AVALIAÇÃO DE DESEMPENHO DE REDES E SISTEMAS

2021.1

Prof. Giozza.

OBJETIVOS

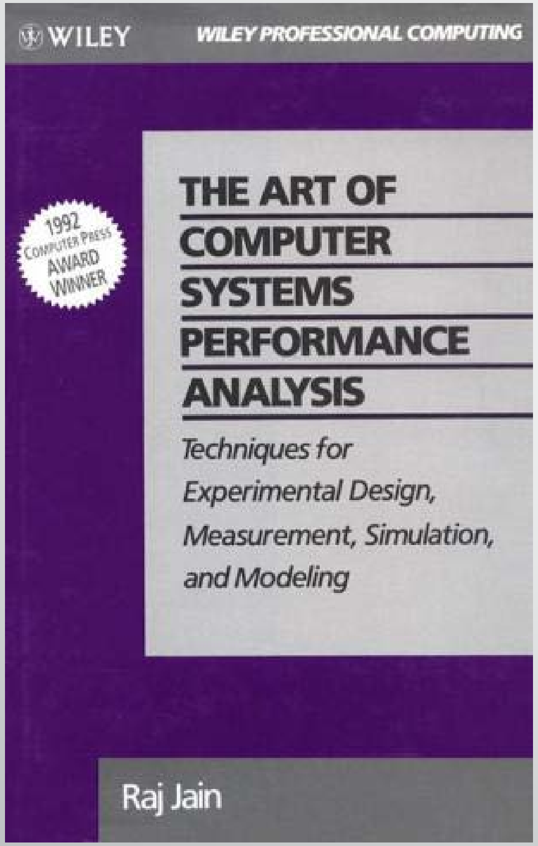
- **Familiarizar o aluno** com os **conceitos de avaliação de desempenho** de sistemas.
- **Torná-lo capaz de**, a partir do conhecimento de um sistema efetuar uma **modelagem** e utilizar **ferramentas para resolução** de modelos.
- **Na resolução de modelos** o aluno deverá poder **identificar** os **resultados** de desempenho e **analisar** o seu **impacto** na **definição/otimização** do sistema.

ROTEIRO DO CURSO

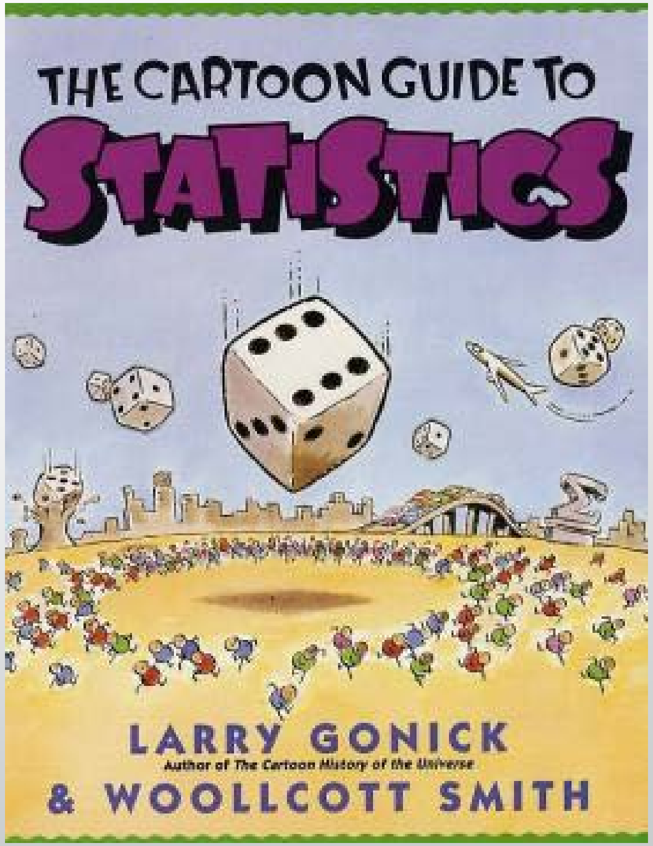
- Introdução à **MODELAGEM DE SISTEMAS**
  - O que é desempenho?
  - Métricas de desempenho
  - Técnicas de Avaliação de Desempenho
- Modelagem por **SIMULAÇÃO DE EVENTOS DISCRETOS**
  - Conceitos
  - Geração de Números e Valores Aleatórios
  - Análise de Resultados de Simulação
  - Ferramentas de Simulação

- **MODELAGEM ANALÍTICA**
  - Revisão de Probabilidade e Estatística
  - Distribuições Comumente Utilizadas
  - Introdução às Filas
- Projeto de **EXPERIMENTOS**
- Estudo de **CASOS**

BIBLIOGRAFIA



......



[Capítulo 1]

INTRODUÇÃO

Capítulo 1

Raj Jain - The Art of Computer Systems Performance Analysis. Wiley, 1991.

INTRODUÇÃO À MODELAGEM DE SISTEMAS

- O que é DESEMPENHO?
- Como MEDIR O DESEMPENHO?
- Quais são as TÉCNICAS DE AVALIAÇÃO DE DESEMPENHO?

O QUE É “DESEMPENHO”?

- Webster’s:
  - *The manner in which a mechanism performs.*
- Aurélio:
  - Mil. Conjunto de características ou de possibilidades de atuação de **uma aeronave**, tais como velocidade de cruzeiro, velocidade de pouso, capacidade de carga, autonomia de vôo, etc.
- Houaiss:
  - **Desempenho**: Maneira como atua ou se comporta alguém ou algo, avaliada em termos de eficiência, de rendimento.
  - **Performance**: conjunto de índices auferidos experimentalmente que define o alcance ideal de algo; desempenho ótimo.

COM O MEDIR O DESEMPENHO?

AUTOMÓVEL

- Velocidade máxima
- Aceleração (tempo para ir de 0 a 100 km/h)
- Espaço de frenagem a uma dada velocidade

MÉTRICAS DE DESEMPENHO?

MÉTRICAS DE DESEMPENHO de Sistemas Computacionais

- **UTILIZAÇÃO**:
  - Fração do tempo em que o recurso permanece ocupado atendendo os pedidos dos usuários
- **TEMPO DE RESPOSTA**:
  - Tempo decorrido entre o pedido e o início/conclusão da realização do serviço.
- **VAZÃO/TAXA *(THROUGHPUT)***
  - Taxa (no tempo) na qual os pedidos são atendidos (servidos) pelo sistema.

VAZÃO *(THROUGHPUT)*

- **Taxa** (no tempo) na qual os pedidos são atendidos (servidos) pelo sistema.
- EXEMPLOS
  - Sistemas em lotes (*batch*): *jobs* por segundo
  - Sistemas interativos: pedidos por segundo
  - CPUs: MIPs ou MFLOPs (instruções ou operações por segundo)
  - **Redes: pacotes por segundo (pps)** ou **bits por segundo (bps)**
  - Sistemas de Processamento de Transações: transações por segundo (TPS)

OUTRAS MÉTRICAS DE DESEMPENHO de Sistemas Computacionais

- **CONFIABILIDADE**
  - Probabilidade de erro
  - Intervalo entre erros
- **DISPONIBILIDADE**
  - Duração do defeito
  - Intervalo entre falhas

BENCHMARKS

- ***BENCHMARKING***
  - processo de **comparação** entre dois ou mais sistemas **através de medições**.
- ***BENCHMARKS***
  - **cargas de trabalho** *(workloads)* utilizadas nestas medições

BENCHMARKS POPULARES

- **“CRIVO DE ERATÓSTENES”**
  - um **algoritmo** e um **método simples e prático** para encontrar **números primos** até um certo valor limite. Segundo a tradição, foi criado pelo matemático grego Eratóstenes, o terceiro bibliotecário-chefe da Biblioteca de Alexandria. Wikipédia
- **“WHETSTONE”**
  - ***benchmark test*** *which attempts to measure the speed and efficiency at which a computer performs* ***floating-point operations****. The result of the test is given in units called kilo-whetstones-per-second or KWIPS. The* ***Whetstone*** *is a synthetic* ***benchmark*** *designed to measure the* ***behavior of scientific programs****.*
- **“LINPACK”**
  - Os benchmarks LINPACK são uma medida da **capacidade de computação de ponto flutuante** de um sistema.
- **SPEC (*Standard Performance Evaluation Corporation*)**

SPEC = *STANDARD PERFORMANCE EVALUATION CORPORATION*

- **SITE**
  - **www.specbench.org**
- Exemplos:
  - **SPEC Cloud IaaS 2018**
    [benchmark info] [published results] [order benchmark]
    *SPEC Cloud IaaS 2018 builds on the original 2016 release, updates metrics, and workloads and adds easier setup. The benchmark stresses **the provisioning, compute, storage, and network resources of infrastructure-as-a-service (IaaS) public and private cloud platforms with multiple multi-instance workloads**. SPEC selected the social media NoSQL database transaction and K-Means clustering using Cassandra and Hadoop as two significant and representative workload types within cloud computing. For use by cloud providers, cloud consumers, hardware vendors, virtualization software vendors, application software vendors, and academic researchers.*
  - **SPEC VIRT_SC 2013**
    [benchmark info] [published results] [support] [order benchmark]
    *SPEC's updated benchmark addressing **performance evaluation of datacenter servers used in virtualized server consolidation**. SPEC VIRT_SC 2013 measures the end-to-end performance of all system components including the hardware, virtualization platform, and the virtualized guest operating system and application software. The benchmark supports hardware virtualization, operating system virtualization, and hardware partitioning schemes.*

- *BENCHMARKING METHODOLOGY WG* (IETF)

BENCHMARKING METHODOLOGY WG (IETF)

- **http://www.ietf.org/html.charters/bmwg-charter.html**
  - *Benchmarking Terminology for Network Interconnection Devices (RFC 1242)*
  - *Benchmarking Terminology for LAN Switching Devices (RFC 2285)*
  - *Terminology for IP Multicast Benchmarking (RFC 2432)*
  - *Benchmarking Methodology for **Network Interconnect Devices** (RFC 2544)*
  - *Terminology for Forwarding Information Base (FIB) based Router Performance (RFC 3222)*
  - *Benchmarking Methodology for **Firewall Performance** (RFC 3511)*
  - *Methodology for **IP Multicast** Benchmarking (RFC 3918)*
  - *Terminology for Benchmarking BGP Device Convergence n the Control Plane (RFC 4098)*
  - ***IPv6** Benchmarking Methodology for Network Interconnect Devices (RFC 5180)*
  - *MPLS Forwarding Benchmarking Methodology for IP Flows (RFC 5695)*
  - ....

TÉCNICAS DE AVALIAÇÃO DE DESEMPENHO

- MEDIÇÕES
- MODELAGEM ANALÍTICA
- SIMULAÇÃO

MEDIÇÕES

- Para efetuarmos medições (como as *Benchmarks*) é preciso termos à disposição ao menos um **protótipo** do sistema.
- Normalmente é difícil comparar alternativas

MODELAGEM ANALÍTICA

- **TEORIA DAS FILAS**
  - Filas associadas a recursos
- Caracterização:
  - Processo de **chegada**
  - Processo de **atendimento**
  - Número de **servidores**
  - **Tamanho** máximo da fila
  - **Política de atendimento** da fila

- É uma **técnica aproximada**
- Aproxima a realidade por um modelo
- Se o modelo for simples e a aproximação boa, é possível **avaliar facilmente compromissos** entre alternativas

SIMULAÇÃO

- **SIMULAÇÃO DE EVENTOS DISCRETOS**
  - Cada **evento** (ex.: chegada de usuário, término de serviço, etc.) é tratado quando do instante de sua ocorrência.
- **Simula o comportamento de um sistema real**
  - Em geral, é possível construir um **modelo** muito **mais próximo da realidade** do que com a teoria das filas

CRITÉRIOS PARA SELEÇÃO DA TÉCNICA DE AVALIAÇÃO

CRITÉRIO	MODELAGEM ANALÍTICA	SIMULAÇÃO	MEDIÇÃO
Estágio	Qualquer	Qualquer	Protótipo
Tempo necessário	Pouco	Médio	Variado
Ferramentas	Analistas	Linguagens de Programação	Instrumentação
Precisão	Pouca	Moderada	Variada
Avaliação de Compromissos	Fácil	Moderada	Difícil
Custo	Baixo	Médio	Alto
Credibilidade	Baixa	Média	Alta

ERROS COMUNS E COMO EVITÁ-LOS
+ Abordagem Sistemática

Capítulo 2

Raj Jain - The Art of Computer Systems Performance Analysis. Wiley, 1991.

ERROS COMUNS

- **Ausência** de Objetivos
  - É importante que o analista compreenda o funcionamento do sistema e identifique o **problema a ser resolvido**.
- Objetivos **Tendenciosos**
  - A **função do analista** de desempenho é a de um **jurado**.
- Abordagem **Não Sistemática**

- Análise **sem Compreender o Problema**
  - "Um **problema bem formulado** já está meio resolvido!"
  - Os modelos são meios para se chegar a **conclusões** e não o resultado final.
- Escolha **incorreta** das **Métricas de Desempenho**
  - Ex.: uso de MIPS para comparar arquiteturas CISC e RISC

ERROS COMUNS (II)

- **Carga de Trabalho** *(Workload)* **Não Representativa**
- Uso de **Técnica de Avaliação Inadequada**
  - Os analistas freqüentemente têm preferência por uma dada técnica de avaliação que eles usam em todo problema de avaliação de desempenho!
- Ignorar **Parâmetros Importantes**
  - Pode tornar inúteis os resultados obtidos.

- **Ignorar Fatores Significativos**
  - Parâmetros cuja variação tem um impacto significativo no desempenho
- **Definição** Inapropriada do **Experimento**
  - Permite observar a interação de parâmetros? Experimentos *full-Factorial* ou *fractional factorial*
- **Nível** Inadequado de **Detalhe**
  - Muitos detalhes x modelo de alto nível
- **Falta** de **Análise dos Resultados**
  - Analistas de desempenho bons em medidas mas sem expertise em análise de dados
- **Análise Incorreta**
  - P. ex., observação apenas de médias, resultados de simulações muito curtas
- **Falta** de **Análise de Sensibilidade**
  - Importância relativa dos diversos parâmetros?

ERROS COMUNS (III)

- **Tratamento Inadequado** de **Pontos Fora da Curva**
  - Ocorrência possível em um sistema real?
- **Assumir** que **Não Haverá Mudanças**
  - Mesma carga e comportamento do sistema no “futuro”?
- **Ignorar** a **Variabilidade**
  - Observar a média é suficiente?
- **Análise** Muito **Complicada**
  - É melhor começar com modelos ou experimentos simples, obter alguns resultados e, depois, introduzir as complicações.

- **Apresentação Imprópria** dos **Resultados**
  - A métrica correta para medir o desempenho de um analista não é o número de análises efetuadas, mas o número de análises que ajudaram na tomada de decisão.
- **Ignorar Aspectos Sociais**
  - Forma de transmitir os resultados de acordo com a audiência
- **Omitir Hipóteses e Limitações**
  - P. ex., estender conclusões para contextos não avaliados

ABORDAGEM SISTEMÁTICA

1. Estabeleça os **Objetivos** e Defina o **Sistema**
2. Liste os **Serviços** e suas **Respostas**
3. Selecione as **Métricas**

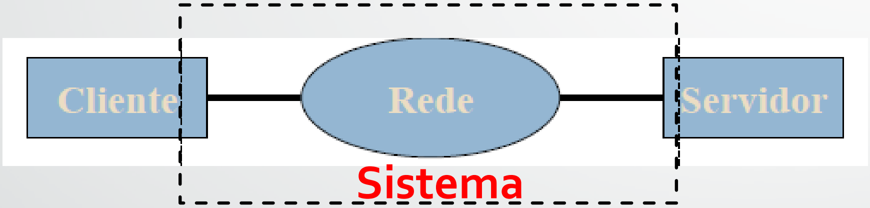
   Em geral estão associadas com velocidade, confiabilidade e disponibilidade dos serviços.
4. Liste os **Parâmetros** (do **Sistema** e da **Carga**)
5. Selecione os **Fatores a Serem Estudados**
6. Selecione a **Técnica de Avaliação**
7. Selecione a **Carga de Trabalho**
8. Planeje os **Experimentos**

   Diferentes combinações de fatores
9. Analise e Interprete os **Dados/Resultados**

   A análise produz apenas resultados e não conclusões!
10. Apresente os **Resultados** e **Conclusões**

ESTUDO DE CASO

- **Problema**:
  - Em um sistema distribuído, nas RPCs (*Remote Procedure Call*) os chamadores ficam bloqueados.
- Definição do **Sistema**:



  - **Canal**: *Procedure* ou *Pipe*
    - Em uma RPC, o programa chamador fica bloqueado, i.e., espera até que o procedimento chamado seja completado e o resultado retornado;
    - Em uma *Pipe*, o chamador não fica bloqueado, i.e., a execução da *pipe* ocorre concorrentemente com a contínua execução do chamador, com os resultados retornando de maneira assíncrona

ESTUDO DE CASO: SERVIÇOS

- Dois **tipos de chamadas** do canal: *Remote Procedure Call* e *Pipe Remoto*
  - Os **recursos usados** pelas **chamadas** dependem:
    - do número de parâmetros passados
    - da ação necessária sobre estes parâmetros.
- **Serviços**:
  - **Transferência de Dados Curtos**
  - **Transferência de Dados Longos**
- **Respostas**:
  - O estudo será limitado à **operação correta** (i.e, sem erros)

ESTUDO DE CASO: MÉTRICAS

- **Métricas**:
  - **Tempo** decorrido **por chamada**
  - **Taxa máxima de chamadas** por **unidade de tempo**
  - **Tempo** de execução da **CPU local** por **chamada**
  - **Tempo** de execução da **CPU remota** por **chamada**
  - **Número de *bytes* enviados** no **enlace** por **chamada**

ESTUDO DE CASO: PARÂMETROS

- **Parâmetros do Sistema** que afetam o desempenho:
  - **Velocidade** da **CPU Local**
  - **Velocidade** da **CPU Remota**
  - Velocidade (**vazão/taxa de transmissão**) da **Rede**
  - *Overhead* **do S.O.** para a **interface com os Canais**
  - *Overhead* **do S.O.** para **a interface com a Rede**
  - **Confiabilidade** da **Rede** (probabilidade de retransmissões)

ESTUDO DE CASO: PARÂMETROS (II)

- **Parâmetros de Carga** que afetam o desempenho:
  - **Intervalo de Tempo** entre **Chamadas Sucessivas**
  - **Número** e **Tamanho dos Parâmetros** das **Chamadas**
  - **Número** e **Tamanho dos Resultados**
  - **Tipo** do **Canal**
  - **Outras Cargas** nas **CPUs Local e Remota**
  - **Outras Cargas** na **Rede**

ESTUDO DE CASO: FATORES

- **Tipo do Canal**:
  - *Pipes* remotos e **RPCs**
- **“Velocidade” (vazão/taxa de tarda Rede**:
  - Duas localizações para os *hosts* remotos: **curta distância** (Campus) e **longa distância**
- **Tamanho dos Parâmetros de Chamada a serem transferidos**:
  - Dois níveis: **pequenos** e **grandes**
- **Número de Chamadas Consecutivas (n)**
  - **n={1, 2, 4, 8, 16, 32, 64, 128, 256, 512 e 1024}**
- Os demais parâmetros permanecerão fixos.

ESTUDO DE CASO: TÉCNICA DE AVALIAÇÃO

- Dado que **existem protótipos** dos dois sistemas de canal: **MEDIÇÕES**
- **MODELAGEM ANALÍTICA** será usada para justificar a consistência dos valores medidos para diferentes parâmetros.

ESTUDO DE CASO: CARGA DE TRABALHO

- A **carga** consistirá de um **programa sintético** gerando os **tipos de chamadas especificadas**.
- O programa irá também **monitorar** os **recursos** consumidos e **registrar** os **resultados** medidos.
  - Pedidos nulos serão utilizados para avaliar os recursos consumidos com a monitoração e registro.

ESTUDO DE CASO:
EXPERIMENTO E ANÁLISE DE DADOS

- **Projeto do Experimento**:
  - Um **projeto fatorial completo** com $2^3 * 11 = $ **88 experimentos** serão utilizados para o estudo inicial.
    - **3 fatores** (canal, velocidade, tamanho) de **dois níveis cada**;
    - **11 quantidades** de chamadas consecutivas.
- **Análise e Interpretação dos Dados**:
  - Será utilizada a **Análise da Variância** para quantificar os efeitos dos três primeiros fatores e **Regressão** para quantificar os efeitos do número *n* de chamadas consecutivas.

ESTUDO DE CASO: APRESENTAÇÃO DOS RESULTADOS

- **Apresentação dos Resultados e Conclusões**:
  - Os **resultados finais** serão **plotados** em **função do tamanho do bloco** ***n*.**
- Este estudo foi efetuado por Glasser e relatado numa Tese de Mestrado em 1987.

SELEÇÃO DE TÉCNICAS E MÉTRICAS

Capítulo 3

Raj Jain - The Art of Computer Systems Performance Analysis. Wiley, 1991.

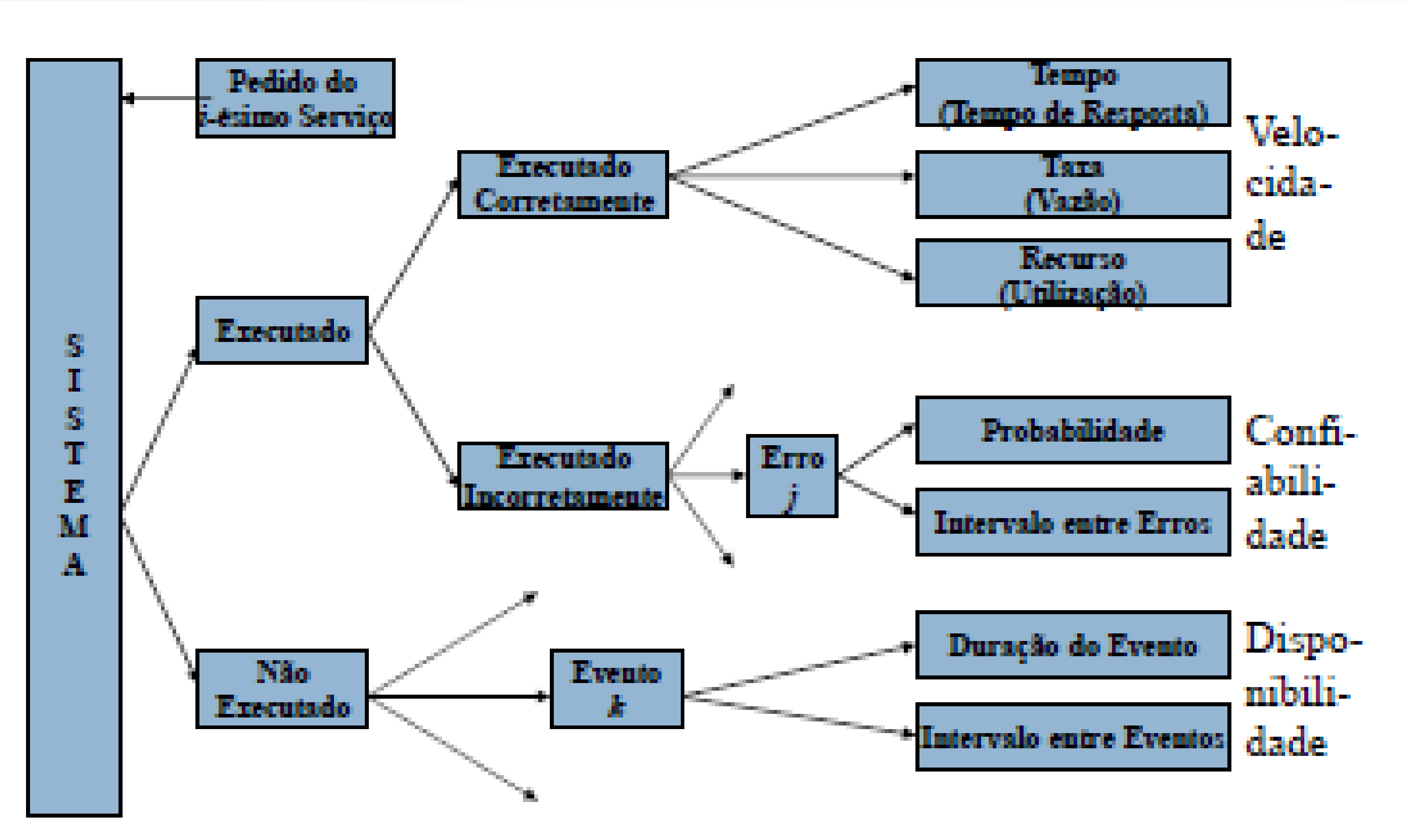
CRITÉRIOS PARA SELEÇÃO DA TÉCNICA DE AVALIAÇÃO

Critério	Modelagem analítica	Simulação	Medição
Estágio	Qualquer	Qualquer	Protótipo
Tempo necessário	Pouco	Médio	Variado
Ferramentas	Analistas	Linguagens de Programação	Instrumentação
Precisão	Pouca	Moderada	Variada
Avaliação de Compromissos	Fácil	Moderada	Difícil
Custo	Baixo	Médio	Alto
Credibilidade	Baixa	Média	Alta

VALIDAÇÃO

- Qualquer Resultado de Avaliação é Suspeito até que seja **Validado!**
- **REGRAS**
  - Não acredite em **resultados de Simulações** até que eles tenham sido validados por Modelagem Analítica ou Medições.
  - Não acredite em **resultados de um Modelo Analítico** até que eles tenham sido validados por Simulações ou Medições.
  - Não acredite em **resultados de Medições** até que eles tenham sido validados por Simulações ou Modelagem Analítica.

POSSÍVEIS RESULTADOS DE UM PEDIDO DE SERVIÇO



MÉTRICAS DE DESEMPENHO

- Para algumas métricas bastam os seus **valores médios**.
  - Mas, não deve ser desprezado o **efeito da variabilidade**.
- Métricas **Individuais** e **Globais**
  - **Utilidade de cada usuário** (e.g., tempo de resposta) ou **utilidade do sistema** como um todo (e.g., utilização de recurso, confiabilidade e disponibilidade).
- **Critérios de seleção** de um subconjunto de métricas:
  - **Baixa variabilidade** (reduz repetições para alcançar um certo nível de confiança; evita-se frações de variáveis)
  - **Não redundância** (comprimento médio de fila = tempo de espera médio)
  - **Completeza** (todas saídas possíveis devem ser refletidas no conjunto de métricas)

ESTUDO DE CASO: Comparação entre Algoritmos de Controle de Congestionamento

- **Sistema**: Rede (sistemas finais interconectados via sistemas intermediários)
- **Serviço**: Transmissão de Pacotes
- **Problema**: congestionamento quando um número de pacotes esperando num sistema intermediário excede a capacidade de buferização do sistema e alguns pacotes devem ser descartados
- **Resultados**:
  - Alguns pacotes são entregues **em ordem** ao destino correto;
  - Alguns pacotes são entregues **fora de ordem** ao destino;
  - Alguns pacotes são entregues **mais de uma vez** ao destino (duplicatas);
  - Alguns pacotes são descartados (**perdidos**) no meio do caminho.

ESTUDO DE CASO: Comparação entre Algoritmos de Controle de Congestionamento

- **Métricas** diretas relativas aos **pacotes entregues em ordem**:
  1. **Tempo de resposta**: atraso na rede para pacotes individuais
  2. **Vazão**: número de pacotes por unidade de tempo
  3. **Tempo de processamento por pacote** no sistema **origem**
  4. **Tempo de processamento por pacote** no sistema **destino**
  5. **Tempo de processamento por pacote** nos sistemas **intermediários**

ESTUDO DE CASO:
Comparação entre Algoritmos de Controle de Congestionamento

- Acrescentando **outras métricas**:

6. **Variabilidade** **do tempo de resposta** (retransmissões desnecessárias)
7. **Probabilidade** de chegada **fora de ordem** (*overhead* adicional)
8. **Probabilidade** de **duplicação** de pacotes
9. **Probabilidade** de **perda** de pacotes
10. **Probabilidade** de **desconexão** (devido a erros/retransmissões excessivas)
11. **Justiça** (função da variabilidade de vazões entre os usuários)

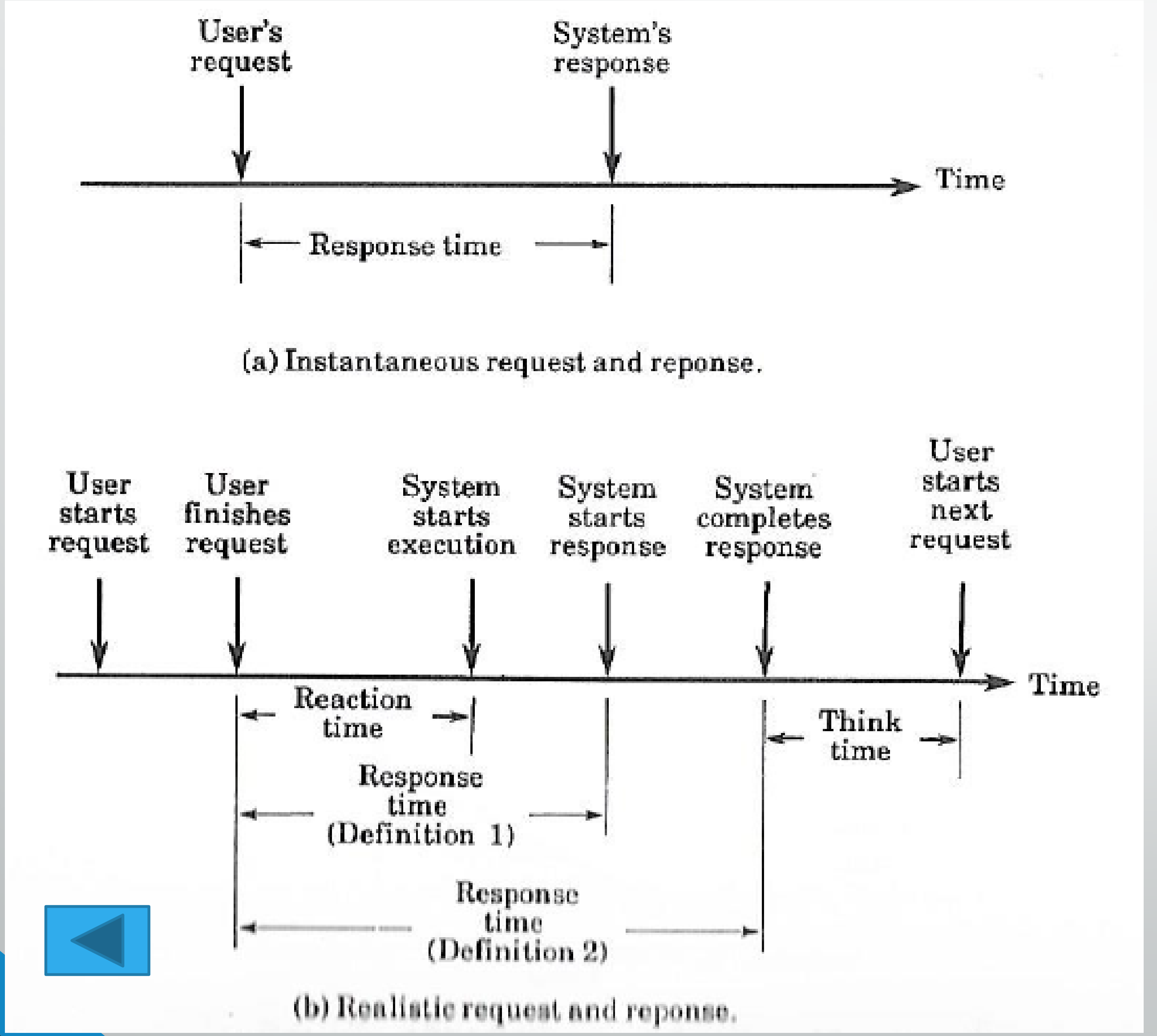
$$f(x_1, x_2, \ldots, x_n) = \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n \sum_{i=1}^{n} x_i^2}$$

MÉTRICAS COMUMENTE UTILIZADAS

- Tempo de Resposta
- Vazão
- Eficiência

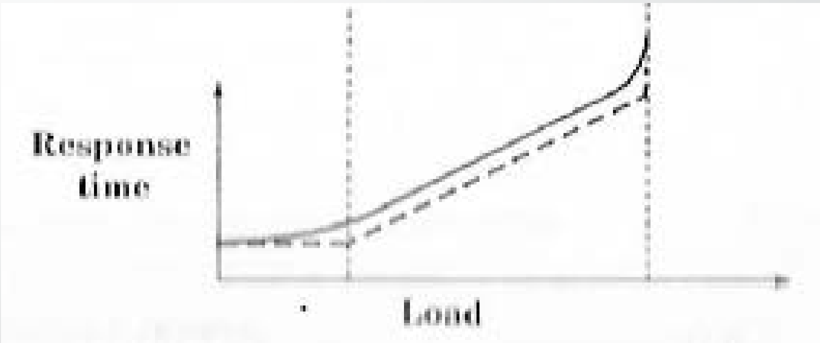
- Outras Métricas

TEMPO DE RESPOSTA



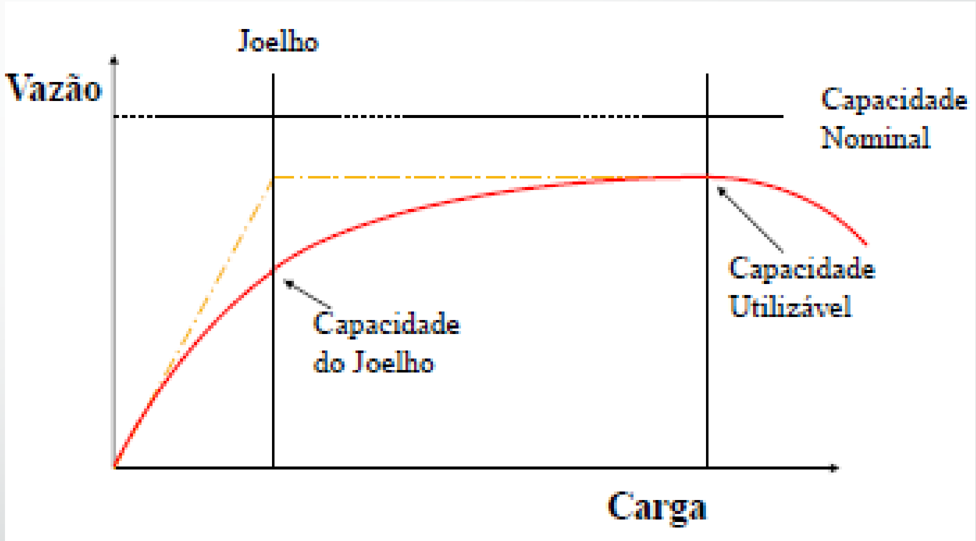
(a) Instantaneous request and reponse.

(b) Realistic request and reponse.



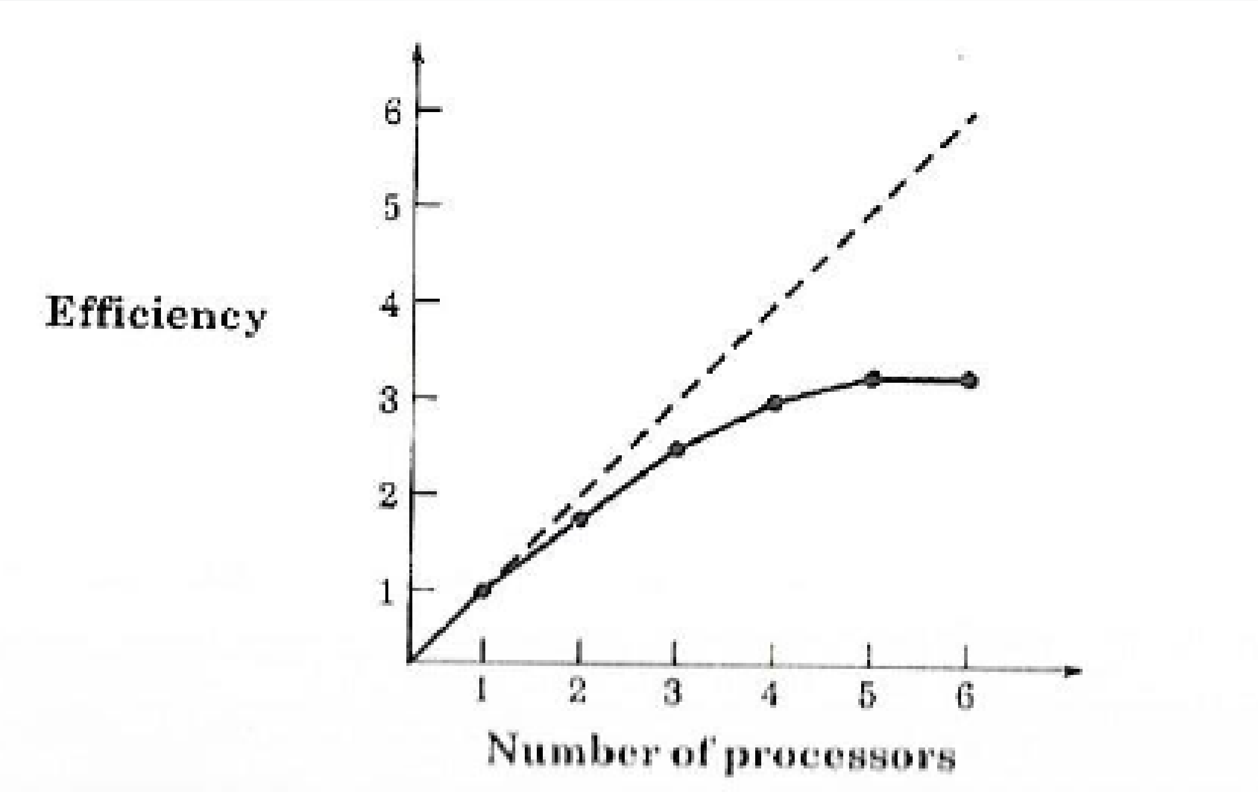
VAZÃO *(THROUGHPUT)*

- Def.: **Taxa** na qual os pedidos são atendidos (servidos) pelo sistema.
- Exemplos:
  - Sistemas em *batch* : jobs por segundo
  - Sistemas interativos: pedidos por segundo
  - CPUs: MIPS ou MFLOPS
  - **Redes: Pacotes por segundo (pps) ou bits por segundo (bps)**
  - Sistemas de Processamento de Transações: Transações por segundo (TPS)



EFICIÊNCIA

- **Razão** entre a **Vazão Máxima** (capacidade utilizável) e a **Capacidade Nominal**.



OUTRAS MÉTRICAS

- **Utilização**: fração do tempo em que o recurso permanece ocupado atendendo os pedidos dos usuários.
- **Confiabilidade**: Probabilidade de Erros ou Intervalo Médio entre Erros.
- **Disponibilidade**: Fração do tempo em que o sistema está disponível para atender os pedidos dos usuários.
  - MTTF = *Mean Time To Failure*
- Relação **Custo/Desempenho**

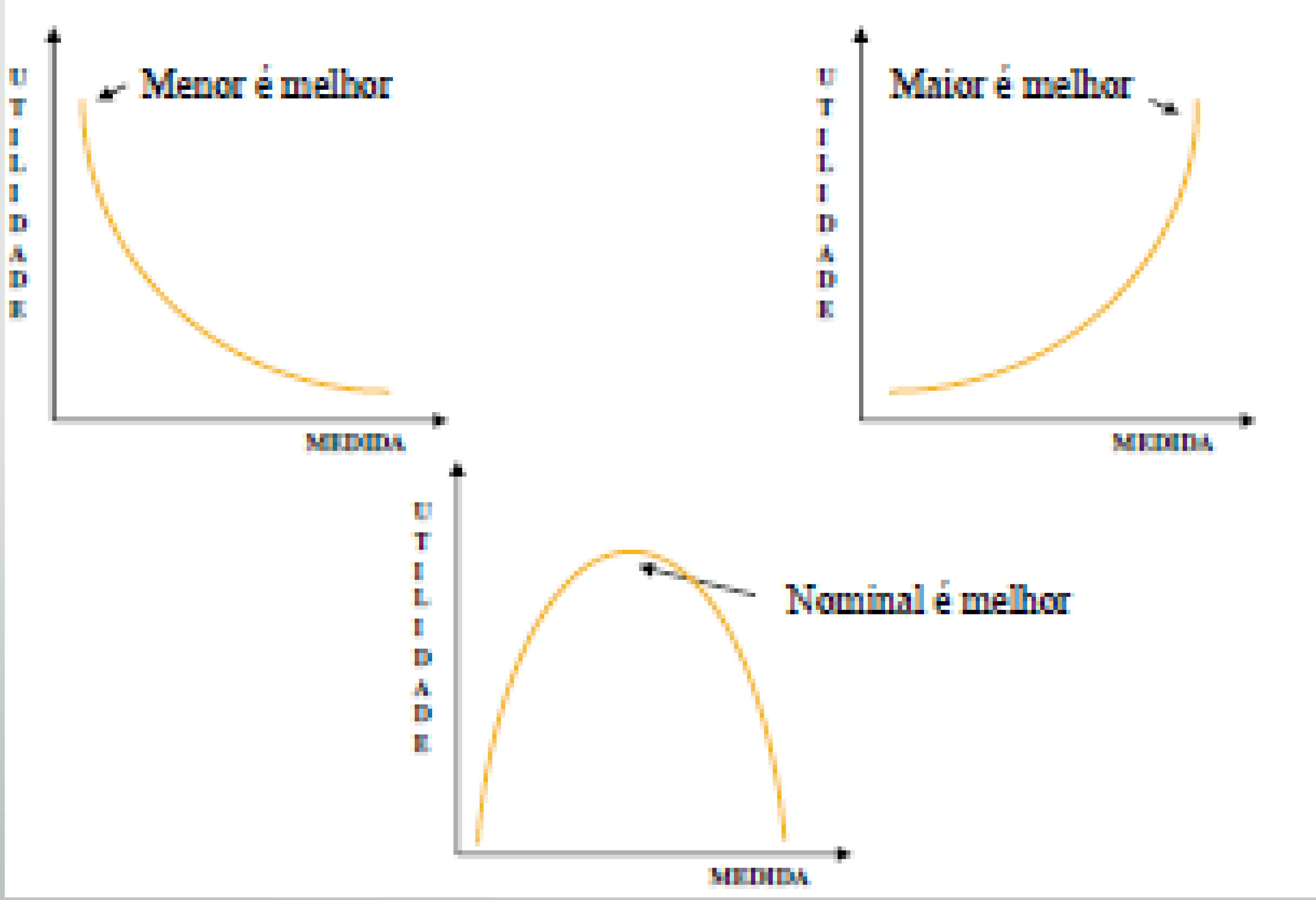
Métricas padronizadas

IPPM – Métricas de Desempenho IP

- http://www.ietf.org/html.charters/ippm-charter.html
- O IPPM WG tem como objetivo desenvolver **métricas padrão** que possam ser aplicadas para medir a qualidade, desempenho e confiabilidade dos **serviços de entrega de dados na Internet**.
- Devem poder ser obtidas por operadores de redes, usuários finais ou grupos independentes de teste.
- Devem fornecer **medidas quantitativas de desempenho**.

- **MÉTRICAS DEFINIDAS**
  - Conectividade [RFC2678]
  - Atraso em um sentido (*one-waydelay*) [RFC2679]
  - Perda em um sentido (*one-wayloss*) [RFC2680]
  - Atraso de ida e volta (*round-tripdelay*) [RFC2681]
  - Perda de ida e volta (*round-triploss*)
  - Variação do atraso [RFC3393]
  - Padrões de perda em um sentido [RFC3357]
  - Capacidade de transporte em lote (*bulk*) [RFC3148]
  - Reordenação de pacotes [RFC 4737]
  - Definições de capacidade da rede [RFC 5136]
  - Duplicação de pacotes em um sentido [RFC 5560]
  - Espacial e Multicast [RFC 5644]

CLASSIFICAÇÃO DA UTILIDADE DAS MÉTRICAS DE DESEMPENHO



ESPECIFICAÇÃO DE REQUISITOS DE DESEMPENHO

Os Requisitos devem ser:

- *SPECIFIC*
- *MEASURABLE*
- *ACCEPTABLE*
- *REALIZABLE*
- *THOROUGH*

- **E**SPECÍFICOS
  - Quantitativos (x qualitativos)
- **M**ENSURÁVEIS
  - Devem poder ser medidos
- **A**CEITÁVEIS
  - Realísticos
- **V**IÁVEIS
  - Devem ser realizáveis
- **A**BRANGENTES
  - Inclusão de **todas** saídas possíveis

ESTUDO DE CASO

- **Problema**: Especificar os requisitos de desempenho de uma LAN de alta velocidade.
- **Serviços**: Transporte de quadros (ou pacotes) até a estação destino.
- **Resultados**:
  - o quadro é **entregue corretamente** ao destino;
  - o quadro é **entregue incorretamente**; ou
  - o quadro **não é entregue**.

ESTUDO DE CASO:
REQUISITOS DE DESEMPENHO

- **Velocidade**: Se o pacote é entregue corretamente, são importantes o tempo para entregá-lo e a taxa de entrega.
  - O **atraso de acesso** em qualquer estação deve ser menor do que 1 segundo.
  - A **vazão média** deve ser de pelo menos 80 Mbps

ESTUDO DE CASO: REQUISITOS DE DESEMPENHO

- **Confiabilidade**
  - Foram considerados importantes **cinco erros diferentes**:
    - A probabilidade de um **bit qualquer estar errado** deve **ser inferior a 10^{-7}**.
    - A probabilidade de **qualquer quadro estar errado** (com indicação de erro) deve ser **menor do que 1%.**
    - A probabilidade de que um **quadro errado seja entregue sem uma indicação de erro** deve ser **inferior a 10^{-15}**.
    - A probabilidade de que um **quadro seja entregue a um destinatário errado** devido a um erro não detectado no endereço de destino deve ser **inferior a 10^{-18}**.
    - A probabilidade de que um **quadro** seja entregue mais de uma vez (**duplicado**) deve ser menor do que 10^{-5}.
  - A **probabilidade** de **perder um quadro** na LAN (devido a todos os tipos de erros) deve ser **inferior a 1%.**

ESTUDO DE CASO:
REQUISITOS DE DESEMPENHO

- **Disponibilidade**:
  - Dois **modos de falha** foram considerados significativos: tempo gasto com **reinicializações** da rede e o tempo gasto com **falhas permanentes**
    - O **tempo médio para inicializar** a LAN deve ser **inferior a 15 mseg**.
    - O **tempo médio entre inicializações** da LAN deve ser de **pelo menos 1 minuto**.
    - O **tempo médio para consertar** uma LAN deve ser **inferior a 1 hora** (as partições da LAN devem estar operacionais durante este período).
    - O **tempo médio entre partições** da LAN deve ser de **pelo menos meia semana**.
- Todos estes valores numéricos foram testados através de **modelagem analítica**, que mostrou que é viável para uma LAN satisfazer estes requisitos.

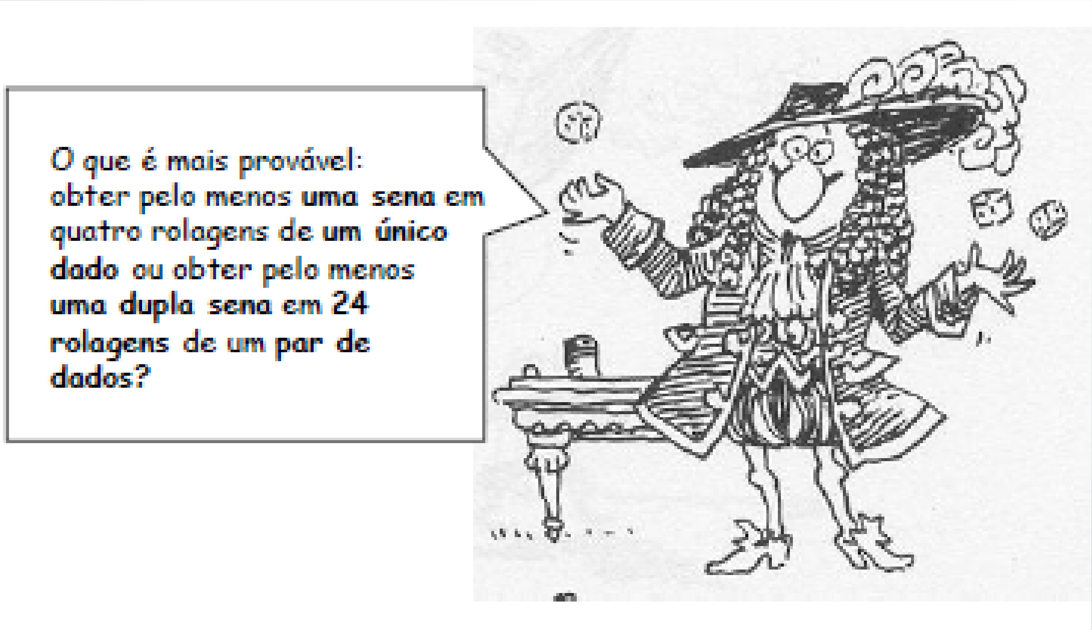
SUMMARIZING MEASURED DATA

Capítulo 12_**PARTE 1**

Raj Jain - The Art of Computer Systems Performance Analysis. Wiley, 1991.

