=gamfit(x); x=sort(x);
plot(x,gamcdf(x,p(1),p(2)))

figure; ecdf(x); hold on; p=wblfit(x);
plot(x,wblcdf(x,p(1),p(2)))

figure; ecdf(x); hold on; p=lognfit(x);
plot(x,logncdf(x,p(1),p(2)))

figure; ecdf(x); hold on; p=expfit(x);
plot(x,expcdf(x,p(1)))
```

- Como este gráfico é geralmente não linear, temos dificuldade para comparar.
  - Temos mais facilidade para comparar quando o gráfico é uma reta.
  - No gráfico *P-P plot* (prob-prob) temos os pontos $(F(x), \hat{F}(x))$ para um conjunto de valores de x.

- No gráfico *Q-Q plot* (quartil-quartil) temos os pontos (x, x') tal que $F(x) = \hat{F}(x')$, para um conjunto de probabilidades $p = F(x)$.
- Os gráficos $P-P$ e $Q-Q$ medem as diferenças nos eixos y e x do gráfico das acumuladas, respectivamente.
  - Portanto, o $P-P$ ajuda a observar diferenças no meio do gráfico, enquanto o $Q-Q$ ajuda a observar diferenças nas caudas.
- Os valores de x e p são distribuídos de tal forma que se as dist $F(x)$ e $\hat{F}(x)$ forem aproximadamente iguais, os gráficos $P-P$ e $Q-Q$ se aproximam de uma reta (mesmo que as dist tenham diferença de localização e escala).

```
x=csvread('service.csv');
p=gamfit(x); y=gamrnd(ones(10000,1)*p(1),p(2));
qqplot(x,y); [eF,x]=ecdf(x); tF=gamcdf(x,p(1),p(2));
figure; plot(tF,eF,'+',[0,1],[0,1])

p=wblfit(x); y=wblrnd(ones(10000,1)*p(1),p(2));
figure; qqplot(x,y); tF=wblcdf(x,p(1),p(2));
figure; plot(tF,eF,'+',[0,1],[0,1])

p=lognfit(x); y=lognrnd(ones(10000,1)*p(1),p(2));
figure; qqplot(x,y); tF=logncdf(x,p(1),p(2));
figure; plot(tF,eF,'+',[0,1],[0,1])

p=expfit(x); y=exprnd(ones(10000,1)*p(1));
figure; qqplot(x,y); tF=expcdf(x,p(1));
figure; plot(tF,eF,'+',[0,1],[0,1])
```

- Podemos também comparar os *box plots*.

