PROBABILIDADE, VARIÁVEIS ALEATÓRIAS, DISTRIBUIÇÃO DE PROBABILIDADES E GERAÇÃO ALEATÓRIA

Conceitos sob a ótica de Avaliação de Desempenho de Sistemas

Marcos Portnoi

Edição 25.4.2007

Universidade Salvador – UNIFACS

2005

Lista de Figuras

Lista de Tabelas

Sumário

Introdução

Este documento apresenta uma revisão de conceitos de probabilidade, distribuição de probabilidades (especificamente para as distribuições de Poisson e Exponencial) e também geração randômica de variáveis aleatórias, baseadas nos inversos das distribuições. O foco do estudo é a Avaliação de Desempenho de Sistemas.

Estatística e Probabilidade

Visão Sistêmica da Estatística

A Figura 1 traz uma síntese de conceitos relacionados à Estatística.

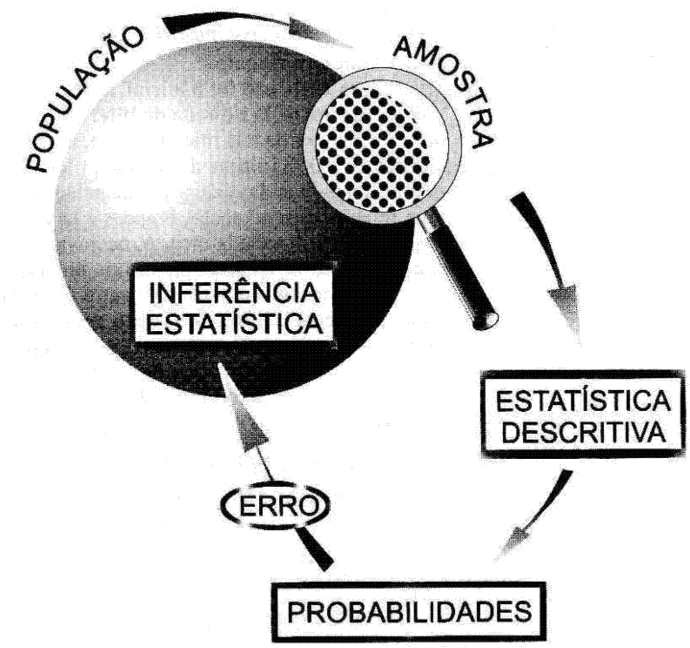


Figura 1: Visão Sistêmica da Estatística.

1. **<u>Objetivo:</u>** a partir de valores obtidos em uma amostra, descreve-se esta e deseja-se caracterizar a população como um todo generalizando observações da amostra.

2. **<u>Estatística Descritiva:</u>** parte da estatística que descreve os aspectos importantes de um conjunto de características observadas.

3. **<u>Inferência Estatística:</u>** parte da estatística que usa uma amostra para fazer generalizações a respeito de aspectos importantes de uma população.

4. **<u>Probabilidade:</u>** número que indica a chance (possibilidade) de determinada situação acontecer.

Definições Básicas

- Um *experimento aleatório* (ε) é o processo de se observar o resultado de um evento não determinístico. Ex.: jogada de um dado ou uma moeda.

- *Resultados elementares* são todos os possíveis resultados de um experimento aleatório. Ex.: para jogada de um dado, os resultados possíveis são 1, 2, 3, 4, 5, 6. Para a jogada de uma moeda, os resultados possíveis são CARA e COROA. Para duas moedas, os resultados possíveis são CARA-CARA, CARA-COROA, COROA-CARA, COROA-COROA.

- O *espaço amostral* {S} é o conjunto de todos os resultados elementares. Ex.: para a jogada de um dado, o espaço amostral S = {1, 2, 3, 4, 5, 6}. Para a jogada de uma moeda, o espaço amostral S = {CARA, COROA}.

- Um *evento* A (relativo a um espaço amostral S particular, associado a um experimento ε) é simplesmente um conjunto de resultados possíveis, ou um conjunto de um grupo de resultados. Ex.: Para o experimento *jogada de dois dados*, o espaço amostral S será igual a todos os resultados possíveis. S = {(1,1), (1,2), (1,3), (1,4), ..., (6,6)}. Se estivermos interessados no *resultado 7* a partir da jogada de dois dados, teremos como possibilidades A = {(1,6), (2,5), (3, 4), (4,3), (5,2), (6,1)}. O conjunto **A** é chamado de **evento obter 7 da jogada de dois dados.** $A \subset S$

Axiomas da Probabilidade

Seja:

- ε → um experimento

- S → o espaço amostral deste experimento

A cada evento **A**, associa-se um número real representado por P(A). Denomina-se P(A) a *probabilidade de A*, e satisfaz as seguintes propriedades:

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

$$P(S) = 1$$

Combinação de Eventos

Dado dois eventos E e F, pode-se obter novos eventos:

- E e F: ocorrência de ambos eventos;

- E ou F: ocorrência de pelo menos um dos eventos;

- não E: o evento E não ocorre.