O QUE É PROBABILIDADE?

- O jogo de dados moderno popularizou-se na Idade Média, a tempo para que um libertino da época, o Cavaleiro De Mere, apresentasse um quebra-cabeças matemático:



- O Cavaleiro considerou que o número médio de rolagens com sucesso seria o mesmo nos dois casos:
  - Chance de obter um seis = 1/6

  Número médio em quatro rolagens= 4 x (1/6) = 2/3

  - Chance de obter uma dupla sena em uma rolagem= 1/36

  Número médio em 24 rolagens= 24 x (1/36) = 2/3

Por que, então, ele **perdia mais freqüentemente** com a segunda jogada???

O QUE É PROBABILIDADE?

- De Mere colocou a questão para o seu amigo, o gênio **BLAISE PASCAL** (1623-1666):

Pascal escreveu a seu colega **PIERRE DE FERMAT**, e após algumas cartas, os dois finalizaram a Teoria da Probabilidade na sua forma moderna - à exceção, claro, dos desenhos.

O QUE É PROBABILIDADE?

- A teoria da probabilidade estuda os **fenômenos aleatórios**.
- Um **modelo probabilístico** consiste de uma lista de todos os possíveis resultados de um experimento e a atribuição de suas respectivas probabilidades.
- Utilizada inicialmente para o estudo de jogos de azar!

O QUE É PROBABILIDADE?

- **DEFINIÇÕES BÁSICAS**
  - Um **experimento aleatório** é o processo de observar o resultado de um evento não determinístico.
  - **Resultados elementares** são todos os possíveis resultados de um experimento aleatório.
  - O **espaço amostral** é o conjunto de todos os resultados elementares.
- **EVENTOS**
  - Um evento *A* (relativo a um particular espaço amostral *S*, associado a um experimento) é simplesmente um conjunto de resultados possíveis.

$$S \supseteq A$$

- **AXIOMAS DA PROBABILIDADE**
  - Seja ε um experimento. Seja *S* um espaço amostral associado a ε . A cada evento *A* associaremos um número real representado por *P(A)* e denominado **probabilidade de** *A*, que satisfaça às seguintes propriedades:

$$0 \leq P(A) \leq 1$$

$$P(S) = 1$$

  - Se A e B forem **mutuamente exclusivos**:

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

O QUE É PROBABILIDADE?

- **ATRIBUIÇÃO DE PROBABILIDADES**
  - **Clássica:** Baseada em idéias de jogos, a hipótese fundamental é que o jogo seja justo e que todos os resultados elementares tenham a mesma probabilidade.
  - **Freqüência Relativa:** Quando um dado experimento pode ser repetido então a probabilidade do evento é a proporção de vezes em que o evento ocorre para um número grande de repetições.
  - **Pessoal:** Avaliação pessoal da verossimilhança de um dado evento

- **COMBINAÇÃO DE EVENTOS**
  - Dados dois eventos *E* e *F*, podemos obter novos eventos:
    - E **e** F : ocorrência de ambos os eventos;
    - E **ou** F : ocorrência de pelo menos um dos eventos;
    - **não** *E* : o evento *E* não ocorre.
- **REGRAS**
  - Adição:
    - $P(E \text{ ou } F) = P(E) + P(F) - P(E \text{ e } F)$
  - Subtração:
    - **P(E) = 1 - P(não E)**

O QUE É PROBABILIDADE?

- **PROBABILIDADE CONDICIONAL**
  - A **probabilidade condicional** de um evento *A*, dado que ocorreu o evento *B*, ou a **probabilidade condicional de *A* dado *B*** é representada simbolicamente por:

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

- **REGRA DA MULTIPLICAÇÃO**
  - $P(E \text{ e } F) = P(E \mid F)\, P(F)$
  - $P(E \text{ e } F) = P(F \mid E)\, P(E)$
  - Se *E* e *F* forem eventos independentes, então:

$$P(E \text{ e } F) = P(E)\, P(F)$$

O QUE É PROBABILIDADE?

- **EVENTOS INDEPENDENTES**
  - Dois eventos são ditos independentes se a **ocorrência de um deles não afeta** de modo algum a **probabilidade do outro**.
  - O conhecimento de que um dos eventos ocorreu não altera de nenhum modo a nossa estimativa da probabilidade do outro evento.

QUEBRA-CABEÇA DE *DE MERE*

- **Como podemos então resolver o quebra-cabeças de De Mere?**
  - O que é mais provável: obter pelo menos **uma sena** em quatro rolagens de **um único dado** ou obter pelo menos **uma dupla sena** em **24 rolagens** de um **par de dados?**
- ***Tente!!!***

QUEBRA-CABEÇA DE *DE MERE*

- Seja *E* o evento de obter pelo menos **uma sena em quatro rolagens** de um único dado. ***P(E) = ?***
  - É mais fácil descrever o complemento de *E*, **não** *E*, ou seja, **não obter nenhuma sena em quatro rolagens do dado**!
  - A probabilidade de não obter uma sena numa rolagem de dado é 5/6. Como as rolagens do dado são independentes, temos que em 4 rolagens:

    $P(\text{não } E) = (5/6)^4 = 0{,}482$ portanto $\mathbf{P(E) = 0{,}518}$

- Seja *F* o evento de obter pelo menos uma **dupla sena em vinte e quatro rolagens** de dois dados. *P(F) = ?*
  - É mais fácil descrever o complemento de *F*, não *F*, ou seja, **não obter nenhuma dupla sena em vinte e quatro rolagens dos dados**!
  - A probabilidade de não obter uma dupla sena numa rolagem de dois dados é 35/36. Como as rolagens dos dados são independentes, temos que:

    $P(\text{não } F) = (35/36)^{24} = 0{,}509$ portanto $\mathbf{P(F) = 0{,}491}$

$\mathbf{P(E) = 0{,}518} > P(F) = 0{,}491$

REGRA DE BAYES

- **Teorema da Probabilidade Total:**

$$P(A) = \sum_{i=1}^{n} P(A|B_i)\ P(B_i)$$

- Aplicando a definição de probabilidade condicional juntamente com este teorema, obtemos:

$$\boxed{P(B_j\,|A) = \frac{P(B_j \cap A)}{P(A)} = \frac{P(A|B_j)P(B_j)}{\sum_i P(A|B_i)P(B_i)}}$$

Regra de Bayes

EXEMPLO

Suponha que uma certa doença rara infecta **uma pessoa a cada 1000 pessoas** de uma população...

E suponha que existe um **teste bom**, mas **imperfeito** para esta doença:

- Se uma pessoa está infectada ele produz um resultado **positivo em 99%** dos casos;
- Mas se uma pessoa não está infectada ele produz um **falso positivo** para **2%** deles.

Se o seu resultado der positivo, **quais são as chances de que você esteja mesmo infectado**?

REGRA DE BAYES: EXEMPLO

- Suponha que uma certa doença rara infecta **uma pessoa a cada 1000 pessoas** de uma população...
- E suponha que existe um **teste bom**, mas **imperfeito** para esta doença:
  - Se uma pessoa está infectada ele produz um resultado **positivo em 99%** dos casos;
  - Mas se uma pessoa não está infectada ele produz um **falso positivo** para **2%** deles.
- Se o seu resultado der positivo, **quais são as chances de que você esteja mesmo infectado**?

- **Eventos**:
  - A : O paciente está infectado (99%)
  - B : O teste dá positivo (98% corretos).
- **Usando a Regra de Bayes**:

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

$$P(A|B) = \frac{0{,}99\ x\ 0{,}001}{0{,}99\ x\ 0{,}001 + 0{,}02\ x\ 0{,}999} = \frac{0{,}00099}{0{,}00099 + 0{,}01998} = 0{,}0472$$

Apenas 5% dos que tiveram resultado positivo estão, de fato, infectados!!!

EXERCÍCIO 12.1

12.1 A distributed system has three file servers, which are chosen independently and with equal probabilities whenever a new file is created. The servers are named A, B, and C. Determine the probabilities of the following events:

a. Server A is selected

b. Server A or B is selected

c. Servers A and B are selected

d. Server A is not selected

e. Server A is selected twice in a row

f. Server selection sequence ABCABCABC is observed (in nine successive file creations)

VARIÁVEIS ALEATÓRIAS

- **Definição**
  - **Regra** que atribui um **valor numérico** a cada possível **resultado de um experimento**.
  - Exemplo:
    - Jogue duas moedas (o experimento aleatório) e registre o **número de caras**: 0, 1 ou 2.
- Notação
  - Usa-se **letras maiúsculas** para a **variável** e letras minúsculas para um valor particular.

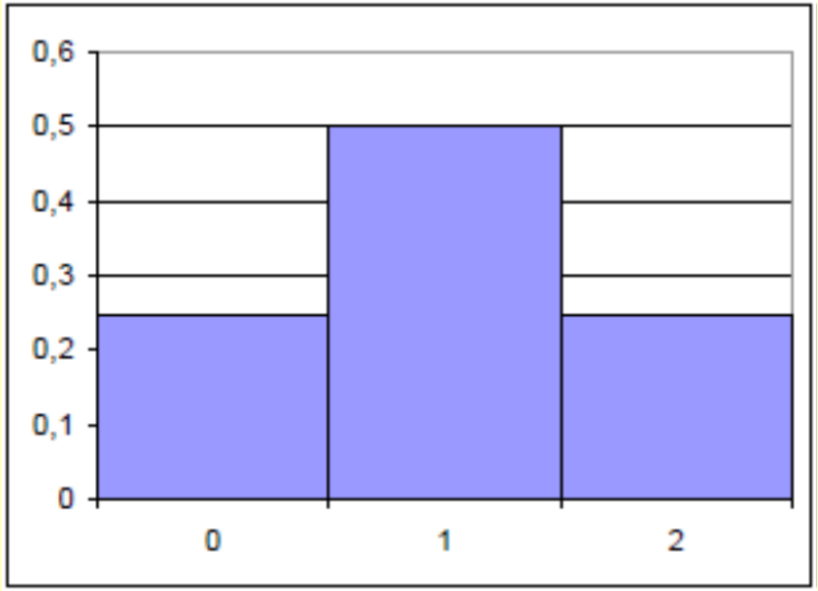
- **Probabilidades Dos Resultados**:

  Pr(X=x) = p(x)

  - Para o exemplo das moedas:

x	0	1	2
$\Pr(X = x)$	$^1/_4$	$^1/_2$	$^1/_4$

- **Histograma**
  - Para cada valor de X, traçamos uma barra com altura igual a *p(x)*:



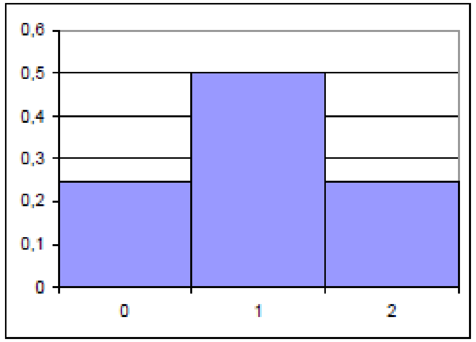
A área total é a soma das probabilidades para todos os resultados, i.e., 1.

VARIÁVEIS ALEATÓRIAS

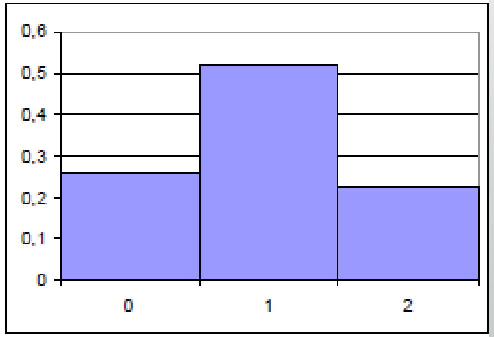
- **Resultado do Lançamento de Moedas**

Modelo Probabilístico		Dados Observados	
$p(x)$	x	n_x	n_x/n
0,25	0	260	0,260
0,50	1	517	0,517
0,25	2	223	0,223
		1.000	

- **Comparação: Modelo x Real**



Histograma
Probabilístico



Histograma da
Freqüência Relativa

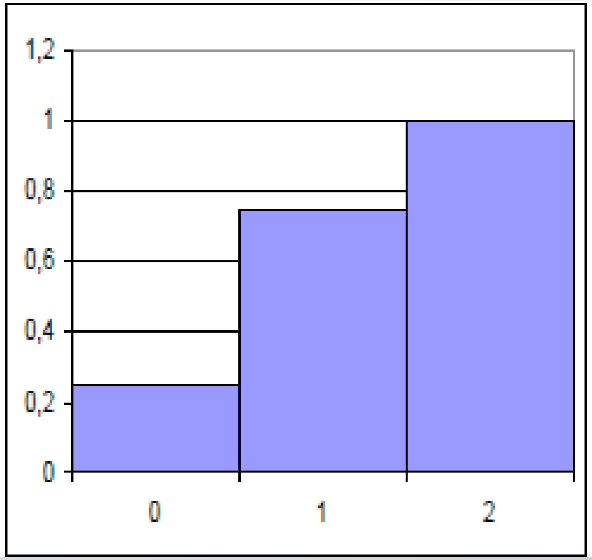
FUNÇÃO DISTRIBUIÇÃO DE PROBABILIDADE

- **CDF (*Cumulative Distribution Function*)** ou PDF (*Probability Distribution Function*)

$$F_X(t) = \Pr(-\infty < X \leq t)$$

$$= \Pr(X \leq t)$$

$$= \sum_{x \leq t} p_x(x)$$



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

$$= F_X(b) - F_X(a)$$

FUNÇÕES DE PROBABILIDADE

- **pdf** (*Probability Density Function*)

$$f(x) = \frac{dF(x)}{dx}$$

- Dada uma **pdf** ***f(x)*** a **probabilidade de *X* se encontrar no intervalo *(x₁ ,x₂)*** pode também ser calculada através de integração

$$P(x_1 < X \leq x_2) = F_X(x_2) - F_X(x_1) = \int_{x_1}^{x_2} f(x)dx$$

- **pmf** (*Probability Mass Function*)

$$f(x_i) = p_i$$

- A **probabilidade de *x* se encontrar no intervalo *(x_1 ,x_2)*** pode também ser calculada através de somas:

$$P(x_1 < X \leq x_2) = F_X(x_2) - F_X(x_1) = \sum_{x_1 < x_i \leq x_2} p_i$$

RESUMINDO DADOS COM UM ÚNICO NÚMERO

- MÉDIA, MEDIANA E MODA
- VARIÂNCIA E DESVIO PADRÃO
- COVARIÂNCIA E CORRELAÇÃO
- QUANTIS

MÉDIA, MEDIANA E MODA

- **Média (Valor Esperado)**

μ = E(x)

para **variáveis discretas**:

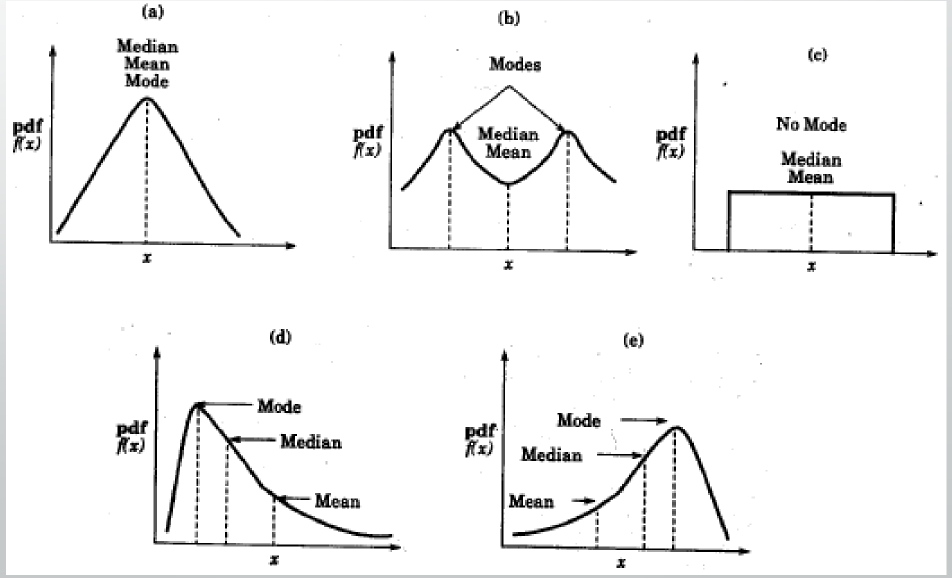
$$=\sum_{i=1}^{n} p_i x_i$$

para **variáveis contínuas**:

$$=\int_{-\infty}^{+\infty} x f(x)\,dx$$

Soma de todos os valores possíveis, **ponderada** pela probabilidade de ocorrência de cada um dos valores.

- **Mediana:** é o posto percentil 50 (ou **quantil 0,5**) de uma variável aleatória.
- **Moda:** é o valor mais provável de uma v.a.
  - Ou seja, é o valor x_i que corresponde à **maior probabilidade** p_i, ou o valor de x para o qual a **pdf** atinge o seu **valor máximo**.



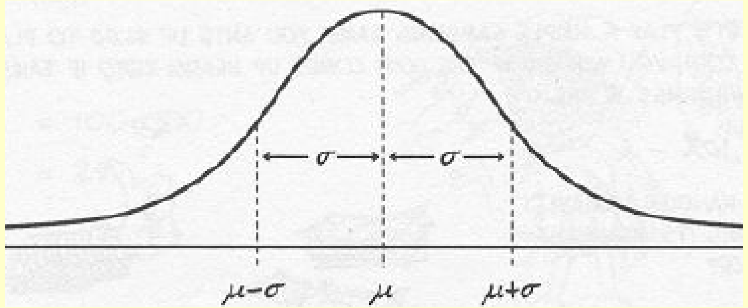
VARIÂNCIA E DESVIO PADRÃO

- A quantidade $(x - \mu)^2$ representa a **distância quadrática entre *x* e a sua média**.
- A **variância** de *x* é o valor esperado desta quantidade:

$$\mathrm{Var}(x) = E[(x-\mu)^2] = \sum_{i=1}^{n} p_i (x_i - \mu)^2$$

$$= \int_{-\infty}^{+\infty} (x - \mu)^2 f(x)\,dx$$

- A **variância** é normalmente denotada por σ^2.
- O **desvio padrão** é a raiz quadrada da variância, denotado por σ.



- **Coeficiente de Variação**

$$\mathrm{C.O.V} = \frac{\text{desvio padrão}}{\text{média}} = \frac{\sigma}{\mu}$$

COVARIÂNCIA E CORRELAÇÃO

- Dadas duas v.a.s X e Y com médias μ_x e μ_y, a **covariância** delas é dada por:

$$\text{Cov}(x,y) = \sigma_{xy}^2 = E[(x-\mu_x)(y-\mu_y)]$$

$$= E(xy) - E(x)E(y)$$

- Para **variáveis independentes** a covariância é **zero**, dado que

Apesar da **independência** sempre implicar em **covariância zero**, **o contrário nem sempre é verdade**.

- Ou simplesmente **correlação** é o **valor normalizado da covariância**

Correlação (x,y) $= \rho_{xy} = \dfrac{\sigma_{xy}^2}{\sigma_x \sigma_y}$

A correlação varia sempre entre **-1 e +1**.

QUANTIS

- O valor *x* no qual a **CDF** corresponde ao valor α é chamado de **α-quantil** ou **100α-percentil**.
- Ele é denotado por x_α

$$P(x \leq x_\alpha) = F(x_\alpha) = \alpha$$

EXEMPLO

$X = 0, 1 \text{ ou } 2$

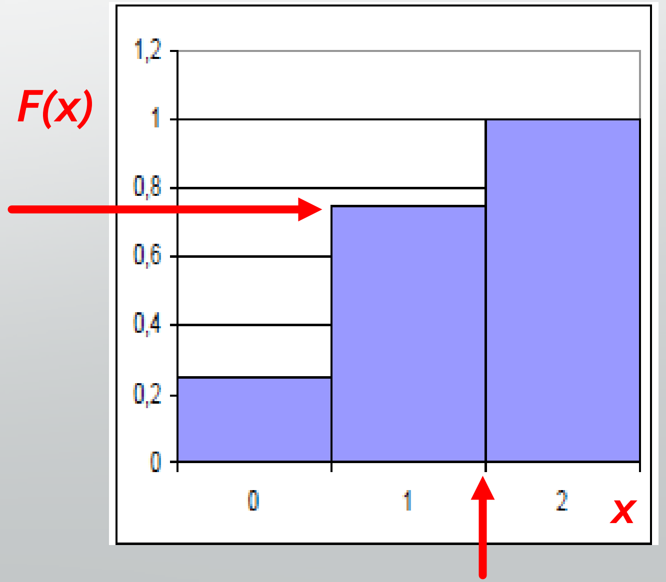
0,75-quantil ou **75% percentil**

$P(x \leq x_{0,75}) = F(x_{0,75}) = 0,75$

$X_{0,75} = 1$

$X_{0,25} = 0$

$X_{100\%} = 2$



MÉDIA E VARIÂNCIA DE SOMAS

- Sejam $x_1, x_2, \ldots, x_k$ *k* **variáveis aleatórias** e $a_1, a_2, \ldots, a_k$ *k* constantes arbitrárias (denominadas de **pesos**), então

$$E(a_1 x_1 + a_2 x_2 + \ldots + a_k x_k) = a_1 E(x_1) + a_2 E(x_2) + \ldots + a_k E(x_k)$$

- Para **variáveis independentes**:

$$\mathrm{Var}(a_1 x_1 + a_2 x_2 + \ldots + a_k x_k) = a_1^2 \mathrm{Var}(x_1) + a_2^2 \mathrm{Var}(x_2) + \ldots + a_k^2 \mathrm{Var}(x_k)$$

DISTRIBUIÇÃO BINOMIAL

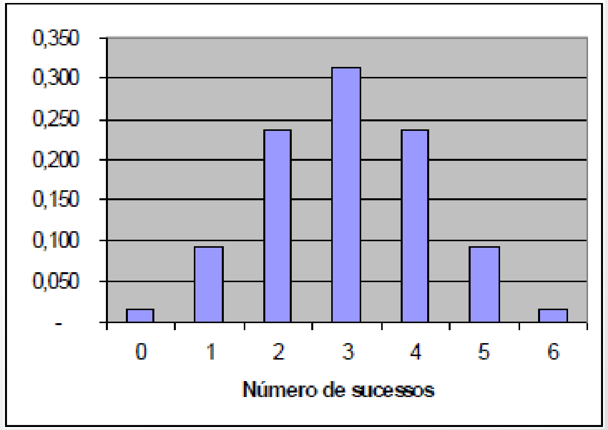
- Suponha que tenhamos um **processo aleatório** com **apenas dois resultados** possíveis: ***sucesso*** ou ***falha***.
- As **tentativas de Bernoulli** são a repetição de um experimento como este, desde que:
  - Haja **apenas dois resultados** em cada tentativa.
  - A **probabilidade de sucesso (*p*)** seja a mesma em cada tentativa.
  - As tentativas sejam **independentes**.
- **Variável Aleatória Binomial**
  - *X* é o **número de sucessos** em *n* tentativas de Bernoulli com probabilidade *p* de sucesso.

$$P(X=k) = \binom{n}{k} p^k (1-p)^{n-k}$$

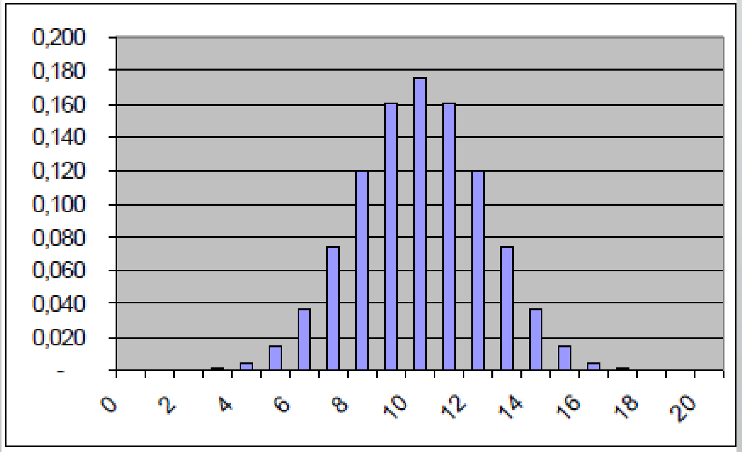
$$\text{onde } \binom{n}{k} = \frac{n!}{k!(n-k)!}$$

Histogramas da **DISTRIBUIÇÃO BINOMIAL**

- **6 jogadas** de moedas, $p = 0,5$

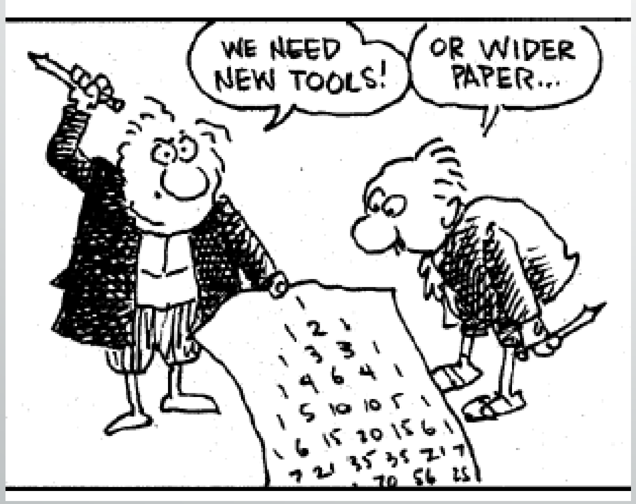


- **20 jogadas** de moedas, $p = 0,5$



DISTRIBUIÇÃO NORMAL

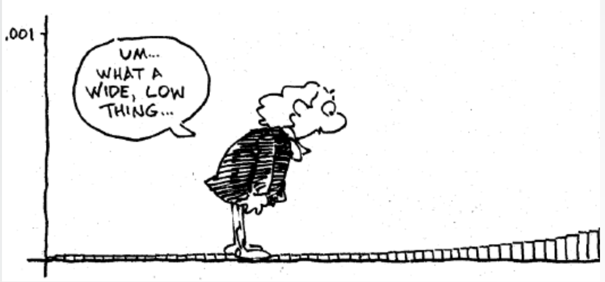
- Mas, calcular estes termos para **grandes valores de n** pode dar muito trabalho... ou pelo menos dava no século 18 quando **James Bernouilli** e **Abraham de Moivre** estavam calculando sem um computador.



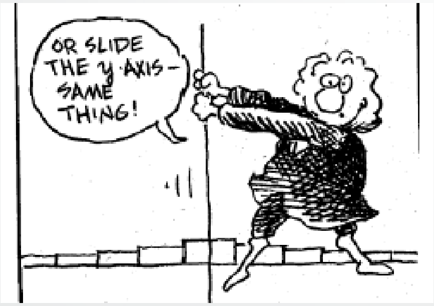
- Utilizando uma ferramenta recém-inventada, o **Cálculo**, **De Moivre** mostrou que para p=0,5, a distribuição normal era bem aproximada por uma função densidade contínua que podia ser descrita de forma bem simples.

DISTRIBUIÇÃO NORMAL

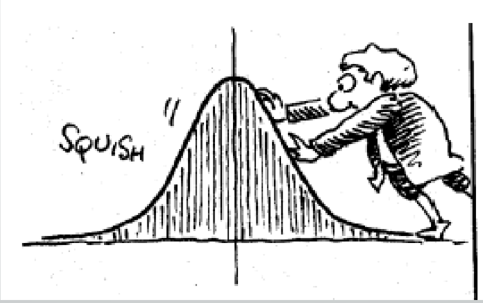
- Imagine a **distribuição binomial com p=0,5** e **n muito grande**, por exemplo, um milhão...



- Agora desloque o gráfico de modo que a **média** seja **zero**.



- Esprema a curva ao longo do eixo x até que o **desvio padrão** seja **1** e estique no eixo y para que a **área** continue sendo **1**.

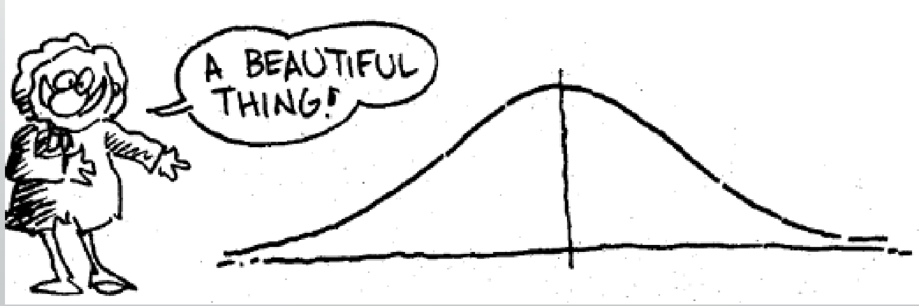


- **DISTRIBUIÇÃO NORMAL UNITÁRIA**

Representada por **N(0,1)** significando média = 0 e variância = 1

O resultado ficou próximo a uma curva **suave, simétrica** e com **forma de sino** que é descrita pela seguinte fórmula:

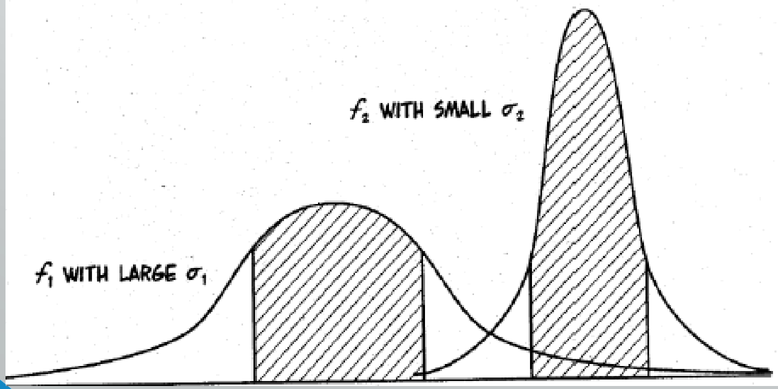
$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$$



DISTRIBUIÇÃO NORMAL

- É a **distribuição mais comumente utilizada** na **análise de dados**.
- Também conhecida como **distribuição gaussiana**.
- A soma de um **grande número de observações independentes de qualquer distribuição** tem uma **distribuição normal**.
- Distribuição $N(\mu, \sigma)$ com **pdf** dada por:

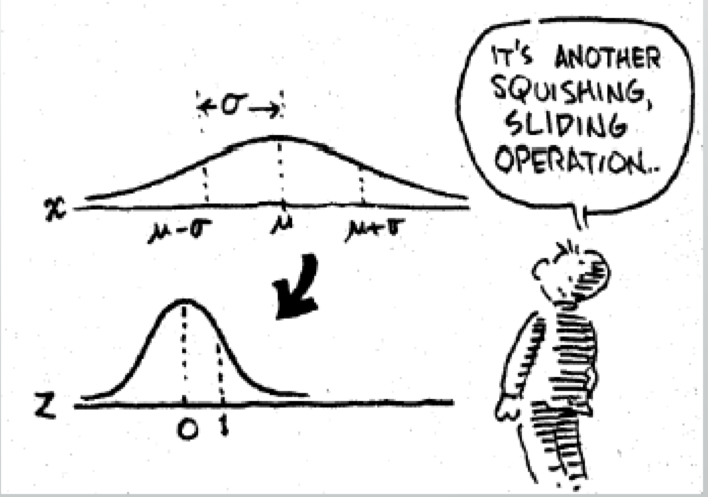
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad \text{para } -\infty \leq x \leq +\infty$$



- A **transformação z**

$$z = \frac{(x-\mu)^2}{\sigma^2}$$

- Muda uma variável aleatória normal com média μ e desvio padrão σ, numa **distribuição normal unitária** N(0,1).



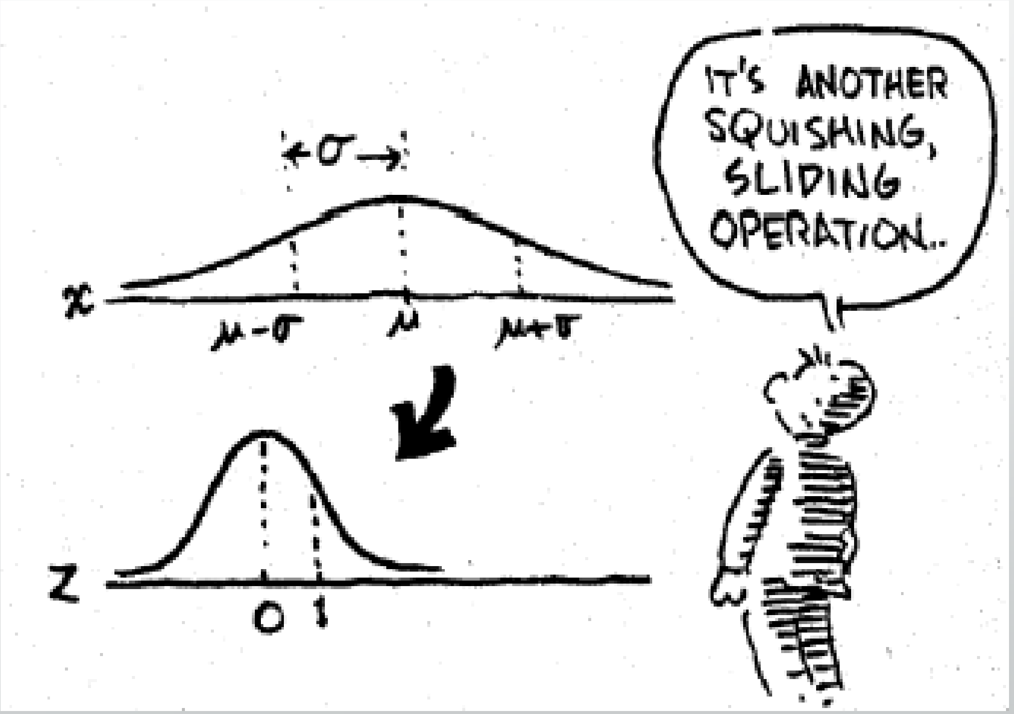
TRANSFORMAÇÃO z

- A **transformação z**

$$z = \frac{(x-\mu)^2}{\sigma^2}$$

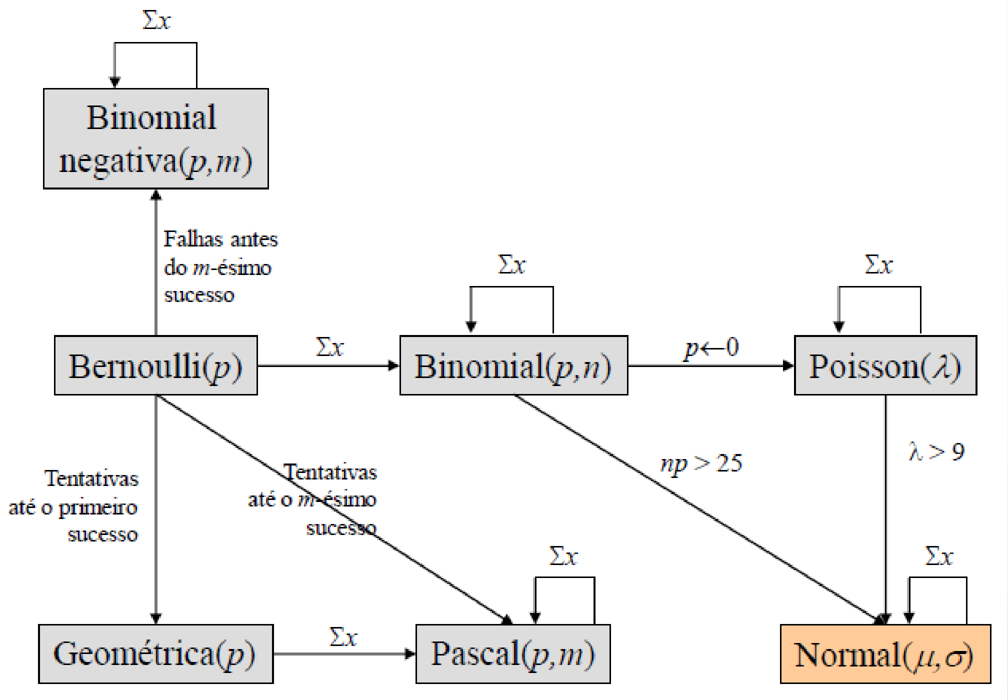
$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$$

- Muda uma variável aleatória normal com média μ e desvio padrão σ, numa **distribuição normal unitária** N(0,1).

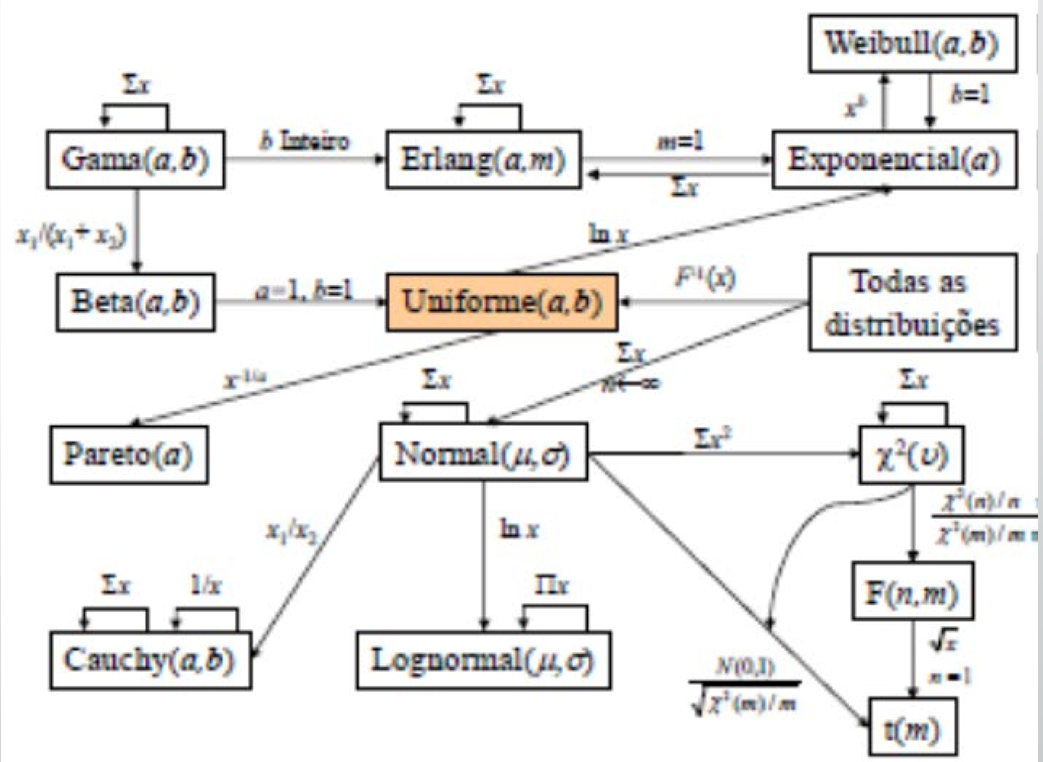


DISTRIBUIÇÕES

DISCRETAS



CONTÍNUAS

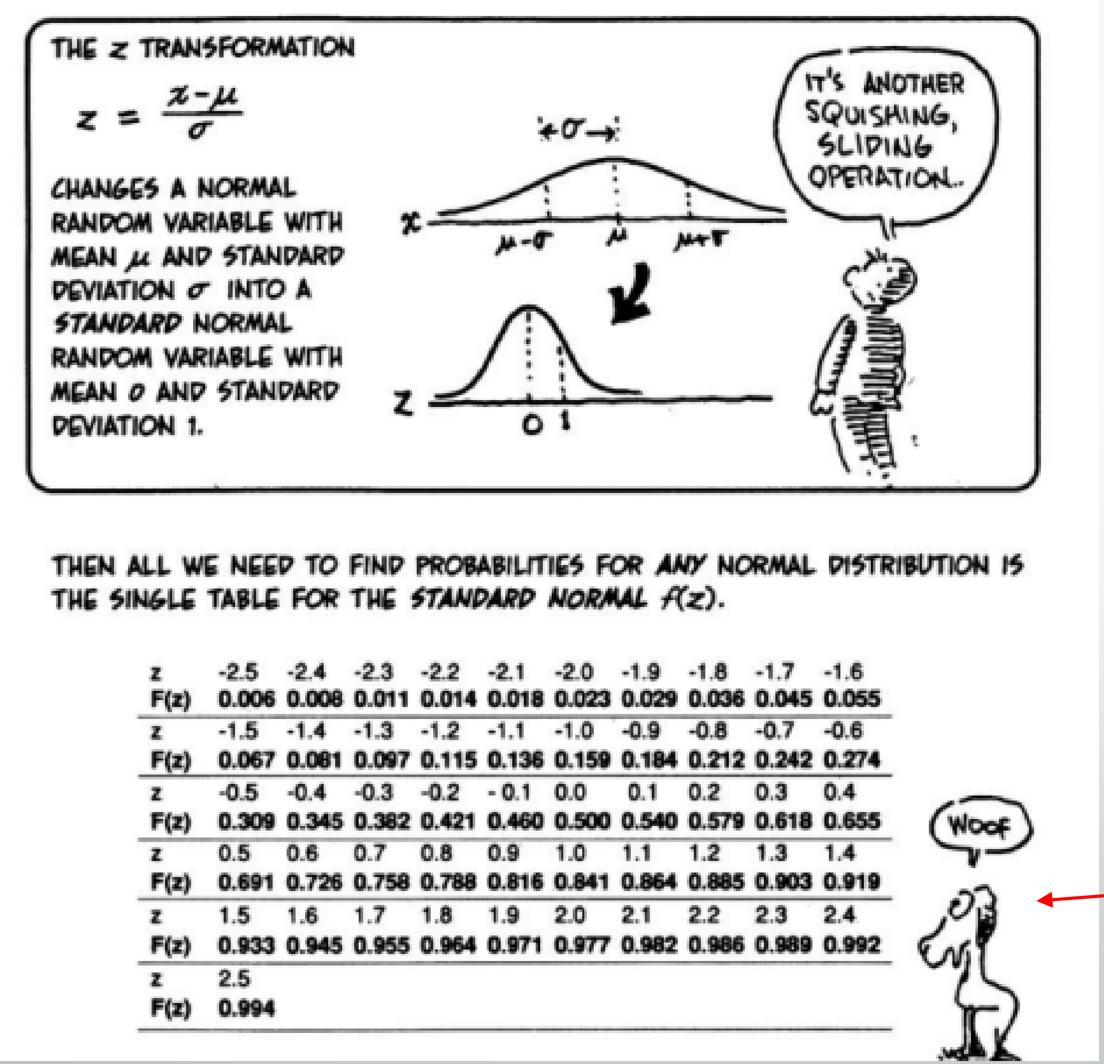


SUMMARIZING MEASURED DATA

Capítulo 12_**PARTE 2**

Raj Jain - The Art of Computer Systems Performance Analysis. Wiley, 1991.

DISTRIBUIÇÃO NORMAL PADRÃO

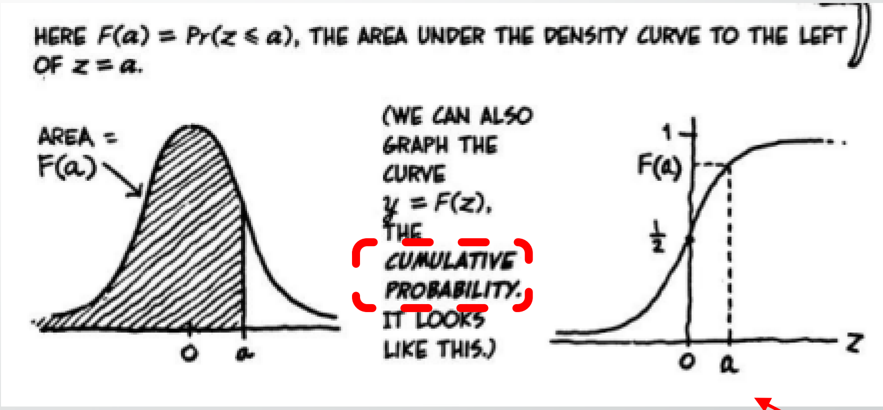

THE z TRANSFORMATION

$$z = \frac{x-\mu}{\sigma}$$

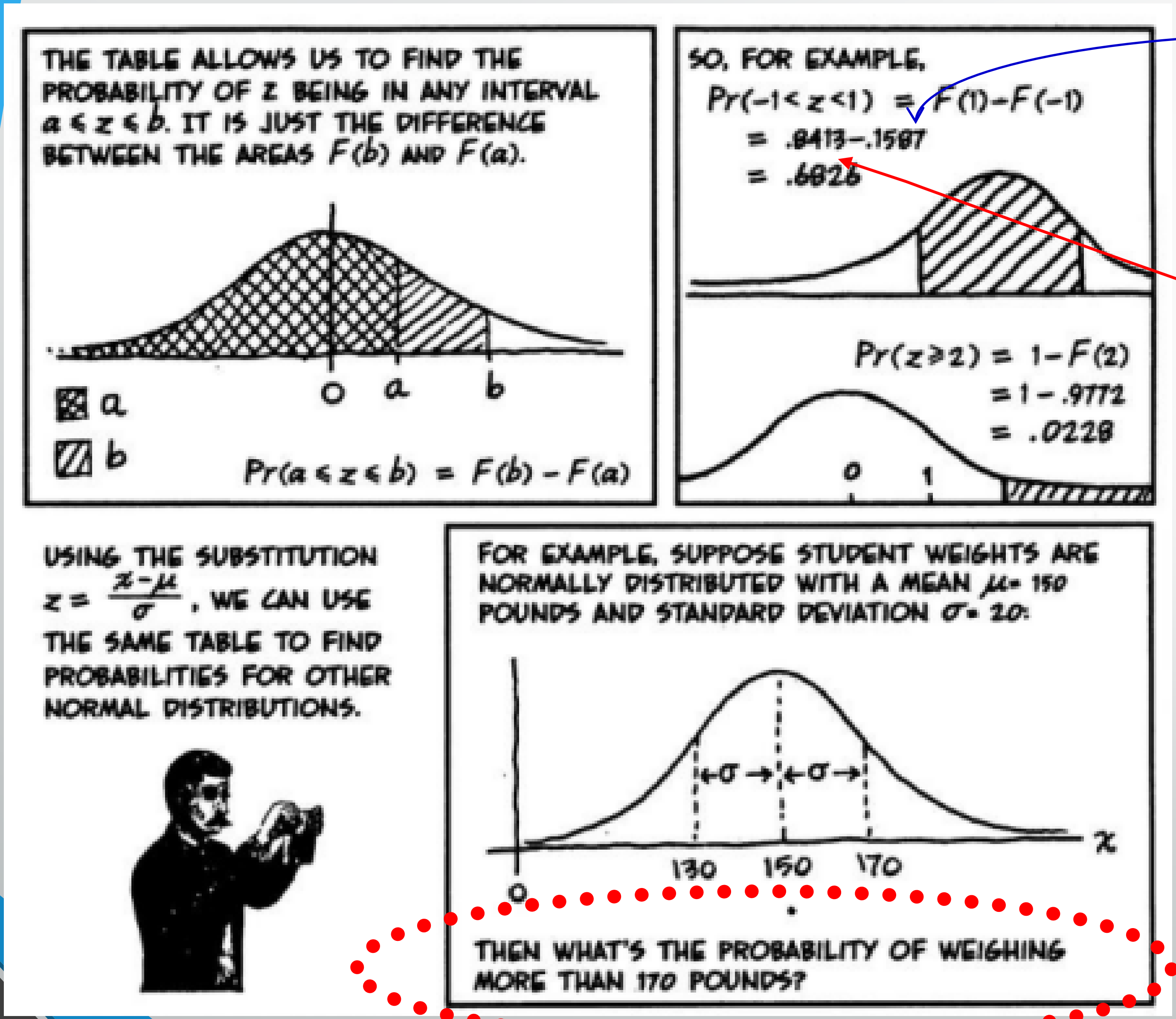
CHANGES A NORMAL RANDOM VARIABLE WITH MEAN μ AND STANDARD DEVIATION σ INTO A *STANDARD* NORMAL RANDOM VARIABLE WITH MEAN 0 AND STANDARD DEVIATION 1.