```
x=csvread('service.csv'); n=length(x);
p=gamfit(x);  y1=gamrnd(p(1)*ones(n,1),p(2));
p=wblfit(x);  y2=wblrnd(p(1)*ones(n,1),p(2));
p=lognfit(x); y3=lognrnd(p(1)*ones(n,1),p(2));
p=expfit(x);  y4=exprnd(p(1)*ones(n,1));
boxplot([x y1 y2 y3 y4],'orientation','horizontal')
```

23.3 Teste de hipótese

- Hipótese
  H_0: $x_1, \ldots, x_n$ são observações i.i.d. da dist $\hat{F}$.
- Vantagem:
  - Forma automática de identificar diferenças grosseiras entre a dist teórica e a observada. Diferenças não provocadas pela flutuação dos dados.
  - Não depende do julgamento do avaliador.
- Desvantagens:
  - Quando temos poucos dados, o teste é pouco sensível (deixa de perceber diferenças, aceitando dist com ajuste ruim).
  - Quando temos muitos dados, grande chance de rejeitar todas as dist teóricas. Por que?
  - O teste de hipótese permite fixar a confiança α em rejeitar uma dist errada, mas não podemos fixar a confiança em aceitar uma dist correta.

23.3.1 Teste Chi-Quadrado

- Este é o teste mais utilizado para determinar se os dados observados atendem a uma determinada distruição.
- Os passos são:
  1. Preparamos um histograma com k intervalos, obtendo assim o percentual de ocorrência o_i em cada intervalo $i = 1, \ldots, k$.
  2. Com base na distruibuição que estamos testando, determinamos o percentual de ocorrências esperado e_i para cada intervalo $i = 1, \ldots, k$.
     - Ex.: dist uniforme, $e_i = 1/k$ para todo i.
  3. Calculamos a estatística
     $$D = \sum_{i=1}^{k} \frac{(o_i - e_i)^2}{e_i},$$
     que tem distruibuição chi-quadrada com $k-1$ graus de liberdade.
  4. Com significância α, se D é menor que $\chi^2_{1-\alpha,k-1}$ (fornecido em softwares ou tabelas), não podemos rejeitar a hipótese de que os números foram originados da distribuição.
- Como o e_i aparece no denominador, erros em intervalos com e_i menores têm mais peso na estatística D.
  - Portanto, o teste funciona melhor se os intervalos do histograma são escolhidos de tal forma que os e_i's sejam iguais.
  - Uma forma aproximada de resolver é agrupar cada intervalo com e_i pequeno com algum intervalo vizinho.
  - Note que no caso da distribuição uniforme, basta utilizar um histograma com intervalos de mesmo tamanho.
- Se r parâmetros da distribuição são estimados utilizando a mesma amostra, o número de graus de liberdades cai de $k-1$ para $k-1-r$.
  - Ex.: se suspeitamos que os dados formam uma normal, e estimamos a médio e o desvio padrão com base na amostra, devemos subtrair 2 graus de liberdade.
  - No caso da distribuição uniforme entre 0 e 1 nenhum parâmetro precisa ser estimado.

- Este teste é mais indicado para distribuições discretas.
  - No caso de distribuições contínuas, o teste chi-quadrado é apenas uma aproximação. E portanto exige mais observações.
  - Pois o teste agrupa (discretiza) os dados em intervalos, unindo valores com probabilidades diferentes (isto pode ser evitado em distribuições discretas).
  - Assim, o nível de significância real apenas se aplica para um número infinito de observações (intervalos).
  - Na prática (amostras finitas), reduzimos este efeito evitando a ocorrência de intervalos com poucas observações: cada intervalo com menos de 5 observações é agrupado com algum intervalo vizinho.
    - Quando o intervalo tem mais observações temos mais confiança da proporção observada.

```
x=csvread('service.csv'); p=gamfit(x);
[h pv]=chi2gof(x,'cdf',@(z)gamcdf(z,p(1),p(2)),
   'nparams',2,'alpha',0.05)

p=expfit(x); [h pv]=chi2gof(x,'cdf',
@(z)expcdf(z,p(1)), 'nparams',2,'alpha',0.05)
```

23.3.2 Teste Kolmogorov-Smirnov

- Passos:
  1. Determine a função de distribuição acumulada observada $F_o(x)$, e a função de distribuição acumulada esperada $F_e(x)$.
  2. Utilizando os n valores da amostra, calcule as estatísticas
     $$K^+ = \sqrt{n} \times \max_x \{F_o(x) - F_e(x)\}$$
     $$K^- = \sqrt{n} \times \max_x \{F_e(x) - F_o(x)\},$$
     que representam as maiores diferenças entre as distribuições para mais e para menos.
  3. Com significância α, não podemos rejeitar a hipótese de que os dados obedecem a distribuição se K^+ e K^- forem menores que $K_{1-\alpha,n}$ (obtido em tabela).
- A função de distribuição acumulada observada $F_o(x)$ é o percentual de observações com valor menor ou igual a x, ou seja:
  - Sejam $x_1 \leq x_2 \leq \ldots \leq x_n$ os elementos da amostra em ordem crescente. Então,
    $$F_o(x) = \begin{cases} 0, & \text{se } x < x_1 \\ i/n, & \text{se } x_i \leq x < x_{i+1}, \quad i = 1, \ldots, n-1 \\ 1, & \text{se } x \geq x_n \end{cases}$$
- Existe uma observação importante no cálculo de K^- (figura 27.1, Jain).
  - Toda função de distribuição acumulada é não decrescente.
  - Como $F_o(x)$ se baseia em um conjunto finito de observações, ela possui incrementos discretos (ou seja, $F_o(x)$ é constante entre observações consecutivas $x_i \leq x < x_{i+1}$).
  - Por outro lado, quando a distribuição é contínua, $F_e(x)$ será contínua.
  - Assim, o $\max_x\{F_e(x) - F_o(x)\}$ para $x_{i-1} \leq x < x_i$ ocorre imediatamente antes de x_i. Ou seja, $\max_{x_{i-1} \leq x < x_i}\{F_e(x) - F_o(x)\}$ converge para $F_e(x_i) - F_o(x_{i-1})$.
- Para a distribuição uniforme entre 0 e 1 temos que $F_e(x) = x$. Portanto,
  $$K^+ = \sqrt{n} \times \max_{i=1,\ldots,n} \left\{ \frac{i}{n} - x_i \right\}$$
  $$K^- = \sqrt{n} \times \max_{i=1,\ldots,n} \left\{ x_i - \frac{(i-1)}{n} \right\}$$
  - Ex.: tabela 27.2, Jain.
- Comparando com o teste Chi-quadrado, concluimos:
  - Ao contrário do teste Chi-quadrado, que é mais apropriado para distribuições discretas e amostras grandes, o teste K-S foi projetado para distribuições contínuas e amostras pequenas.
  - O teste K-S compara as distribuições acumuladas (teórica e observada), enquanto o teste Chi-quadrado compara as densidades de probabilidades.
  - Ao contrário do teste Chi-quadrado, o teste K-S não faz agrupamento de observações. Neste sentido, o teste K-S faz melhor uso dos dados.
  - A escolha dos tamanhos dos intervalos é um problema do teste Chi-quadrado (não existe regras bem definidas para isso). Esta escolha afeta o resultado.
  - O teste Chi-quadrado é sempre aproximado, enquanto o teste K-S é exato sempre que os parâmetros da distribuição são conhecidos.