Regras

Adição:

- P(E ou F) = P(E) + P(F) – P(E e F)

É preciso eliminar as ocorrências de ambos eventos simultaneamente, caso contrário contar-se-á estes eventos duas vezes (uma vez para a contagem do conjunto E, outra vez para a contagem do conjunto F) (Figura 2).

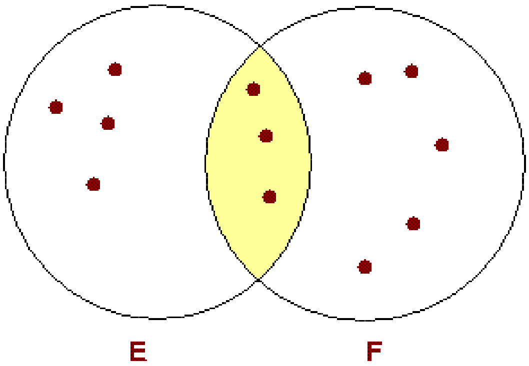


Figura 2: Probabilidade combinada de eventos. Para a probabilidade de E ou F, os eventos ambos simultaneamente têm de ser subtraidos.

Se E e F forem eventos independentes, então P(E ou F) = P(E) + P(F) (pois P(E e F) = 0). Exemplo: Probabilidade de se obter o total 5 **ou** o total 7 na jogada de dois dados. A = {(1,4), (1,6), (2,3), (2,5), ...}. O evento *obter 5* é independente do evento *obter 7*, pois ambos não podem ocorrer ao mesmo tempo em uma única jogada de dados (mutuamente excludentes).

Subtração:

- P(E) = 1 – P(não E)

Probabilidade Condicional

A probabilidade condicional de um evento E, dado que ocorreu um evento F, ou a probabilidade condicional de E dado F, é representada simbolicamente por:

$$P(E \mid F) = \frac{P(E \cap F)}{P(F)} \qquad (2)$$

Ex.: Calcular P(E|F)

- E: evento “obter o total 8 com um par de dados”
- F: evento “obter 5 na jogada do primeiro dado”

Solução:

- Número de elementos do espaço amostral = 6 x 6 = 36.
- P(F) = 1/6.
- $E \cap F$ = evento "obter 5 na primeira jogada e o total 8". Só pode ocorrer se se obtiver o par (5,3). Logo, P(E e F) = 1/36.
- $P(E \mid F) = \frac{1/36}{1/6} = 6/36 = 1/6$

<u>Regras da Multiplicação</u>

- P(E e F) = P(E|F) . P(F)
- P(E e F) = P(F|E) . P(E)

Se E e F forem eventos independentes, então P(E e F) = P(E).P(F). Exemplo: Probabilidade de se obter 5 na primeira jogada de um dado e um 3 na segunda jogada. O evento *obter 5* é independente do evento *obter 3*, pois a ocorrência de um evento nada garante ou informa para a ocorrência do outro evento.

Então: P(E|F) = 1/6 (não importa que F aconteceu) = P(E) (definição formal para eventos independentes)

Donde vem que, se E e F são independentes, então P(E|F) = P(E) (definição formal) e

P(E e F) = P(E|F) . P(F) = P(E) . P(F)

Variável Aleatória

<u>Definição</u>

Regra que atribui um valor numérico a cada possível resultado de um experimento. A Figura 3 ilustra o processo.

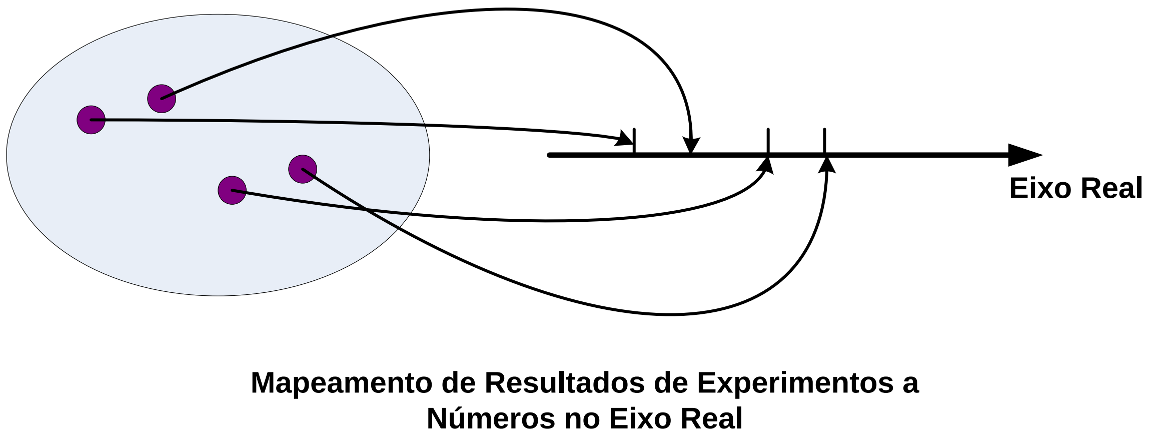


Figura 3: Mapeamento de eventos em números.