THEN ALL WE NEED TO FIND PROBABILITIES FOR *ANY* NORMAL DISTRIBUTION IS THE SINGLE TABLE FOR THE *STANDARD NORMAL* $f(z)$.

z	-2.5	-2.4	-2.3	-2.2	-2.1	-2.0	-1.9	-1.8	-1.7	-1.6
F(z)	0.006	0.008	0.011	0.014	0.018	0.023	0.029	0.036	0.045	0.055
z	-1.5	-1.4	-1.3	-1.2	-1.1	-1.0	-0.9	-0.8	-0.7	-0.6
F(z)	0.067	0.081	0.097	0.115	0.136	0.159	0.184	0.212	0.242	0.274
z	-0.5	-0.4	-0.3	-0.2	- 0.1	0.0	0.1	0.2	0.3	0.4
F(z)	0.309	0.345	0.382	0.421	0.460	0.500	0.540	0.579	0.618	0.655
z	0.5	0.6	0.7	0.8	0.9	1.0	1.1	1.2	1.3	1.4
F(z)	0.691	0.726	0.758	0.788	0.816	0.841	0.864	0.885	0.903	0.919
z	1.5	1.6	1.7	1.8	1.9	2.0	2.1	2.2	2.3	2.4
F(z)	0.933	0.945	0.955	0.964	0.971	0.977	0.982	0.986	0.989	0.992
z	2.5									
F(z)	0.994									



EXEMPLO



THE SINGLE TABLE FOR THE *STANDARD NORMAL* $f(z)$.

z	-2.5	-2.4	-2.3	-2.2	-2.1	-2.0	-1.9	-1.8	-1.7	-1.6
F(z)	0.006	0.008	0.011	0.014	0.018	0.023	0.029	0.036	0.045	0.055
z	-1.5	-1.4	-1.3	-1.2	-1.1	-1.0	-0.9	-0.8	-0.7	-0.6
F(z)	0.067	0.081	0.097	0.115	[illegible]	0.159	0.184	0.212	0.242	0.274
z	-0.5	-0.4	-0.3	-0.2	-0.1	0.0	0.1	0.2	0.3	0.4
F(z)	0.309	0.345	0.382	0.421	0.460	[illegible]	0.540	0.579	0.618	0.655
z	0.5	0.6	0.7	0.8	0.9	1.0	1.1	1.2	1.3	1.4
F(z)	0.691	0.726	0.758	0.788	[illegible]	[illegible]	0.864	0.885	0.903	0.919
z	1.5	1.6	1.7	1.8	1.9	2.0	2.1	2.2	2.3	2.4
F(z)	[illegible]	0.945	0.955	0.964	0.971	0.977	0.982	0.986	0.989	0.992
z	2.5									
F(z)	0.994									

➡ **Calcule !**

A.1 AREA OF THE UNIT NORMAL DISTRIBUTION

Table A.1 lists area between 0 and z. For example, the area between $z = 0$ and $z = 1.03$ is 0.3485. Due to symmetry of the normal distribution, the area between $z = 0$ and $z = -1.03$ is also the same.

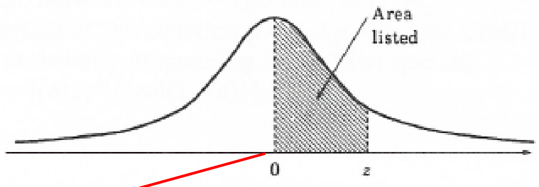
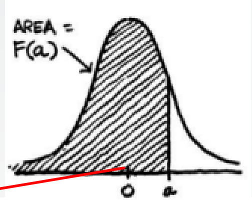


TABLE A.1 Area of the Unit Normal Distribution

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.0279	0.0319	0.0359
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.1480	0.1517
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736	0.1772	0.1808	0.1844	0.1879
0.5	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.2190	0.2224
0.6	0.2257	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2517	0.2549
0.7	0.2580	0.2611	0.2642	0.2673	0.2703	0.2734	0.2764	0.2794	0.2823	0.2852
0.8	0.2881	0.2910	0.2939	0.2967	0.2995	0.3023	0.3051	0.3078	0.3106	0.3133
0.9	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.3340	0.3365	0.3389
1.0	0.3413	0.3438	0.3461	0.3485	0.3508	0.3531	0.3554	0.3577	0.3599	0.3621
1.1	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749	0.3770	0.3790	0.3810	0.3830
1.2	0.3849	0.3869	0.3888	0.3907	0.3925	0.3944	0.3962	0.3980	0.3997	0.4015
1.3	0.4032	0.4049	0.4066	0.4082	0.4099	0.4115	0.4131	0.4147	0.4162	0.4177
1.4	0.4192	0.4207	0.4222	0.4236	0.4251	0.4265	0.4279	0.4292	0.4306	0.4319
1.5	0.4332	0.4345	0.4357	0.4370	0.4382	0.4394	0.4406	0.4418	0.4429	0.4441
1.6	0.4452	0.4463	0.4474	0.4484	0.4495	0.4505	0.4515	0.4525	0.4535	0.4545
1.7	0.4554	0.4564	0.4573	0.4582	0.4591	0.4599	0.4608	0.4616	0.4625	0.4633
1.8	0.4641	0.4649	0.4656	0.4664	0.4671	0.4678	0.4686	0.4693	0.4699	0.4706
1.9	0.4713	0.4719	0.4726	0.4732	0.4738	0.4744	0.4750	0.4756	0.4761	0.4767
2.0	0.4772	0.4778	0.4783	0.4788	0.4793	0.4798	0.4803	0.4808	0.4812	0.4817
2.1	0.4821	0.4826	0.4830	0.4834	0.4838	0.4842	0.4846	0.4850	0.4854	0.4857
2.2	0.4861	0.4864	0.4868	0.4871	0.4875	0.4878	0.4881	0.4884	0.4887	0.4890
2.3	0.4893	0.4896	0.4898	0.4901	0.4904	0.4906	0.4909	0.4911	0.4913	0.4916
2.4	0.4918	0.4920	0.4922	0.4925	0.4927	0.4929	0.4931	0.4932	0.4934	0.4936
2.5	0.4938	0.4940	0.4941	0.4943	0.4945	0.4946	0.4948	0.4949	0.4951	0.4952
2.6	0.4953	0.4955	0.4956	0.4957	0.4959	0.4960	0.4961	0.4962	0.4963	0.4964
2.7	0.4965	0.4966	0.4967	0.4968	0.4969	0.4970	0.4971	0.4972	0.4973	0.4974
2.8	0.4974	0.4975	0.4976	0.4977	0.4977	0.4978	0.4979	0.4979	0.4980	0.4981
2.9	0.4981	0.4982	0.4982	0.4983	0.4984	0.4984	0.4985	0.4985	0.4986	0.4986
3.0	0.4987	0.4987	0.4987	0.4988	0.4988	0.4989	0.4989	0.4989	0.4990	0.4990



z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5753
0,2	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141
0,3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,6480	0,6517
0,4	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808	0,6844	0,6879
0,5	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,7190	0,7224
0,6	0,7257	0,7291	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486	0,7517	0,7549
0,7	0,7580	0,7611	0,7642	0,7673	0,7704	0,7734	0,7764	0,7794	0,7823	0,7852
0,8	0,7881	0,7910	0,7939	0,7967	0,7995	0,8023	0,8051	0,8078	0,8106	0,8133
0,9	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289	0,8315	0,8340	0,8365	0,8389
1,0	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8554	0,8577	0,8599	0,8621
1,1	0,8643	0,8665	0,8686	0,8708	0,8729	0,8749	0,8770	0,8790	0,8810	0,8830
1,2	0,8849	0,8869	0,8888	0,8907	0,8925	0,8944	0,8962	0,8980	0,8997	0,9015
1,3	0,9032	0,9049	0,9066	0,9082	0,9099	0,9115	0,9131	0,9147	0,9162	0,9177
1,4	0,9192	0,9207	0,9222	0,9236	0,9251	0,9265	0,9279	0,9292	0,9306	0,9319
1,5	0,9332	0,9345	0,9357	0,9370	0,9382	0,9394	0,9406	0,9418	0,9429	0,9441
1,6	0,9452	0,9463	0,9474	0,9484	0,9495	0,9505	0,9515	0,9525	0,9535	0,9545
1,7	0,9554	0,9564	0,9573	0,9582	0,9591	0,9599	0,9608	0,9616	0,9625	0,9633
1,8	0,9641	0,9649	0,9656	0,9664	0,9671	0,9678	0,9686	0,9693	0,9699	0,9706
1,9	0,9713	0,9719	0,9726	0,9732	0,9738	0,9744	0,9750	0,9756	0,9761	0,9767
2,0	0,9772	0,9778	0,9783	0,9788	0,9793	0,9798	0,9803	0,9808	0,9812	0,9817
2,1	0,9821	0,9826	0,9830	0,9834	0,9838	0,9842	0,9846	0,9850	0,9854	0,9857
2,2	0,9861	0,9864	0,9868	0,9871	0,9875	0,9878	0,9881	0,9884	0,9887	0,9890
2,3	0,9893	0,9896	0,9898	0,9901	0,9904	0,9906	0,9909	0,9911	0,9913	0,9916
2,4	0,9918	0,9920	0,9922	0,9925	0,9927	0,9929	0,9931	0,9932	0,9934	0,9936
2,5	0,9938	0,9940	0,9941	0,9943	0,9945	0,9946	0,9948	0,9949	0,9951	0,9952
2,6	0,9953	0,9955	0,9956	0,9957	0,9959	0,9960	0,9961	0,9962	0,9963	0,9964
2,7	0,9965	0,9966	0,9967	0,9968	0,9969	0,9970	0,9971	0,9972	0,9973	0,9974
2,8	0,9974	0,9975	0,9976	0,9977	0,9977	0,9978	0,9979	0,9979	0,9980	0,9981
2,9	0,9981	0,9982	0,9982	0,9983	0,9984	0,9984	0,9985	0,9985	0,9986	0,9986
3,0	0,9987	0,9987	0,9987	0,9988	0,9988	0,9989	0,9989	0,9989	0,9990	0,9990

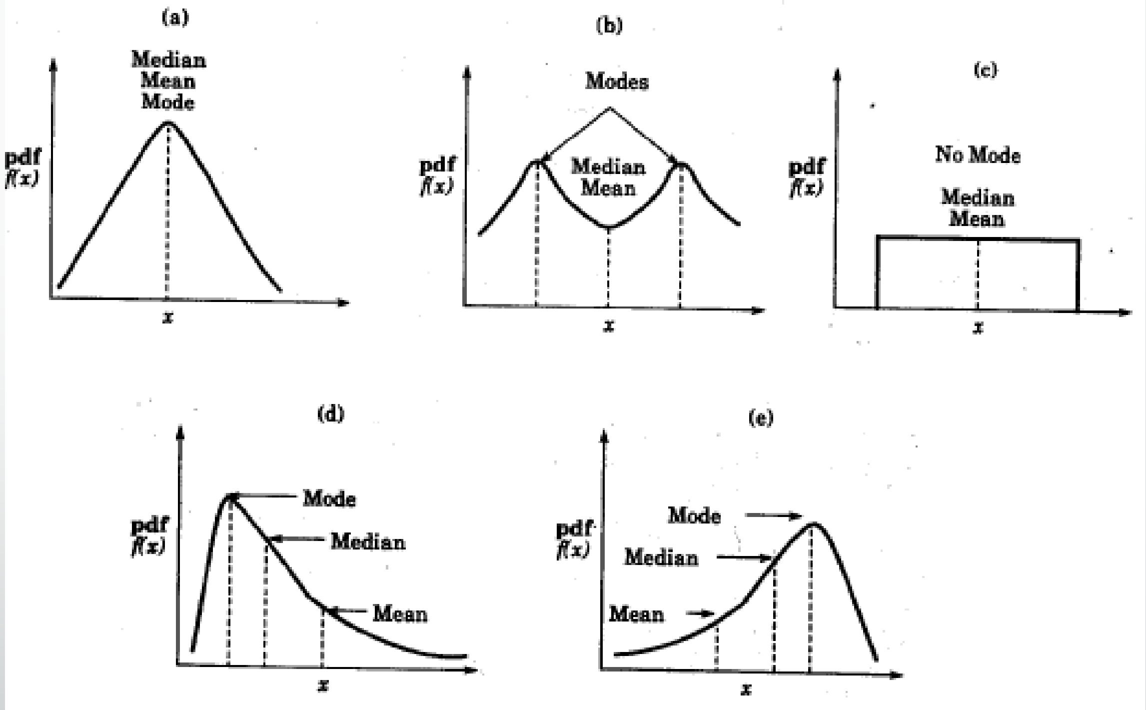
RAZÕES DA POPULARIDADE DA DISTRIBUIÇÃO NORMAL

- A **soma de** ***n*** **variáveis normais independentes é uma variável normal**.
- A **soma de um grande número de observações independentes** de **qualquer distribuição** tende a uma **distribuição normal**:

Teorema do limite central.

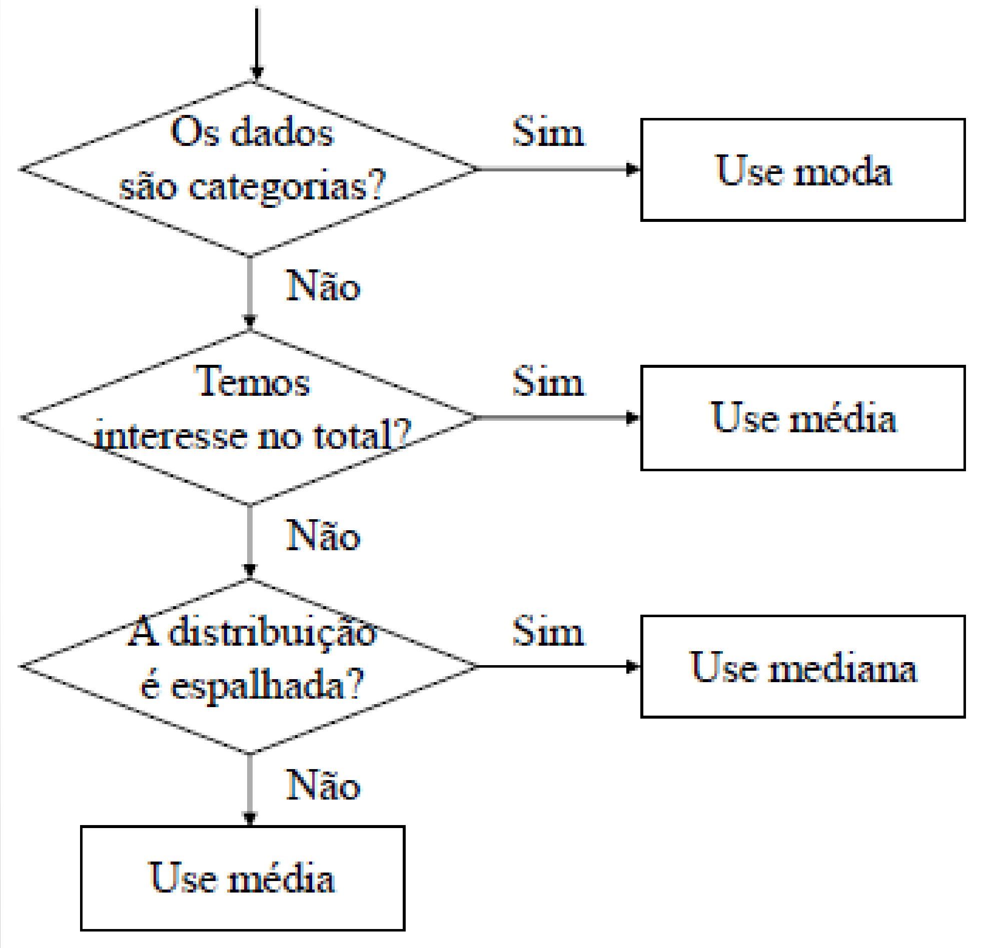
MEDIDAS DE TENDÊNCIA CENTRAL

- **Média aritmética**
  - obtida através da **soma** de todas as **observações** e **dividindo** esta soma pelo **número de observações** da amostra.
- **Mediana**
  - obtida ordenando-se as **observações** em **ordem crescente** e tomando a observação que se encontra no **meio da série**.
- **Moda**
  - **escore** ou **categoria** que, numa distribuição, ocorre **com mais freqüência**.



ESCOLHA DA MEDIDA DE TENDÊNCIA CENTRAL

- **Média**
  - muito afetada por **valores extremos** *(outliers)*
  - dá o **mesmo peso** a cada observação
  - **propriedade linear**: média da soma é a soma das médias.
- **Mediana**
  - exige uma **ordenação**
- **Moda**
  - pode ser obtida para **qualquer conjunto de dados**



ESCOLHA DA MEDIDA DE TENDÊNCIA CENTRAL

- **Recurso mais utilizado** do sistema
  - recursos são categorias, portanto deve-se utilizar a **moda**.
- **Intervalo entre chegadas**
  - o tempo total é de interesse, portanto deve-se utilizar a **média**.
- **Carga** de um computador
  - é preferível usar a **mediana** devido ao espalhamento da distribuição.

MAU USO DAS MÉDIAS

- Usar a média de **valores significativamente diferentes**
  - não é muito útil dizer que o tempo médio de CPU por transação é 505 mseg quando as duas medidas observadas foram 10 e 1000 mseg!
- Usar a média sem levar em conta o **espalhamento da distribuição**

	Sistema A	Sistema B
	10	5
	9	5
	11	5
	10	4
	10	31
Soma	50	50
Média	10	10
Típico	10	5

- Multiplicar as médias para obter a **Média de um produto**
  - se x e y forem correlacionadas => $E(x,y) \neq E(x)E(y)$
- Efetuar a média de **frações com bases diferentes**.

EXERCÍCIO 12.7

12.7 The execution times of queries on a database is normally distributed with a mean of 5 seconds and a standard deviation of 1 second. Determine the following:

a. What is the probability of the execution time being more than 8 seconds?

b. What is the probability of the execution time being less than 6 seconds?

c. What percentage of responses will take between 4 and 7 seconds?

d. What is the 95-percentile execution time?

μ = 5 s
σ = 1 s

EXERCÍCIO 12.7

12.7 The execution times of queries on a database is normally distributed with a mean of 5 seconds and a standard deviation of 1 second. Determine the following:

a. What is the probability of the execution time being more than 8 seconds?

b. What is the probability of the execution time being less than 6 seconds?

c. What percentage of responses will take between 4 and 7 seconds?

d. What is the 95-percentile execution time?

THE SINGLE TABLE FOR THE *STANDARD NORMAL* $f(z)$.

z	-2.5	-2.4	-2.3	-2.2	-2.1	-2.0	-1.9	-1.8	-1.7	-1.6
F(z)	0.006	0.008	0.011	0.014	0.018	0.023	0.029	0.036	0.045	0.055
z	-1.5	-1.4	-1.3	-1.2	-1.1	-1.0	-0.9	-0.8	-0.7	-0.6
F(z)	0.067	0.081	0.097	0.115	0.136	0.159	0.184	0.212	0.242	0.274
z	-0.5	-0.4	-0.3	-0.2	-0.1	0.0	0.1	0.2	0.3	0.4
F(z)	0.309	0.345	0.382	0.421	0.460	0.500	0.540	0.579	0.618	0.655
z	0.5	0.6	0.7	0.8	0.9	1.0	1.1	1.2	1.3	1.4
F(z)	0.691	0.726	0.758	0.788	0.816	0.841	0.864	0.885	0.903	0.919
z	1.5	1.6	1.7	1.8	1.9	2.0	2.1	2.2	2.3	2.4
F(z)	0.933	0.945	0.955	0.964	0.971	0.977	0.982	0.986	0.989	0.992
z	2.5									
F(z)	0.994									

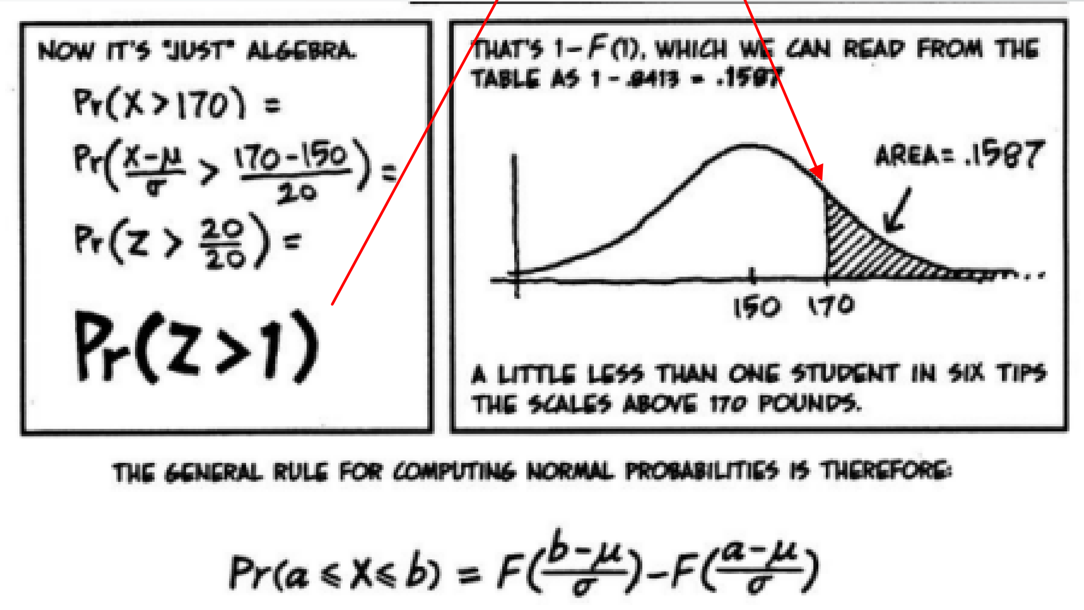
μ = 5 s

σ = 1 s

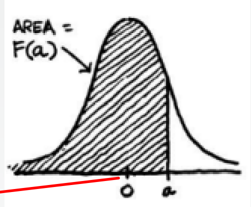
a) ????

b) ????

Tente !



EXERCÍCIO 12.7



12.7 The execution times of queries on a database is normally distributed with a mean of 5 seconds and a standard deviation of 1 second. Determine the following:

a. What is the probability of the execution time being more than 8 seconds?

b. What is the probability of the execution time being less than 6 seconds?

c. What percentage of responses will take between 4 and 7 seconds?

d. What is the 95-percentile execution time?

μ = 5 s ; σ = 1 s

a) $z = (8 - 5)/1 = 3$
$Pr(z > 3) = 1 - 0{,}9987 = \mathbf{0{,}0013}$ (~ 0,1 %)

b) $z = (6 - 5)/1 = 1$
$Pr(z < 1) = \mathbf{0{,}8413}$ (~ 84%)

c) $z1 = (7-5)/1 = 2$
$z2 = (4-5)/1 = -1$
$F(2) - F(-1) = 0{,}9772 - (0{,}8413 - 0{,}5) = \mathbf{0{,}6359}$

d) $F(x_\alpha) = 0{,}95$
$x_\alpha = \mathbf{6{,}65}$

z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5753
0,2	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141
0,3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,6480	0,6517
0,4	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808	0,6844	0,6879
0,5	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,7190	0,7224
0,6	0,7257	0,7291	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486	0,7517	0,7549
0,7	0,7580	0,7611	0,7642	0,7673	0,7704	0,7734	0,7764	0,7794	0,7823	0,7852
0,8	0,7881	0,7910	0,7939	0,7967	0,7995	0,8023	0,8051	0,8078	0,8106	0,8133
0,9	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289	0,8315	0,8340	0,8365	0,8389
1,0	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8554	0,8577	0,8599	0,8621
1,1	0,8643	0,8665	0,8686	0,8708	0,8729	0,8749	0,8770	0,8790	0,8810	0,8830
1,2	0,8849	0,8869	0,8888	0,8907	0,8925	0,8944	0,8962	0,8980	0,8997	0,9015
1,3	0,9032	0,9049	0,9066	0,9082	0,9099	0,9115	0,9131	0,9147	0,9162	0,9177
1,4	0,9192	0,9207	0,9222	0,9236	0,9251	0,9265	0,9279	0,9292	0,9306	0,9319
1,5	0,9332	0,9345	0,9357	0,9370	0,9382	0,9394	0,9406	0,9418	0,9429	0,9441
1,6	0,9452	0,9463	0,9474	0,9484	0,9495	0,9505	0,9515	0,9525	0,9535	0,9545
1,7	0,9554	0,9564	0,9573	0,9582	0,9591	0,9599	0,9608	0,9616	0,9625	0,9633
1,8	0,9641	0,9649	0,9656	0,9664	0,9671	0,9678	0,9686	0,9693	0,9699	0,9706
1,9	0,9713	0,9719	0,9726	0,9732	0,9738	0,9744	0,9750	0,9756	0,9761	0,9767
2,0	0,9772	0,9778	0,9783	0,9788	0,9793	0,9798	0,9803	0,9808	0,9812	0,9817
2,1	0,9821	0,9826	0,9830	0,9834	0,9838	0,9842	0,9846	0,9850	0,9854	0,9857
2,2	0,9861	0,9864	0,9868	0,9871	0,9875	0,9878	0,9881	0,9884	0,9887	0,9890
2,3	0,9893	0,9896	0,9898	0,9901	0,9904	0,9906	0,9909	0,9911	0,9913	0,9916
2,4	0,9918	0,9920	0,9922	0,9925	0,9927	0,9929	0,9931	0,9932	0,9934	0,9936
2,5	0,9938	0,9940	0,9941	0,9943	0,9945	0,9946	0,9948	0,9949	0,9951	0,9952
2,6	0,9953	0,9955	0,9956	0,9957	0,9959	0,9960	0,9961	0,9962	0,9963	0,9964
2,7	0,9965	0,9966	0,9967	0,9968	0,9969	0,9970	0,9971	0,9972	0,9973	0,9974
2,8	0,9974	0,9975	0,9976	0,9977	0,9977	0,9978	0,9979	0,9979	0,9980	0,9981
2,9	0,9981	0,9982	0,9982	0,9983	0,9984	0,9984	0,9985	0,9985	0,9986	0,9986
3,0	0,9987	0,9987	0,9987	0,9988	0,9988	0,9989	0,9989	0,9989	0,9990	0,9990

MÉDIA GEOMÉTRICA

- A **média geométrica** é utilizada se o **produto das observações** for uma quantidade de interesse.
- Calculada através de:

$$\dot{x} = \left(\prod_{i=1}^{n} x_i \right)^{1/n}$$

- Outras medidas que trabalham de forma multiplicativa:
  - Percentual de **melhora de desempenho** entre **versões sucessivas**
  - **Taxa média de erro por etapa** em um caminho de múltiplas etapas numa rede

EXEMPLO 2.12

- Os **melhoramentos de desempenho** na última versão das sete camadas de um *novo* **protocolo de rede** foram **medidos separadamente para cada uma das camadas**:

Camada do Protocolo	Melhoramento do Desempenho (%)
7	18
6	13
5	11
4	8
3	10
2	28
1	5

- Calcule o **melhoramento médio por camada**.

Melhoramento médio por camada

$=\{(1{,}18)(1{,}13)(1{,}11)(1{,}08)(1{,}10)(1{,}28)(1{,}05)\}^{1/7}-1$

$=$ **0,13**

Portanto, o melhoramento médio por camada é de **13%**.

EXEMPLO 12.3

- A utilização da CPU de um sistema medida em cinco intervalos diferentes resultou em:

Duração da Medição	Ocupação da CPU (%)
1	45
1	45
1	45
1	45
100	20
Soma	200%
Média	~~200/5 ou 40%~~

- A utilização média **não é 40%** pois as bases (denominadores) das frações (tempos totais) não são comparáveis.

- A **utilização média** é obtida através do cálculo do tempo total em que a CPU esteve ocupada e do tempo total e da divisão dos dois:

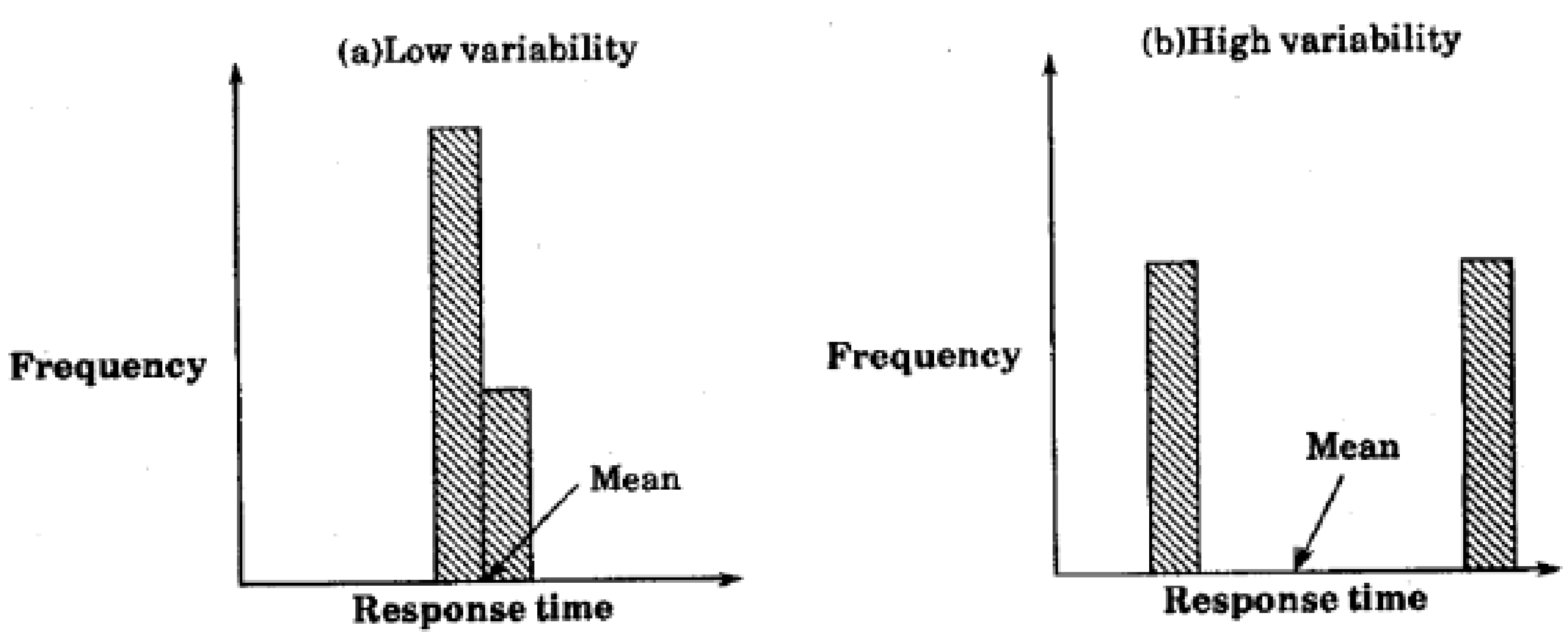
Utilização Média da CPU =

$$\frac{soma\ do\ tempo\ ocupado\ da\ CPU}{soma\ da\ duração\ das\ medições}$$

$$= \frac{0{,}45+0{,}45+0{,}45+0{,}45+20}{1+1+1+1+100}$$ **= 21%**

VARIABILIDADE

- **Tempos de resposta** para dois sistemas com **mesma média**



- **Qual deles você prefere?**

MEDIDAS DE VARIABILIDADE

- Ou "**ÍNDICES DE DISPERSÃO**":
  - Amplitude total
  - Variância ou Desvio Padrão
  - Postos percentil 10 e 90
  - Metade da distância interquartílica
  - Desvio Médio absoluto

AMPLITUDE TOTAL

- É a **diferença entre o maior e o menor escore** da distribuição.
- É **simples** mas extremamente dependente dos valores extremos:
  - o mínimo pode ser zero e o máximo um ponto atípico, fora da curva
- É útil apenas se houver uma boa razão para acreditar que a **variável** seja **limitada**.

VARIÂNCIA

- A **variância de uma amostra** de *n* observações é calculada da seguinte forma:

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

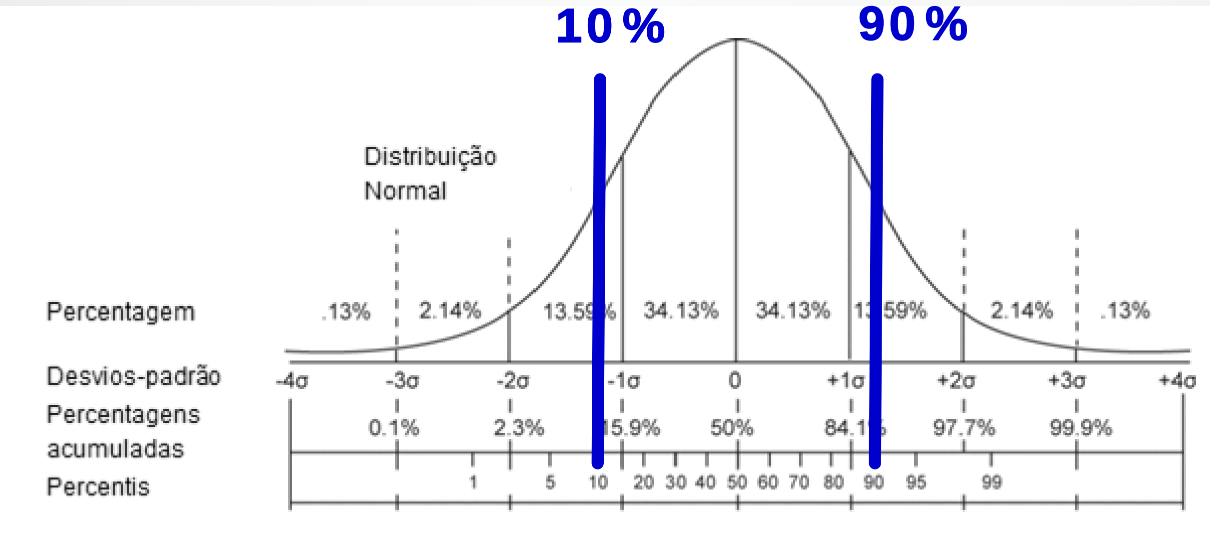
- O **desvio padrão de uma amostra** é a raiz quadrada da variância da amostra.

- **Coeficiente de Variação**

$$\textbf{C.O.V} = \frac{desvio\ padrão}{média} = \boldsymbol{\frac{\sigma}{\mu}}$$

POSTOS PERCENTIL 10 E 90

- Semelhante à Amplitude Total, mas **funciona mesmo que a variável não seja limitada**.

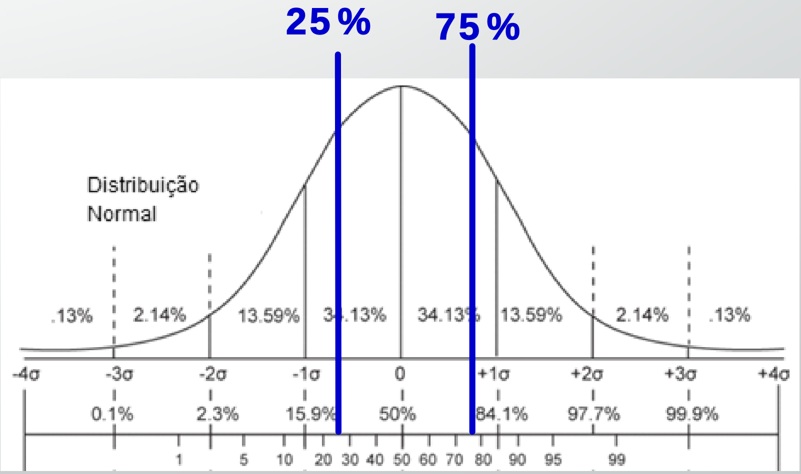


METADE DA DISTÂNCIA INTERQUARTÍLICA

- A **distância interquartílica** é a **diferença** entre o **terceiro** e o **primeiro quartil**.

- **SIQR** ***(Semi-Interquartil Range)***

$$\mathrm{SIQR} = \frac{Q_3 - Q_1}{2} = \frac{x_{0,75} - x_{0,25}}{2}$$



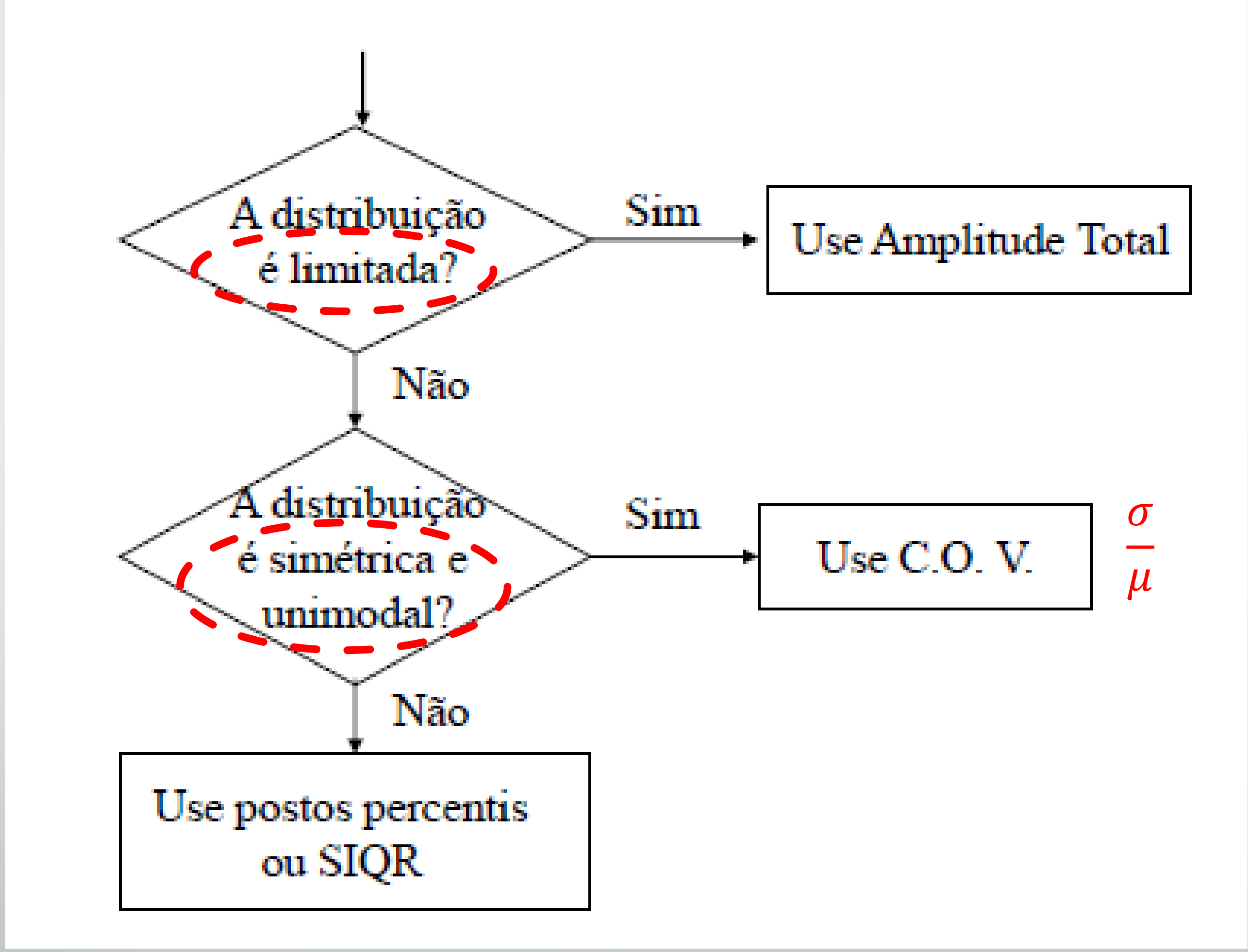
DESVIO MÉDIO ABSOLUTO

- Calculado através de:

$$\text{Desvio médio absoluto} = \frac{1}{n}\sum_{i=1}^{n}|x_i - \bar{x}|$$

- Vantagem principal sobre o desvio padrão: **não faz produtos nem extrai raiz quadrada**.

SELEÇÃO DA MEDIDA DE VARIABILIDADE



EXEMPLO 12.4

- Em um **experimento**, **repetido 32 vezes**, os **tempos medidos** de CPU foram:

{3,1; 4,2; 2,8; 5,1; 2,8; 4,4; 5,6; 3,9; 3,9; 2,7; 4,1; 3,6; 3,1; 4,5; 3,8; 2,9; 3,4; 3,3; 2,8; 4,5; 4,9; 5,3; 1,9; 3,7; 3,2; 4,1; 5,1; 3,2; 3,9; 4,8; 5,9; 4,2}

Qual é a variabilidade dos tempos:

- em postos percentis 10 e 90?

- em SIQR?

EXEMPLO 12.4

- O **conjunto ordenado** é:

{1,9; 2,7; 2,8; **2,8**; 2,8; 2,9; 3,1; 3,1; **3,2**; 3,2; 3,3; 3,4; 3,6; 3,7; 3,8; 3,9; 3,9; 3,9; 4,1; 4,1; 4,2 ; 4,2; 4,4; 4,5; **4,5**; 4,8; 4,9; 5,1; **5,1**; 5,3; 5,6; 5,9}

- O **posto percentil 10** é dado por [1+(31)(0,10)]= **4º Elemento** = 2,8
- O **posto percentil 90** é dado por [1+(31)(0,90)]= **29º Elemento** = 5,1

Portanto

Distância = 5,1 – 2,8 = **2,3**

- **Q_1** é dado por [1+(31)(0,25)] = **9º Elemento** = 3,2
- **Q_3** é dado por [1+(31)(0,75)] = **24º Elemento** = 4,5

Portanto,

SIQR = $\frac{Q_3 - Q_1}{2} = \frac{4,5 - 3,2}{2} =$ **0,65**

DETERMINAÇÃO DA DISTRIBUIÇÃO DOS DADOS

- O modo mais fácil é fazer um gráfico com o **histograma** das observações.
  - Usando, por exemplo, a ferramenta de análise de dados-histograma do Excel!
- O maior problema é determinar o **tamanho de cada classe** (célula).
  - Se qualquer classe tiver menos do que 5 observações, deve-se aumentar o tamanho das classes ou usar um histograma com classes de tamanhos variáveis.

GRÁFICO QUANTIL-QUANTIL

- Para **pequenas amostras** o melhor é fazer um **gráfico dos quantis** observados em relação ao **quantil teórico**.
- Se a distribuição da amostra corresponder à distribuição teórica, o **gráfico quantil-quantil deve ser linear**.
- Os **quantis da distribuição teórica** são obtidos através de **transformação inversa da CDF**:

$$x_i = F^{-1}(q_i)$$

Distribuição	CDF $F(x)$	Inversa
Exponencial	$1-e^{-x/a}$	$-a\ln(u)$
Valor Extremo	$1-e^{-e^{(x-a)/b}}$	$a+b\ln\ln(u)$
Geométrica	$1-(1-p)^x$	$\left\lceil \dfrac{\ln(u)}{\ln(1-p)} \right\rceil$
Logística	$1-\dfrac{1}{1+e^{(x-\mu)/b}}$	$\mu-b\ln\left(\dfrac{1}{u}-1\right)$
Pareto	$1-x^{-a}$	$1/u^{1/a}$
Weibull	$1-e^{(x/a)^b}$	$a(\ln u)^{1/b}$

- Para a **distribuição normal unitária** *N(0,1)* utiliza-se freqüentemente a seguinte **aproximação**:

$$x_i = 4{,}91\left[q_i^{0{,}14} - (1-q_i)^{0{,}14}\right]$$

EXEMPLO 12.5

- O **erro de modelagem** (diferença entre valores medidos e valores previstos por um modelo de distribuição normal) para **8 predições de um modelo** foram os seguintes:

-0,04; -0,19; 0,14; -0,09; -0,14; 0,19; 0,04 e 0,09.

i	*qi=(i-0,5)/n*	*yi*	*xi*
1	0,0625	-0,19	-1,535
2	0,1875	-0,14	-0,885
3	0,3125	-0,09	-0,487
4	0,4375	-0,04	-0,157
5	0,5625	0,04	0,157
6	0,6875	0,09	0,487
7	0,8125	0,14	0,885
8	0,9375	0,19	1,535

A distribuição dos valores medidos corresponde a uma distribuição normal?

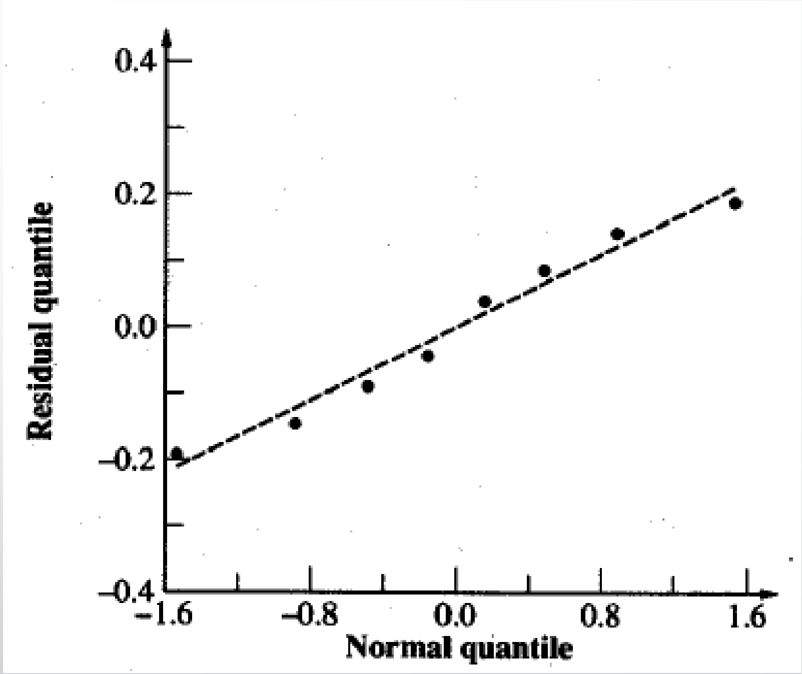
EXEMPLO 12.5

- O **erro de modelagem** (diferença entre valores medidos e valores previstos por um modelo) para 8 predições de um modelo foram os seguintes:

-0,04; -0,19; 0,14; -0,09; -0,14; 0,19; 0,04 e 0,09.

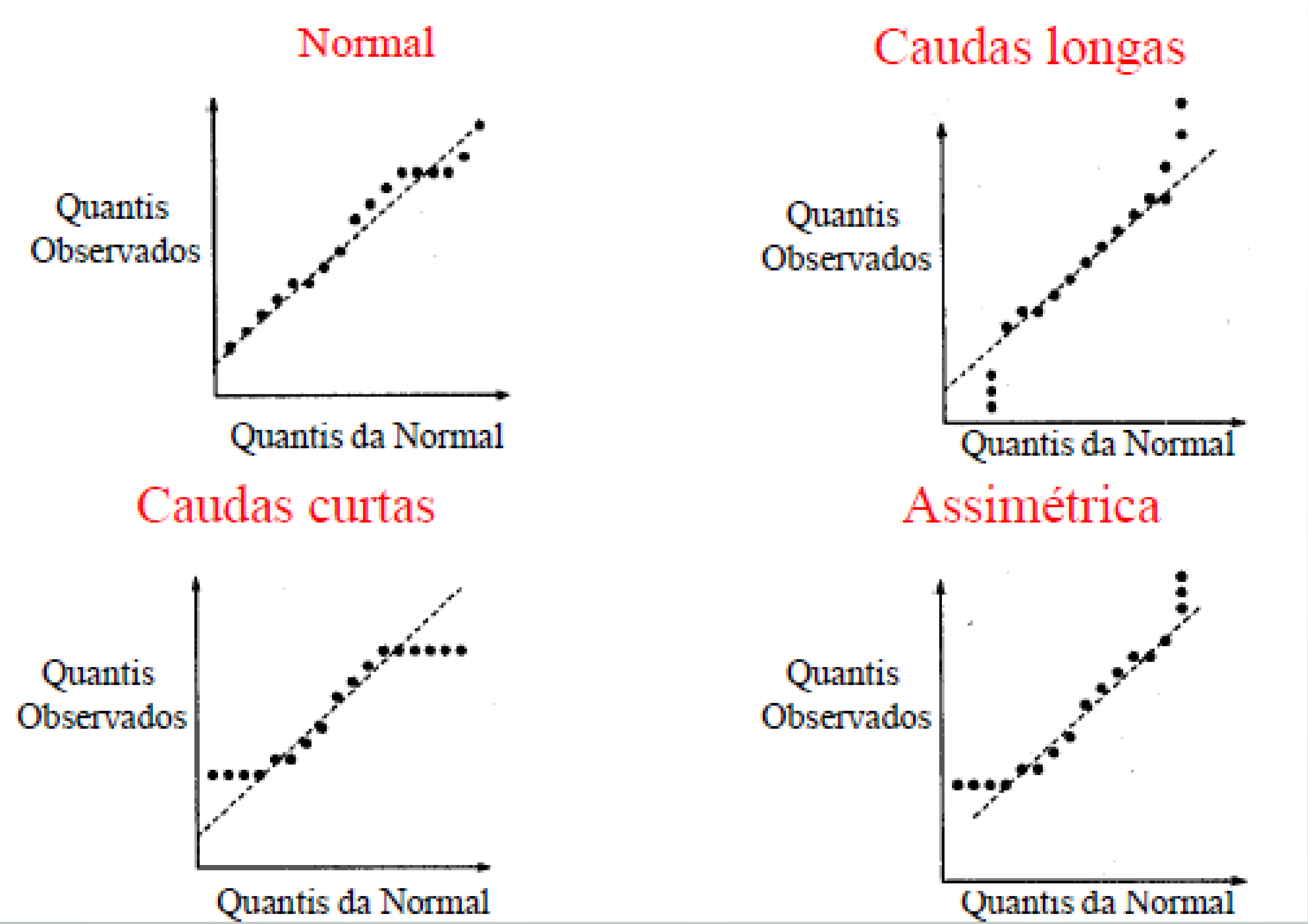
i	*qi=(i-0,5)/n*	*yi*	*xi*
1	0,0625	-0,19	-1,535
2	0,1875	-0,14	-0,885
3	0,3125	-0,09	-0,487
4	0,4375	-0,04	-0,157
5	0,5625	0,04	0,157
6	0,6875	0,09	0,487
7	0,8125	0,14	0,885
8	0,9375	0,19	1,535

$$x_i = 4{,}91\left[q_i^{0{,}14} - (1-q_i)^{0{,}14}\right]$$



Os erros aparentam ser distribuídos normalmente!