```
x=csvread('service.csv'); p=gamfit(x);
y=gamrnd(p(1)*ones(10000,1),p(2));
[h pv]=kstest2(x,y,0.05)

p=expfit(x); y=exprnd(p(1)*ones(10000,1));
[h pv]=kstest2(x,y,0.05)
```

24 Distribuições deslocadas e truncadas

- Várias dist permitem valores muito próximos de zero, mas nem sempre isso é válido na prática.
  - Ex.: Podemos assumir que não faz sentido tempo de atendimento no banco inferior a 10 segundos.
- Podemos facilmente inserir um parâmetro de localização nas dist que não possuem, mas isso pode tornar o estimador MLE difícil de ser obtido.
  - Trocamos na função de prob x por $x - \tilde{\gamma}$.
- Uma estratégia simples é utilizar o estimador proposto em [Dubey67]:

$$\tilde{\gamma} = \frac{x_{(1)}x_{(n)} - x_{(2)}^2}{x_{(1)} + x_{(n)} - 2x_{(2)}},$$

  onde $x_{(i)}$ é a i-ésima observação em ordem crescente.
  - Subtraimos de $\tilde{\gamma}$ cada observação, e estimamos os outros parâmetros.

```
x=csvread('service.csv');p=gamfit(x),x=sort(x);
g=(x(1)*x(end)-x(2)*x(2))/(x(1)+x(end)-2*x(2)),
x=x-g; p=gamfit(x)
```

- Em alguns casos podemos assumir que a variável aleatória não assume valor maior que U.
- É fácil adaptar o gerador de números aleatórios para ficarem restritos a um intervalo (veremos depois).
  - Note que quando truncamos a dist no intervalo $[L, U]$, dividimos a densidade de probabilidade por $\int_L^U f(x)dx$.

25 Seleção de distribuição na ausência de dados

- Algumas vezes o sistema não existe, ou o tempo não é suficiente para ajustar todas as dist necessárias.
- Utilizando dist triangular:
  - Podemos perguntar a um especialista quais seriam os valores mínimo a, máximo b e mais comum c da v.a..
- Utilizando a dist beta:
  - Perguntamos ao especialista, além dos valores mínimo a, máximo b e mais comum c, qual seria o valor esperado μ da v.a..
  - Neste caso, temos os seguintes parâmetros para a dist beta:

$$\tilde{\alpha_1} = \frac{(\mu - a)(2c - a - b)}{(c - \mu)(b - a)}$$

$$\tilde{\alpha_2} = \frac{(b - \mu)\tilde{\alpha_1}}{(\mu - a)}$$

  - Note que quando média μ é maior que a moda c, a dist tem assimetria a direita.