DESVIOS DA DISTRIBUIÇÃO NORMAL



Box 12.1 Summarizing Observations

Given: A sample $\{x_1, x_2, \ldots, x_n\}$ of n observations.

1. Sample arithmetic mean: $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$
2. Sample geometric mean: $\dot{x} = \left(\prod_{i=1}^{n} x_i\right)^{1/n}$
3. Sample harmonic mean: $\ddot{x} = \frac{n}{\frac{1}{x_1} + \frac{1}{x_2} + \cdots + \frac{1}{x_n}}$
4. Sample median: $\begin{cases} x_{((n-1)/2)} & \text{if } n \text{ is odd} \\ 0.5(x_{(n/2)} + x_{((1+n)/2)}) & \text{otherwise} \end{cases}$

   Here $x_{(i)}$ is the ith observation in the sorted set.
5. Sample mode = observation with the highest frequency (for categorical data).
6. Sample variance: $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$
7. Sample standard deviation: $s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2}$
8. Coefficient of variation $= s/\bar{x}$
9. Coefficient of skewness $= \frac{1}{ns^3}\sum_{i=1}^{n}(x_i - \bar{x})^3$
10. Range: Specify the minimum and maximum.
11. Percentiles: 100p-percentile $x_p = x_{(\lfloor 1+(n-1)p \rfloor)}$.
12. Semi-interquartile range SIQR $= \frac{Q_3 - Q_1}{2} = \frac{x_{0.75} - x_{0.25}}{2}$
13. Mean absolute deviation $= \frac{1}{n}\sum_{i=1}^{n}|x_i - \bar{x}|$

COMPARING SYSTEMS USING SAMPLE DATA

Capítulo 13

Raj Jain - The Art of Computer Systems Performance Analysis. Wiley, 1991.

AMOSTRAS E POPULAÇÕES

- **População** ou **Universo**: todo o conjunto de possíveis resultados.
- **Amostra:** número menor de resultados extraídos da população.
- A idéia é **generalizar** (conclusões) **da amostra** (o grupo pequeno) **para a população toda** (o grupo maior), da qual essa mesma amostra foi extraída.

- Uma **amostra** é apenas um **exemplo!**
  - As palavras inglesas *sample* (amostra) e *example* (exemplo) têm origem na palavra francesa *essample.*
- **Um exemplo não é suficiente para provar uma teoria**!
- Queremos fazer afirmações probabilísticas sobre a **região em que as características do sistema residem**:

>>> INTERVALO DE CONFIANÇA

MÉDIA DA POPULAÇÃO E DAS AMOSTRAS

- Suponha que escrevemos um programa para gerar diversos **milhões de números aleatórios** com **média** μ e **desvio padrão** σ.
- Em seguida colocamos estes números numa urna e extraímos uma **amostra** formada por n observações.
- Suponha que a amostra $\{x_1, x_2, \ldots, x_n\}$ possua uma **média**.
- A **média da amostra** provavelmente é **diferente** da **média da população**!

- Em muitos problemas da vida real desconhecemos as **características da população** (p. ex., a média da população).
  - A **média de uma amostra** é usada como uma **estimativa da média da população**.
- As **características da população** são chamadas de **parâmetros** enquanto que as **estimativas das amostras** são chamadas de **estatísticas**.
  - Os **parâmetros são constantes**, enquanto que as **estatísticas são variáveis aleatórias**.

INTERVALO DE CONFIANÇA

- **Cada média de uma amostra** é uma **estimativa da média da população**.
  - Dadas *k* **amostras**, teremos *k* **estimativas**, todas **diferentes**.
- Como obter uma **estimativa da média da população** a partir destas *k* **estimativas**?
  - O melhor que podemos fazer é obter **limites probabilísticos**!

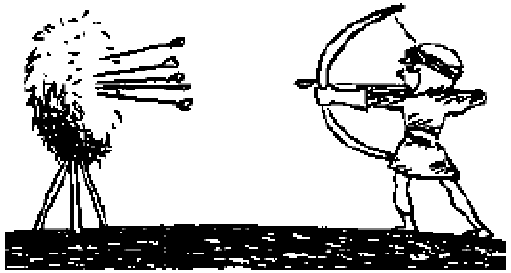
Probabilidade $\{c_1 \leq \mu \leq c_2\} = 1 - \alpha$

- O **intervalo** (c_1, c_2) é chamado de **Intervalo de Confiança (IC) da média da população**.
- α é o **nível de significância**.
- **100(1-α) é o nível de confiança.**
- **1-α** é o **coeficiente de confiança.**

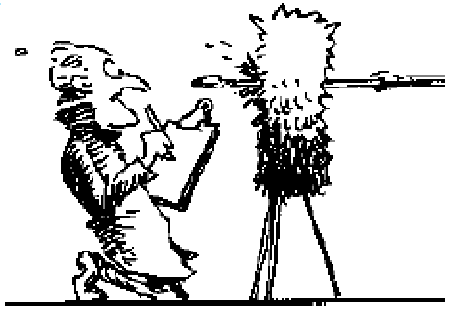
$\alpha = 0{,}1 \Rightarrow$ **nível de confiança = 90%**

INTERVALO DE CONFIANÇA (2)

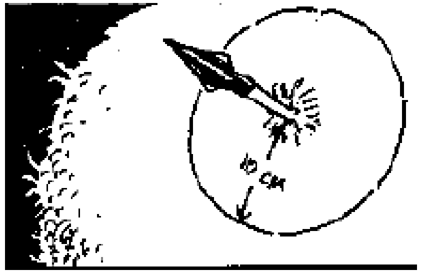
CONSIDERE UMA ARQUEIRA ATIRANDO EM UM ALVO. SUPONHA QUE ELA ACERTA NO CENTRO COM RAIO DE 10 CM 95% DAS VEZES. OU SEJA, ERRA APENAS UMA VEZ A CADA 20 TENTATIVAS.



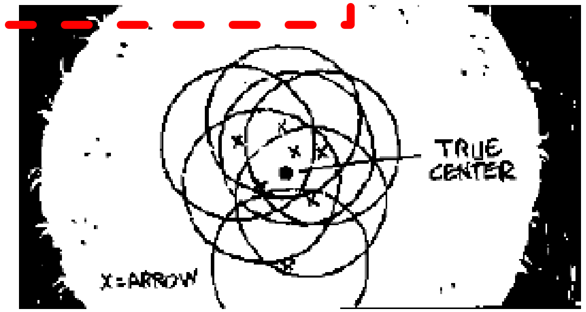
SENTADO ATRÁS DO ALVO ENCONTRA-SE UM BRAVO DETETIVE, QUE NÃO VÊ ONDE ESTÁ O CENTRO. A ARQUEIRA ATIRA A PRIMEIRA FLECHA..



CONHECENDO O NÍVEL DA HABILIDADE DA ARQUEIRA, O DETETIVE DESENHA UM CÍRCULO COM 10 CM DE RAIO AO REDOR DA FLECHA. ELE TEM **95% DE CONFIANÇA** DE QUE O SEU CÍRCULO INCLUI O CENTRO DO ALVO!



ELE RACIOCINOU QUE SE DESENHASSE CÍRCULOS COM 10 CM DE RAIO AO REDOR DE **MUITAS FLECHAS**, OS SEUS CÍRCULOS INCLUIRIAM O CENTRO DO ALVO EM 95% DOS CASOS.



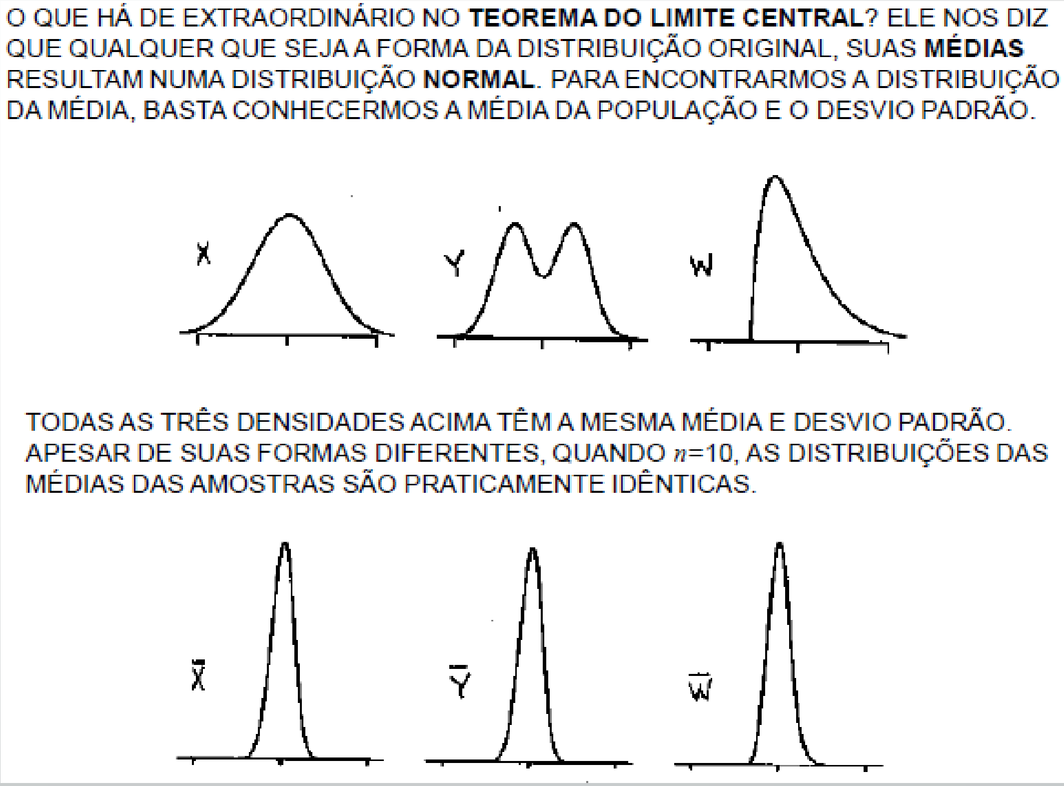
INTERVALO DE CONFIANÇA (3)

- Uma forma de determinar o **intervalo de confiança de 90 %** seria utilizar os **postos percentis 5 e 95** das médias das amostras, **como limites**.
- Por exemplo, poderíamos fazer ***k* amostras**, calcular a média de cada amostra, e tomar o **[1+0,05(*k*-1)]-ésimo** e o **[1+0,95(*k*-1)]-ésimo** elemento do **conjunto ordenado**.
- **Felizmente não é preciso obter tantas amostras**!

TEOREMA DO LIMITE CENTRAL

- Se as **observações** em uma dada amostra $\{x_1, x_2, \dots, x_n\}$ forem **independentes** e extraídas de uma **mesma população** com **média** μ e **desvio padrão** σ, então a **média da amostra** para um **número** n **grande** de observações terá uma **distribuição aproximadamente normal** com **média** μ e **desvio padrão** $\sigma/\sqrt{n}$:

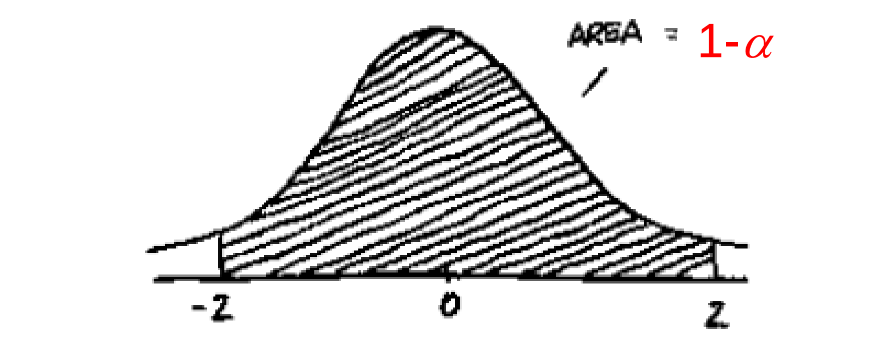
$$\bar{x} \sim N\left(\mu, \underbrace{\sigma/\sqrt{n}}_{\text{Erro padrão}}\right)$$



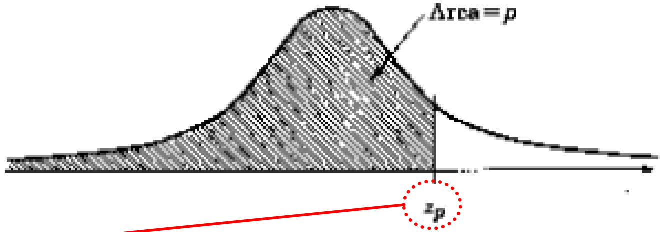
CÁLCULO DO INTERVALO DE CONFIANÇA

- Usando o **teorema do limite central**, um **intervalo de confiança** de **100(1-α)%** para a **média da população** seria dado por

$$\left(\bar{x} - z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}\ ,\bar{x} + z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}\right)$$



Extrato da
Tabela A.2
(ref. Jain)



p	0.000	0.001	0.002	0.003	0.004	0.005	0.006	0.007	0.008	0.009
0.90	1.282	1.287	1.293	1.299	1.305	1.311	1.317	1.323	1.329	1.335
0.91	1.341	1.347	1.353	1.359	1.366	1.372	1.379	1.385	1.392	1.398
0.92	1.405	1.412	1.419	1.426	1.433	1.440	1.447	1.454	1.461	1.468
0.93	1.476	1.483	1.491	1.499	1.506	1.514	1.522	1.530	1.538	1.546
0.94	1.555	1.563	1.572	1.580	1.589	1.598	1.607	1.616	1.626	1.635
0.95	1.645	1.655	1.665	1.675	1.685	1.695	1.706	1.717	1.728	1.739
0.96	1.751	1.762	1.774	1.787	1.799	1.812	1.825	1.838	1.852	1.866
0.97	1.881	1.896	1.911	1.927	1.943	1.960	1.977	1.995	2.014	2.034
0.98	2.054	2.075	2.097	2.120	2.144	2.170	2.197	2.226	2.257	2.290

p	0.0000	0.0001	0.0002	0.0003	0.0004	0.0005	0.0006	0.0007	0.0008	0.0009
0.995	2.576	2.583	2.590	2.597	2.605	2.612	2.620	2.628	2.636	2.644

EXEMPLO 13.1

- **Exemplo 12.4**: em um experimento, repetido **32 vezes**, os tempos medidos de CPU foram:

{3,1; 4,2; 2,8; 5,1; 2,8; 4,4; 5,6; 3,9; 3,9; 2,7; 4,1; 3,6; 3,1; 4,5; 3,8; 2,9; 3,4; 3,3; 2,8; 4,5; 4,9; 5,3; 1,9; 3,7; 3,2; 4,1; 5,1; 3,2; 3,9; 4,8; 5,9; 4,2}

- Para a **amostra do Exemplo 12.4** temos:
  - **média** μ = **?**
  - **desvio padrão** σ = **?**
  - n = **32**
  - **erro padrão** = $\sigma/\sqrt{n}$ = ?

EXEMPLO 13.1

- **Exemplo 12.4**: em um experimento, repetido **32 vezes**, os tempos medidos de CPU foram (ordenados):

{1,9; 2,7; 2,8; 2,8; 2,8; 2,9; 3,1; 3,1; 3,2; 3,2; 3,3; 3,4; 3,6; 3,7; 3,8; 3,9; 3,9; 3,9; 4,1; 4,1; 4,2; 4,2; 4,4; 4,5; 4,5; 4,8; 4,9; 5,1; 5,1; 5,3; 5,6; 5,9}

- Para a **amostra do Exemplo 12.4** temos:
  - **Média =** (1,9+2,7+2,8+2,8+2,8+2,9+3,1+3,1+3,2+3,2+3,3+3,4+3,6+3,7+3,8+3,9+3,9+3,9+4,1+4,1+4,2+4,2+4,4+4,5+4,5+4,8+4,9+5,1+5,1+5,3+5,6+5,9)/32 = 121,9/32 = μ = **3,80**
  - **Variância de uma amostra de "n" observações =**

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

$[(1,9-3,8)^2+(2,7-3,8)^2+(2,8-3,8)^2+(2,8-3,8)^2+(2,8-3,8)2+(2,9-3,8)^2+(3,1-3,8)^2+(3,1-3,8)^2+(3,2-3,8)^2+(3,2-3,8)^2+(3,3-3,8)^2+(3,4-3,8)^2+(3,6-3,8)^2+(3,7-3,8)^2+(3,8-3,8)^2+(3,9-3,8)^2+(3,9-3,8)^2+(3,9-3,8)^2+(4,1-3,8)^2+(4,1-3,8)^2+(4,2-3,8)^2+(4,2-3,8)^2+(4,4-3,8)^2+(4,5-3,8)^2+(4,5-3,8)^2+(4,8-3,8)^2+(4,9-3,8)^2+(5,1-3,8)^2+(5,1-3,8)^2+(5,3-3,8)^2+(5,6-3,8)^2+(5,9-3,8)^2]/31$ **=**

$[(-1,9)^2+(-1,1)^2+(-1,0)^2+(-1,0)^2+(-1,0)^2+(-0,9)^2+(-0,7)^2+(-0,7)^2+(-0,6)^2+(-0,6)^2+(-0,5)^2+(-0,4)^2+(-0,2)^2+(-0,1)^2+(0) +(0,1)^2 +(0,1)^2 +(0,1)^2 +(0,3)^2 +(0,3)^2 +(0,4)^2 +(0,4)^2 +(0,6)^2 +(0,7)^2 +(0,7)^2 +(1,0)^2 +(1,1)^2 +(1,3)^2+(1,3)^2+(1,5)^2+(1,8)^2+(2,1)^2]$ /31 =

[3,61 + 1,21 + 1,0 + 1,0 + 1,0 + 0,81 + 0,49 + 0,49 + 0,36 + 0,36 + 0,25 + 0,16 + 0,04 + 0,01 + 0 + 0,01 + 0,01 + 0,01 + 0,09+ 0,09 + 0,16+ 0,16 + 0,36 +0,49 +0,49 + 1,0 + 1,21 +1,69 +1,69 +2,25 +3,24 +4,41]/32 = 28,15/31 = 0,90

- **desvio padrão**

  $\sigma = \sqrt{0,90} =$ **0,95**

- **número de observações** n **=32**
- **erro padrão**

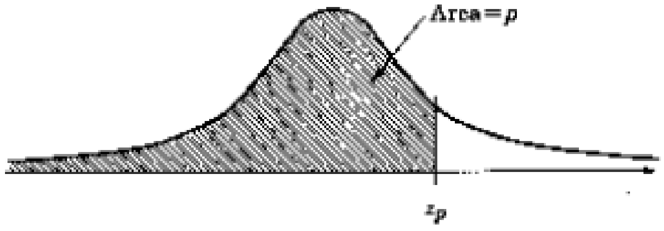
  $\sigma/\sqrt{n}$ = 0,168

EXEMPLO 13.1

- **Exemplo 12.4**: em um experimento, repetido **32 vezes**, os tempos medidos de CPU foram (ordenados):

{1,9; 2,7; 2,8; 2,8; 2,8; 2,9; 3,1; 3,1; 3,2; 3,2; 3,3; 3,4; 3,6; 3,7; 3,8; 3,9; 3,9; 3,9; 4,1; 4,1; 4,2; 4,2; 4,4; 4,5; 4,5; 4,8; 4,9; 5,1; 5,1; 5,3; 5,6; 5,9}

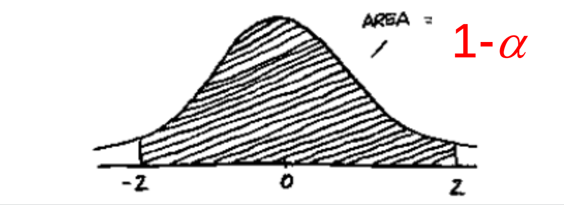
Extrato da **Tabela A.2** (ref. Jain)



p	0.000	0.001	0.002	0.003	0.004	0.005	0.006	0.007	0.008	0.009
0.90	1.282	1.287	1.293	1.299	1.305	1.311	1.317	1.323	1.329	1.335
0.91	1.341	1.347	1.353	1.359	1.366	1.372	1.379	1.385	1.392	1.398
0.92	1.405	1.412	1.419	1.426	1.433	1.440	1.447	1.454	1.461	1.468
0.93	1.476	1.483	1.491	1.499	1.506	1.514	1.522	1.530	1.538	1.546
0.94	1.555	1.563	1.572	1.580	1.589	1.598	1.607	1.616	1.626	1.635
0.95	1.645	1.655	1.665	1.675	1.685	1.695	1.706	1.717	1.728	1.739
0.96	1.751	1.762	1.774	1.787	1.799	1.812	1.825	1.838	1.852	1.866
0.97	1.881	1.896	1.911	1.927	1.943	1.960	1.977	1.995	2.014	2.034
0.98	2.054	2.075	2.097	2.120	2.144	2.170	2.197	2.226	2.257	2.290
p	0.0000	0.0001	0.0002	0.0003	0.0004	0.0005	0.0006	0.0007	0.0008	0.0009
0.995	2.576	2.583	2.590	2.597	2.605	2.612	2.620	2.628	2.636	2.644

- Para a **amostra do Exemplo 12.4** temos **média** μ = **3,80**; **desvio padrão** σ =**0,95**; n =**32;** e $\sigma/\sqrt{n}$ =**0,168** .
- **Qual é o Intervalo de confiança** de **90%** para a **média ?**

$$(c_1\, ,\, c_2) = \left(\bar{x} - z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}\ \ , \bar{x} + z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}\right)$$

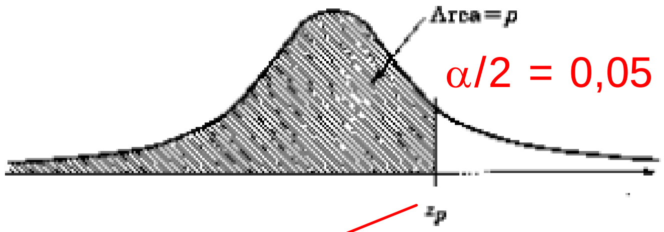


EXEMPLO 13.1

- **Exemplo 12.4**: em um experimento, repetido **32 vezes**, os tempos medidos de CPU foram (ordenados):

{1,9; 2,7; 2,8; 2,8; 2,8; 2,9; 3,1; 3,1; 3,2; 3,2; 3,3; 3,4; 3,6; 3,7; 3,8; 3,9; 3,9; 3,9; 4,1; 4,1; 4,2; 4,2; 4,4; 4,5; 4,5; 4,8; 4,9; 5,1; 5,1; 5,3; 5,6; 5,9}

Extrato da **Tabela A.2** (ref. Jain)

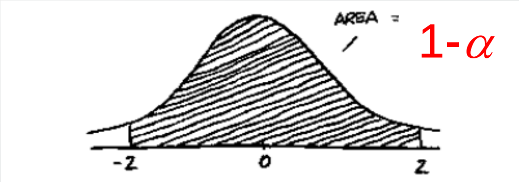


p	0.000	0.001	0.002	0.003	0.004	0.005	0.006	0.007	0.008	0.009
0.90	1.282	1.287	1.293	1.299	1.305	1.311	1.317	1.323	1.329	1.335
0.91	1.341	1.347	1.353	1.359	1.366	1.372	1.379	1.385	1.392	1.398
0.92	1.405	1.412	1.419	1.426	1.433	1.440	1.447	1.454	1.461	1.468
0.93	1.476	1.483	1.491	1.499	1.506	1.514	1.522	1.530	1.538	1.546
0.94	1.555	1.563	1.572	1.580	1.589	1.598	1.607	1.616	1.626	1.635
0.95	1.645	1.655	1.665	1.675	1.685	1.695	1.706	1.717	1.728	1.739
0.96	1.751	1.762	1.774	1.787	1.799	1.812	1.825	1.838	1.852	1.866
0.97	1.881	1.896	1.911	1.927	1.943	1.960	1.977	1.995	2.014	2.034
0.98	2.054	2.075	2.097	2.120	2.144	2.170	2.197	2.226	2.257	2.290

p	0.0000	0.0001	0.0002	0.0003	0.0004	0.0005	0.0006	0.0007	0.0008	0.0009
0.995	2.576	2.583	2.590	2.597	2.605	2.612	2.620	2.628	2.636	2.644

- Para a **amostra do Exemplo 12.4** temos **média** μ = **3,80**; **desvio padrão** σ =**0,95**; n =**32;** e $\sigma/\sqrt{n}$ =**0,168** .
- **Intervalo de confiança** de **90%** para a **média**

$\alpha = 0,1 \Rightarrow \alpha/2 = 0,05$



$$\left(\bar{x} - z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}, \bar{x} + z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}\right)$$

$= \mathbf{3,80} \pm (1,645)(0,95)/\sqrt{32} = \mathbf{(3,52\ ;\ 4,07)}$

EXEMPLO 13.1

- **Exemplo 12.4**: em um experimento, repetido **32 vezes**, os tempos medidos de CPU foram (ordenados):

{1,9; 2,7; 2,8; 2,8; 2,8; 2,9; 3,1; 3,1; 3,2; 3,2; 3,3; 3,4; 3,6; 3,7; 3,8; 3,9; 3,9; 3,9; 4,1; 4,1; 4,2; 4,2; 4,4; 4,5; 4,5; 4,8; 4,9; 5,1; 5,1; 5,3; 5,6; 5,9}

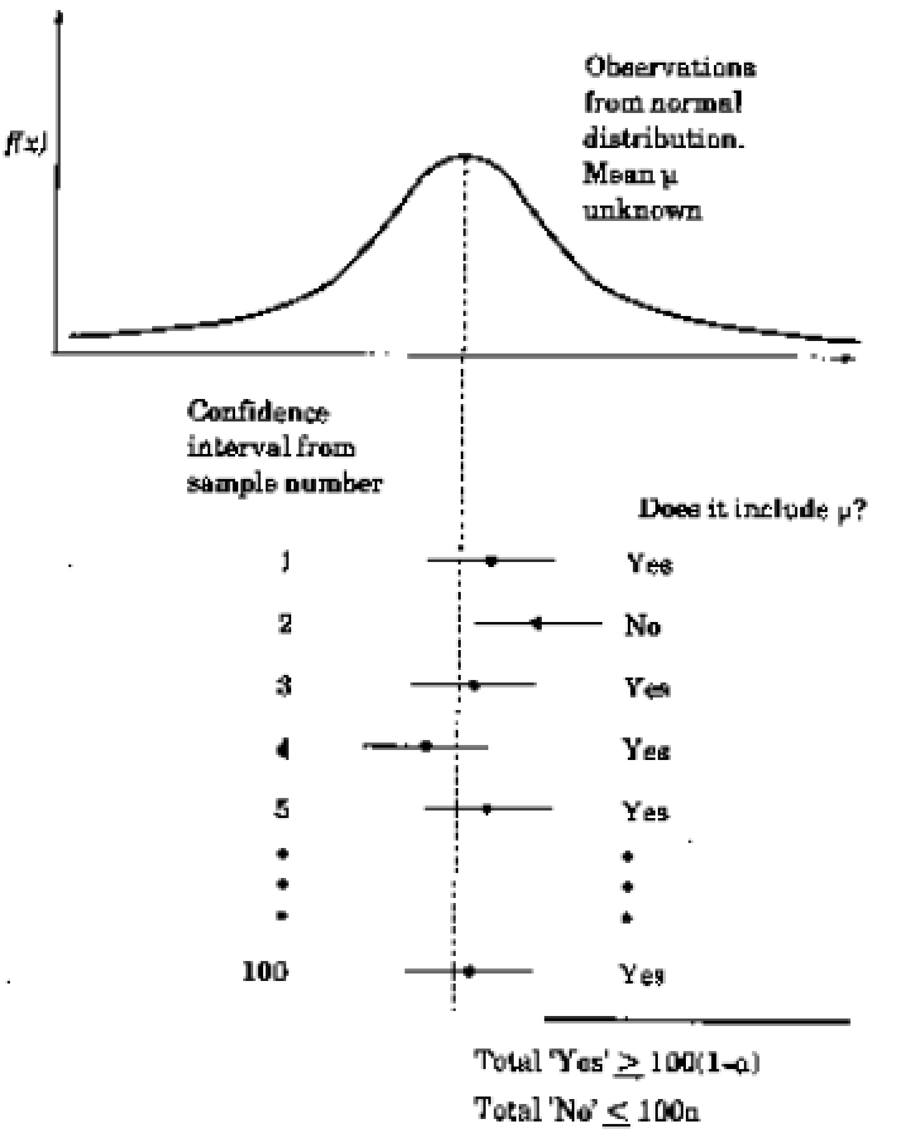
- Para a **amostra do Exemplo 12.4** temos **média** μ = **3,80**; **desvio padrão** σ =**0,95** e n =**32**.
  - **Intervalo de confiança** de **90%** para a **média**

    $= 3{,}80 \pm (1{,}645)(0{,}95)/\sqrt{32} =$ **(3,52 ; 4,07)**
  - **Intervalo de confiança** de **95%** para a **média**

    $= 3{,}80 \pm (1{,}960)(0{,}95)/\sqrt{32} =$ **(3,47 ; 4,33)**
  - **Intervalo de confiança** de **99%** para a **média**

    $= 3{,}80 \pm (2{,}576)(0{,}95)/\sqrt{32} =$ **(3,36 ; 4,43)**

SIGNIFICADO DO INTERVALO DE CONFIANÇA



Como melhorar a confiança?

QUAL É O NÍVEL DE CONFIANÇA A SER USADO?

- Apesar de usarmos **ICs de 90 ou 95%**, eles não precisam ser sempre assim tão altos!
- A **escolha** depende da **perda** caso o parâmetro se encontre fora da faixa do IC e do **ganho** caso ele se encontre dentro da faixa do IC.
- Se a **perda for grande** comparada com o ganho, os **níveis de confiança** devem ser **altos**.
- Se a **perda for desprezível** comparada com o ganho, um **nível de confiança baixo** é **suficiente**.

- Considere, por exemplo, uma loteria na qual o **bilhete** custa **um dólar** mas paga **5 milhões para o vencedor**.
- Suponha que a probabilidade de vencer é 10^{-7}
- Para vencer a loteria com **90%** de confiança seria necessário **comprar 9 milhões de bilhetes**!
- Neste caso, um nível de confiança baixo como 0,01% (**bem mais barato**) já seria suficiente.

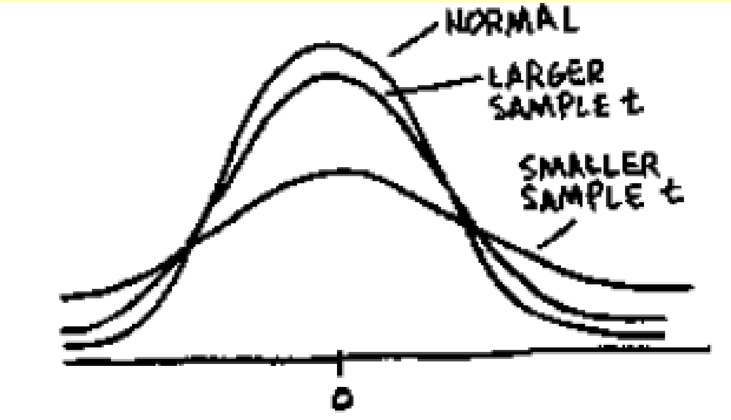
INTERVALO DE CONFIANÇA PARA PEQUENAS AMOSTRAS

- Os resultados anteriores aplicam-se **somente** a **amostras maiores do que 30**.
- Para **amostras menores** as observações devem ser extraídas de uma **população distribuída normalmente**.
- Para estas amostras, o **IC é dado** por:

$$\left(\bar{x} - \underbrace{t_{[1-\alpha/2;n-1]}}_{} \frac{\sigma}{\sqrt{n}}, \bar{x} + t_{[1-\alpha/2;n-1]} \frac{\sigma}{\sqrt{n}}\right)$$

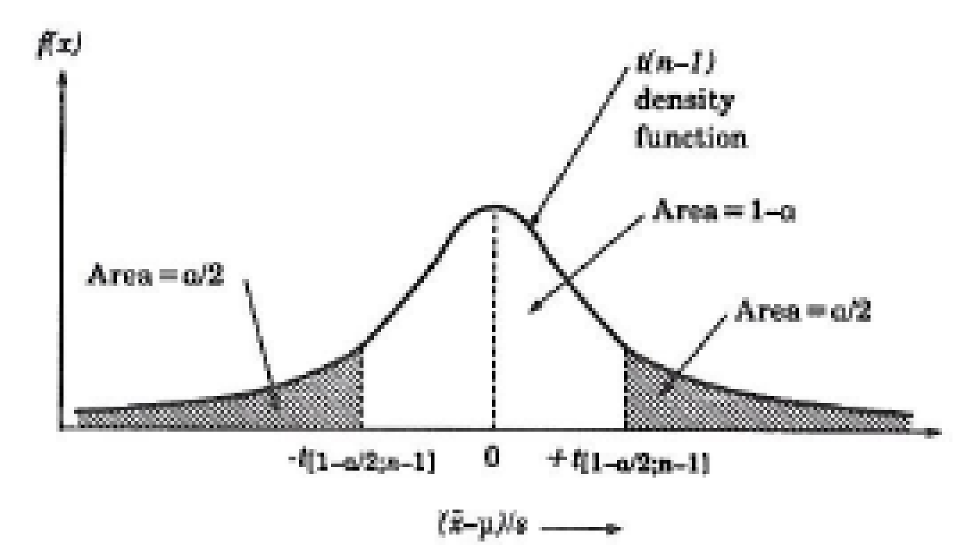
Quantil (1-α/2) de uma variável *t* com *n*-1 graus de liberdade

- **Distribuição *t* de Student**
- Inventada por William Gosset, sob o pseudônimo de "Student"
- *t* é **mais espalhada** do que *z*
- O grau de espalhamento depende do **tamanho da amostra**:



FUNÇÃO DENSIDADE *t(n-1)*

- **Função densidade *t(n-1)***



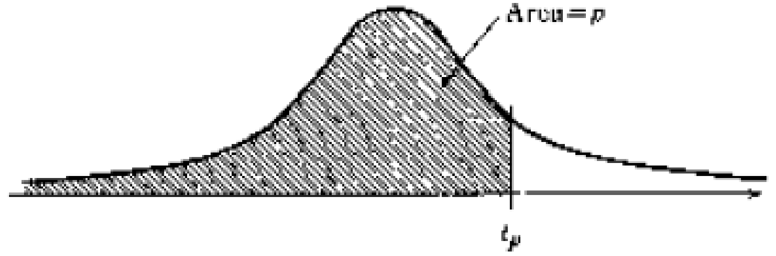


TABLE A.4 Quantiles of the *t* Distribution

	p							
n	0.6000	0.7000	0.8000	0.9000	0.9500	0.9750	0.9950	0.9995
1	0.325	0.727	1.377	3.078	6.314	12.706	63.657	636.619
2	0.289	0.617	1.061	1.886	2.920	4.303	9.925	31.599
3	0.277	0.584	0.978	1.638	2.353	3.182	5.841	12.924
4	0.271	0.569	0.941	1.533	2.132	2.776	4.604	8.610
5	0.267	0.559	0.920	1.476	2.015	2.571	4.032	6.869
6	0.265	0.553	0.906	1.440	1.943	2.447	3.707	5.959
7	0.263	0.549	0.896	1.415	1.895	2.365	3.499	5.408
8	0.262	0.546	0.889	1.397	1.860	2.306	3.355	5.041
9	0.261	0.543	0.883	1.383	1.833	2.262	3.250	4.781
10	0.260	0.542	0.879	1.372	1.812	2.228	3.169	4.587

EXEMPLO 13.2

Exemplo 12.5

O **erro de modelagem** (diferença entre valores medidos e valores previstos por um **modelo de distribuição normal**) para 8 predições de um modelo foram os seguintes:

{-0,04; -0,19; 0,14; -0,09; -0,14; 0,19; 0,04 e 0,09}

- Considere os dados de erros do Exemplo 12.5, que possui uma **distribuição normal**.
- Os **oito valores** dos erros são: -0,04; -0,19; 0,14; -0,09; -0,14; 0,19; 0,04 e 0,09.
- média = ?
- desvio padrão = ?
- n = **8**
- **IC** de **90%** **= ?**

EXEMPLO 13.2

- Considere os dados de erros do Exemplo 12.5, que possui uma **distribuição normal**.
- Os oito valores dos erros são: -0,04; -0,19; 0,14; -0,09; -0,14; 0,19; 0,04 e 0,09.
- A **média** destes valores é **0**, o **desvio padrão** da amostra é **0,138** e n = **8**
- O IC de **90%** seria então:

Tabela A4 $\Rightarrow t_{[0,95;7]} = 1,895$

$$0 \pm 1,895 \times 0,138/\sqrt{8} = 0 \pm 0,0926$$

$$= (-0,0926\ ;\ 0,0296)$$

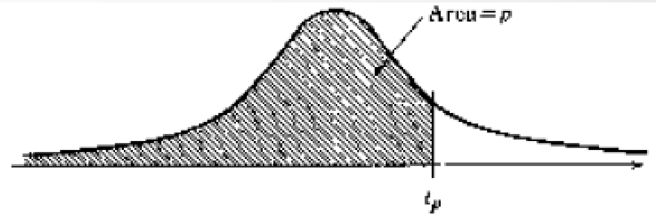
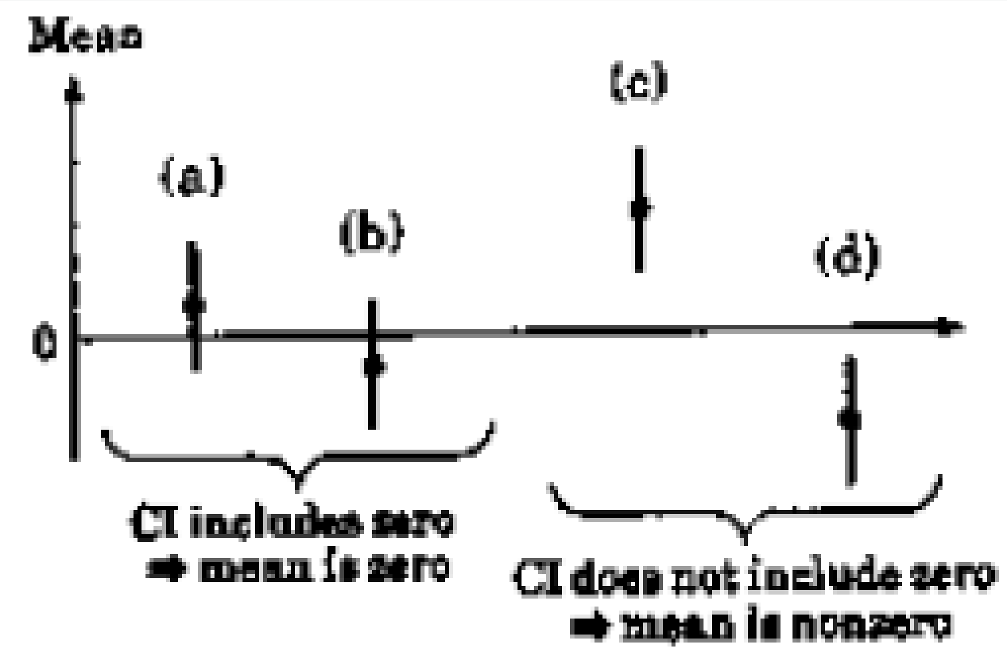


TABLE A.4 Quantiles of the t Distribution

n	0.6000	0.7000	0.8000	0.9000	0.9500	0.9750	0.9950	0.9995
1	0.325	0.727	1.377	3.078	6.314	12.706	63.657	636.619
2	0.289	0.617	1.061	1.886	2.920	4.303	9.925	31.599
3	0.277	0.584	0.978	1.638	2.353	3.182	5.841	12.924
4	0.271	0.569	0.941	1.533	2.132	2.776	4.604	8.610
5	0.267	0.559	0.920	1.476	2.015	2.571	4.032	6.869
6	0.265	0.553	0.906	1.440	1.943	2.447	3.707	5.959
7	0.263	0.549	0.896	1.415	1.895	2.365	3.499	5.408
8	0.262	0.546	0.889	1.397	1.860	2.306	3.355	5.041
9	0.261	0.543	0.883	1.383	1.833	2.262	3.250	4.781
10	0.260	0.542	0.879	1.372	1.812	2.228	3.169	4.587

TESTE DE MÉDIA ZERO

- Um uso comum de ICs é para determinar **se um valor medido é significativamente diferente de zero**, para um certo nível de confiança.
- O **teste** consiste em **obter o IC** e simplesmente **verificar se 0 pertence a este intervalo**.



EXEMPLO 13.3

- A diferença de tempo de execução de duas implementações do mesmo algoritmo foi medida em sete cargas de trabalho semelhantes.
- As diferenças obtidas foram: {1,5; 2,6; -1,8; 1,3; -0,5; 1,7; 2,4}.
- Podemos dizer com **99% de confiança** que **uma implementação é superior à outra?**

- Tamanho da amostra ***n* = 7**
- **Média** = 7,20/7 = 1,03
- **Variância** da amostra =(0,22+2,46+8,00+0,07+2,34+0,45+1,88)/6 = 2,57
- **Desvio padrão** da amostra = 1,60
- **IC** = 1,03 $\pm$ *t* x 1,60/$\sqrt{7}$ = 1,03 $\pm$ 0,605 *t*
- 100(1-α)=99 α=0,01 **1-α/2=0,995**
- Da Tabela A.4 $\Rightarrow$ $t_{[0,995;6]}$ **=3,707**
- **IC de 99%** = (-1,21 ; 3,27)
- Inclui 0, portanto não podemos dizer com 99% de confiança que a diferença média seja significativamente diferente de 0.

COMPARAÇÃO ENTRE DUAS ALTERNATIVAS

- Comparação de dois sistemas submetidos a cargas bem semelhantes.
- Consideraremos dois casos:
  - **Observações casadas**: correspondência unívoca entre os testes efetuados em cada sistema.
    - Efetua-se a diferença e faz-se o **teste para média zero**.
  - **Observações não casadas:** não há correspondência entre as observações das duas amostras.
    - **Teste *t***

EXEMPLO 13.5

Observações casadas

- Foram usadas **6 cargas de trabalho** semelhantes em **dois sistemas**. Os resultados obtidos foram {(5,4;19,1),(16,6;3,5),(0,6;3,4),(1,4;2,5), (0,6;3,6),(7,3;1,7)}.
- **Algum** destes sistemas **é melhor do que o outro**?
  - As diferenças de desempenho correspondem à nossa amostra:

  {-13,7; 13,1; -2,8; -1,1; -3,0; 5,6}

- **Média** da amostra = -0,32
- **Variância** da amostra = 81,62
- **Desvio padrão** da amostra = 9,03
- **Intervalo de confiança** para a média =

  $-0{,}32 \pm t\sqrt{(81{,}62/6)} = -0{,}32 \pm t\,(3{,}69)$

- O **quantil 0,95** de uma variável *t* com **5 graus de liberdade** é 2,015
- **IC de 90%**

  $-0{,}32 \pm (2{,}015)(3{,}69) = (-7{,}75\ ;\ 7{,}11)$

- O **IC inclui zero** => Os dois sistemas não são diferentes!

TESTE *t*

- Calcule a **médias** das amostras:

$$\bar{x}_a = \frac{1}{n_a}\sum_{i=1}^{n_a} x_{ia} \qquad \bar{x}_b = \frac{1}{n_b}\sum_{i=1}^{n_b} x_{ib}$$

- Calcule os **desvios padrão** das amostras:

$$s_a = \left\{ \frac{\left(\sum_{i=1}^{n_a} x_{ia}^2\right) - n_a \bar{x}_a^2}{n_a - 1} \right\}^{1/2}$$

$$s_b = \left\{ \frac{\left(\sum_{i=1}^{n_b} x_{ib}^2\right) - n_b \bar{x}_b^2}{n_b - 1} \right\}^{1/2}$$

- Calcule a **diferença média**: $\bar{x}_a$-$\bar{x}_b$
- Calcule o **desvio padrão da diferença** média:

$$s = \sqrt{\frac{s_a^2}{n_a} + \frac{s_b^2}{n_b}}$$

- Calcule o **número efetivo de graus de liberdade**:

$$\upsilon = \frac{\left(s_a^2 / n_a + s_b^2 / n_b\right)^2}{\frac{1}{n_a+1}\left(\frac{s_a^2}{n_a}\right)^2 + \frac{1}{n_b+1}\left(\frac{s_b^2}{n_b}\right)^2} - 2$$

TESTE *t*

- Calcule o **I.C. para a diferença média**:

$$(\bar{x}_a - \bar{x}_b) \pm t_{[1-\alpha/2;\nu]}\, s$$

- Se o **I.C. incluir zero**, a diferença não é significativa com um nível de confiança de 100(1-a)%.

- Se o **I.C. não inclui zero**, então o **sinal da diferença média** indica que sistema é o melhor.

EXEMPLO 13.6

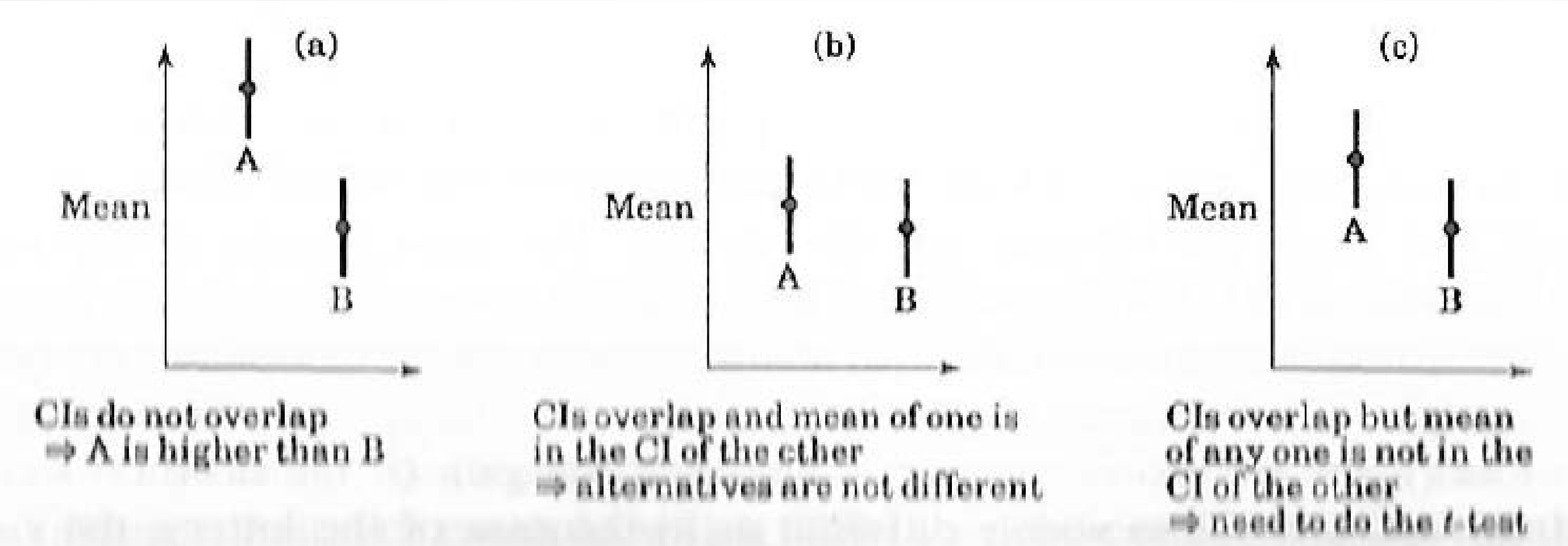
- O **tempo de processador** consumido na execução de uma determinada tarefa foi medido em dois sistemas.
- Os tempos no **sistema A** foram {5,36; 16,57; 0,62; 1,41; 0,64; 7,26}
- Os tempos no **sistema B** foram {19,12; 3,52; 3,38; 2,50; 3,60; 1,74}
- **Os dois sistemas são significativamente diferentes?**

Sistema A	Sistema B
Média = 5,31	Média = 5,64
Variância = 37,92	Variância = 44,11
$n_a = 6$	$n_b = 6$

- **Diferença média** = -0,33
- **Desvio padrão** da diferença média = 3,698
- **Número efetivo de graus** de liberdade = 11,921
- **Quantil 0,95** de uma variável t com **12 graus de liberdade** = 1,71
- **IC de 90%** para a diferença = (-6,92 ; 6,26)
- O IC inclui zero, portanto, **neste nível de confiança** os dois **sistemas não são diferentes!**

TESTE VISUAL (aproximado)

- Calcula-se os **I.C.s** para cada um dos **sistemas separadamente**.
- **Compara-se** os valores obtidos:



TESTE VISUAL (aproximado)

- Para os dados do **Exemplo 13.6**,
- Valor *t* com cinco graus de liberdade e 90% de confiança = 2,015
- Intervalo de confiança de 90% da média de A =

  $5{,}31 \pm (2{,}015)\sqrt{(37{,}92/6)} = (0{,}24\ ;\ 10{,}38)$
- Intervalo de confiança de 90% da média de B =

  $5{,}64 \pm (2{,}015)\sqrt{(44{,}11/6)} = (0{,}18\ ;\ 11{,}10)$
- Os **ICs se sobrepõem** e a **média de um está incluída no IC do outro**.
- Portanto, **os dois sistemas não são diferentes neste nível de confiança!**

TESTE DE HIPÓTESES *vs.* ICs

- Um **teste de hipóteses** normalmente produz uma **resposta booleana**.
- O uso de **IC** não dá apenas este resultado mas **dá informações sobre a possível faixa de valores para o parâmetro**.
  - Um **IC estreito** indica que o parâmetro foi estimado com um **alto grau de precisão**.
  - Os ICs dizem-nos não apenas o que dizer mas também o quão alto devemos falar.
    - Exemplo: (-100,100) e (-1,1)
      - **Ambos os intervalos incluem 0** !

ICs UNILATERAIS

- Até agora tratamos de ICs bilaterais, mas algumas vezes desejamos comparar apenas unilateralmente.
- Por exemplo, pode-se querer testar a **hipótese de que a média seja maior do que um dado valor**.
- Neste caso, o IC seria dado por:

$$\left(\bar{x} - t_{[1-\alpha;n-1]} \frac{s}{\sqrt{n}}, \bar{x} \right)$$

- Analogamente, o **IC superior para a média da população** seria dado por:

$$\left(\bar{x}, \bar{x} + t_{[1-\alpha;n-1]} \frac{s}{\sqrt{n}} \right)$$

EXEMPLO 13.8

- Foram medidos os intervalos entre falhas de dois sistemas A e B.
- A média e o desvio padrão destes intervalos são:

Sistema	Número	Média	Desvio Padrão
A	972	124,10	198,20
B	153	141,47	226,11

- O **sistema A é mais susceptível a falhas do que o sistema B ?**

- A diferença **média** é -17,37
- O **desvio padrão** da diferença é 19,35
- O **número efetivo de graus de liberdade** é 191,05
- Dado que o **número de graus de liberdade** é **maior do que 30**, usaremos os **quantis da normal unitária**.
- Como o IC é unilateral usaremos $z_{0,90}$ = 1,28 para calculá-lo com confiança de **90%**:

(-17,37 ; -17,37+1,28x19,35) = (-17,37; 7,40)

- Como o **IC inclui zero**, rejeitamos a hipótese de que o sistema A seja mais susceptível a falhas que o sistema B.

ICs PARA PROPORÇÕES

- **Proporções** são **probabilidades associadas a categorias**.
- ICs para proporções podem ser calculados da seguinte forma:

Proporção da amostra = $p = \frac{n_1}{n}$

IC para proporções = $p \pm z_{1-\alpha/2}\sqrt{\frac{p(1-p)}{n}}$

- Expressões válidas apenas para ***np* maior ou igual a 10.**

EXEMPLO 13.9

Se 10 entre 1000 páginas impressas numa laser são ilegíveis, então a proporção de páginas ilegíveis é caracterizada da seguinte forma:

Proporção da amostra = $p = \frac{10}{1000} = 0{,}01$

IC para proporções = $0{,}01 \pm z_{1-\alpha/2}\sqrt{\frac{0{,}01(0{,}99)}{1000}}$
$= 0{,}01 \pm 0{,}003z$

- IC de 90%=(0,005 ; 0,015)
- IC de 95%=(0,004 ; 0,016)

Com **90% de confiança** podemos afirmar que **0,5 a 1,5%** das páginas impressas são **ilegíveis**.

DETERMINAÇÃO DO TAMANHO DA AMOSTRA

- Quanto **maior** for a **amostra**, **maior** será a **confiança** associada.
- No entanto, para se obter amostras maiores é preciso mais esforço e recursos.
- O **objetivo** do analista é encontrar o **menor tamanho de amostra que forneça a precisão e confiança desejadas**.
- São feitas **medições iniciais** para **estimar a variância** que será utilizada na determinação do comprimento da amostra necessário.

- **Tamanho da amostra para médias**
  - Suponha que desejamos estimar o **desempenho médio** de um sistema com uma **precisão de r %** e um **nível de confiança de 100(1-α)%**.
  - O **número de observações *n*** necessárias para atingir este objetivo pode ser determinado da seguinte forma:

$$\bar{x} \mp z\frac{s}{\sqrt{n}} = \bar{x}\left(1 \mp \frac{r}{100}\right)$$

$$z\frac{s}{\sqrt{n}} = \bar{x}\frac{r}{100}$$

$$n = \left(\frac{100zs}{r\bar{x}}\right)^2$$

EXEMPLO 13.11

- Baseado em um teste preliminar, a **média de amostras de tempo de resposta** é 20 segundos e o **desvio padrão** é 5.
- **Quantas repetições** são necessárias para conseguir um tempo de resposta com **precisão de 1 segundo** com **95% de confiança?**

Precisão requerida $= p = \frac{1}{20} = 0{,}05 = 5\%$

$\bar{x} = 20$, $s = 5$, $z = 1{,}960$ (@Tab. A2) e $r = 5$

$$n = \left(\frac{(100)(1{,}960)(5)}{(5)(20)}\right)^2 = (9{,}8)^2 = 96{,}04$$

Um total de 97 observações é necessário!

TAMANHO DA AMOSTRA PARA PROPORÇÕES

- **IC para proporções**

$$= p \pm z_{1-\alpha/2}\sqrt{\frac{p(1-p)}{n}}$$

- Para **precisão** de $p \pm r$ tem-se:

$$p \mp r = p \mp z\sqrt{\frac{p(1-p)}{n}}$$

$$r = z\sqrt{\frac{p(1-p)}{n}}$$

$$n = z^2\frac{p(1-p)}{r^2}$$

EXEMPLO 13.12

Uma medida preliminar com uma impressora a laser mostra **1 entre 10.000 páginas** impressas **ilegíveis**.

- Quantas páginas devem ser observadas para conseguir uma **precisão** de **1 por milhão** com **95% de confiança**?

$p = 10^{-4}$, $r = 10^{-6}$, z (@Tab. A2) = 1,960

$$n = (1{,}960)^2\,\frac{10^{-4}(1-10^{-4})}{(10^{-6})^2} = 384.160.000$$

Portanto mais de 384 milhões de páginas devem ser observadas!

TAMANHO DA AMOSTRA PARA COMPARAÇÃO DE DUAS ALTERNATIVAS

- O requisito de **não-sobreposição de ICs** permite calcular o tamanho de amostra requerido para comparar duas alternativas.
- **EXEMPLO 13.13**: Dois algoritmos de encaminhamento de pacotes foram observados. Medidas preliminares mostram que o **algoritmo A** apresenta uma **perda de pacotes de 0,5%** e o **algoritmo B** perda de **0,6%**. Quantos pacotes precisam ser observados para afirmar com **95% de confiança** que o algoritmo A é melhor do que o algoritmo B?
- **IC do do algoritmo A** = $0{,}005 \pm 1{,}960\left(\frac{0{,}005(1-0{,}005)}{n}\right)^{1/2}$
- **IC do do algoritmo B** = $0{,}006 \pm 1{,}960\left(\frac{0{,}006(1-0{,}006)}{n}\right)^{1/2}$
- Para não sobreposição dos ICs , o limite superior do menor IC deve estar abaixo do limite inferior do maior IC

$$0{,}005 \pm 1{,}960\left(\frac{0{,}005(1-0{,}005)}{n}\right)^{1/2} \leq 0{,}006 \pm 1{,}960\left(\frac{0{,}006(1-0{,}006)}{n}\right)^{1/2}$$

$n \geq 84.340$

Box 13.1 Confidence Intervals

1. Given: A sample $\{x_1, x_2, \ldots, x_n\}$ of n observations.
   $\bar{x}$ = sample mean; s = sample standard deviation

   (a) Standard error of the sample mean: $\sigma_{\bar{x}} = \frac{s}{\sqrt{n}}$
   (b) $100(1-\alpha)\%$ two-sided confidence interval for the mean:
   $\bar{x} \mp z_{1-\alpha/2} s/\sqrt{n}$
   If $n \leq 30^{\dagger}$: $\bar{x} \mp t_{[1-\alpha/2;n-1]} s/\sqrt{n}$
   (c) $100(1-\alpha)\%$ one-sided confidence interval for the mean:
   $(\bar{x}, \bar{x} + z_{1-\alpha} s/\sqrt{n})$ or $(\bar{x} - z_{1-\alpha} s/\sqrt{n}, \bar{x})$
   If $n \leq 30^{\dagger}$: $(\bar{x}, \bar{x} + t_{[1-\alpha;n-1]} s/\sqrt{n})$ or $(\bar{x} - t_{[1-\alpha;n-1]} s/\sqrt{n}, \bar{x})$

2. To compare two systems using unpaired observations:

   (a) The standard error of the mean difference: $s = \sqrt{\frac{s_a^2}{n_a} + \frac{s_b^2}{n_b}}$
   (b) The effective number of degrees of freedom:

   $$\nu = \frac{(s_a^2/n_a + s_b^2/n_b)^2}{\frac{1}{n_a+1}\left(\frac{s_a^2}{n_a}\right)^2 + \frac{1}{n_b+1}\left(\frac{s_b^2}{n_b}\right)^2} - 2$$

   (c) The confidence interval for the mean difference:
   $(\bar{x}_a - \bar{x}_b) \mp t_{[1-\alpha/2;\nu]} s$

3. If n_1 of the n observations belong to a certain class, the following statistics can be reported for the class:

   (a) Proportion of the observations in the class: $p = \frac{n_1}{n}$
   (b) $100(1-\alpha)\%$ two-sided confidence interval for the proportion‡:
   $p \mp z_{1-\alpha/2}\sqrt{\frac{p(1-p)}{n}}$
   (c) $100(1-\alpha)\%$ one-sided confidence interval for the proportion‡:

   $$\left(p, p + z_{1-\alpha}\sqrt{\frac{p(1-p)}{n}}\right) \quad \text{or} \quad \left(p - z_{1-\alpha}\sqrt{\frac{p(1-p)}{n}}, p\right)$$

† Only for samples from normal populations.
‡ Provided $np \geq 10$.

SIMULAÇÃO: QUESTÕES FUNDAMENTAIS

- Quais são os **erros mais comuns** em simulação e **por que muitas simulações falham**?
- Que **linguagem** deveria ser usada para desenvolver um **modelo de simulação**?
- Quais são os diversos **tipos de simulação**?
- Como programar **(agendar/*escalonar*) eventos** numa simulação?
- Como **verificar** e **validar** um **modelo**?
- Como determinar se a simulação atingiu um **estado permanente**?

- Por **quanto tempo** devemos executar uma simulação?
- Como **gerar números aleatórios** uniformemente distribuídos?
- Como **verificar** se um determinado **gerador de números aleatórios** é bom?
- Como selecionar **sementes** para um **gerador de números aleatórios**?
- Como gerar **variáveis aleatórias** de acordo com uma **determinada distribuição**?
- Que **distribuições** devem ser **utilizadas** e quando?

SIMULAÇÃO

- Técnica **útil** para a análise de **desempenho de sistemas computacionais**.
- Especialmente:
  - se o **sistema não** estiver **disponível**,
  - para **prever o desempenho** de diversas alternativas,
  - facilidade de **efetuar comparações** para uma maior **variedade de cargas e de ambientes**.

ERROS COMUNS: Nível Inadequado de Detalhes

- Nem sempre um modelo mais detalhado é um melhor modelo.
- **Problemas** com **demasiados detalhes**:
  - necessita de **mais tempo** para ser desenvolvido
  - **aumenta a probabilidade de erros** e o tempo gasto em identificá-los
  - **maior tempo de execução**
  - necessita de um **conhecimento mais detalhado dos parâmetros de entrada**, que se não estiverem disponíveis, tornam o **modelo impreciso**

- É melhor **começar com um modelo menos detalhado**, obter alguns resultados, **estudar as sensibilidades** e **introduzir maiores detalhes** nas áreas que causem maior impacto nos resultados.

ERROS COMUNS (II)

USO DE LINGUAGEM INADEQUADA

- **linguagens de simulação** necessitam de **menos tempo para o desenvolvimento do modelo** e facilitam a **verificação e análise estatística**.
- **linguagens de propósito geral** são mais portáteis e possibilitam um **maior controle sobre a eficiência e tempo de execução** da simulação.

FALHA NA VERIFICAÇÃO OU VALIDAÇÃO

- **Modelos não verificados:** erros de programação podem tornar os **resultados sem sentido**.
- **Modelos inválidos:** o modelo pode não representar corretamente o sistema real por causa de **hipóteses incorretas**. É preciso que o modelo seja validado.

ERROS COMUNS (III)

CONDIÇÕES INICIAIS E TEMPO DE SIMULAÇÃO

- **Tratamento inadequado das condições iniciais:** a parte inicial de uma simulação geralmente não é representativa do comportamento do sistema em **regime permanente**.
- **Simulações demasiado curtas:** os **resultados** são muito **dependentes das condições iniciais** e podem não ser representativos do sistema real.

GERADOR DE NÚMEROS ALEATÓRIOS RUINS

- **Geradores de números aleatórios ruins:** é mais seguro **utilizar geradores conhecidos** que já foram vastamente analisados.
- **Seleção inadequada das sementes*:** o uso inadequado de sementes dos geradores de números aleatórios pode levar, por exemplo, a **correlações indesejadas**.

*Semente = primeiro número aleatório na sequência

OUTRAS CAUSAS DE ERROS

- Estimativa inadequada do **tempo necessário**
- Falta de **objetivo claro**
- Mistura incompleta de **talentos essenciais** na equipe
  - Liderança de projeto
  - Modelagem e Estatística
  - Programação
  - Conhecimento do sistema que está sendo modelado

- Nível inadequado de **participação do usuário**
  - Identificação precoce de *bugs* no modelo; modelo em sintonia com o sistema
- **Documentação** Obsoleta ou Inexistente
  - Modelos e sistema em contínua evolução
- Incapacidade de **Gerenciar o Desenvolvimento** de um Programa de Computador Grande e Complexo
  - Ferramentas de engenharia de software para gerenciar o desenvolvimento de modelos de simulação complexos
- **Resultados Misteriosos**
  - *Bugs* no programa, hipóteses inválidas ou falta de conhecimento do sistema real

LISTA DE VERIFICAÇÕES

- **Antes de desenvolver** uma simulação
- **Durante o desenvolvimento** da simulação
- **Após** a simulação **estar sendo executada**

LISTA DE VERIFICAÇÕES

- Verificações **antes de desenvolver uma simulação**:
  - O **objetivo da simulação** está claramente especificado?
  - O **nível de detalhes do modelo** está adequado para o objetivo?
  - O **time de simulação** inclui pessoas com experiência de liderança de projetos, modelagem, programação e sistemas computacionais?
  - Foi alocado **tempo suficiente** para este projeto?

LISTA DE VERIFICAÇÕES

- Verificações **durante o desenvolvimento**:
  - O **gerador de números aleatórios** que está sendo utilizado na simulação, foi **testado** em relação à **uniformidade e independência**?
  - O **modelo** está sendo **revisado periodicamente** com o usuário final?
  - O **modelo** está **documentado**?

LISTA DE VERIFICAÇÕES

- Verificações depois que a **simulação** está **sendo executada**:
  - A **duração da simulação** está apropriada?
  - Foram removidos os **transientes iniciais**?
  - O **modelo** foi completamente **verificado**?
  - O **modelo** foi **validado** antes da utilização de seus resultados?
  - Se há **resultados inesperados**, eles foram **validados**?
  - As **sementes** foram escolhidas de modo que as seqüências de **valores aleatórios não se sobreponham**?

TERMINOLOGIA

(Exemplo: Simulação de escalonamento de CPU)

- **Variáveis de estado:** definem o **estado do sistema**.
  - Se uma simulação for parada no meio, pode ser reiniciada se e só se os valores de todas as variáveis de estado forem conhecidos
  - Exemplo: comprimento da fila de pacotes.
- **Evento: mudança no estado do sistema**.
  - Exemplos:
    - **chegada** de um *job*
    - início de uma **nova execução**
    - **partida** do *job*

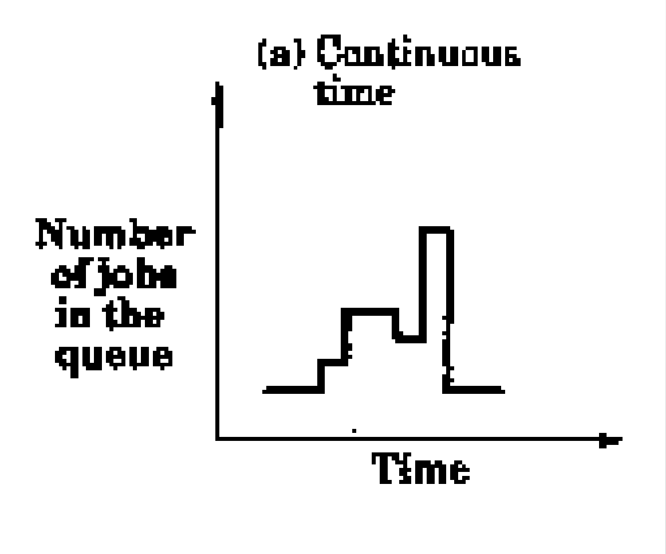
MODELOS DE TEMPO

- **MODELO DE TEMPO CONTÍNUO:**

o **estado** do sistema está definido em **todos os instantes**.

Exemplo:

Modelo de escalonamento de CPU.

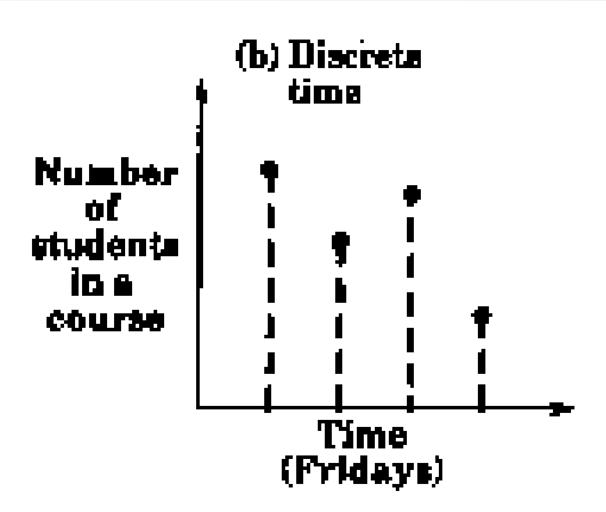


- **MODELO DE TEMPO DISCRETO:**

o **estado** do sistema está definido apenas em **instantes particulares**.

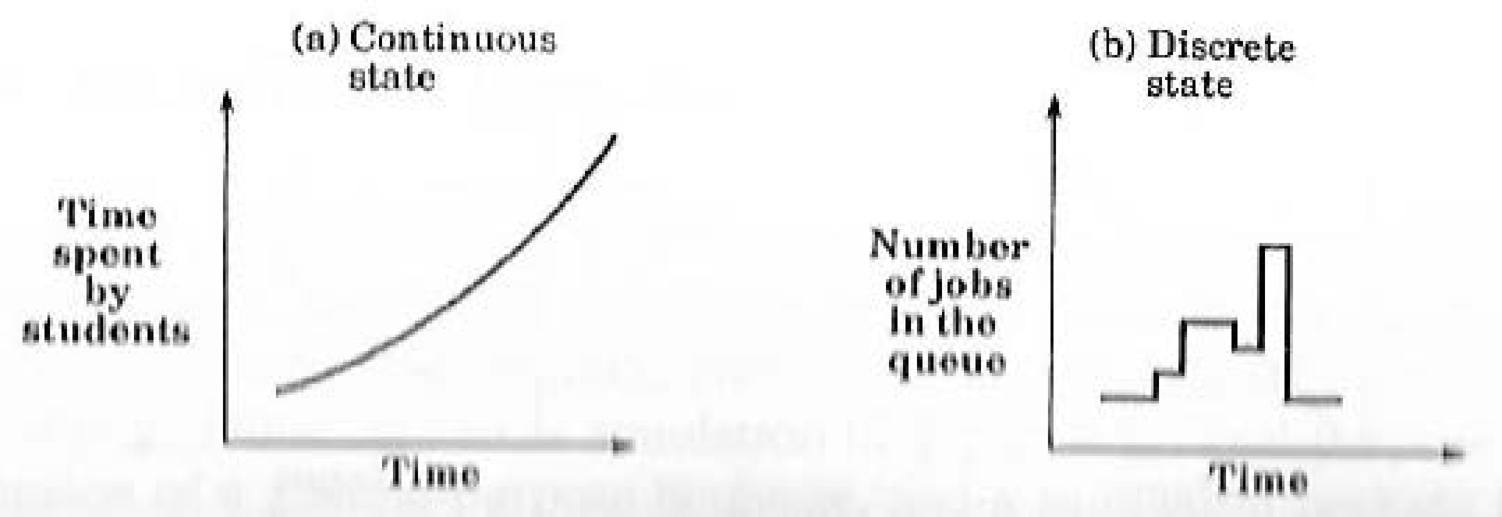
Exemplo:

Número de estudantes que assistem ADRS às sextas.



MODELOS DE ESTADO

- **Modelos de ESTADO CONTÍNUO e de ESTADO DISCRETO:** dependendo das **variáveis de estado** serem **contínuas** ou **discretas**.
  - Exemplo: Tempo gasto estudando uma determinada matéria vs Número de estudantes.

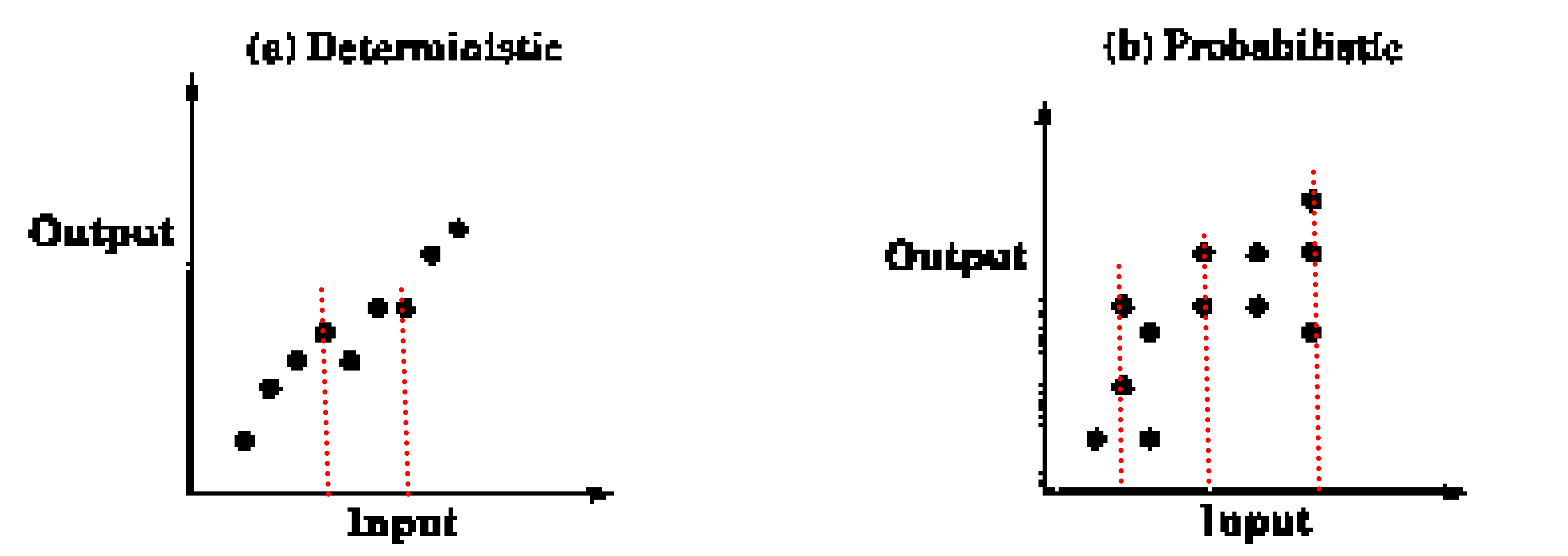


COMBINAÇÕES DE MODELOS

- Modelo de **ESTADO DISCRETO** = Modelo de **EVENTOS DISCRETOS**
- Modelo de **ESTADO CONTÍNUO** = Modelo de **EVENTOS CONTÍNUOS**
- Continuidade de tempo ≠ Continuidade de estado
- **Combinações possíveis de modelos**
  - estado discreto/tempo discreto
  - estado discreto/tempo contínuo
  - estado contínuo/tempo discreto
  - estado contínuo/tempo contínuo

MODELOS DETERMINÍSTICOS E PROBABILÍSTICOS

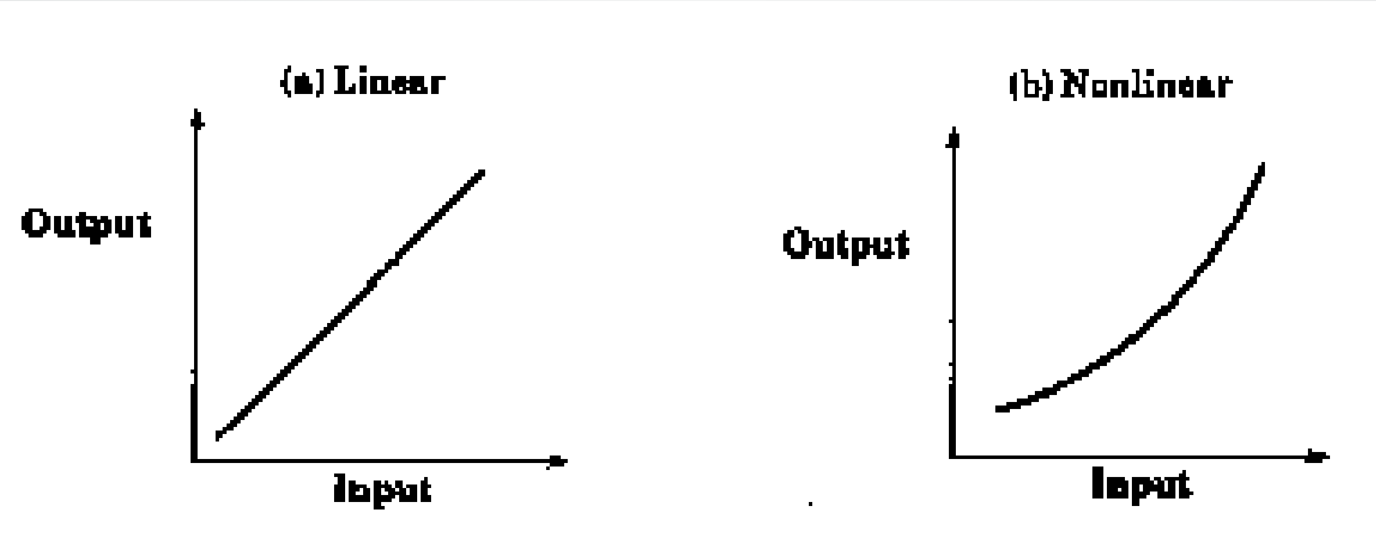
- Modelos **DETERMINÍSTICOS** e **PROBABILÍSTICOS:** nos modelos determinísticos, os resultados podem ser previstos com certeza.



TERMINOLOGIA

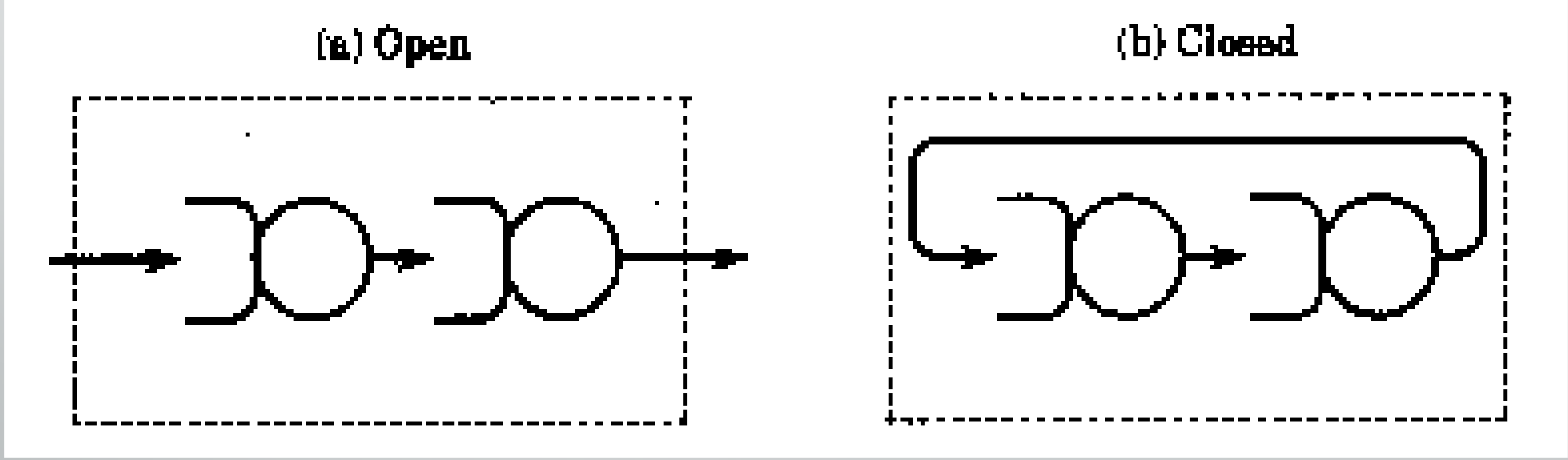
- **Modelos ESTÁTICOS e DINÂMICOS:** modelos estáticos são aqueles nos quais o **tempo** não é uma variável.
  - Exemplo: $E = mc^2$ vs Modelo de Escalonamento de CPU
- **Modelos LINEARES e NÃO-LINEARES:**

Saída = *f* (Entrada)



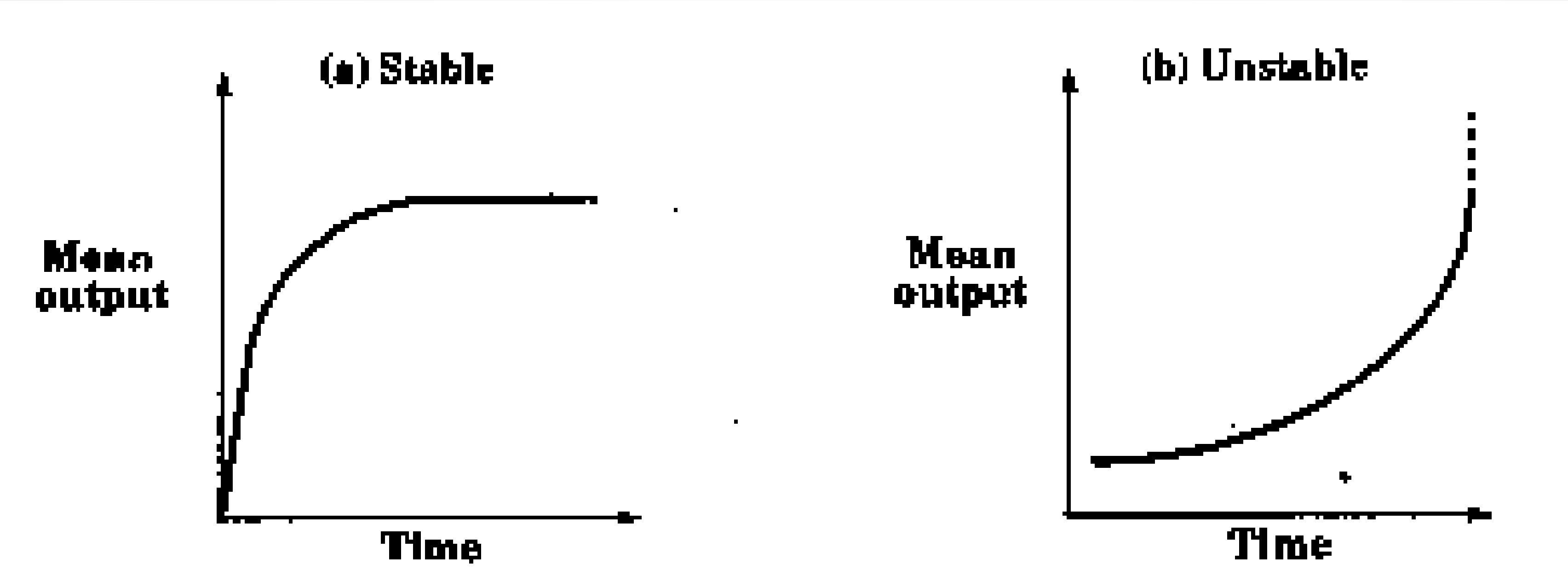
TERMINOLOGIA

- **Modelos** **ABERTOS** **e** **FECHADOS:** nos modelos abertos a entrada é externa ao modelo e independente do mesmo.



TERMINOLOGIA

- **Modelos ESTÁVEIS e INSTÁVEIS:**
  - Estável ⇒ atinge **estado permanente**
  - Instável ⇒ **muda continuamente** de comportamento.



MODELOS DE SISTEMAS COMPUTACIONAIS

- **Tempo contínuo**
- **Estados discretos**
- **Probabilístico**
- **Dinâmico**
- **Não-linear**
- **Aberto ou fechado**
- **Estável ou instável**

SELEÇÃO DE UMA LINGUAGEM PARA SIMULAÇÃO

- Linguagem de simulação
- Linguagem de propósito geral
- Extensão de uma linguagem de propósito geral
- Pacote de simulação

LINGUAGENS DE SIMULAÇÃO

- Economizam **tempo de desenvolvimento**
  - Exemplos: SIMULA e SIMSCRIPT
- **Recursos embutidos** para:
  - avançar o tempo
  - programar eventos
  - manipulação de entidades
  - geração de valores aleatórios
  - coleta de dados estatísticos
  - geração de relatórios
- Sobra mais tempo para questões específicas do sistema
- **Código modular**, bastante legível

LINGUAGEM DE PROPÓSITO GERAL

- **Familiaridade** do analista
- Grande **disponibilidade**
- **Início imediato**
- **Tempo gasto** com o **desenvolvimento de rotinas para tratamento de eventos e geração de valores aleatórios**
- Outras questões: **Eficiência**, **Flexibilidade**, **Portabilidade**

Recomendação: Aprenda pelo menos uma linguagem de simulação.

EXTENSÃO DE UMA LINGUAGEM DE PROPÓSITO GERAL

- Exemplos: GASP (para FORTRAN), SMPL (para C), SimPy (Python), SystemC (C++)
  - Coleção de **rotinas** para tratar **tarefas de simulação**
  - **Compromisso** entre eficiência, flexibilidade e portabilidade.

PACOTES DE SIMULAÇÃO

- Exemplos: *OPNET Modeler* (comercial), ns-2, **ns-3**...
  - Diálogo de **entrada**
  - **Biblioteca** de estruturas de dados, rotinas e algoritmos
  - Grande **economia de tempo**
  - Inflexível $\Rightarrow$ **força** o analista fazer **simplificações**

TIPOS DE LINGUAGENS E DE SIMULAÇÕES

- **Linguagens de simulação contínua**:
  - Exemplos: CSMP *(Continuous System Modeling Program*), DYNAMO (*open source*)
    - CSMP $\Rightarrow$ solução numérica de Equações Diferenciais
    - Usadas em engenharia química (i.e., processos industriais contínuos)
- **Linguagens de simulação** de **eventos discretos**:
  - **ns-3** (*open source*), SimPy (*open source*), SIMULA (*open source*) e GPSS (comercial)
  - ver também:

  https://pt.wikipedia.org/wiki/Simula%C3%A7%C3%A3o_de_eventos_discretos

  https://en.wikipedia.org/wiki/List_of_discrete_event_simulation_software
- **Combinadas**:
  - SIMSCRIPT (www.simscript.com) e GASP (*Gaussian Stochastic Process Emulation*).
  - Permitem simulações discretas, contínuas ou combinadas.

TIPOS DE SIMULAÇÕES

- **Emulação**: utilizando *hardware* ou *firmware*
  - Exemplos: emulador de terminal, emulador de processador
  - Envolve basicamente questões de projeto de hardware
- **Simulação de Monte Carlo**
- **Simulação Dirigida por Traces**
- **Simulação de Eventos Discretos**

SIMULAÇÃO DE MONTE CARLO

- Origem: homenagem ao Conde Montgomery de Carlo, jogador e gerador de números aleatórios italiano (1792-1838).
- Características principais
  - Simulação **estática** (sem eixo do tempo)
  - Para **modelar fenômenos probabilísticos**
  - Necessita de **números pseudo-aleatórios**
  - Usado para **avaliar expressões não probabilísticas** usando métodos probabilísticos.
    - Exemplo: **cálculo de integral**

EXEMPLO DE SIMULAÇÃO DE MONTE CARLO

- Deseja-se obter o **resultado da seguinte integral**:

$$I = \int_0^2 e^{-x^2} dx$$

$$x \sim \text{Uniforme}(0,2)$$

$$\text{Função densidade } f(x) = \frac{1}{2} \quad \text{iff } 0 \le x \le 2$$

$$y = 2e^{-x^2}$$

- O **valor esperado de y** é:

$$E(y) = \int_0^2 2e^{-x^2} f(x) dx$$

$$= \int_0^2 2e^{-x^2} \frac{1}{2} dx$$

$$= \int_0^2 e^{-x^2} dx$$

$$= I$$

- Portanto, **I pode ser calculada** através da **geração de números aleatórios** x_i, **cálculo de** y_i, e **tomando a média** como a seguir:

$$x_i \sim \text{Uniforme}(0,2)$$

$$y_i = 2e^{-x_i^2}$$

$$I = E(y) = \frac{1}{n} \sum_{i=1}^{n} y_i$$

SIMULAÇÃO DIRIGIDA POR *TRACES*

- **Simulação dirigida por *Traces*** = utiliza os ***traces* como entrada**
  - *Trace* = registro de eventos de um sistema ordenado de acordo com o tempo
- Utilizada para analisar ou ajustar algoritmos de gerenciamento de recursos:
  - paginação, caches, escalonamento de CPU, prevenção de *deadlock*, alocação dinâmica de memória.

VANTAGENS DAS SIMULAÇÕES DIRIGIDAS POR *TRACES*

- **Credibilidade**: fácil de convencer os outros membros da equipe
- **Validação fácil**: basta comparar o resultado de simulação com o medido
- **Carga de trabalho precisa**: preserva correlações e interferências
- **Compromissos detalhados**: Carga de trabalho detalhada $\Rightarrow$ permite estudar pequenas alterações nos algoritmos
- **Menor aleatoriedade**: *Trace* $\Rightarrow$ entrada determinística $\Rightarrow$ Menor número de repetições para um determinado nível de confiança
- **Comparação Justa**: Melhor do que entrada aleatória
- **Semelhança com a Implementação Real**: o modelo dirigido por *traces* é bastante semelhante ao sistema real $\Rightarrow$ permite compreender a complexidade da implementação de um algoritmo

DESVANTAGENS DAS SIMULAÇÕES DIRIGIDAS POR *TRACES*

- **Complexidade**: simulação mais detalhada
- **Representatividade**: a carga de trabalho varia com o tempo e com o equipamento
- **Finitude**: poucos minutos de *trace* enchem um disco
- **Ponto único de validação**: um t*race* = um ponto
  - Um algoritmo melhor com um *trace* pode não ser o melhor com outro
- **Detalhes**: alto nível de detalhes
  - em geral longas sequências exigindo processamento de cada elemento do *trace*
- **Compromissos**: é difícil alterar a carga de trabalho

SIMULAÇÃO DE EVENTOS DISCRETOS

- Concentração de substâncias químicas $\Rightarrow$ Simulação de eventos contínuos
- **Número de *jobs*** (ou pacotes) $\Rightarrow$ **Eventos discretos**
- Estados discretos $\neq$ tempo discreto

COMPONENTES DA SIMULAÇÃO DE EVENTOS DISCRETOS

- **Escalonador de Eventos**
  - Programa o evento *X* para o instante *T*
  - Congela o evento *X* durante o intervalo de tempo *dt*
  - Cancela um evento *X* previamente programado
  - Congela o evento *X* indefinidamente (até ser programado por outro evento)
  - Programa um evento congelado indefinidamente

- **Relógio de Simulação** e **Mecanismo de Avanço do Tempo**
  - Tempo de simulação = Variável global
  - Avanço do tempo
    - abordagem baseada em **unidades de tempo**
    - abordagem dirigida a **eventos** (mais comum)
- **Variáveis de Estado do Sistema**
  - Global = número de na fila
  - Local = tempo de CPU para um *job*

COMPONENTES DA SIMULAÇÃO DE EVENTOS DISCRETOS

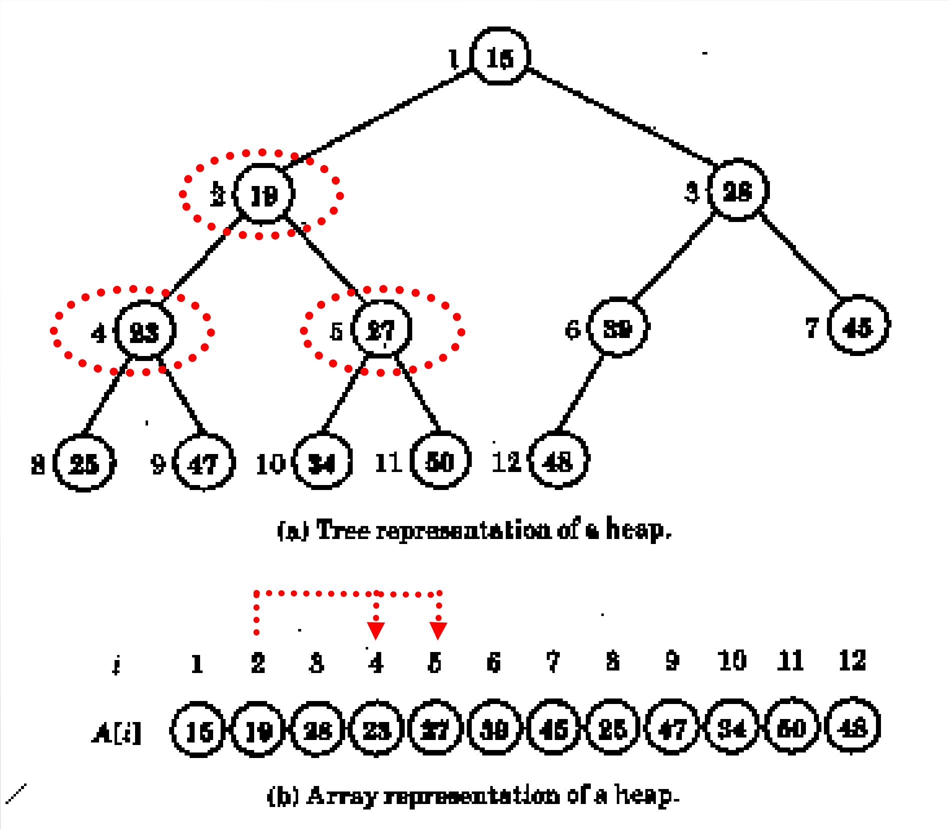
- **Rotinas associadas aos eventos**: Uma para cada evento.
  - Exemplo: **chegada** de *jobs*, **escalonamento** de *jobs* e **partida** de *jobs*
- **Rotinas de Entrada**
  - Obtenção dos parâmetros do modelo com o usuário (e.g., demanda média de CPU por *job*)
  - Variação dos parâmetros dentro de uma certa faixa.
- **Gerador de Relatórios**
  - Cálculo e apresentação do resultado final
- **Rotinas de Inicialização**
  - Atribui o **estado inicial**. Inicializa as **sementes**.
    - Sugere-se diferentes rotinas para o início de uma simulação, início de uma iteração e início de uma repetição
- **Rotinas de *Trace***
  - Podem ser **ligadas** ou **desligadas**
- **Gerenciamento Dinâmico de Memória**
  - Coleta de Lixo
- **Programa Principal**
  - Chama as rotinas de entrada, inicializa a simulação, executa várias iterações, e finalmente, chama as rotinas de saída

ALGORITMOS DE TRATAMENTO DE CONJUNTO DE EVENTOS

- **Conjunto de Eventos** = **Lista ordenada do registro dos eventos futuros**
- **Operações Básicas**:
  - Inserção de um **novo evento**
  - Descoberta e remoção do **próximo evento** a ser executado
- **Estruturas de Dados**:
  - Lista Ligada Ordenada
  - Lista Linear Indexada
  - Estruturas em Árvores
    - ***Heap* (árvore binária)**

HEAP

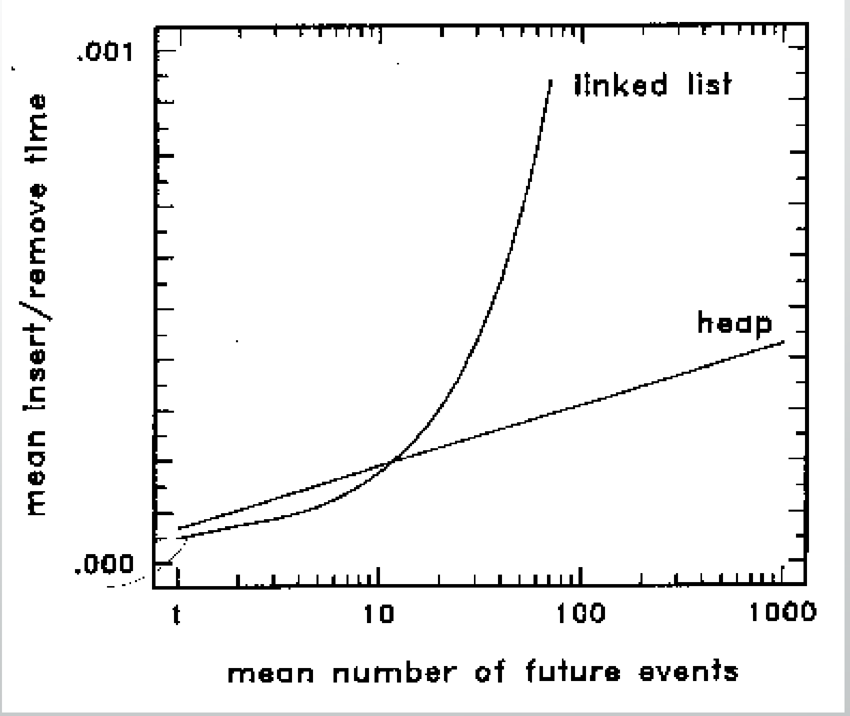
- O **evento** é um **nó de uma árvore binária**
  - Cada nó possui até dois filhos
  - O instante de ocorrência do evento associado a cada nó é menor do que o de qualquer de seus filhos.
- ***Heaps*** podem ser armazenadas como vetores:
  - Os filhos de um nó na posição *i* encontram-se nas posições **2*i*** e **2*i* + 1**.