```
x=csvread('service.csv'); a=min(x), b=max(x),
c=mode(x), u=mean(x),
a1=(u-a)*(2*c-a-b)/((c-u)*(b-a)),
a2=(b-u)*a1/(u-a), [h z]=hist(x,15);
h=h/length(x); bar(z,h); hold on; y=a:.01:b;
plot(y,(z(2)-z(1))*betapdf((y-a)/(b-a),a1,a2)/(b-a))
```

26 Verificando homogeneidade entre amostras

- O que fazer quando coletamos dados durante vários período, e achamos que pode haver diferença nas dist destes períodos?
  - Por exemplo, o movimento no banco pode depender do dia do mês.
- Kruskal-Wallis propuseram um teste de hipótese para verificar se k amostras têm a mesma dist.
  - Seja x_{ij} a j-ésima observação da amostra i, n_i o número de observações da i-ésima amostra, $n = \sum_{i=1}^{k} n_i$, $R(x_{ij})$ o rank da observação (i, j), e $R_i = \sum_{j=1}^{n_i} R(x_{i,j})$.
  - O rank de uma observação é a posição dela quando ordenamos todas as observações (todas as amostras juntas).
  - Se todas as amostras têm a mesma dist, então estatística abaixo tem dist chi-quadrada com $k-1$ graus de liberdade:

$$T = \frac{12}{n(n+1)} \sum_{i=1}^{k} \frac{R_i^2}{n_i} - 3(n+1).$$

  - Portanto, rejeitamos (com confiança α) a hipótese de que as dist são iguais se $T > \chi^2_{k-1,1-\alpha}$ (ponto crítico da chi-quadrada).

```
clear all; ni=1000;
x=[exprnd(ones(1,4*ni)) exprnd(2*ones(1,ni))];
[y i]=sort(x); R(i)=1:length(i);
R=reshape(R,ni,5)'; Ri=sum(R'); n=5*ni;
T=(12/(n*(n+1)))*(Ri*Ri')/ni - 3*(n+1),
chi2inv(1-0.05,5-1)
```

27 Ajuste para dados não homogêneos

- O que fazer quando a dist muda com o tempo?
- Uma possibilidade é particionar o tempo em intervalos e ajustar uma dist para cada intervalo.
  - Desvantagem: muitos parâmetros para ajustar (perdemos graus de liberdade).
- Se for possível modelar os dados como um *processo de Poisson não homogêneo*, basta estimar como a taxa de ocorrência do evento evolui com o tempo.
- A chegada de clientes obedeçe um processo de Poisson quando:
  1. Os clientes chegam um por vez.
  2. O número de clientes que chegam no intervalo $(t, t+s]$ independe do número de clientes que chegam no intervalo $(0, t]$.
  3. A dist do número de clientes que chegam no intervalo $(t, t+s]$ independe de t.
  - Quando clientes abandonam o sistema por estar muito cheio, estamos violando ao condição (2).
  - A condição (3) indica que a taxa de chegada é constante, o que normalmente é válido apenas para intervalos pequenos de observação.
  - Quando estas condições são satisfeitas, o número de chegadas em uma unidade de tempo tem dist Poisson, onde λ é a taxa de chegada (chegadas por unidade de tempo).
    - Além disso, o tempo entre chegadas tem dist exponencial com média $1/\lambda$.
    - Para saber a dist do número de chegadas em s unidades de tempo, basta substituir λ por $s\lambda$ na densidade de probabilidade da Poisson.
- Quando a chegada satisfaz as condições (1) e (2), mas não necessariamente a condição (3), temos um processo de Poisson não homogêneo.
  - Neste caso, a taxa de chegada $\lambda(t)$ é função do tempo t.
  - A probabilidade de observar x chegadas no intervalo $(t, t+s]$ vale

$$p(x,t,s) = \frac{e^{-b(t,s)}b(t,s)^x}{x!},$$

    onde $b(t,s) = \int_t^{t+s} \lambda(x)dx$.
- Na prática, vamos dividir o tempo de observação em intervalos e estimar uma taxa constante em cada intervalo.
  - O número de intervalos segue as mesmas recomendações utilizadas para construir um bom histograma.
  - A estimativa da taxa de chegada em cada intervalo é o número de observações neste intervalo, dividido pela duração do intervalo.