CRITÉRIOS DE ESCOLHA DA ESTRUTURA MAIS APROPRIADA

- Qual a alternativa **mais eficiente**?
  - **Lista ligada simples**: menos do que **20 eventos**
  - **Listas lineares indexadas**: **entre 20 e 120 eventos**
  - *Heaps* : **mais do que 120 eventos**

- **Tempos Médios de Inserção e Remoção**



ANÁLISE DE RESULTADOS DE SIMULAÇÃO

Capítulo 25

Raj Jain - The Art of Computer Systems Performance Analysis. Wiley, 1991.

ROTEIRO

I. Técnicas de verificação do modelo
II. Técnicas de validação do modelo
III. Remoção de transientes
IV. Simulações terminais
V. Critério de parada: estimativa da variância
VI. Redução da variância

VERIFICAÇÃO vs. VALIDAÇÃO DO MODELO

- **Verificação** => Depuração
- **Validação** => Modelo = Mundo Real
- Quatro possibilidades:
  - Não verificado, Inválido
  - Não verificado, Válido
  - Verificado, Inválido
  - Verificado, Válido

VERIFICAÇÃO BÁSICA

- MEDIDAS *ANTIBUGS*
  - Inclusão de **verificações** tais como:
    - Σ probabilidades = 1
    - *Jobs* remanescentes = gerados - servidos
- VARRIDA ESTRUTURADA
  - **Explique o código** a outra pessoa ou grupo de pessoas.
  - Funciona mesmo se a pessoa estiver dormindo!

I. TÉCNICAS DE VERIFICAÇÃO DO MODELO

- **Modelos Determinísticos**:
  - Use valores constantes
- Execução de **Casos Simplificados**:
  - Apenas um pacote
  - Apenas uma fonte
  - Apenas um nó intermediário

- *Trace*
- Apresentação de Gráficos *On-line*
- Teste de Continuidade
- Outras Técnicas

TRACE

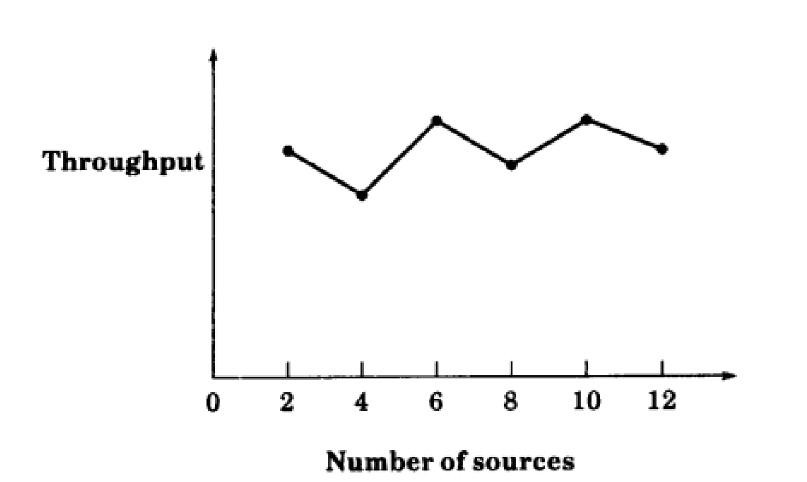
- ***Trace*** = Lista ordenada de eventos e variáveis
- Diversos níveis de detalhes:
  - *Trace* de **eventos**
  - *Trace* de **procedimentos**
  - *Trace* de **variáveis**
- O usuário seleciona o nível de detalhes
- Inclui chave de ligar e desligar

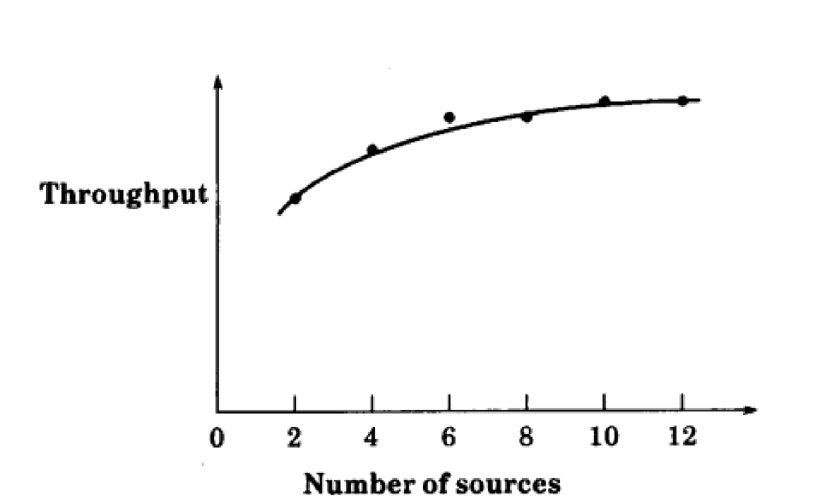
APRESENTAÇÃO DE GRÁFICOS *ON-LINE*

- Torna a **simulação interessante**
- **Ajuda a “vender”** os resultados
- **Mais compreensível** do que o *trace*

TESTE DE CONTINUIDADE

- Execute para **diferentes valores de parâmetros de entrada**
  - Pequena mudança na entrada => pequena mudança na saída
- **Antes**: **Depois**:





OUTRAS TÉCNICAS DE VERIFICAÇÃO

- **Testes degenerativos**:
  - Tente usar configurações e cargas extremas:
    - Uma CPU, nenhum disco, etc.
- **Testes de consistência**:
  - Resultados semelhantes para entradas que têm o mesmo efeito:
    - Quatro usuários a 100Mbps vs. Dois a 200 Mbps
  - Construa uma biblioteca de testes de continuidade, degeneração e consistência
- **Independência das sementes**:
  - Resultados semelhantes para sementes diferentes

II. TÉCNICAS DE VALIDAÇÃO DO MODELO

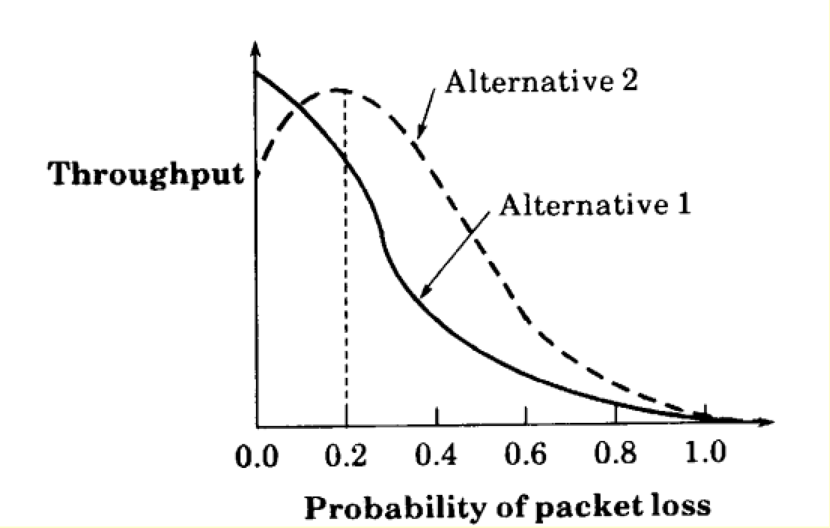
- **Técnicas de validação** para um problema podem não se aplicar a outro problema
- **Aspectos chaves a serem validados**:
  1. **Hipóteses**
  2. **Valores dos parâmetros de entrada e distribuições**
  3. **Valores de saída e conclusões**
- **Fontes de validação**:
  1. **Intuição do especialista**
  2. **Medidas de um sistema real**
  3. **Resultados teóricos**

Total de **possibilidades** de **testes** de validação = **9** (i.e., 3 x 3)

INTUIÇÃO DO ESPECIALISTA

- Forma **mais prática e comum**
- **Especialistas** = envolvidos no projeto, arquitetura, implementação, análise, marketing ou manutenção do sistema
- **Seleção** = função do estágio do ciclo de vida
  - Apresente hipóteses, entradas e saídas
  - É melhor validar um de cada vez
  - Observe se os especialistas podem distinguir os resultados do modelo das medições de sistemas reais.

- Exemplo de **problemas causados por hipóteses inválidas** que são facilmente detectadas por especialistas



MEDIDAS DE UM SISTEMA REAL

- **Compare** as hipóteses, entradas e saídas com o **mundo real**
- Freqüentemente é **inviável ou muito caro**
- Mesmo **uma ou duas medidas contribuem** para a validação

RESULTADOS TEÓRICOS

- **Resultados Analíticos** = Simulação
  - Usados também para validar os resultados analíticos
  - Ambos podem ser inválidos
- Use **teoria em conjunto com a intuição** de um especialista.
  - Ex.: Use a teoria para uma configuração maior
- Um modelo **"completamente validado"** é um **mito**
  - Podemos apenas mostrar que o modelo não é inválido para algumas das situações comparadas

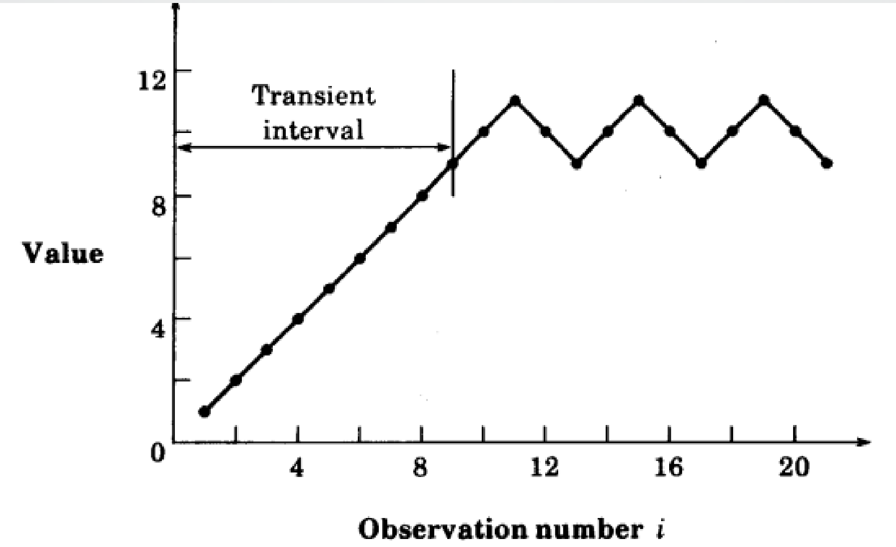
III. REMOÇÃO DE TRANSIENTES

- Geralmente estamos interessados no **desempenho do sistema em regime permanente**
  - Remova a parte inicial
- **Técnicas**: Não há uma definição exata => **Heurísticas**:
  1. **Execuções longas**
     - Desperdiça recursos; **difícil de assegurar** que é **longa o bastante**
  2. **Inicialização apropriada**
     - Inicie num **estado próximo ao estado permanente** esperado, isso **reduz** o comprimento e o efeito do **estado transiente**
  3. **Truncagem**
  4. **Descarte dos dados iniciais**
  5. **Média móvel de replicações independentes**
  6. **Médias de Lotes (*Batches*)**

TRUNCAGEM

- Assume que a variabilidade é menor no estado permanente.
- Plote o máx-min de $n - l$ observações para $l = 1, 2, \ldots$
- Quando a $(l+1)$-ésima observação não for nem um mínimo nem um máximo $\Rightarrow$ encerrou o estado transiente.

Exemplo: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 10, 9, 10, 11, 10, 11, 10, 9, 10, 11, 10, 9,...
Com $l = 9$, Range $= (9, 11)$, próxima observação $= 10$



DESCARTE DOS DADOS INICIAIS

- Descarte algumas **observações iniciais**
- Calcule a **média**
  - **Nenhuma mudança** => **Estado permanente**
- Use **diversas replicações** para **suavizar a média**
  - *m* **replicações** cada uma de tamanho *n*
  - x_{ij} = *j–ésima* observação da *i-ésima* replicação

DESCARTE DOS DADOS INICIAIS

1. Obtenha uma trajetória média, tomando a média entre as replicações:

$$\bar{x}_j = \frac{1}{m}\sum_{i=1}^{m} x_{ij} \text{ para } j = 1, 2, \ldots, n$$

2. Obtenha a média global:

$$\bar{\bar{x}} = \frac{1}{n}\sum_{j=1}^{n} \bar{x}_j$$

Faça $l = 1$ e prossiga para o próximo passo.

3. Descarte as primeiras l observações e obtenha uma média global a partir dos restantes $n - l$ valores:

$$\bar{\bar{x}}_l = \frac{1}{n-l}\sum_{j=l+1}^{n} \bar{x}_j$$

4. Calcule a mudança relativa:

$$\text{Mudança relativa} = \frac{\bar{\bar{x}}_l - \bar{\bar{x}}}{\bar{\bar{x}}}$$

5. Repita os passos 3 e 4 variando l de 1 a $n - 1$
6. Plote a média global e a mudança relativa
7. Valor de l no joelho = comprimento do intervalo transiente.

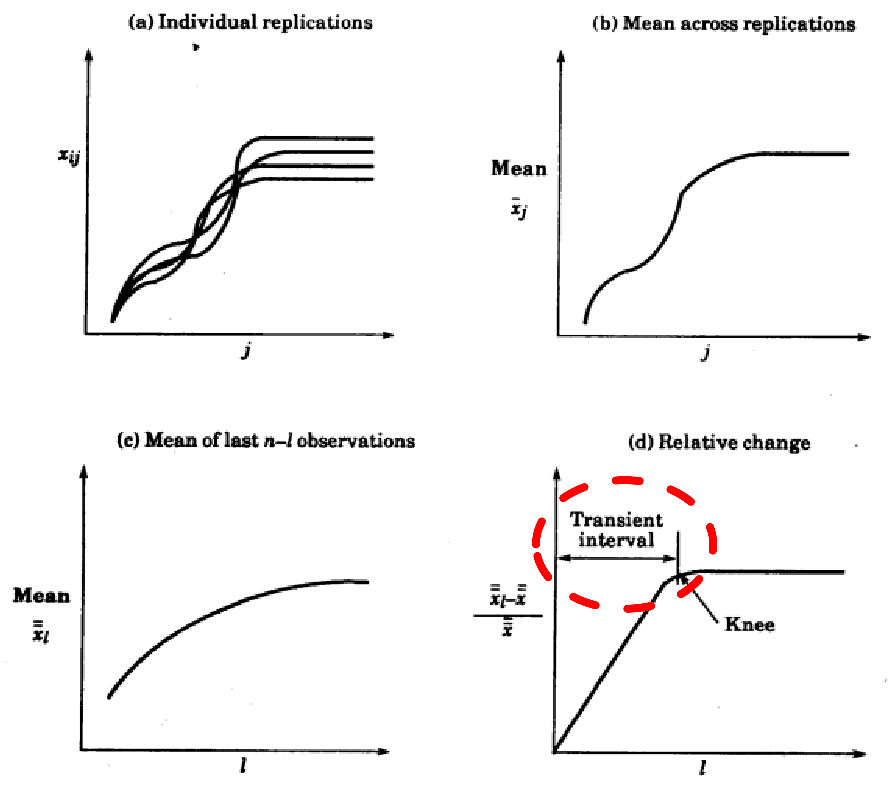


FIGURE 25.8 Initial data deletion method.

m replicações cada uma de tamanho *n*

x_{ij} = *j–ésima* observação da *i-ésima* replicação

MÉDIA MÓVEL DE REPLICAÇÕES INDEPENDENTES

1. Obtenha uma trajetória média tomando a média entre replicações:

$$\bar{x}_j = \frac{1}{m}\sum_{i=1}^{m} x_{ij} \text{ para } j = 1, 2, ..., n$$

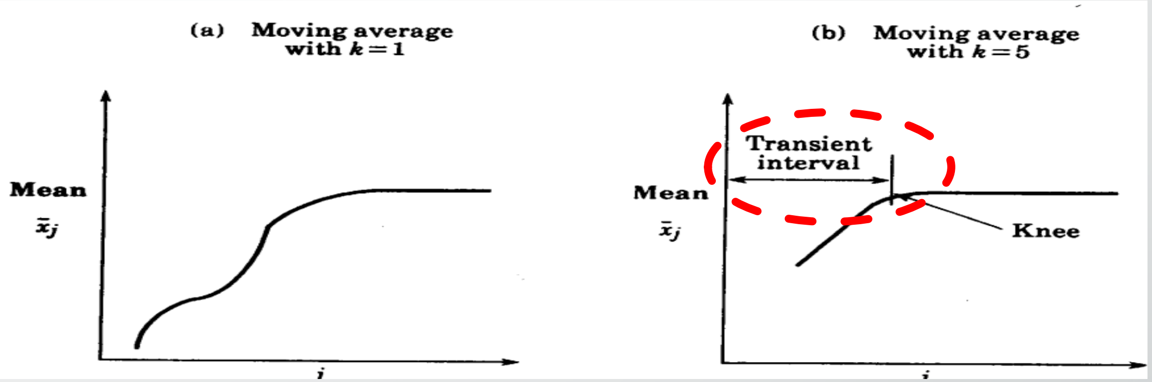
Faça $k = 1$ e prossiga para o próximo passo.

2. Plote uma trajetória da média móvel de $2k + 1$ valores sucessivos:

$$\bar{\bar{x}}_j = \frac{1}{2k+1}\sum_{l=-k}^{k} \bar{x}_{j+l} \text{ para } j = k+1, k+2, ..., n-k$$

3. Repita o passo 2, com $k = 2, 3, ...$ até que a curva fique suave.

4. O valor de j no joelho corresponde ao comprimento da fase transiente.



m replicações cada uma de tamanho *n*
x_{ij} = *j–ésima* observação da *i-ésima* replicação

IV. SIMULAÇÕES TERMINAIS

- Temos **interesse no desempenho do transiente**
  - Ex.: Tráfego de redes
- **Sistemas que são desligados**
  - Não é necessária a remoção de transientes.
- **Condições finais**:
  - Pode ser necessário excluir dos resultados a parte final da simulação
  - Técnicas semelhantes às de remoção de transientes.

V. CRITÉRIO DE PARADA: ESTIMATIVA DA VARIÂNCIA

- Execute até que o intervalo de confiança seja estreito o bastante

$$\bar{x} \pm z_{1-\alpha/2}\mathrm{Var}(\bar{x})$$

- Para observações independentes:

$$\mathrm{Var}(\bar{x}) = \frac{\mathrm{Var}(x)}{n}$$

- Independência não é aplicável a muitas simulações. Grande tempo de espera para o i-ésimo job $\Rightarrow$ Grande tempo de espera para o $(i+1)$-ésimo job

- Para observações correlacionadas:

$$\text{Variância real} \gg \frac{\mathrm{Var}(x)}{n}$$

- Soluções
  - Replicações Independentes
  - Média de Lotes (*Batches*)
  - Método Regenerativo

REPLICAÇÕES INDEPENDENTES

- Assume que as **médias de replicações independentes são independentes**
- Conduza m replicações de tamanho $n_0 + n$ cada

1. Calcule a **média para cada replicação**:

$$\bar{x}_i = \frac{1}{n} \sum_{j=n_0+1}^{n_0+n} x_{ij},\ i = 1, 2, \ldots, m$$

2. Calcule a **média geral para todas as replicações**:

$$\bar{\bar{x}} = \frac{1}{m} \sum_{i=1}^{m} \bar{x}_i$$

REPLICAÇÕES INDEPENDENTES

3. Calcule a **variância das médias das replicações**:

$$\text{Var}(\bar{x}) = \frac{1}{m-1}\sum_{i=1}^{m}\left(\bar{x}_i - \bar{\bar{x}}\right)^2$$

O **intervalo de confiança** para a **resposta média** é:

$$\left[\bar{\bar{x}} \mp z_{1-\alpha/2}\text{Var}(\bar{x})\right]$$

- Mantenha as replicações longas para evitar desperdícios.
- Normalmente bastam **dez replicações**.

VI. REDUÇÃO DA VARIÂNCIA

- Reduz a variância através do **controle das cadeias de números aleatórios**
  - Introduz correlação em observações sucessivas
- Problema
  - o uso sem o devido cuidado pode não dar certo, levando a aumentar a variância.
- Apenas para **analistas com conhecimentos sofisticados de estatística**
  - Não é recomendado para iniciantes!!!

EXERCÍCIO 25.1

- Imagine que **você** foi chamado como um **especialista** para rever um estudo de simulação. Quais dos seguintes **resultados de simulação** você consideraria como **não-intuitivo** e gostaria de tê-los validado cuidadosamente?
    1. A vazão de um sistema aumenta à medida que a sua carga aumenta.
    2. A vazão de um sistema diminui à medida que a sua carga aumenta.
    3. O tempo de resposta de um sistema aumenta à medida que a carga aumenta.
    4. O tempo de resposta de um sistema diminui à medida que a carga aumenta.
    5. A taxa de perda de um sistema diminui à medida que a carga aumenta.

EXERCÍCIO 25.2

25.2 Find the duration of the transient interval for the following sample:

11, 4, 2, 6, 5, 7, 10, 9, 10, 9, 10, 9, 10, ...

Does the method of truncation give the correct result in this case?

GERAÇÃO DE NÚMEROS ALEATÓRIOS

Capítulo 26

Raj Jain - The Art of Computer Systems Performance Analysis. Wiley, 1991.

GERAÇÃO DE NÚMEROS ALEATÓRIOS

- **Número aleatório** = Uniforme(0,1)
- **Variável aleatória** = Outras distribuições = Função(número aleatório uniforme)

UM GERADOR DE AMOSTRAS

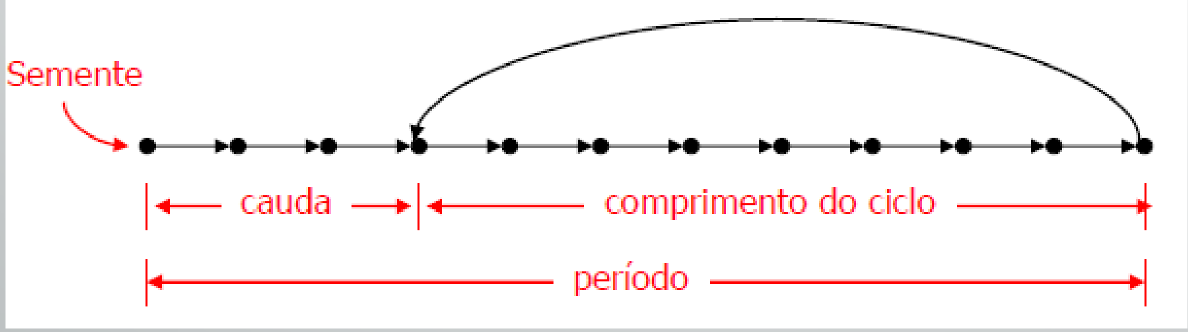
- Por exemplo, $x_n = f(x_{n-1}, x_{n-2}, \ldots)$

  $x_n = (5x_{n-1} + 1) \bmod 16$
- Começando com $x_0 = 5$

  $x_1 = (5x5+1) \bmod 16 = 26 \bmod 16 = 10$
- Os primeiros 32 números obtidos desta forma são:
  - 10, 3, 0, 1, 6, 15, 12, 13, 2, 11, 8, 9, 14, 7, 4, 5
  - 10, 3, 0, 1, 6, 15, 12, 13, 2, 11, 8, 9, 14, 7, 4, 5
- Dividindo os x's por 16 teríamos:
  - 0,6250; 0,1875; 0,0000; 0,0625; 0,3750, 0,9375; 0,7500; 0,8125; 0,1250; 0,6875; 0,5000; 0,5625; 0,8750; 0,4375; 0,2500; 0,3125;
  - 0,6250; 0,1875; 0,0000; 0,0625; 0,3750, 0,9375; 0,7500; 0,8125; 0,1250; 0,6875; 0,5000; 0,5625; 0,8750; 0,4375; 0,2500; 0,3125;

TERMINOLOGIA

- **Semente x_0**: valor usado para iniciar a seqüência
- **Pseudo-Aleatório**:
  - Apesar de determinístico passaria nos **testes de aleatoriedade**
  - Completamente aleatório: Irrepetível
- Alguns geradores não repetem uma parte inicial da sequência
  - Comprimento do ciclo, Cauda, **Período**



PROPRIEDADES DESEJÁVEIS DE UM BOM GERADOR

- Deve ser **eficiente computacionalmente**.
- O **período** deve ser **longo**.
- **Valores sucessivos** devem ser **independentes** e **uniformemente distribuídos**.

TIPOS DE GERADORES DE NÚMEROS ALEATÓRIOS

I. **Geradores Congruo-Lineares**
II. **Geradores de Tausworthe**
III. **Geradores de Fibonacci estendidos**
IV. **Geradores Combinados**

I. GERADORES CONGRUO-LINEARES

- Descobertos por **D.H. Lehmer em 1951**: Os resíduos de potências sucessivas de um número têm boas propriedades aleatórias.

$$x_n = a^n \bmod m$$

- Ou de forma equivalente,

$$\mathbf{x_n = ax_{n-1} \bmod m}$$

a = multiplicador

m = módulo

- **Lehmer** escolheu os seguintes valores: a = 23 e m = 10^8+1. Bom para o **ENIAC**: máquina com 8 dígitos decimais.

- Generalização:

$$\mathbf{x_n = (ax_{n-1} + b) \bmod m}$$

- Pode ser analisado facilmente utilizando a **teoria das congruências**
  - Geradores Congruo-Lineares Mistos ou **Geradores Congruo-Lineares (LCG)**
    - Misto = possui tanto uma multiplicação por *a* como uma adição de *b*.

SELEÇÃO DOS PARÂMETROS DO LCG

- ***a, b*** e ***m*** afetam o **período** e a **autocorrelação**
- O módulo ***m*** deve ser **grande**.
  - O período nunca será maior do que m.
- Para que o cálculo do módulo seja eficiente, ***m*** deve ser uma **potência de 2**
  - mod *m* pode ser obtido por truncagem.

- Se ***b*** for **não-nulo**, o período máximo *m* será obtido se e só se:
  - Os inteiros ***m* e *b* forem primos** entre si, isto é, não possuam nenhum fator comum além de 1.
  - Todo número primo que for um fator de *m* deve ser também um fator de *a*-1.
  - Se *m* for múltiplo de 4, *a*-1 também deve ser múltiplo de 4.
- **Todas estas condições são satisfeitas se $m = 2^k$, $a = 4c+1$ e *b* for ímpar.**
  - Onde, *c*, *b* e *k* são inteiros positivos.

LCG

PERÍODO *vs* AUTOCORRELAÇÃO

- Um **gerador** que possua um **período máximo** é chamado de **gerador de período completo**.

$$x_n = \left((2^{34}+1)x_{n-1}+1\right) \bmod 2^{35}$$

$$x_n = \left((2^{18}+1)x_{n-1}+1\right) \bmod 2^{35}$$

- é **preferível** aquele que exibir **baixa autocorrelação entre números sucessivos**.
  - Ambos os geradores têm o **mesmo período** completo, mas o primeiro tem uma **correlação** de 0,25 entre x_{n-1} e x_n, enquanto que o segundo tem uma correlação desprezível de menos do que 2^{-18}.

LCG MULTIPLICATIVOS

- LCG Multiplicativo ($b = \mathbf{0}$)

$$x_n = a x_{n-1} \bmod m$$

- Dois tipos:

$$\mathbf{m = 2^k}$$

$$\mathbf{m \neq 2^k}$$

LCG MULTIPLICATIVO COM $m = 2^k$

- **$m = 2^k$**:
  - **divisão trivial**
  - **período máximo** possível **2^{k-2}** = (1/4) de 2^k
- Este período é obtido se:
  - o **multiplicador *a*** for da forma $8i \pm 3$ e
  - a **semente inicial** for um **inteiro ímpar**
- Um quarto do período máximo possível pode não ser tão pequeno.
- Os bits menos significativos dos números aleatórios obtidos utilizando-se LCGs com $m=2^k$ têm padrão cíclico.

EXEMPLO 26.1

$$x_n = 5x_{n-1} \bmod 2^5$$

- Usando uma semente **ímpar** $x_0=1$:
  - Seq. ?
  - **período = ?**
- Com $x_0=2$ (semente **par**)
  - Seq. ?
  - **período = ?**
- Caso multiplicador não da forma 8i±3:

$$x_n = 7x_{n-1} \bmod 2^5$$

- Com semente $x_0=1$:
  - Seq. ?
  - **período = ?**

LCG MULTIPLICATIVO COM $m = 2^k$

- **$m = 2^k$**:
  - **divisão trivial**
  - **período máximo** possível **2^{k-2}** $= (1/4)$ de 2^k
- Este período é obtido se:
  - o **multiplicador *a*** for da forma $8i \pm 3$ e
  - a **semente inicial** for um **inteiro ímpar**
- Um quarto do período máximo possível pode não ser tão pequeno.
- Os bits menos significativos dos números aleatórios obtidos utilizando-se LCGs com $m=2^k$ têm padrão cíclico.

EXEMPLO 26.1

$$x_n = 5x_{n-1} \bmod 2^5$$

- Usando uma semente **ímpar** $x_0=1$:
  - 5, 25, 29, 17, 21, 9, 13, 1, 5, ...
  - **período = 8** = 32/4
- Com $x_0=2$ (semente **par**) a seqüência é: 10, 18, 26, 2, 10, ...
  - aqui o **período é apenas 4**.
- Multiplicador não da forma $8i\pm3$:

$$x_n = 7x_{n-1} \bmod 2^5$$

- Usando uma semente $x_0=1$, obtemos a seqüência:

  7, 17, 23, 1, 7, ...
  - O **período é apenas 4**

LCG MULTIPLICATIVO COM $m \neq 2^k$

- **Módulo *m*** = **número primo**
  - Com um multiplicador *a* adequado, **período = *m*-1**
  - Período máximo possível = **m**
- Gerador de **período completo** se e só se o **multiplicador *a*** for uma ***raiz primitiva*** do **módulo *m***.
- **a** é uma raiz primitiva de *m* se e só se a^n **mod *m*** $\neq$ **1** para *n* = 1, 2,..., *m*-2.

EXEMPLO 26.2

$$x_n = 3x_{n-1} \bmod 31$$

- Usando uma semente $x_0 = 1$:
  - 1, 3, 9, 27, 19, 26, 16, 17, 20, 29, 25, 13, 8, 24, 10, 30, 28, 22, 4, 12, 5, 15, 14, 11, 2, 6, 18, 23, 7, 21, 1, ...
  - O **período** é **30**
    - **3** é uma **raiz primitiva** de **31**
- Com um **multiplicador** ***a* = 5** : 1, 5, 25, 1, ...
  - O **período** é **apenas 3** !
    - 5 não é uma raiz primitiva de 31
    - 5^3 mod 31 = 125 mod 31 = 1
- **Raízes primitivas de 31** = 3, 11, 12, 13, 17, 21, 22 e 24.

MÉTODO DE SCHRAGE

- Assume que:
  - Não haja erros de arredondamento
  - Usa aritmética inteira sem *overflows*
  - Não dá para fazer em BASIC
  - Produto ax_{n-1} > Maior inteiro ⇒ *Overflow*
- Identidade: $ax \bmod m = g(x) + mh(x)$

  $g(x) = a(x \bmod q) - r(x \operatorname{div} q)$

  $h(x) = (x \operatorname{div} q) - (ax \operatorname{div} m)$

  $q = m \operatorname{div} a$, $r = m \bmod a$

  'A div B' = divisão de A por B e truncar o resultado.
- Para todos os x's na faixa 1, 2,..., m-1, o cálculo de $g(x)$ envolve apenas números menores do que m-1.
- Se r < q, h(x) é 0 ou 1, e isto pode ser derivado a partir de g(x);
  - $h(x)$ é 1 se e só se $g(x)$ for negativo.

EXEMPLO 26.3

$$x_n = 7^5 x_{n-1} \bmod (2^{31} - 1)$$

- 2^{31}-1 = 2.147.483.647 = número primo
- 7^5 = 16.807 é uma das suas 534.600.000 raízes primitivas
- O produto ax_{n-1} pode ser tão grande quanto 16.807 x 2.147.483.647 ≈1,03 x 2^{45}.
- Necessita de inteiros com 46 bits
  - a= 16.807
  - m= 2.147.483.647
  - q= m div a = 2.147.483.647 div 16.807 = 127.773
  - r= m mod a = 2.147.483.647 mod 16.807 = 2.836
- Para uma implementação correta, $x_0 = 1 \Rightarrow x_{10000}$= 1.043.618.065.

IMPLEMENTAÇÃO

COM ARITMÉTICA INTEIRA

```
FUNCTION Random(VAR x:INTEGER): REAL;
    CONST
        a = 16807; (* Multiplicador *)
        m = 2147483647; (* Modulo *)
        q = 127773; (* m div a *)
        r = 2836; (* m mod a *)
    VAR
        x_div_q, x_mod_q, x_new: INTEGER;
    BEGIN
        x_div_q := x DIV q;
        x_mod_q := x MOD q;
        x_new := a*x_mod_q -r*x_div_q;
        IF x_new > 0 THEN x := x_new ELSE x :=
        x_new + m;
        Random := x/m;
    END;
```

COM ARITMÉTICA REAL

```
FUNCTION Random(VAR x:DOUBLE): DOUBLE;
    CONST
        a = 16807.0D0; (* Multiplicador *)
        m = 2147483647.0D0; (* Modulo *)
        q = 127773.0D0; (* m div a *)
        r = 2836.0D0; (* m mod a *)
    VAR
        x_div_q, x_mod_q, x_new: DOUBLE;
    BEGIN
        x_div_q := TRUNC(x/q);
        x_mod_q := x -q*x_div_q;
        x_new := a*x_mod_q -r*x_div_q;
        IF x_new >= 0.0D0 THEN x := x_new ELSE x
        := x_new + m;
        Random := x/m;
    END;
```

II. GERADORES DE TAUSWORTHE

- Necessitamos de longos números aleatórios para **aplicações de criptografia**.
- Gera **seqüência aleatória de dígitos binários** (0 ou 1)
- **Divide a seqüência** em **cadeias de comprimento desejado**

- Proposto por **Tausworthe** (1965)
  - $b_n = c_{q-1}b_{n-1} \oplus c_{q-2}b_{n-2} \oplus c_{q-3}b_{n-3} \oplus \ldots \oplus c_0 b_{n-q}$
    - c_i e b_i são variáveis binárias com valores 0 ou 1
    - $\oplus$ é a operação ou-exclusivo (adição mod 2)
- Usa os **últimos *q* bits da seqüência**
  - seqüência autoregressiva de ordem *q* ou AR(*q*)
- Um gerador AR(*q*) pode ter um **período máximo** de $\mathbf{2^q - 1}$.

GERADORES DE TAUSWORTHE

- D **= operador de atraso** tal que $Db(n) = b(n+1)$:

  $D^q b(i\text{-}q) = c_{q\text{-}1} D^{q\text{-}1} b(i\text{-}q) + c_{q\text{-}2} D^{q\text{-}2} b(i\text{-}q) + \ldots + c_0 b(i\text{-}q)$ mod 2

  ou

  $D^q - c_{q\text{-}1} D^{q\text{-}1} - c_{q\text{-}2} D^{q\text{-}2} - \ldots - c_0 = 0$ mod 2

  ou

  $D^q + c_{q\text{-}1} D^{q\text{-}1} + c_{q\text{-}2} D^{q\text{-}2} + \ldots + c_0 = 0$ mod 2

- **polinômio característico**

  $x^q + c_{q\text{-}1} x^{q\text{-}1} + c_{q\text{-}2} x^{q\text{-}2} + \ldots + c_0 = 0$

- O **período** é o menor inteiro positivo n para o qual x^n-1 **é divisível** pelo **polinômio característico**.
- O **período máximo** possível com um polinômio de ordem q é 2^q-1.
- Os polinômios que atingem este período são chamados de **polinômios primitivos**

EXEMPLO 26.4

$x^7 + x^3 + 1$

- Usando o operador *D* no lugar de *x* :
  - $D^7b(n) + D^3b(n) + b(n) = 0 \bmod 2$
- Ou:
  - $b_{n+7} + b_{n+3} + b_n = 0 \bmod 2 \quad n = 0,1,2,\ldots$
- ou usando o **operador ou-exclusivo**
  - $b_{n+7} \oplus b_{n+3} \oplus b_n = 0 \quad n = 0,1,2,\ldots$
- Ou:
  - $b_{n+7} = b_{n+3} \oplus b_n \quad n = 0,1,2,\ldots$
- Substituindo ***n* -7** por ***n*** :

$b_n = b_{n-4} \oplus b_{n-7} \quad n=7,8,9,\ldots$

Começando com $b_0 = b_1 = \ldots = b_6 = 1$:

$b_7 = b_3 \oplus b_0 = 1 \oplus 1 = 0$

$b_8 = b_4 \oplus b_1 = 1 \oplus 1 = 0$

$b_9 = b_5 \oplus b_2 = 1 \oplus 1 = 0$

$b_{10} = b_6 \oplus b_3 = 1 \oplus 1 = 0$

$b_{11} = b_7 \oplus b_4 = 0 \oplus 1 = 1$

...

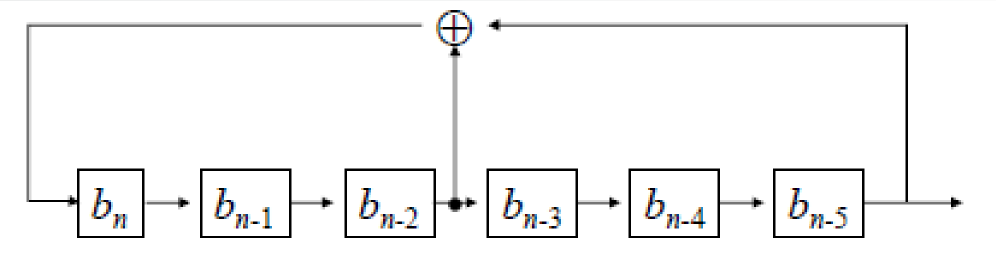
- A seqüência completa é:

  1111111 0000111 0111100 1011001 0010000 0010001 0011000 1011101 0110110 0000110 0110101 0011100 1111011 0100001 0101011 1110100 1010001 1011100 0111111 1000011 1000000.

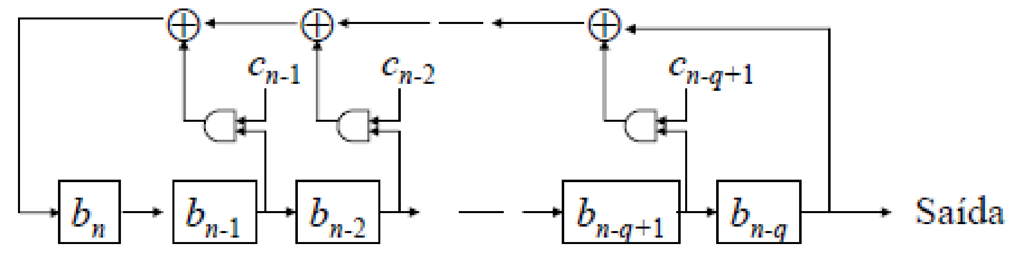
- Período = 127 ou 2^7-1 bits
- O polinômio $\mathbf{x^7 + x^3 + 1}$ é um **polinômio primitivo**.

REGISTRADOR COM DESLOCAMENTO COM REALIMENTAÇÃO LINEAR

$x^5 + x^3 + 1 \Rightarrow b_n = b_{n-2} \oplus b_{n-5}$



generalizando



As portas AND não são necessárias se os c_i *'s* forem conhecidos.

GERANDO U(0,1)

- Divida a seqüência em grupos sucessivos de ***s*** bits e use os primeiros ***l*** bits de cada grupo como uma fração binária:

  $x_n = \mathbf{0,}b_{sn}\ b_{sn+1}\ b_{sn+2}\ b_{sn+3} \ldots b_{sn+l-1}$

- Ou equivalentemente:

$$x_n = \sum_{j=1}^{l} 2^{-j} b_{sn+j-1}$$

  - Onde, ***s*** é uma constante maior ou igual a ***l*** e é **primo** em relação a $\mathbf{2^q-1}$.
  - $s \geq l \Rightarrow x_n$ e x_j para $n \neq j$ não têm nenhum bit em comum.
  - Serem primos entre si garante um período completo de 2^q-1 para x_n

EXEMPLO 26.5

$$b_n = b_{n-4} \oplus b_{n-7}$$

- O período $2^7-1 = 127$ é primo.
- $l = 8$, $s = 8$

$x_0 = 0{,}11111110_2 = 0{,}99219_{10}$

$x_1 = 0{,}00011101_2 = 0{,}11328_{10}$

$x_2 = 0{,}11100101_2 = 0{,}89453_{10}$

$x_3 = 0{,}10010010_2 = 0{,}29688_{10}$

$x_4 = 0{,}00000100_2 = 0{,}36328_{10}$

$x_5 = 0{,}01001100_2 = 0{,}42188_{10}$

....

PROPRIEDADES DOS GERADORES DE TAUSWORTHE

- Os **números de l bits** têm as seguintes **propriedades**:

  1. A média da seqüência é um meio:

     $$E[x_n] \approx 1/2$$

  2. A variância da seqüência é um dozeavos:

     $$Var[x_n] \approx 1/12$$

  3. A correlação serial é zero:

     $$Corre[x_n, x_{n+s}] = 0$$

     para $0 < |s| < \dfrac{2^q - 1 - l}{l}$

  4. A **seqüência** é ***k*-distribuída** para todos os *k*'s até $\lfloor q/l \rfloor$. Isto significa que **cada *k*-tupla** de **números de l bits** aparece $\mathbf{2^{q-kl}}$ **vezes** no período completo, à exceção da tupla com todos os valores zero que aparece uma vez a menos.

Usando $k = 1$ e $l = 1$

- a seqüência de bits contém $\mathbf{2^{q-1}}$ uns e $\mathbf{2^{q-1}-1}$ zeros
- Se deslocássemos uma janela de comprimento q na seqüência, cada uma das k-tuplas diferentes de zero apareceriam exatamente uma vez num período completo.

III. GERADORES DE FIBONACCI ESTENDIDOS

- **Seqüência de Fibonacci**:
  $\mathbf{x_n = x_{n-1} + x_{n-2}}$
  - **Números aleatórios**: $x_n = x_{n-1} + x_{n-2} \bmod m$
    - Alta correlação serial
    - Péssimas propriedades aleatórias
- **Combinando** o **quinto** e o **décimo-sétimo** valores mais recentes:

  $x_n = x_{n-5} + x_{n-17} \bmod 2^k$
  - Este gerador passa em muitos testes estatísticos e recomenda-se a sua implementação como segue, usando 17 posições de memória $L[1],\ldots, L[17]$

- **Inicialização**: preencha as posições com 17 inteiros, nem todos pares, e crie dois ponteiros *i* e *j* com valores 17 e 5, respectivamente.
- Em cada chamada sucessiva faça:

```
x := L[i] + L[j];
L[i] := x;
i := i -1; IF i = 0 THEN i := 17;
j := j -1; IF j = 0 THEN j := 17;
Return x;
```

- A operação de adição na primeira linha deste gerador é automaticamente módulo 2^k em máquinas com *k* bits com aritmética de complemento de 2.

- O **período do gerador** é $2^k(2^{17}-1)$.
- Para k = 8, 16 e 32, este período é $1,6\times10^7$, $4,3\times10^9$ e $2,8\times10^{14}$, respectivamente.
- O período é **consideravelmente maior** do que o que é possível com LCGs.

IV. GERADORES COMBINADOS

1. **Adição de números aleatórios** obtidos por dois ou mais geradores.

$$w_n = (x_n + y_n) \bmod m$$

Por exemplo, **L'Ecuyer** (**1986**):

$$x_n = 40014 x_{n-1} \bmod 2147483563$$

$$y_n = 40692 y_{n-1} \bmod 2147483399$$

Isto produz:

$$w_n = (x_n - y_n) \bmod 2147483562$$

com **período** = $\mathbf{2{,}3 \times 10^{18}}$

Para **computadores de 16 bits**:

$w_n = 157\ w_{n-1} \bmod 32363$; $x_n = 146\ x_{n-1} \bmod 31727$; $y_n = 142\ y_{n-1} \bmod 31657$

Isto produz:

$v_n = (w_n - x_n + y_n) \bmod 32362$ com **período** = $\mathbf{8{,}1 \times 10^{12}}$

2. Fazendo o **ou-exclusivo de números aleatórios** obtidos por **dois ou mais geradores**.

3. **Embaralhamento**. Use uma seqüência como índice para decidir qual de diversos números gerados pela segunda seqüência deve ser retornado.

Algoritmo M:

a) Preencha um vetor de tamanho, por exemplo, 100.

b) Gere um novo y_n (entre 0 e m-1)

c) índice $i = 1 + 100\ y_n / m$

d) O i-ésimo elemento do vetor é retornado como o próximo número aleatório

e) Um novo valor de x_n é gerado e guardado na i-ésima posição.

Não dá para saltar uma longa subseqüência como é normalmente necessário em simulações com múltiplos *streams*.

COLETÂNEA DE GERADORES DE NÚMEROS ALEATÓRIOS

- **LCG multiplicativo** bastante **popular**:

  $x_n = 7^5 x_{n-1} \bmod (2^{31} - 1)$
- Usado no:
  - Sistema SIMPL/I (IBM 1972),
  - Sistema APL da IBM (Katzan 1971),
  - Sistema Operacional PRIMOS da Prime Computer (1984) e
  - Biblioteca científica do IMSL (1980)
- $2^{31} - 1$ é um número primo e 7^5 é uma de suas raízes primitivas
  - **Período** completo de $2^{31} - 2$.
- Este gerador foi bastante analisado e foi comprovado que é bom.
  - Os seus **bits de ordem mais baixa** são **uniformemente distribuídos**.

- **Busca exaustiva de Fishman e Moore** (1986) para $m = 2^{31} - 1$:
  - $x_n = 48271\, x_{n-1} \bmod (2^{31} - 1)$
  - $x_n = 69621\, x_{n-1} \bmod (2^{31} - 1)$
- SIMSCRIPT II.5 e FORTRAN do DEC-20:
  - $x_n = 630360016\, x_{n-1} \bmod (2^{31} - 1)$

ESCOLHA DA SEMENTE

- Simulações com **multi-seqüências**: necessitam de **mais de uma cadeia de números aleatórios**.

Fila única = **Duas cadeias**:

- Intervalo entre chegadas e tempos de serviço aleatórios

1. **Não use zero**. Pode ser usada com LCGs. Mas, LCGs multiplicativos ou um LCG de Tausworthe ficarão presos em zero.

2. **Evite valores pares**. Para LCGs multiplicativos com módulo $m = 2^k$, a **semente deve ser ímpar**. É melhor evitar geradores que possuam muitas restrições sobre os valores das sementes ou cujo desempenho (período e aleatoriedade) dependam do valor da semente.

3. **Não subdivida uma cadeia**. Não use uma única sequência $\{u_1, u_2, \ldots\}$ gerada p. ex. pela semente u_0 para todas as variáveis: u_1 para gerar intervalos entre chegadas, u_2 para gerar o tempo de serviço $\Rightarrow$ **Forte correlação**

4. **Use cadeias que não se sobreponham**.
   Superposição $\Rightarrow$ Correlação.

   Mesma raiz $\Rightarrow$ **mesma cadeia**!

   Use p. ex. u_0 para sequência 1, $u_{10.000}$ para sequência 2, $u_{20.000}$ para sequência 3 etc

5. **Reutilize sementes** em **replicações sucessivas**:

   a seqüência pode continuar a partir do último número gerado.

6. **Não utilize** sementes aleatórias tais como a **hora do dia**:

   Não dá para ser reproduzida. Não dá para garantir ausência de superposição.

7. **Escolha $\{u_0, u_{100.000}, u_{200.000}, \ldots\}$** como sementes para sequências distintas

   como calcular (para LCGs) o n-ésimo número de uma sequência sem gerar os valores intermediários

$$x_n = a^n x_0 + \frac{c(a^n - 1)}{a - 1} \bmod m$$

MATERIAL MAIS RECENTE

GERADORES MERSENNE TWISTER

- **MERSENNE TWISTER (MT)**, proposto em **1998**, possui um **período** de **2^{19937}-1**
- Dados estão **equidistribuídos** em **623 dimensões** para uma **precisão de 32 bits**.

http://www.math.sci.hiroshima-u.ac.jp/~m-mat/MT/emt.html

- Disponível no **OMNet++**

https://docs.omnetpp.org/tutorials/

- **SIMD-oriented FAST MERSENNE TWISTER (SFMT)**, variante do Mersenne Twister proposta em **2006**:

http://www.math.sci.hiroshima-u.ac.jp/~m-mat/MT/SFMT/index.html

- Gerador do tipo ***Linear Feedbacked Shift Register*** (LFSR)
  - gera um **inteiro pseudo-aleatório** de **128-bits** a cada etapa.
- Projetado para usar o paralelismo das CPUs modernas, tais como pipeline multiestágios e instruções SIMD
- Dá suporte a saídas de inteiros de 32 e 64-bit, assim como números de ponto flutuante com precisão dupla.

TRUE RANDOM NUMBER SERVICE

- **Aleatoriedade** obtida **do ruído atmosférico !**

  https://www.random.org/

  - *RANDOM.ORG offers true random numbers to anyone on the Internet.*
  - *The **randomness comes from atmospheric noise**, which for many purposes is **better than the pseudo-random number algorithms typically used in computer programs**.*
  - *People use RANDOM.ORG for holding drawings, lotteries and sweepstakes, to drive online games, for scientific applications and for art and music.*

GERADOR QUÂNTICO

- Artigo: *Random numbers certified by Bellâ€™s theorem*
  Autores: S. Pironio, A. AcÃn, S. Massar, A. Boyer de la Giroday, D. N. Matsukevich, P. Maunz, S. Olmschenk, D. Hayes, L. Luo, T. A. Manning, C. Monroe
  Revista: Nature
  Data: 15 April 2010
  Vol.: 464, 1021-1024
  DOI: 10.1038/nature09008
- https://www.inovacaotecnologica.com.br/noticias/noticia.php?artigo=gerador-quantico-numeros-aleatorios&id=010150100422#.XxoUl-dv_IU

GERADOR DO NS-3

- **GERADOR MRG32k3a**
  - *L'Ecuyer, P. - Good parameters and implementations for combined multiple recursive random number generators.Oper.Res.47(1) 159–164, 1999.*
  - Descrição detalhada em: *http://www.iro.umontreal.ca/~lecuyer/myftp/papers/streams00.pdf*.
- **PRINCIPAIS CARACTERÍSTICAS**
  - Provê **$1,8 \times 10^{19}$ sequências** de números aleatórios independentes, cada uma consistindo de **$2,3 \times 10^{15}$ subsequências**.
  - Cada **subsequência** tem um **período** (i.e., o número de números aleatórios antes da sobreposição) de **$7,6 \times 10^{22}$**.
  - O **período** do gerador inteiro é **$3,1 \times 10^{57}$**.

Tem **dois componentes de ordem 3** e no **passo *n***, seu **estado** é o **par de vetores**

$$s_{1,n} = (x_{1,n}, x_{1,n+1}, x_{1,n+2})$$
$$s_{2,n} = (x_{2,n}, x_{2,n+1}, x_{2,n+2})$$

que evoluem de acordo com as recorrências lineares

$$x_{1,n} = (1403580 \times x_{1,n-2} - 810728 \times x_{1,n-3}) \bmod m_1$$
$$x_{2,n} = (527612 \times x_{2,n-1} - 1370589 \times x_{2,n-3}) \bmod m_2$$

onde

$$m_1 = 2^{32} - 209 = 4294967087$$
$$m_2 = 2^{32} - 22853 = 4294944443$$

e sua saída u_n é definida por

$$z_n = (x_{1,n} - x_{2,n}) \bmod 4294967087$$

$$u_n = \begin{cases} z_n/4294967088 & \text{if } z_n > 0 \\ 4294967087 / 4294967088 & \text{if } z_n = 0 \end{cases}$$

Seu **período** é

$$\rho = (m_1^3 - 1)(m_2^3 - 1) / 2 \approx 2^{191} \approx \mathbf{3,1 \times 10^{57}}$$

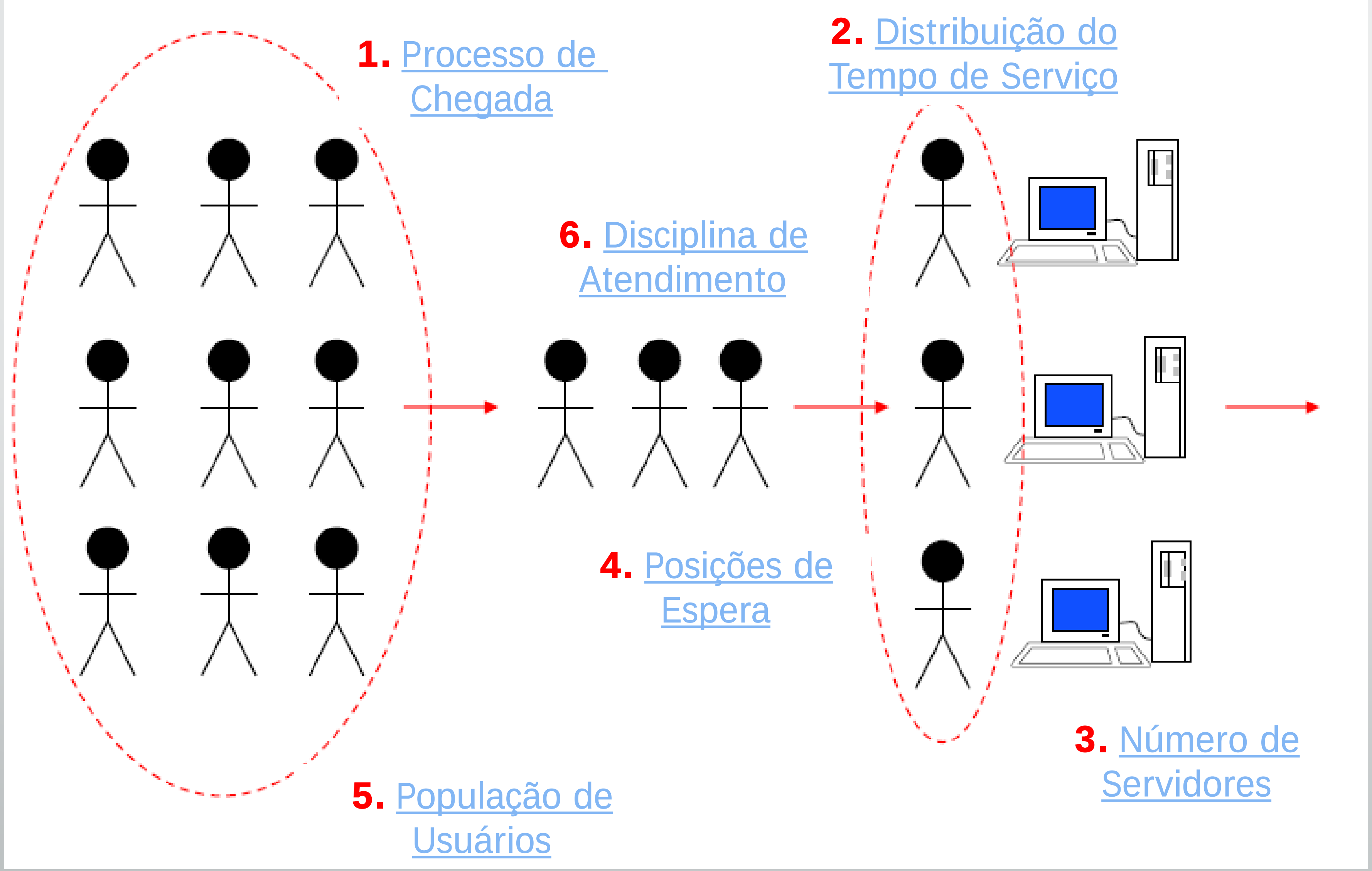
*The parameters have been chosen so that the **period is long**, a **fast implementation** is available (in floating point arithmetic), and the generator **performs well with respect to the spectral test** in up to (at least) **45 dimensions**.*

INTRODUÇÃO À TEORIA DAS FILAS

Capítulo 30

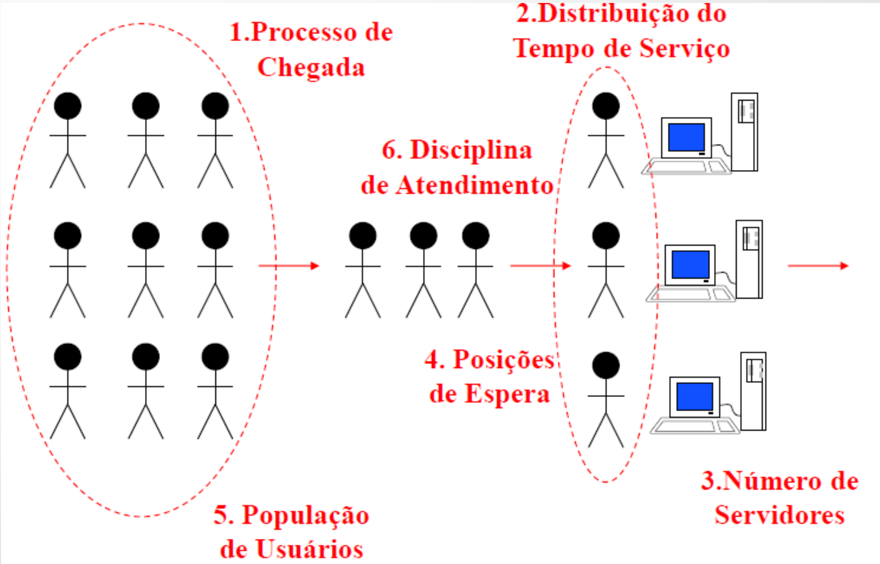
Raj Jain - The Art of Computer Systems Performance Analysis. Wiley, 1991.

COMPONENTES BÁSICOS DE UMA FILA



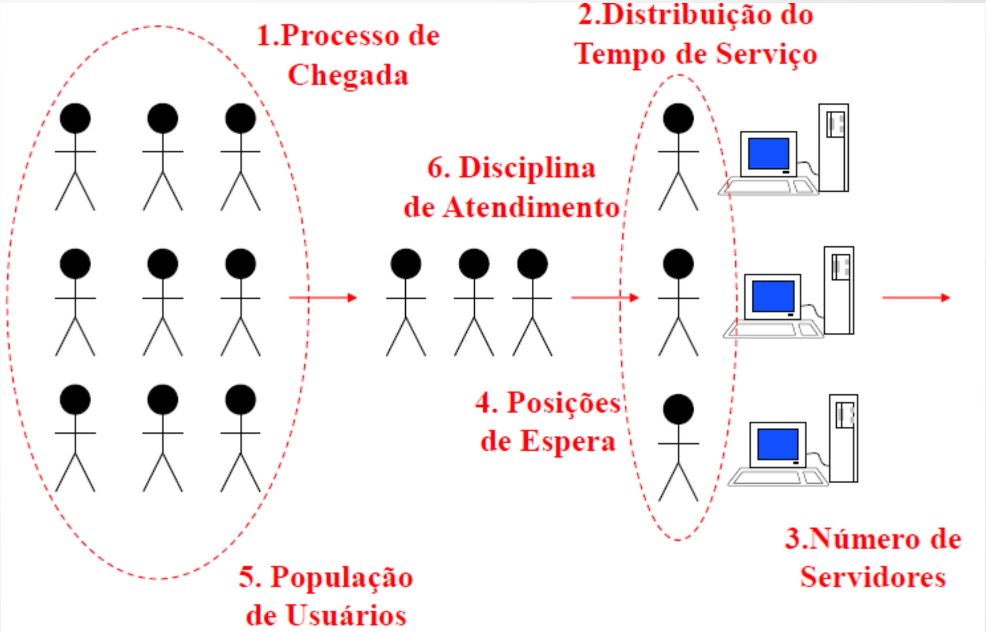
1. PROCESSO DE CHEGADA

- Se os usuários chegam nos instantes $t_1, t_2, \ldots, t_j,$ então as variáveis aleatórias $t_j = t_j - t_{j-1}$ são chamadas de **intervalos entre chegadas**.
  - Assume-se, em geral, que a **seqüência dos** t_j são **v.a.s** independentes e identicamente distribuídas (**IID**).
- Exemplo:
  - chegadas de **Poisson**
    - **intervalos entre chegadas** são **IID** e **exponencialmente** distribuídos.
  - Outras distribuições: Erlang, Hiperexponencial, etc.



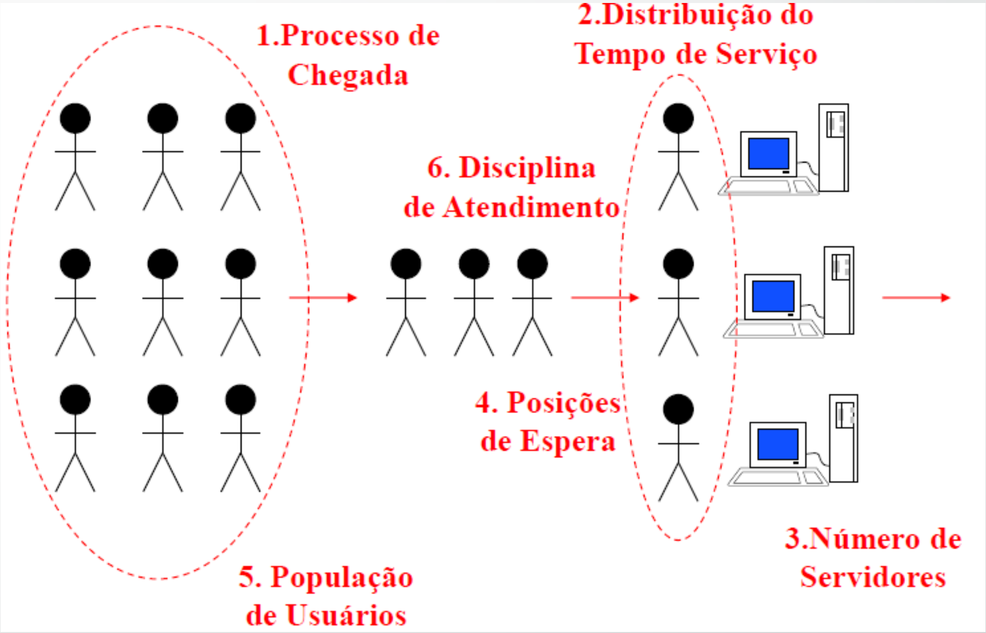
2. DISTRIBUIÇÃO DO TEMPO DE SERVIÇO

- **Tempo de serviço:** tempo que o usuário gasta no servidor.
  - É comum assumir que os tempos de serviço sejam v.a.s IID.
- Distribuições utilizadas: **exponencial**, Erlang, hiper-exponencial e geral.



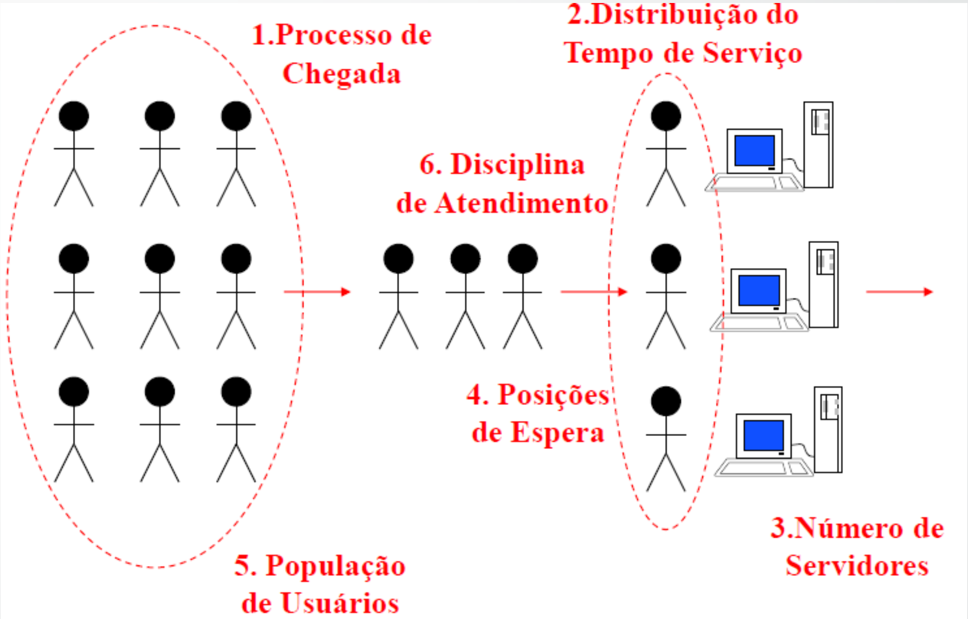
3. NÚMERO DE SERVIDORES:

- **Número de servidores**: para uma mesma fila.



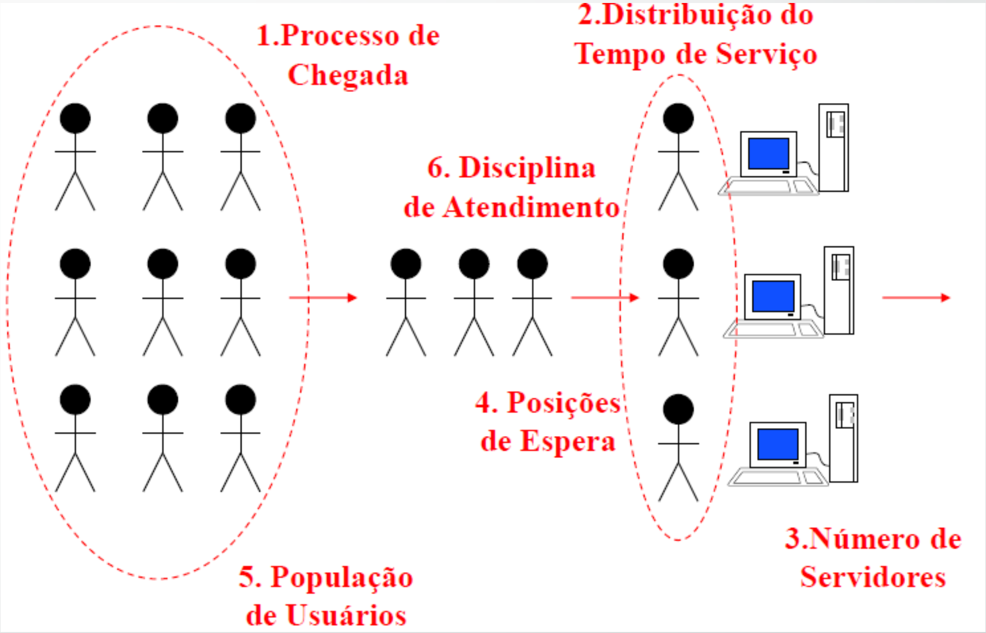
4. POSIÇÕES DE ESPERA

- **Capacidade do sistema:** número máximo de usuários que podem ser acomodados no sistema (em **espera na fila +** em **atendimento** pelos servidores).
- Se a **capacidade do sistema for grande**, é mais fácil analisá-lo com a **hipótese** de que a **fila** seja **infinita**.



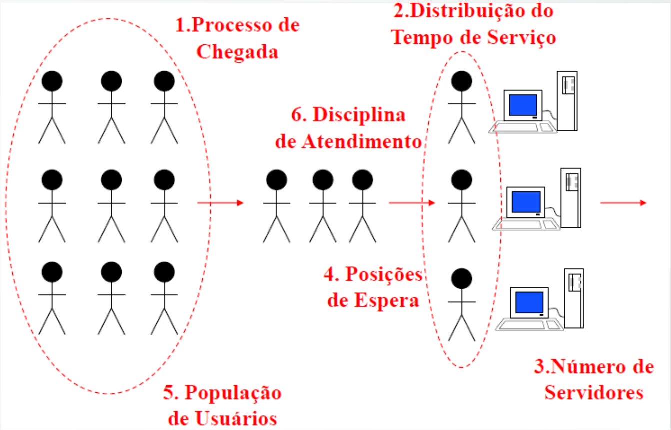
5. POPULAÇÃO DE USUÁRIOS

- **Tamanho da população:** número potencial de usuários.
  - Se o **tamanho for grande** é mais fácil analisar o sistema com a **hipótese** de que a **população** seja **infinita**.



6. DISCIPLINA DE ATENDIMENTO

- **Disciplina de atendimento**: **ordem** na qual os **usuários são atendidos**.



- **FCFS:** ***First Come, First Served***
- LCFS: *Last Come, First Served*
- LCFS-PR: *Last Come, First Served with Preempt and Resume*
- RR: *Round-Robin*
- PS: *Processor Sharing*
- IS: *Infinite Server*

- SPT: *Shortest Processing Time first*
- SRPT: *Shortest Remaining Processing Time first*
- SEPT: *Shortest Expected Processing Time first*
- SERPT: *Shortest Expected Remaining Processing Time first*
- ...

Notação de Kendall

NOTAÇÃO DE KENDALL

- *A/S/m/B/K/SD*
  - *A* é a distribuição do intervalo entre chegadas,
  - *S* é a distribuição do tempo de serviço,
  - *m* é o número de servidores,
  - *B* é o número de *buffers* (capacidade do sistema),
  - *K* é o tamanho da população
  - *SD* é a disciplina de atendimento.

- **Representação das Distribuições**
  - *M* Exponencial (M de *Memoryless*)
  - E_k Erlang com parâmetro k
  - H_k Hiperexponencial com parâmetro k
  - *D* Determinística
  - *G* Geral
  - $M^{[x]}$ Exponencial com chegada em bloco *(bulk)* de tamanho *x*

Exemplos

M/M/3/20/1500/FCFS

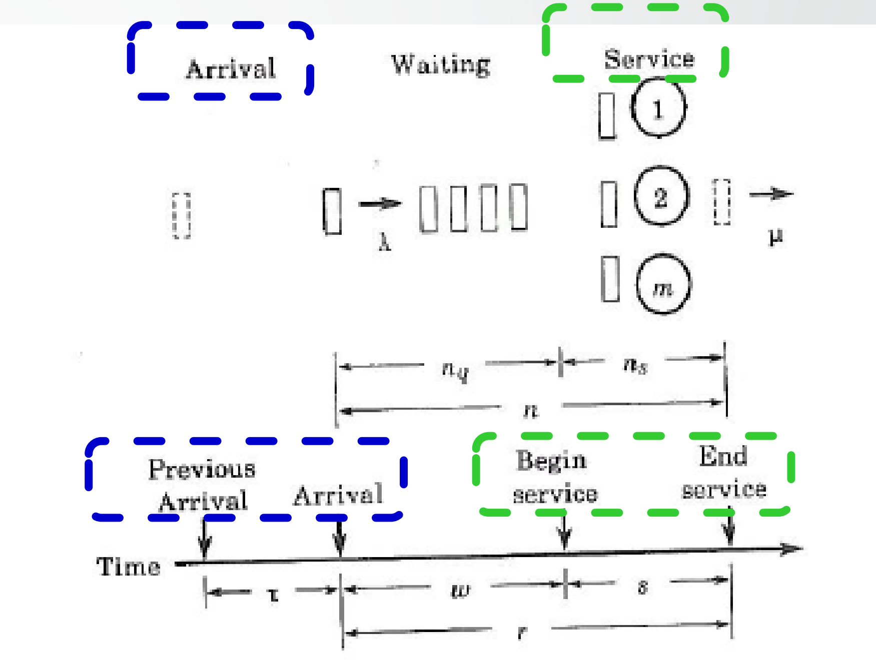
- O **intervalo entre chegadas** sucessivas é distribuído **exponencialmente**.
- Os **tempos de serviço** são **exponencialmente** distribuídos.
- Há **três servidores**.
- O sistema possui **capacidade** *(buffers)* para **20 usuários**.
  - Isto é, **3 usuários** em **atendimento** e **17 esperando na fila** por serviço. Enquanto o número de usuários estiver em seu valor máximo, 20, todos os usuários que chegarem serão perdidos até que o comprimento da fila diminua.
- Há um **total de 1500 usuários** que podem ser atendidos.
- A **disciplina de atendimento** é ***first come, first served***.

OUTROS EXEMPLOS

- $M/M/1/\infty/\infty/FCFS \Leftrightarrow M/M/1$
- $G/G/1/\infty/\infty/FCFS \Leftrightarrow G/G/1$
- $M^{[x]}/M/1$
- $M/G^{[x]}/m$

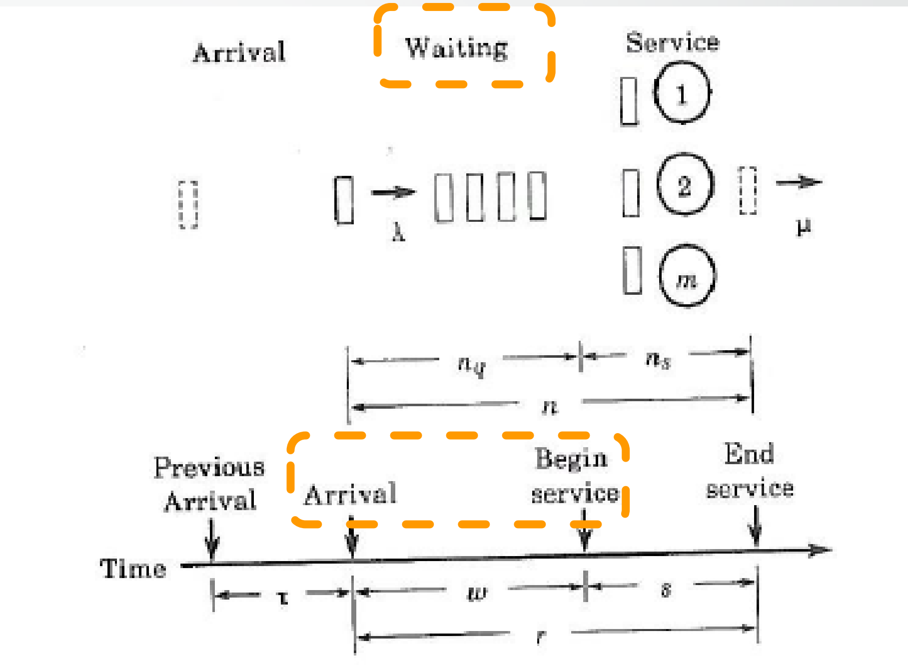
REGRAS VÁLIDAS PARA TODAS AS FILAS

- τ = **intervalo entre duas chegadas sucessivas**.
- λ = **taxa média de chegadas** = $\mathbf{1/E[\tau]}$
  - Em alguns sistemas, esta taxa pode ser uma função do estado do sistema. Por exemplo, ela pode depender do número de usuários que já se encontram no sistema.
- s = **tempo de serviço por usuário**.
- μ = **taxa média de serviço por servidor** = $\mathbf{1/E[s]}$
  - A **taxa total de serviço** para **m servidores** é **mμ**



REGRAS VÁLIDAS PARA TODAS AS FILAS

- n = **número de usuários no sistema**.
  - também chamado de **comprimento da fila**. Note que inclui os usuários que já **estão sendo atendidos**, assim como os que **estão esperando na fila**.
- n_q = **número de usuários esperando** para serem atendidos.
  - este número é **sempre menor do que** n , dado que não inclui os usuários que estão sendo atendidos.
- n_s = **número de usuários em atendimento**.
- r = **tempo de resposta** ou tempo no sistema.
  - Inclui tanto o tempo de espera como o tempo em atendimento.
- w = **tempo de espera**
  - intervalo de tempo entre o instante de chegada e o instante em que iniciou a ser atendido.
- **Todas estas variáveis** (à exceção de λ e μ) são **variáveis aleatórias**.



RELACIONAMENTO ENTRE AS VARIÁVEIS

- **Condição de estabilidade:**
  - A taxa média de chegadas deve ser menor do que a taxa média de atendimento:

    $\lambda < m\mu$
  - **Válida apenas para filas infinitas.** Com filas finitas o sistema nunca é instável.
- **Número no Sistema *versus* Número na Fila:**

  $n = n_q + n_s$

  $E[n] = E[n_q] + E[n_s]$

  Se a taxa de atendimento em cada servidor for **independente do número de usuários na fila**, temos:

  $Cov(n_q, n_s) = 0$ e $Var[n] = Var[n_q] + Var[n_s]$

- **Número *versus* Tempo:** Se os usuários não forem perdidos por causa de *buffers* insuficientes, temos que:
  - **Número médio de usuários no sistema** = taxa de chegada x tempo médio de resposta
  - **Número médio de usuários na fila** = taxa de chegada x tempo médio de espera
  - Lei de Little
- **Tempo no Sistema *versus* Tempo na Fila:**

  $r = w + s$; $E[r] = E[w] + E[s]$

  Se a taxa de atendimento em cada servidor for **independente do número de usuários na fila**, temos:

  $Cov(w, s) = 0$ e $Var[r] = Var[w] + Var[s]$

LEI DE LITTLE

- Suponha que um sistema é monitorado durante um intervalo de tempo ***T***, e registra-se **o tempo de chegada** e o **tempo de partida** de cada *job* individual (Fig. 30.4).
  - Se ***T*** é grande, o **número de chegadas** será aproximadamente igual ao número de partidas = ***N***

taxa de chegadas = total de chegadas/tempo total = **N/T**

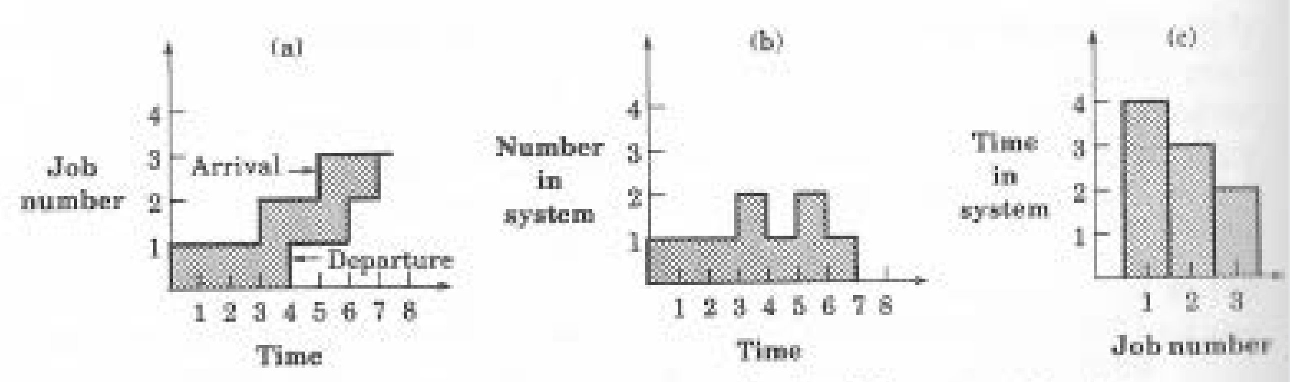


FIGURE 30.4 Three ways to plot the arrival and departure time data.

- ***J*** é a área hachuriada em cada uma das três figuras, representando o **tempo total de todos os *jobs*** dentro do sistema
- Da Fig. 30.4c:

  tempo médio de um *job* no sistema = J/N
- Da Fig. 30.4b:

  número médio de *jobs* no sistema = **J/T**

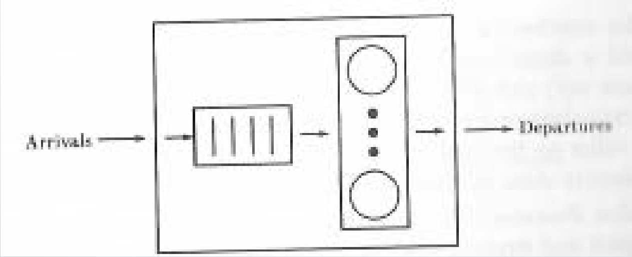
  $= \frac{N}{T} \times \frac{J}{N}$

  = **taxa de chegadas x tempo médio gasto pelo *job* no sistema**

LEI DE LITTLE

Número médio de usuários no sistema = taxa de chegada x tempo médio de resposta

- Provada pela primeira vez por Little (1961), aplica-se a **qualquer sistema ou parte do mesmo** no qual o **número de usuários que entram no sistema** é igual ao **número de usuários que terminam o atendimento**.



- **Tamanho médio da fila** = **taxa de chegadas x tempo médio de espera**
- Número médio em atendimento = taxa de chegadas x tempo médio de serviço

- EXEMPLO 30.3: O **tempo médio de atendimento de pedidos de I/O** em um servidor de disco é **100 ms**. A **taxa de I/O** é cerca de **100 pedidos por segundo**. Qual é o **número médio de pedidos** no servidor de disco?
  - Usando a lei de Little,

  Número médio no servidor de disco = taxa de chegadas x tempo de resposta

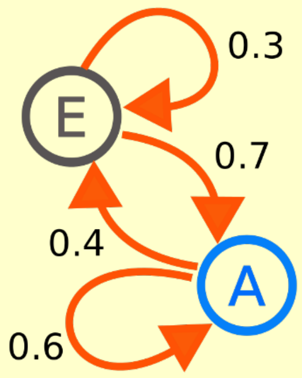
  = (100 pedidos/s) x 0,1 segundo

  = **10 pedidos**

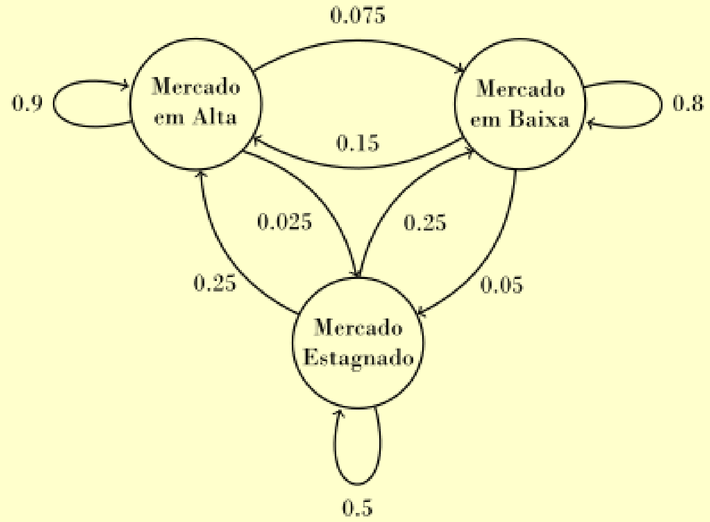
PROCESSOS ESTOCÁSTICOS

- **Processos estocásticos**
  - funções aleatórias **dependentes do tempo**
  - úteis para representar o **estado de sistemas de filas**
- Exemplos:
  - *n(t)* = número de *jobs* na CPU de um sistema computacional.

- **Processos de Estado Discreto e de Estado Contínuo**: dependendo dos valores que as suas variáveis de estados podem assumir, um processo estocástico de **estado discreto** é também chamado de **cadeia estocástica**.
- **Processos de Markov**: se os **estados futuros do sistema forem independentes do passado** e dependerem exclusivamente do estado atual. Um processo de Markov de **estado discreto** é também chamado de **cadeia de Markov.**



Cadeia de Markov simples com 2 estados

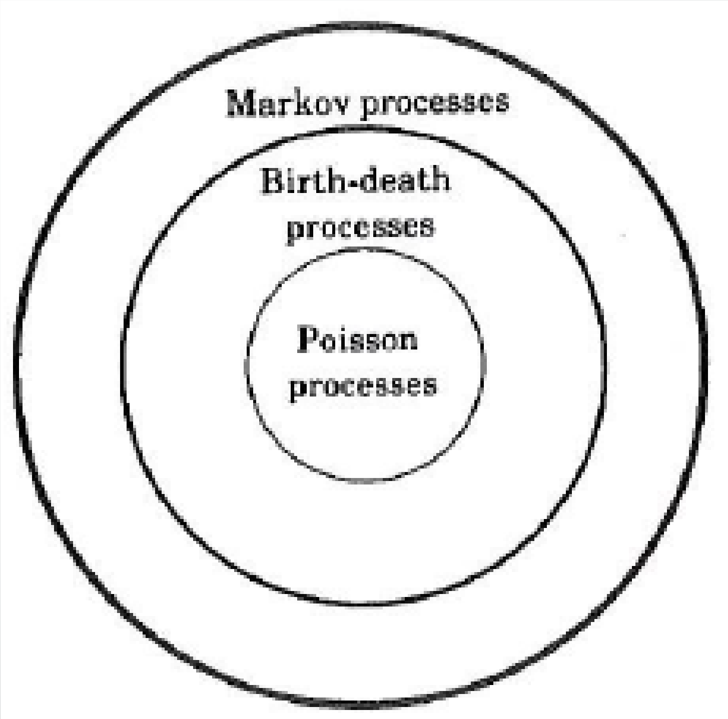


Cadeia de Markov com 3 estados

Fonte: pt.wikipedia.org

PROCESSOS DE MARKOV

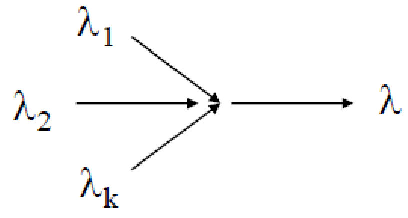
- **Processos de Markov**: se os **estados futuros do sistema forem independentes do passado** e dependerem exclusivamente do estado atual. Um processo de Markov de **estado discreto** é também chamado de **cadeia de Markov**.
- **Processos de Nascimento e Morte**: **processos de Markov** de espaço discreto no qual **as transições entre estados** estão **restritas a estados vizinhos**.
- **Processos de Poisson**: para **intervalos entre chegadas IID** com **distribuição exponencial**.



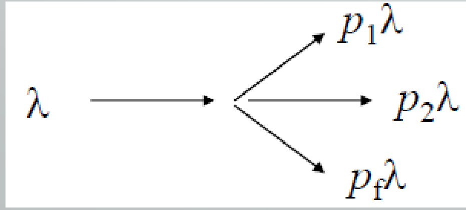
PROPRIEDADES DOS FLUXOS DE POISSON

- **Superposição** de fluxos de Poisson:

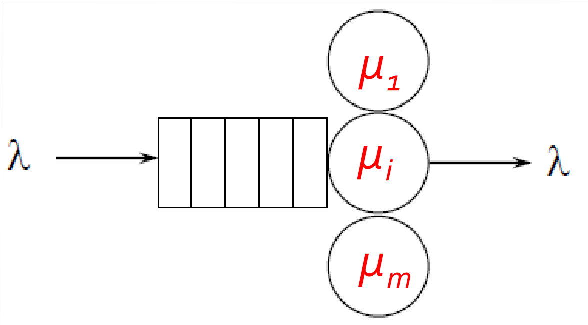
$$\lambda = \sum_{i=1}^{k} \lambda_i$$



- **Divisão**: se um fluxo de Poisson for dividido em f subfluxos com probabilidade p_i de um usuário seguir o subfluxo i, então, cada subfluxo é também um fluxo de Poisson com **taxa média $p_i\lambda$**.



- Se as **chegadas a uma fila** com ***m* servidores** e tempos de serviço exponenciais forem **Poisson com taxa média λ**, então as **partidas também são Poisson**, com **mesma média λ** (desde que $\lambda < \Sigma\mu_i$).



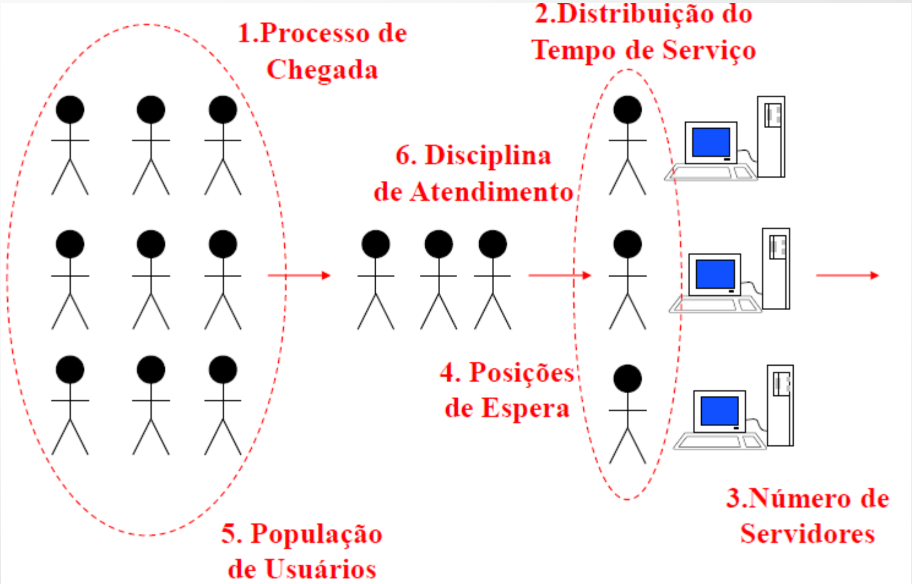
ANÁLISE DE UMA FILA ÚNICA

Capítulo 31

Raj Jain - The Art of Computer Systems Performance Analysis. Wiley, 1991.

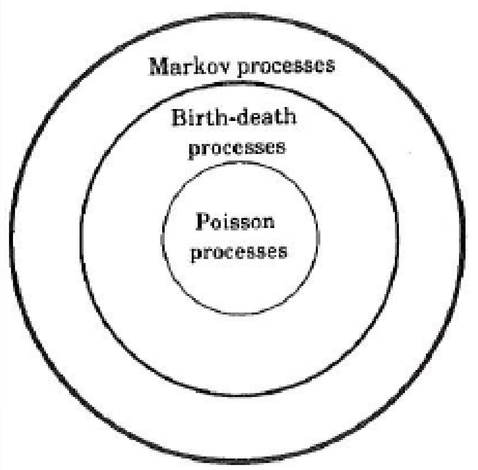
FILA ÚNICA

- O **modelo** de filas **mais simples** contém **apenas uma fila**!
  - Pode ser usado para analisar **recursos individuais em sistemas** de computação.
- **Muitas filas** podem ser **modeladas** como **processos estocásticos de nascimento e morte**.



PROCESSOS ESTOCÁSTICOS

- **Processos estocásticos**
  - funções aleatórias **dependentes do tempo**
  - úteis para representar o **estado de sistemas de filas**

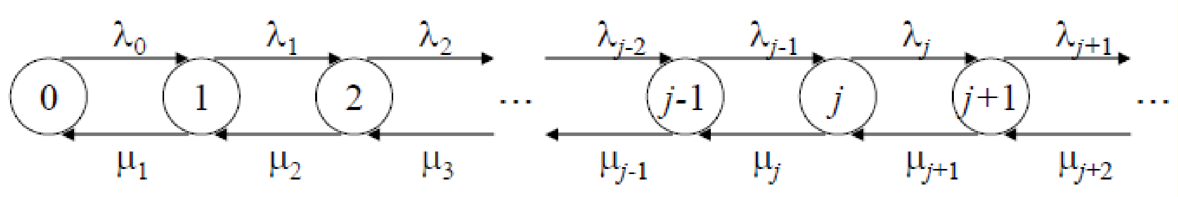


- **Processos de Markov**: se os **estados futuros do sistema forem independentes do passado** e dependerem exclusivamente do estado atual. Um processo de Markov de **estado discreto** é também chamado de **cadeia de Markov**.
  - **Processos de Nascimento e Morte**: **processos de Markov** de espaço discreto no qual **as transições entre estados** estão **restritas a estados vizinhos**.
    - **Processos de Poisson**: para **intervalos entre chegadas IID** com **distribuição exponencial**.

PROCESSOS DE NASCIMENTO E MORTE

- Um **processo de nascimento e morte** é útil para **modelar sistemas** nos quais os ***jobs*** chegam **um de cada vez** (e não em lotes).
  - P. ex.: *job* = pacote chegando individualmente para ser processado por um roteador.
- O **estado do sistema** pode ser representado pelo **número de jobs *n*** no sistema.
- A **chegada de um novo *job*** (**nascimento**) leva para o **estado** para ***n* +1**
- A **partida de um *job*** (**morte**) leva o **estado** para ***n* -1**

DIAGRAMA DE ESTADOS



- Taxa de **novas chegadas**: λ_n
- Taxa de **serviço/atendimento**: μ_n
- Assumimos que tanto os intervalos entre chegadas como os tempos de serviço são **exponencialmente distribuídos** (processo de **Poisson**).
- **Probabilidade de estar em um estado ?**

PROBABILIDADE DOS ESTADOS

- (Teorema 31.1) A **probabilidade em** **regime permanente** p_n que um **processo de nascimento e morte** esteja no **estado** n é dada por:

$$p_n = \frac{\lambda_0 \lambda_1 \lambda_2 \ldots \lambda_{n-1}}{\mu_1 \mu_2 \mu_3 \ldots \mu_n} p_0 \;, \quad n = 1,2,\ldots,\infty$$

$$p_n = p_0 \prod_{j=0}^{n-1} \frac{}{}$$

onde p_0 é a probabilidade de que o sistema se encontre no **estado 0** (vazio).

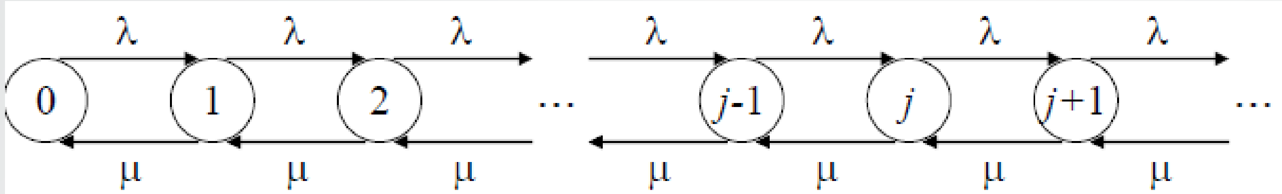
- Usando a condição adicional de que a **soma de todas as probabilidades deve ser 1**, obtemos:

$$\boldsymbol{p_0} = \frac{\mathbf{1}}{\mathbf{1} + \sum_{n=1}^{\infty} \prod_{j=0}^{n-1} [\lambda_j / \mu_{j+1}]}$$

A partir destas probabilidades podemos calcular muitas outras medidas de desempenho.

[Exemplo com Fila M/M/1]

FILA *M/M/1*



- Modelada como um **processo de nascimento e morte** onde:

$$\lambda_n = \lambda \text{ , } n = 0,1,2\ldots,\infty$$

$$\mu_n = \mu \text{ , } n = 1,2\ldots,\infty$$

- Utilizando o Teorema 31.1, obtemos:

$$p_n = \left(\frac{\lambda}{\mu}\right)^n p_0 \text{ , } n = 1,2\ldots,\infty$$

$$\frac{\lambda}{\mu} = \rho = \textbf{intensidade de tráfego}$$

- Portanto:

$$p_n = \rho^n p_0 \quad \text{, } n = 1,2\ldots,\infty$$

- Como a soma das probabilidades deve dar 1:

$$p_0 = \frac{1}{1+\rho+\rho^2+\cdots+\rho^\infty} = \mathbf{1\text{-}\rho}$$

- Substituindo na expressão para p_n :

$$p_n = (1-\rho)\rho^n \text{ , } n = 0,1,2,\ldots,\infty$$

- Note que p_n é **geometricamente distribuída**.

OUTRAS PROPRIEDADES DA FILA *M/M/1*

- **Utilização do servidor** (probabilidade de ter **um ou mais** *jobs* no sistema):

  $U = 1 - p_0 = \rho$

  Para M/M/1 ser **estável** $\rho < 1$

- **Número médio** de *jobs* no sistema:

  $E[n] = \sum_{n=1}^{\infty} n p_n = \sum_{n=1}^{\infty} n(1-\rho)\rho^n = \frac{\rho}{1-\rho}$

- **Variância** do número de *jobs* no sistema:

  $Var[n] = E[n^2] - (E[n])^2 = \left(\sum_{n=1}^{\infty} n^2(1-\rho)\rho^n\right) - E[n]^2 = \frac{\rho}{(1-\rho)^2}$

- **Probabilidade de ter *n* ou mais *jobs* no sistema**:

  $P(\geq n \text{ jobs no sistema}) = \sum_{j=n}^{\infty} p_j = \sum_{j=n}^{\infty} (1-\rho)\rho^j = \rho^n$

- O **Tempo médio de resposta** pode ser calculado usando a **Lei de Little**:

  $E[n] = \lambda E[r]$

  $E[r] = \frac{E[n]}{\lambda} = \left(\frac{\rho}{1-\rho}\right)\frac{1}{\lambda} = \frac{1/\mu}{1-\rho}$

OUTRAS PROPRIEDADES DA FILA *M/M/1*

- A função distribuição cumulativa (**CDF**) do **tempo de resposta** *r* é dada por:

$$F(r) = 1 - e^{-r\mu(1-\rho)}$$

  - Note que o **tempo de resposta** é distribuído **exponencialmente**.

- **Posto percentil *q*** :

$$1 - e^{-rq\mu(1-\rho)} = \frac{q}{100}$$

$$r_q = \frac{1}{\mu(1-\rho)} \ln\left(\frac{100}{100-q}\right)$$

- Do mesmo modo, pode ser mostrado que a **CDF** do **tempo de espera** *w* é dada por:

$$F(w) = 1 - \rho e^{-w\mu(1-\rho)}$$

  - Esta é uma distribuição **exponencial truncada**.

- O seu **Posto percentil *q*** é dado por:

$$w_q = \frac{1}{\mu(1-\rho)} \ln\left(\frac{100\rho}{100-q}\right)$$

Esta fórmula se aplica **apenas se *q* for maior do que 100(1-ρ).** Todos os postos percentis mais baixos são 0.

OUTRAS PROPRIEDADES DA FILA *M/M/1*

- **Número médio** de *jobs* na **fila**:

  $E[n_q] = \sum_{n=1}^{\infty}(n-1)p_n = \sum_{n=1}^{\infty}(n-1)(1-\rho)\rho^n = \frac{\rho^2}{1-\rho}$

- **Número médio** de *jobs* em **serviço**:

  $E[n_s] = E[n] - E[n_q] = \rho$

- O servidor é dito **ocioso** quando não houver nenhum *job* no sistema; em todos os demais momentos ele é dito **ocupado**.
- O intervalo de tempo entre dois intervalos ociosos sucessivos é denominado de **período ocupado**.

- Probabilidade de ter *k jobs* **na fila**:

$$P(n_q = k) = \begin{cases} 1-\rho^2, & k=0 \\ (1-\rho)\,\rho^{k+1}, & k>0 \end{cases}$$

- Probabilidade de ter *n jobs* **em serviço** em um período ocupado:

$$P(n_s = n) = \frac{1}{n}\binom{2n-2}{n-1}\frac{\rho^{n-1}}{(1+\rho)^{2n-1}}$$

M/M/1

1. Parameters:
   λ = arrival rate in jobs per unit time
   μ = service rate in jobs per unit time
2. Traffic intensity: $\rho = \lambda/\mu$
3. Stability condition: Traffic intensity ρ must be less than 1.
4. Probability of zero jobs in the system: $p_0 = 1-\rho$
5. Probability of n jobs in the system: $p_n = (1-\rho)\rho^n,\ n = 0,1,\ldots,\infty$
6. Mean number of jobs in the system: $E[n] = \rho/(1-\rho)$
7. Variance of number of jobs in the system: $\mathrm{Var}[n] = \rho/(1-\rho)^2$
8. Probability of k jobs in the queue:

$$P(n_q = k) = \begin{cases} 1-\rho^2, & k = 0 \\ (1-\rho)\rho^{k+1}, & k > 0 \end{cases}$$

9. Mean number of jobs in the queue: $E[n_q] = \rho^2/(1-\rho)$
10. Variance of number of jobs in the queue:
    $\mathrm{Var}[n_q] = \rho^2(1+\rho-\rho^2)/(1-\rho)^2$
11. Cumulative distribution function of the response time:
    $F(r) = 1-e^{-r\mu(1-\rho)}$
12. Mean response time: $E[r] = (1/\mu)/(1-\rho)$
13. Variance of the response time: $\mathrm{Var}[r] = \dfrac{1/\mu^2}{(1-\rho)^2}$
14. q-Percentile of the response time: $E[r]\ln[100/(100-q)]$
15. 90-Percentile of the response time: $2.3E[r]$
16. Cumulative distribution function of waiting time:
    $F(w) = 1-\rho e^{-\mu w(1-\rho)}$
17. Mean waiting time: $E[w] = \rho\dfrac{1/\mu}{1-\rho}$
18. Variance of the waiting time: $\mathrm{Var}[w] = (2-\rho)\rho/[\mu^2(1-\rho)^2]$
19. q-Percentile of the waiting time: $\max\left(0, \dfrac{E[w]}{\rho}\ln[100\rho/(100-q)]\right)$
20. 90-Percentile of the waiting time: $\max\left(0, \dfrac{E[w]}{\rho}\ln[10\rho]\right)$
21. Probability of finding n *or more* jobs in the system: ρ^n
22. Probability of serving n jobs in one busy period:

$$\frac{1}{n}\binom{2n-2}{n-1}\frac{\rho^{n-1}}{(1+\rho)^{2n-1}}$$

23. Mean number of jobs served in one busy period: $1/(1-\rho)$
24. Variance of number of jobs served in one busy period:
    $\rho(1+\rho)/(1-\rho)^3$
25. Mean busy period duration: $1/[\mu(1-\rho)]$
26. Variance of the busy period: $1/[\mu^2(1-\rho)^3]-1/[\mu^2(1-\rho)^2]$

EXEMPLO 31.1

- Medições efetuadas em um **roteador** indicaram que os pacotes chegam a uma **taxa média de 125 pacotes por segundo (pps)** e o roteador leva aproximadamente **2 milisegundos para encaminhá-los**.
- Analise o roteador usando um **modelo M/M/1**.
  1. Qual é a **utilização** do roteador?
  2. Qual é o **número médio de pacotes** no roteador?
  3. Qual é **tempo médio gasto por pacote** no roteador?
  4. Qual é a **probabilidade de estouro do buffer** se o roteador tiver apenas **13 buffers**?
  5. **Quantos buffers** seriam necessários para mantermos as **perdas** de pacotes **abaixo de um pacote por milhão**?

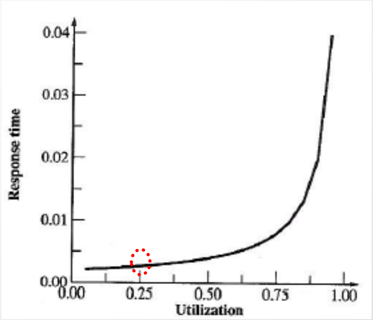
EXEMPLO 31.1

- Medições efetuadas em um **roteador** indicaram que os pacotes chegam a uma **taxa média** de **125 pacotes por segundo (pps)** e o roteador leva aproximadamente **2 milisegundos para encaminhá-los**. Analise o roteador usando um **modelo M/M/1**.

1. Qual é a **utilização** do roteador?
2. Qual é o **número médio de pacotes** no roteador?
3. Qual é **tempo médio gasto por pacote** no roteador?
4. Qual é a **probabilidade de estouro do buffer** se o roteador tiver apenas **13 buffers**?
5. **Quantos buffers** seriam necessários para mantermos as **perdas** de pacotes **abaixo de um pacote por milhão**?

- **Taxa de chegada** λ = **125 pps**
- **Taxa de serviço** μ = (1/ 0,002) = **500 pps**
- **Utilização do roteador**

  $\rho = \lambda \,/\, \mu = \mathbf{0{,}25}$



- **Probabilidade** de termos "*n*" **pacotes** no roteador =

  $(1-\rho)\,\rho^{n} = 0{,}75\,(0{,}25)^{n}$

- **Número médio de pacotes** no roteador = $\frac{\rho}{1-\rho}$ =

  (0,25/0,75) = **0,33**

- **Tempo médio gasto** no roteador = $\frac{1/\mu}{1-\rho} = \frac{1/500}{1-0{,}25}$ =

  2,66 mseg

EXEMPLO 31.1

- Medições efetuadas em um **roteador** indicaram que os pacotes chegam a uma **taxa média** de **125 pacotes por segundo (pps)** e o roteador leva aproximadamente **2 milisegundos para encaminhá-los**. Analise o roteador usando um **modelo M/M/1**.

1. Qual é a **utilização** do roteador?
2. Qual é o **número médio de pacotes** no roteador?
3. Qual é **tempo médio gasto por pacote** no roteador?
4. Qual é a **probabilidade de estouro do buffer** se o roteador tiver apenas **13 buffers**?
5. **Quantos buffers** seriam necessários para mantermos as **perdas** de pacotes **abaixo de um pacote por milhão**?

- **Probabilidade de estouro do *buffer***

  = P (**13 ou mais pacotes** no roteador)

  = ρ^{13} = $(0{,}25)^{13}$ = **$1{,}49 \times 10^{-8}$**

  $\approx$ 15 pacotes por bilhão de pacotes

- Para limitar a **probabilidade de perda** menor do que **10^{-6}** ,

  $\rho^n \leq 10^{-6}$

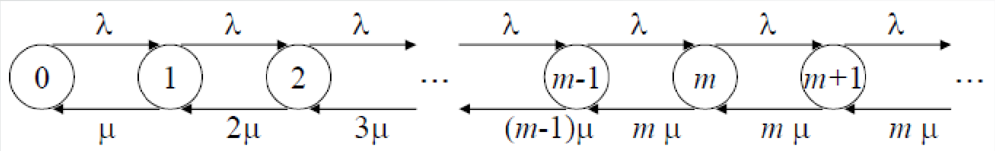
  $n > \log(10^{-6})/\log(0{,}25) = 9{,}96$

  ou seja, precisamos de cerca de **10 buffers**.

- Estes **resultados** são **aproximados**, pois teríamos que modelar o roteador como uma fila M/M/1/B.

FILA *M/M/m*

- ***m* servidores idênticos**, mas quando todos eles estão ocupados os *jobs* são **mantidos** numa **única fila**.



- Modelada como um **processo de nascimento e morte** onde:

$$\lambda_n = \lambda \quad , \; n = 0,1,2,\ldots,\infty$$

$$\mu_n = \begin{cases} n\mu \,, & n = 1,2,\ldots,m\text{-}1 \\ m\mu \,, & n = m, m+1,\ldots,\infty \end{cases}$$

- Utilizando o Teorema 31.1, obtemos a **probabilidade** do sistema estar no **estado *n*** :

$$p_n = \begin{cases} \dfrac{\lambda^n}{n!\,\mu^n}\, p_0, & n = 1,2,\ldots,m-1 \\ \dfrac{\lambda^n}{m!\,m^{n-m}\,\mu^n}\, p_0, & n = m, m+1,\ldots,\infty \end{cases}$$

- Em termos da **intensidade de tráfego** $\rho = \lambda / m\mu$:

$$p_n = \begin{cases} \dfrac{(m\rho)^n}{n!}\, p_0, & n = 1,2,\ldots,m-1 \\ \dfrac{\rho^n m^m}{m!}\, p_0, & n = m, m+1,\ldots,\infty \end{cases}$$

FILA *M/M/m*

- A **probabilidade** de ter **0 *jobs*** no sistema é calculada pela relação:

$$\sum_{0}^{\infty} p_n = 1$$

Isto dá:

$$p_0 + p_0 \sum_{n=1}^{m-1} \frac{(m\rho)^n}{n!} + p_0 \frac{(m\rho)^m}{m!} \sum_{n=m}^{\infty} \rho^{n-m} = 1$$

- Ou

$$\boldsymbol{p_0} = \left[1 + \frac{(m\rho)^m}{m!\,(1-\rho)} + \sum_{n=1}^{m-1} \frac{(m\rho)^n}{n!}\right]^{-1}$$

FÓRMULA C DE ERLANG

- A **probabilidade** de que um *job* tenha que **esperar na fila** é dada por:

$$C(m,\rho) = P\,(\geq m \text{ jobs}) = p_m + p_{m+1} + p_{m+2} + \dots$$

$$p_0 + p_0 \sum_{n=1}^{m-1} \frac{(m\rho)^n}{n!} + p_0 \frac{(m\rho)^m}{m!} \sum_{n=m}^{\infty} \rho^{n-m} = 1$$

$$C(m,\rho) = p_0 \frac{(m\rho)^m}{m!} \sum_{n=m}^{\infty} \rho^{n-m} = \boldsymbol{p_0 \frac{(m\rho)^m}{m!(1-\rho)}}$$

- O **número médio** de ***jobs* em espera**

$$E\,[n_q] = \sum_{n=m+1}^{\infty} (n-m) p_n$$

$$= p_0 \frac{(m\rho)^m}{m!} \sum_{n=m+1}^{\infty} (n-m) \rho^{n-m}$$

$$E[n_q] = p_0 \frac{(m\rho)^m \rho}{m!(1-\rho)^2} = \boldsymbol{\frac{\rho C(m,\rho)}{1-\rho}}$$

FILA *M/M/m*

- **Número médio** de ***jobs* em serviço**

  $E[n_s] = \sum_{n=1}^{m-1} np_n + \sum_{n=m}^{\infty} mp_n$

$$
\begin{aligned}
&= 1p_0\frac{(m\rho)}{1!} + 2p_0\frac{(m\rho)^2}{2!} + \cdots + (m-1)p_0\frac{(m\rho)^{m-1}}{(m-1)!} \\
&+ m(p_m + p_{m+1} + p_{m+2} + \cdots +) \\
&= m\rho\left(p_0 + p_0\frac{(m\rho)}{1!} + p_0\frac{(m\rho)^2}{2!} + \cdots + p_0\frac{(m\rho)^{m-2}}{(m-2)!}\right) + mC(m,\rho) \\
&= m\rho(p_0 + p_1 + p_2 + \cdots + p_{m-2}) + mC(m,\rho) \\
&= m\rho[1 - p_{m-1} - C(m,\rho)] + mC(m,\rho) \\
&= m\rho - m\rho p_{m-1} + mC(m,\rho)(1-\rho)
\end{aligned}
$$

- $E[n_s] = m\rho$ dado que

  $mC(m, \rho)(1-\rho) = mp_m = m\rho p_{m-1}$

- **Número esperado** de ***jobs* no sistema**:

$$E[n] = E[n_q] + E[n_s]$$

$$= \frac{\rho C(m,\rho)}{1-\rho} + m\rho$$

- Pode-se mostrar que a **variância de n** e n_q são dadas por:

$$\mathrm{Var}[n] = m\rho + \rho C(m,\rho)\left[\frac{1+\rho-\rho C(m,\rho)}{(1-\rho)^2} + m\right]$$

$$\mathrm{Var}[n_q] = \frac{C(m,\rho)\rho[1+\rho-\rho C(m,\rho)]}{(1-\rho)^2}$$

UTILIZAÇÃO E TEMPOS MÉDIOS DA FILA M/M/m

- Se observarmos o sistema por um **tempo longo**, por exemplo ***T* segundos**,
  - o **número total de** ***jobs*** que chegarão e receberão serviço será λT
  - o **tempo total ocupado** dos ***m*** **servidores** para atender estes *jobs* será $\lambda T/\mu$
  - A **utilização** de **cada servidor** será:

$$\boldsymbol{U} = \frac{tempo\ ocupado\ por\ servidor}{tempo\ total}$$

$$= \frac{(\lambda T/\mu)/m}{T} = \frac{\lambda}{m\mu} = \rho$$

- **Tempo médio de resposta**, usando a **Lei de Little**:

$$E[r] = \frac{E[n]}{\lambda} = \frac{1}{\mu} + \frac{C(m,\rho)/m\mu}{1-\rho} = \frac{1}{\mu}\left(1 + \frac{C(m,\rho)}{m(1-\rho)}\right)$$

- Do mesmo modo, o **tempo médio de espera** é dado por:

$$E[w] = \frac{E[n_q]}{\lambda} = \frac{C(m,\rho)}{m\mu(1-\rho)}$$

FDPs DOS TEMPOS NO SISTEMA

- **TEMPO DE RESPOSTA**

$$F[r] = \begin{cases} 1 - e^{-\mu r} - \dfrac{C(m,\rho)}{1 - m + m\rho} e^{-m\mu(1-\rho)r} - e^{-\mu r}, & \rho \neq (m-1)/m \\ 1 - e^{-\mu r} - C(m,\rho)\mu r e^{-\mu r}, & \rho = (m-1)/m \end{cases}$$

- O tempo de resposta ***r*** **não é exponencialmente distribuído** a menos que *m*=1.
- Em geral, o **coeficiente de variação** (C.O.V), isto é, a fração entre o desvio padrão e a média, de ***r*** é menor do que 1.

- **TEMPO DE ESPERA**
- A **função distribuição de probabilidade do tempo de espera** é dada por:

$$F(w) = 1 - C(m,\rho)e^{-m\mu(1-\rho)w}$$

- Dado que ***w*** possui uma função **distribuição exponencial truncada**, o **posto percentil *q*** pode ser calculado do seguinte modo:

$$w_q = \max\left\{0, \frac{1}{m\mu(1-\rho)} \ln\left(\frac{100C(m,\rho)}{100-q}\right)\right\}$$

obs.: esta fórmula se aplica apenas se **q** for maior do que **100(1-ρ).** Todos os postos percentis mais baixos são 0.

$M/M/m$

1. Parameters:
   λ = arrival rate in jobs per unit time
   μ = service rate in jobs per unit time
   m = number of servers
2. Traffic intensity: $\rho = \lambda/(m\mu)$
3. The system is stable if the traffic intensity ρ is less than 1.
4. Probability of zero jobs in the system:

$$p_0 = \left[1 + \frac{(m\rho)^m}{m!(1-\rho)} + \sum_{n=1}^{m-1} \frac{(m\rho)^n}{n!}\right]^{-1}$$

5. Probability of n jobs in the system:

$$p_n = \begin{cases} p_0 \dfrac{(m\rho)^n}{n!}, & n < m \\ p_0 \dfrac{\rho^n m^m}{m!}, & n \geq m \end{cases}$$

6. Probability of queueing:

$C(m,\rho) \Rightarrow$ $$\varrho = P(\geq m \text{ jobs}) = \frac{(m\rho)^m}{m!(1-\rho)} p_0$$

   In the remaining formulas below we will use ϱ as defined here.
7. Mean number of jobs in the system: $E[n] = m\rho + \rho\varrho/(1-\rho)$
8. Variance of number of jobs in the system:

$$\text{Var}[n] = m\rho + \rho\varrho\left[\frac{1+\rho-\rho\varrho}{(1-\rho)^2} + m\right]$$

9. Mean number of jobs in the queue: $E[n_q] = \rho\varrho/(1-\rho)$
10. Variance of number of jobs in the queue:
    $\text{Var}[n_q] = \varrho\rho(1+\rho-\varrho\rho)/(1-\rho)^2$
11. Average utilization of each server: $U = \lambda/(m\mu) = \rho$
12. Cumulative distribution function of response time:

$$F(r) = \begin{cases} 1 - e^{-\mu r} - \dfrac{\varrho}{1-m+m\rho} e^{-m\mu(1-\rho)r} - e^{-\mu r}, & \rho \neq (m-1)/m \\ 1 - e^{-\mu r} - \varrho\mu r e^{-\mu r}, & \rho = (m-1)/m \end{cases} \qquad r > 0$$

13. Mean response time:

$$E[r] = \frac{1}{\mu}\left(1 + \frac{\varrho}{m(1-\rho)}\right)$$

14. Variance of the response time:

$$\text{Var}[r] = \frac{1}{\mu^2}\left[1 + \frac{\varrho(2-\varrho)}{m^2(1-\rho)^2}\right]$$

15. Cumulative distribution function of waiting time:
    $F(w) = 1 - \varrho e^{-m\mu(1-\rho)w}$
16. Mean waiting time: $E[w] = E[n_q]/\lambda = \varrho/[m\mu(1-\rho)]$
17. Variance of the waiting time: $\text{Var}[w] = \varrho(2-\varrho)/[m^2\mu^2(1-\rho)^2]$
18. q-Percentile of the waiting time: $\max\left(0, \dfrac{E[w]}{\varrho}\ln\dfrac{100\varrho}{100-q}\right)$.
19. 90-Percentile of the waiting time: $\dfrac{E[w]}{\varrho}\ln(10\varrho)$

Once again, ϱ in these formulas is the probability of m or more jobs in the system: $\varrho = [(m\rho)^m/\{m!(1-\rho)\}]p_0$. For $m = 1$, ϱ is equal to ρ and all of the formulas become identical to those for M/M/1 queues.

EXEMPLO 31.2

- **Alunos chegam** no laboratório num **processo de Poisson** com taxa média igual a **10 por hora**.
- Cada aluno **usa um computador** em **média por 20 minutos** segundo uma **distribuição exponencial**.
- O CPD tem **5 computadores**.
- Alguns alunos reclamaram que o **tempo de espera** é **muito longo!**
- Analise usando um **modelo de filas** **M/M/5** e calcule:
  - **tempo de espera** na **fila**
  - **posto percentil 90%** do **tempo de espera**

EXEMPLO 31.2

- O laboratório pode ser modelado por um **sistema de filas M/M/5** com:
  - Taxa de chegada λ = 1/6 por minuto
  - Taxa de serviço μ = 1/20 por minuto
  - Intensidade de tráfego $\rho = \frac{\lambda}{m\mu} = \frac{0,167}{5x0,05} = 0,67$
- A **probabilidade** de que **todos os computadores** estejam **ociosos** é

$$p_0 = \left[1 + \frac{(5\times 0,67)^5}{5!(1-0,67)} + \frac{(5\times 0,67)^1}{1!} + \frac{(5\times 0,67)^2}{2!} + \frac{(5\times 0,67)^3}{3!} + \frac{(5\times 0,67)^4}{4!}\right]^{-1} = 0,0318$$

- A **probabilidade** de que **todos os computadores** estejam **ocupados** é

$$C(m,\rho) = \frac{(m\rho)^m}{m!(1-\rho)} p_0 = \frac{(5\times 0,67)^5}{5!(1-0,67)} \times 0,0318 = 0,33$$

- A **utilização média** dos computadores é ρ = 0,67
- O **número médio** de alunos **no laboratório** (no sistema) é

$$E[n] = m\rho + \frac{\rho C(m,\rho)}{1-\rho} = 5\times 0,67 + \frac{0,67\times 0,33}{1-0,67} = 4,0$$

- O **número médio** de alunos **esperando na fila** é

$$E[n_q] = \frac{\rho C(m,\rho)}{1-\rho} = \frac{0,67\times 0,33}{1-0,67} = 0,65$$

EXEMPLO 31.2

- O **número médio** de alunos **usando os computadores** (em serviço) é

$$E[n_s] = E[n] - E[n_q] = 4 - 0{,}65 = 3{,}35$$

- A **média** e a **variância** do **tempo gasto no laboratório** é

$$E[r] = \frac{1}{\mu}\left(1 + \frac{C(m,\rho)}{m(1-\rho)}\right) = \frac{1}{0{,}05}\left(1 + \frac{0{,}33}{5(1-0{,}67)}\right) = 24$$

$$\mathrm{Var}[r] = \frac{1}{\mu^2}\left(1 + \frac{C(m,\rho)(2 - C(m,\rho))}{m^2(1-\rho)^2}\right)$$

$$= \frac{1}{0{,}05^2}\left(1 + \frac{0{,}33(2-0{,}33)}{5^2(1-0{,}67)^2}\right) = 479$$

- Portanto, cada aluno gasta uma média de **24 minutos** no laboratório:
  - 20 minutos trabalhando no computador e
  - **4 minutos esperando na fila**.
- Podemos comprovar este tempo de espera através da fórmula:

$$E[w] = \frac{C(m,\rho)}{m\mu(1-\rho)} = \frac{0{,}33}{5 \times 0{,}05 \times (1 - 0{,}67)} = 4$$

- O **posto percentil 90 %** do **tempo de espera** é

$$\max\left\{0, \frac{E[w]}{C(m,\rho)} \ln(10C(m,\rho))\right\} = \max\left\{0, \frac{4}{0{,}33} \ln(10 \times 0{,}33)\right\} = 14$$

- Portanto, apenas **10 % dos alunos** têm que esperar **mais do que 14 minutos!**

EXEMPLO 31.3

- Os estudantes querem limitar o seu **tempo de espera** para uma **média de 2 minutos** e **menos do que 5 minutos em 90%** dos casos.
- **Isto é viável**?
- Se for, **quantos computadores seriam necessários**?
- Vamos analisar o sistema com m= 6,7,... computadores mantendo a mesma **taxa de chegada** e **taxa de atendimento** de λ **=0,167** e μ **=0,05**, respectivamente.

EXEMPLO 31.3

- Os estudantes querem limitar o seu **tempo de espera** para uma **média de 2 minutos** e **menos do que 5 minutos em 90%** dos casos.
- **Isto é viável**?
- Se for, **quantos computadores seriam necessários**?
- Vamos analisar o sistema com m= 6,7,... computadores mantendo a mesma **taxa de chegada** e **taxa de atendimento** de λ **=0,167** e μ**=0,05**, respectivamente.

- Com m **=6 computadores**, temos:

Intensidade de tráfego $\rho = \frac{\lambda}{m\mu} = \frac{0{,}167}{6x0{,}05} = 0{,}556$

Prob. de todos computadores ociosos p_o = 0,0346

Prob. de todos computadores ocupados C(m, ρ)=0,15

Tempo médio de espera $E[w]$ **= 1,1 minutos**

- O **posto percentil 90** do **tempo de espera** é

$$\max\left\{0, \frac{1{,}1}{0{,}15}\ln(10\times 0{,}15)\right\} = \max\{0;3{,}0\} = 3{,}0$$

- Portanto, **basta mais um computador** para atender à demanda dos alunos!

m x $M/M/1$ vs. $M/M/m$

- Com m **servidores**, devemos:
  - manter **filas separadas** para cada servidor ou
  - devemos manter **uma única fila** para todos os servidores?
- Para **chegadas** de **Poisson** e **tempos de serviço exponenciais**:
  - m **filas M/M/1** com **taxa de chegada** λ/m
  - **uma única fila M/M/m** com **taxa de chegada** λ

Mostre que **uma única fila é melhor**!

m x $M/M/1$ vs. $M/M/m$

- Com ***m* servidores**, devemos:
  - manter **filas separadas** para cada servidor ou
  - devemos manter **uma única fila** para todos os servidores?
- Para chegadas de Poisson e tempos de serviço exponenciais:
  - m filas M/M/1 com taxa de chegada λ/m
  - uma fila M/M/m com taxa de chegada λ
- **Uma única fila é melhor**, vejamos o exemplo.

- Considere o que teria acontecido se os **5 computadores** do Exemplo 31.2 estivessem localizados em **5 diferentes localidades do campus**, necessitando, portanto de uma **fila separada** para cada um:
- Usando m=1, λ =0,167/5=0,0333 e μ=0,05, temos:

Intensidade de tráfego $\rho = \frac{0,0333}{0,05} = 0,67$

$$E[r] = \frac{1/\mu}{1-\rho} = \frac{1/0,05}{1-0,067} = 60$$

$$\mathrm{Var}[r] = \frac{1/\mu^2}{(1-\rho)^2} = \frac{1/0,05^2}{(1-0,067)^2} = 3600$$

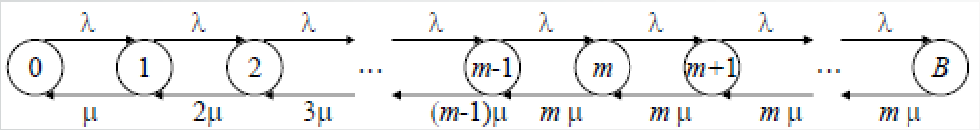
vs.

M/M/5
24

479

Fila M/M/m/B (*Buffers* finitos)

- **Número de *buffers B* (≥ m)** é **finito**.
- Quando todos os *buffers* estiverem ocupados novas chegadas são perdidas.



- Modelada como um **processo de nascimento e morte** onde:

$$\lambda_n = \lambda, \qquad n = 0,1,2,\ldots,B-1$$

$$\mu_n = \begin{cases} n\mu, & n = 1,2,\ldots,m-1 \\ m\mu, & n = m, m+1,\ldots,B \end{cases}$$

- Utilizando o Teorema 31.1, obtemos a **probabilidade** do sistema estar no **estado *n*** :

$$p_n = \begin{cases} \dfrac{\lambda^n}{n!\,\mu^n}\, p_0, & n = 1,2,\ldots,m-1 \\ \dfrac{\lambda^n}{m!\,m^{n-m}\,\mu^n}\, p_0, & n = m, m+1,\ldots,B \end{cases}$$

- Em termos da intensidade de tráfego $\rho = (\lambda / m\mu)$:

$$p_n = \begin{cases} \dfrac{(m\rho)^n}{n!}\, p_0, & n = 1,2,\ldots,m-1 \\ \dfrac{\rho^n m^m}{m!}\, p_0, & n = m, m+1,\ldots,B \end{cases}$$

Fila M/M/m/B (*Buffers* finitos)

- A **probabilidade de ter 0 *jobs*** no sistema é calculada pela relação:

$$\sum_{n=0}^{B} p_n = 1$$

- Isto dá:

$$p_0 + p_0 \sum_{n=1}^{m-1} \frac{(m\rho)^n}{n!} + p_0 \frac{(m\rho)^m}{m!} \sum_{n=m}^{B} \rho^{n-m} = 1$$

- Ou

$$p_0 = \left[1 + \frac{(1-\rho^{B-m+1})(m\rho)^m}{m!(1-\rho)} + \sum_{n=1}^{m-1} \frac{(m\rho)^n}{n!} \right]^{-1}$$

- **Números Médios** de *jobs*

$$E[n] = \sum_{n=1}^{B} n p_n$$

$$E[n_q] = \sum_{n=m+1}^{B} (n-m) p_n$$

- A **variância** e **outras estatísticas** de n e n_q podem ser calculados de forma semelhante.

Fila M/M/m/B (*Buffers* finitos)

- **TAXA EFETIVA DE CHEGADAS**
  - Todas as chegadas que acontecem enquanto o **sistema está cheio** *(n=B)* são perdidas.
  - A **taxa de** ***jobs*** que **efetivamente entram no sistema** é dada por:

$$\lambda' = \sum_{n=0}^{B-1} \lambda p_n = \lambda \sum_{n=0}^{B-1} p_n = \lambda(1 - p_B)$$

  - A diferença **λ-λ'= λ p_B** representa a **taxa de perda de pacotes**.

- **TEMPOS MÉDIOS**
  - **Tempo médio de resposta**, usando a Lei de Little:

$$E[r] = \frac{E[n]}{\lambda'} = \frac{E[n]}{\lambda(1 - p_B)}$$

  - Do mesmo modo, o **tempo médio de espera** é dado por:

$$E[w] = \frac{E[n_q]}{\lambda'} = \frac{E[n_q]}{\lambda(1 - p_B)}$$

Fila M/M/m/B (*Buffers* finitos)

- **UTILIZAÇÃO**
  - Se observarmos o sistema por um **tempo longo**, por exemplo ***T* segundos**,
  - o **número total de *jobs*** que chegarão e **receberão serviço** será $\lambda' T$.
  - O **tempo total ocupado** dos *m* **servidores** para atender estes *jobs* será $\lambda' T/\mu$
  - A **utilização** será:

$$\mathbf{U} = \frac{\text{tempo ocupado por servidor}}{\text{tempo total}}$$

$$= \frac{\left(\lambda' T/\mu\right)/m}{T} = \frac{\lambda'}{m\mu} = \rho\,(1-p_B)$$

- **FÓRMULA B (DE PERDAS) DE ERLANG**
  - A **probabilidade de perda** num sistema **M/M/m/m** onde o número de *buffers* é exatamente igual ao número de servidores (**B=m**) é dada por:

$$p_m = \frac{(m\rho)^m}{m!} p_0 = \frac{(m\rho)^m / m!}{\sum_{j=0}^{m} [(m\rho)^j / j!]}$$

M/M/m/B

1. Parameters:
   λ = arrival rate in jobs per unit time
   μ = service rate in jobs per unit time
   m = number of servers
   B = number of buffers, $B \geq m$
2. Traffic intensity: $\rho = \lambda/(m\mu)$
3. The system is always stable: $\rho < \infty$
4. Probability of zero jobs in the system:

$$p_0 = \left[1 + \frac{(1-\rho^{B-m+1})(m\rho)^m}{m!(1-\rho)} + \sum_{n=1}^{m-1} \frac{(m\rho)^n}{n!}\right]^{-1}$$

   For $m = 1$:

$$p_0 = \begin{cases} \dfrac{1-\rho}{1-\rho^{B+1}}, & \rho \neq 1 \\ \frac{1}{B+1}, & \rho = 1 \end{cases}$$

5. Probability of n jobs in the system:

$$p_n = \begin{cases} \dfrac{1}{n!}(m\rho)^n p_0, & 0 \leq n < m \\ \dfrac{m^m \rho^n}{m!} p_0, & m \leq n \leq B \end{cases}$$

6. Mean number of jobs in the system: $E[n] = \sum_{n=1}^{B} n p_n$
   For $m = 1$:

$$E[n] = \frac{\rho}{1-\rho} - \frac{(B+1)\rho^{B+1}}{1-\rho^{B+1}}$$

7. Mean number of jobs in the queue: $E[n_q] = \sum_{n=m+1}^{B}(n-m)p_n$
   For $m = 1$:

$$E[n_q] = \frac{\rho}{1-\rho} - \rho\frac{1+B\rho^B}{1-\rho^{B+1}}$$

8. Effective arrival rate in the system: $\lambda' = \sum_{n=0}^{B-1} \lambda p_n = \lambda(1-p_B)$
9. Average utilization of each server: $U = \lambda'/(m\mu) = \rho(1-p_B)$
10. Mean response time: $E[r] = E[n]/\lambda' = E[n]/[\lambda(1-p_B)]$
11. Mean waiting time: $E[w] = E[r] - 1/\mu = E[n_q]/[\lambda(1-p_B)]$
12. The loss rate is given by λp_B jobs per unit time.
13. For an M/M/m/m queue, the probability of a full system is given by

$$p_m = \frac{(m\rho)^m/m!}{\sum_{j=0}^{m} \frac{(m\rho)^j}{j!}}$$

M/M/1/B

1. Parameters:
   λ = arrival rate in jobs per unit time
   μ = service rate in jobs per unit time
   B = number of buffers
2. Traffic intensity: $\rho = \lambda/\mu$
3. The system is always stable: $\rho < \infty$
4. Probability of zero jobs in the system:

$$p_0 = \begin{cases} \frac{1-\rho}{1-\rho^{B+1}}, & \rho \neq 1 \\ \frac{1}{B+1}, & \rho = 1 \end{cases}$$

5. Probability of n jobs in the system:

$$p_n = \begin{cases} \frac{1-\rho}{1-\rho^{B+1}}\rho^n, & \rho \neq 1 \quad 0 \leq n \leq B \\ \frac{1}{B+1}, & \rho = 1 \\ 0 & n > B \end{cases}$$

6. Mean number of jobs in the system:

$$E[n] = \frac{\rho}{1-\rho} - \frac{(B+1)\rho^{B+1}}{1-\rho^{B+1}}$$

7. Mean number of jobs in the queue:

$$E[n_q] = \frac{\rho}{1-\rho} - \rho\frac{1+B\rho^B}{1-\rho^{B+1}}$$

8. Effective arrival rate in the system: $\lambda' = \sum_{n=0}^{B-1} \lambda p_n = \lambda(1-p_B)$
9. Mean response time: $E[r] = E[n_q]/\lambda' = E[n]/[\lambda(1-p_B)]$
10. Mean waiting time: $E[w] = E[r] - 1/\mu = E[n_q]/[\lambda(1-p_B)]$

EXEMPLO 31.5

- Considere de novo o **roteador do Exemplo 31.1**.
- Vamos analisar o roteador assumindo que possua **apenas 2 buffers**.
  - A **taxa média de chegadas** e a **taxa média de serviço**, são **125 pps** e **500 pps**, respectivamente.
- Neste caso **M/M/m/B** temos $\lambda = 125$, $\mu = 500$, $m = 1$ e $B = 2$.
- **Intensidade de tráfego** $\rho = \frac{\lambda}{m\mu} =$

$$\rho = \frac{125}{1x500} = 0{,}25$$

EXEMPLO 31.5

- **M/M/m/B** temos λ **=125**, μ **=500**, m**=1** e B **=2**.
- **Intensidade de tráfego** $\rho = \frac{\lambda}{m\mu} = \frac{125}{1x500} = 0{,}25$

- Para n=1,2,...,B os p_n são:

$$p_1 = \rho p_0 = 0{,}25\, p_0$$
$$p_2 = \rho^2 p_0 = 0{,}25^2\, p_0 = 0{,}0625\, p_0$$

- Depois, p_0 é determinado a partir da soma de todas as probabilidades:

$$p_0 + p_1 + p_2 = 1 \Rightarrow p_0 + 0{,}25 p_0 + 0{,}0625 p_0 = 1 \Rightarrow$$
$$p_0 = \frac{1}{1 + 0{,}25 + 0{,}0625} = 0{,}76$$

- Portanto:

$$p_1 = 0{,}25 p_0 = 0{,}19$$
$$p_2 = 0{,}0625 p_0 = 0{,}0476$$

Probabilidade de 2 pacotes no roteador

Probabilidade de estouro do buffer p_3 = 0,01172

$$E[n] = \sum_{n=1}^{B} n p_n = 1 \times 0{,}19 + 2 \times 0{,}0476 = 0{,}29$$

$$E[n_q] = \sum_{n=m+1}^{B} (n-m) p_n = (2-1) \times 0{,}0476 = 0{,}0476$$

$$\lambda' = \lambda(1 - p_B) = 125(1 - p_2) = 125(1 - 0{,}0476) = 119 \text{ pps}$$

Taxa de perda de pacotes $\lambda - \lambda' = 125 - 119 = 6$ pps

$$E[r] = \frac{E[n]}{\lambda'} = \frac{0{,}29}{119} = 2{,}40 \times 10^{-3} \text{ segundos}$$

$$E[w] = \frac{E[n_q]}{\lambda'} = \frac{0{,}0476}{119} = 4{,}0 \times 10^{-4} \text{ segundos}$$

EXEMPLO 31.5

- A **variância** e **outras estatísticas** para o **número de *jobs* no sistema** podem também ser calculadas dado que conhecemos a função probabilidade de massa completa.
- Por exemplo,

$$\begin{aligned}\mathrm{Var}[n] &= E[n^2] - (E[n])^2 \\ &= (1^2 \times 0{,}19 + 2^2 \times 0{,}0476) - (0{,}29)^2 = 0{,}2963\end{aligned}$$

OUTROS SISTEMAS DE FILAS

- M/G/1
- M/D/1
- M/G/$\propto$
- G/M/1
- G/G/m

M/G/1

Box 31.5 M/G/1 Queue

1. Parameters:
   λ = arrival rate in jobs per unit time
   $E[s]$ = mean service time per job
   C_s = coefficient of variation of the service time
2. Traffic intensity: $\rho = \lambda E[s]$
3. The system is stable if the traffic intensity ρ is less than 1.
4. Probability of zero jobs in the system: $p_0 = 1 - \rho$
5. Mean number of jobs in the system: $E[n] = \rho + \rho^2(1 + C_s^2)/[2(1-\rho)]$
   This equation is known as the Pollaczek-Khinchin (P-K) mean-value formula. Note that the mean number in the queue grows linearly with the variance of the service time distribution.
6. Variance of number of jobs in the system:
$$\mathrm{Var}[n] = E[n] + \lambda^2 \mathrm{Var}[s] + \frac{\lambda^3 E[s^3]}{3(1-\rho)} + \frac{\lambda^4 (E[s^2])^2}{4(1-\rho)^2}$$
7. Mean number of jobs in the queue: $E[n_q] = \rho^2(1 + C_s^2)/[2(1-\rho)]$
8. Variance of number of jobs in the queue: $\mathrm{Var}[n_q] = \mathrm{Var}[n] - \rho + \rho^2$
9. Mean response time:
   $E[r] = E[n]/\lambda = E[s] + \rho E[s](1 + C_s^2)/[2(1-\rho)]$

Box 31.5 Continued

10. Variance of the response time:
    $\mathrm{Var}[r] = \mathrm{Var}[s] + \lambda E[s^3]/[3(1-\rho)] + \lambda^2 (E[s^2])^2/[4(1-\rho)^2]$
11. Mean waiting time: $E[w] = \rho E[s](1 + C_s^2)/[2(1-\rho)]$
12. Variance of the waiting time: $\mathrm{Var}[w] = \mathrm{Var}[r] - \mathrm{Var}[s]$
13. Idle time distribution: $F(I) = 1 - e^{-\lambda I}$. The idle time is exponentially distributed.
14. Mean number of jobs served in one busy period: $1/(1-\rho)$
15. Variance of number of jobs served in one busy period:
    $\rho(1-\rho) + \lambda^2 E[s^2]/(1-\rho)^3$
16. Mean busy period duration: $E[s]/(1-\rho)$
17. Variance of the busy period: $E[s^2]/(1-\rho)^3 - (E[s])^2/(1-\rho)^2$

For last come, first served (LCFS) or service in, random order (SIRO), the expressions for $E[n]$ and $E[r]$ are the same as above for FCFS. The variance expressions are different:

$$\mathrm{Var}[r_{\mathrm{SIRO}}] = \mathrm{Var}[s] + \frac{2\lambda E[s^3]}{3(1-\rho)(2-\rho)} + \frac{\lambda^2(2+\rho)(E[s^2])^2}{4(1-\rho)^2(2-\rho)}$$

$$\mathrm{Var}[r_{\mathrm{LCFS}}] = \mathrm{Var}[s] + \frac{\lambda E[s^3]}{3(1-\rho)^2} + \frac{\lambda^2(1+\rho)(E[s^2])^2}{4(1-\rho)^3}$$

Notice that $\mathrm{Var}[r_{\mathrm{FCFS}}] \le \mathrm{Var}[r_{\mathrm{SIRO}}] \le \mathrm{Var}[r_{\mathrm{LCFS}}]$

M/D/1

Box 31.7 M/D/1 Queue

1. Parameters:
   λ = arrival rate in jobs per unit time
   $E[s]$ = service time per job, s is constant
   Substituting $E[s^k] = (E[s])^k$, $k = 2, 3, \ldots$, in the results for M/G/1, we obtain the results listed here for M/D/1.
2. Traffic intensity: $\rho = \lambda E[s]$
3. The system is stable if the traffic intensity is less than 1.
4. Probability of n jobs in system:

$$p_n = \begin{cases} 1-\rho, & n = 0 \\ (1-\rho)(e^\rho - 1), & n = 1 \\ (1-\rho)\sum_{j=0}^{n} \dfrac{(-1)^{n-j}(j\rho)^{n-j-1}(j\rho+n-j)e^{j\rho}}{(n-j)!}, & n \geq 2 \end{cases}$$

5. Mean number of jobs in the system: $E[n] = \rho + \rho^2/[2(1-\rho)]$
6. Variance of number of jobs in the system:
   $\text{Var}[n] = E[n] + \rho^3/[3(1-\rho)] + \rho^4/[4(1-\rho)^2]$
7. Cumulative distribution function for response time:

$$F(r) = p_n \frac{(r - nE[s])}{E[s]} + \sum_{j=0}^{n-1} p_j, \qquad r \geq E[s] \text{ and } n = \left\lfloor \frac{r}{E[s]} \right\rfloor$$

8. Mean response time: $E[r] = E[s] + \rho E[s]/[2(1-\rho)]$
9. Variance of response time:
   $\text{Var}[r] = \rho(E[s])^2/[3(1-\rho)] + \rho^2(E[s])^2/[4(1-\rho)^2]$
10. Mean number of jobs in the queue:

$$E[n_q] = \rho^2/[2(1-\rho)]$$

11. Variance of number of jobs in the queue:

$$\text{Var}[n_q] = \rho^2 + \frac{\rho^2}{2(1-\rho)} + \frac{\rho^3}{3(1-\rho)} + \frac{\rho^4}{4(1-\rho)^2}$$

12. Mean waiting time: $E[w] = \rho E[s]/[2(1-\rho)]$
13. Variance of waiting time: $\text{Var}[w] = \text{Var}[r]$
14. Probability of serving n jobs in one busy period:

$$P(n) = \frac{(n\rho)^{n-1}}{n!} e^{-n\rho}$$

15. The cumulative distribution function of the busy period:

$$F(b) = \sum_{j=1}^{n} \frac{(j\rho)^{j-1}}{j!} e^{-j\rho}, \qquad n = \left\lfloor \frac{b}{E[s]} \right\rfloor$$

Here, $\lfloor x \rfloor$ is the largest integer not exceeding x.

$M/G/\infty$

Box 31.8 M/G/∞ Queue

1. Parameters:
   λ = arrival rate in jobs per unit time
   $E[s]$ = mean service time per jobs
2. Traffic intensity: $\rho = \lambda E[s]$
3. The system is always stable: $\rho < \infty$ is less than 1.
4. Probability of no jobs in the system: $p_0 = e^{-\rho}$
5. Probability of n jobs in the system: $p_n = (e^{-\rho}/n!)\rho^n$, $n = 0, 1, \ldots, \infty$
6. Mean number of jobs in the system: $E[n] = \rho$
7. Variance of number of jobs in the system: $\text{Var}[n] = \rho$
8. The number of jobs in the queue is always zero since there is no queueing. Thus, $E[n_q] = 0$.
9. Response time is equal to the service time. Therefore, it has the same distribution as the service time.
10. Mean response time: $E[r] = E[s]$

For the special case of the M/M/∞ queue, substitute $E[s] = 1/\mu$ in these results.

G/M/1

Box 31.9 G/M/1 Queue

1. Parameters:
   $E[\tau]$ = mean interarrival time
   μ = service rate in jobs per unit time
   If $L_\tau(\zeta)$ is the Laplace transform of the probability density function of the interarrival time τ, then let ϱ be the solution of the equation $\varrho = L_\tau(\mu - \varrho\mu)$.
2. Traffic intensity: $\rho = 1/(E[\tau]\mu)$
3. The system is stable if the traffic intensity ρ is less than 1.
4. Probability of zero jobs in the system: $p_0 = 1 - \rho$
5. Probability of n jobs in the system: $p_n = \rho\varrho^{n-1}(1-\varrho)$, $n = 1, 2, \ldots, \infty$
6. Mean number of jobs in the system: $E[n] = \rho/(1-\varrho)$
7. Variance of number of jobs in the system:
   $\text{Var}[n] = \rho(1 + \varrho - \rho)/(1-\varrho)^2$
8. Cumulative distribution function of the response time:
   $F(r) = 1 - e^{(1-\varrho)\mu r}$, $r \geq 0$
9. Mean response time: $E[r] = 1/[\mu(1-\varrho)]$
10. Variance of response time: $\text{Var}[r] = [1/\{\mu(1-\varrho)\}]^2$
11. Probability distribution function of the waiting time:
    $F(w) = 1 - \varrho e^{(1-\varrho)\mu w}$, $w \geq 0$
12. Mean waiting time: $E[w] = \varrho/[\mu(1-\varrho)]$
13. Variance of waiting time: $\text{Var}[w] = (2-\varrho)\varrho/[\mu^2(1-\varrho)^2]$
14. q-Percentile of the response time: $E[r]\ln[100/(100-q)]$
15. 90-Percentile of the response time: $E[r]\ln[10] = 2.3E[r]$
16. q-Percentile of the waiting time: $\max(0, (E[w]/\varrho)\ln[100\varrho/(100-q)])$. At low traffic intensities, the second term in this expression can be negative. The correct q-percentile for those cases is 0.
17. 90-Percentile of the waiting time: $\max(0, (E[w]/\varrho)\ln(10\varrho))$. At low traffic intensities, the second term in this expression can be negative. The correct 90-percentile for those cases is 0.
18. Probability of finding n or more jobs in the system: $\varrho^n(1-\rho)/(1-\varrho)$

For Poisson arrivals, $\varrho = \rho$, and all formulas become identical to those for M/M/1 queues.

$G/G/m$

Box 31.10 G/G/m Queue

1. Parameters:
   $E[\tau]$ = mean interarrival time
   λ = arrival rate, $= 1/E[\tau]$
   $E[s]$ = mean service time per job
   μ = service rate, $= 1/E[s]$
2. Traffic intensity: $\rho = \lambda/(m\mu)$
3. The system is stable if traffic intensity ρ is less than 1.
4. Mean number of jobs in service: $E[n_s] = m\rho$
5. Mean number of jobs in the system: $E[n] = E[n_q] + m\rho$
6. Variance of number of jobs in the system: $\text{Var}[n] = \text{Var}[n_q] + \text{Var}[n_s]$
7. Mean response time: $E[r] = E[w] + E[s]$. Alternately, $E[r] = E[n]/\lambda$
8. Variance of the response time: $\text{Var}[r] = \text{Var}[w] + \text{Var}[s]$
9. Mean waiting time: $E[w] = E[n_q]/\lambda$

EXERCÍCIO 31.1

- Considere um sistema com **servidor único** com **chegadas desencorajadas**, ou seja, a **taxa de chegada** é de apenas λ**/(n+1)** quando houver *n jobs* no sistema. Assuma que os intervalos entre chegadas assim como os tempos de serviço sejam identicamente e independentemente distribuídos com uma **distribuição exponencial**. Usando uma modelagem através de um **processo de nascimento e morte** para este sistema, desenvolva expressões para:
  - **probabilidade** de encontrar *n* **jobs no sistema**
  - **probabilidade** do sistema estar **ocioso**
  - **número médio de** *jobs* no sistema
  - **taxa efetiva de chegadas**
  - **tempo médio de resposta**