Expressando em Números os Resultados do Experimento

- Em alguns experimentos, os resultados elementares podem ter uma representação simbólica, não numérica. Ex.: para a jogada de uma moeda, os resultados podem ser CARA (H) ou COROA (T).

- Gera-se uma dificuldade em se tratar aspectos relacionados aos eventos, como por exemplo as probabilidades em forma de gráficos e eventos do tipo "número de caras em 3 jogadas de uma moeda", já que os experimentos são representados por símbolos.

A Função "Variável Aleatória"

Seja X um valor numérico, cujo valor depende do resultado do experimento. Se X associa um resultado a um número, então X é uma *função* cujo *domínio* é o *conjunto de resultados* e cuja *imagem* é o *conjunto dos números reais*. Essa função X é conhecida pelo nome de *Variável Aleatória*.

- Desta forma, pode-se escrever os *resultados de um experimento aleatório* através de números, ao invés de palavras ou símbolos, possibilitando um tratamento matemático facilitado.

- No cálculo de probabilidades, estudam-se as VAs (Variáveis Aleatórias) e calculam-se as probabilidades associadas a elas. Uma medida de probabilidade é associada ao espaço amostral por meio de uma variável aleatória X. A medida pode ser um número, uma área, um volume.

- Na Estatística Descritiva, constrói-se uma Tabela de Freqüência, na qual uma freqüência absoluta (e também uma freqüência relativa) é associada a cada valor (ver exemplo das duas moedas, adiante). Pode-se fazer o mesmo com o cálculo de probabilidades, originando uma tabela que associa a cada valor, sua probabilidade de ocorrência. Esta tabela é denominada *Distribuição de Probabilidade*.

<u>Exemplo de VA: Lançamento de duas moedas</u>

Seja o lançamento de duas moedas simultaneamente. Os resultados elementares deste experimento podem ser listados na Tabela 1 adiante (toma-se *cara* como H e *coroa* como T, de *head* e *tail*, respectivamente).

Tabela 1: Resultados do experimento "lançamento de duas moedas".

Resultado	**Símbolo**
cara-cara	HH
cara-coroa	HT
coroa-cara	TH
coroa-coroa	TT

O espaço amostral é representado conforme a Figura 4.

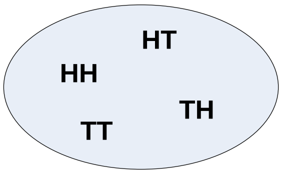


Figura 4: Espaço amostral do lançamento de duas moedas.

Pode-se estabelecer a probabilidade de ocorrência dos eventos, conforme a Tabela 2.

Tabela 2: Probabilidade de ocorrência dos resultados do lançamento de duas moedas.

Resultado	**Símbolo**	**Probabilidade de Ocorrência**
cara-cara	HH	0,25
cara-coroa	HT	0,25
coroa-cara	TH	0,25
coroa-coroa	TT	0,25

→<u>Questão</u>: Como transformar a representação simbólica destes eventos em números reais?

Defina-se a VA (X) que representa o *número de caras* do experimento. A Tabela 3 mostra os valores tomados por X, de acordo com os resultados.

Tabela 3: Variável aleatória "número de caras" no lançamento de duas moedas.

Resultado	**X = número de caras**
HH	2
HT	1
TH	1
TT	0

A variável aleatória X pode, portanto, assumir os valores 0, 1 e 2. A probabilidade associada ao valor de X=0 é 0,25, de X=1 é 0,25 + 0,25 (pois concentra as possibilidades CARA-COROA ***ou*** COROA-CARA), e de X=2 é 0,25. Agora, constrói-se a tabela de probabilidades para X (Tabela 4).

Tabela 4: Probabilidade relacionada à variável aleatória "número de caras".

x	$P(X = x)$
0	¼=0,25
1	¼+¼=½=0,5
2	¼=0,25
	$\sum = 1$

Pode-se construir um gráfico (Figura 5) relacionando a probabilidade acumulada com os valores de X. A função correspondente, F(x), é conhecida como *Função Distribuição de Probabilidade*, ou PDF (*Probability Distribution Function*), ou ainda, Cumulativa. A função que associa as probabilidades a cada valor individual de X, f(x), é conhecida como *Função Massa de Probabilidade*, ou pmf (*Probability Mass Function*), que só é definida para valores discretos.

X	P(X=x)	F(x)
0	0,25	0,25
1	0,5	0,75
2	0,25	1

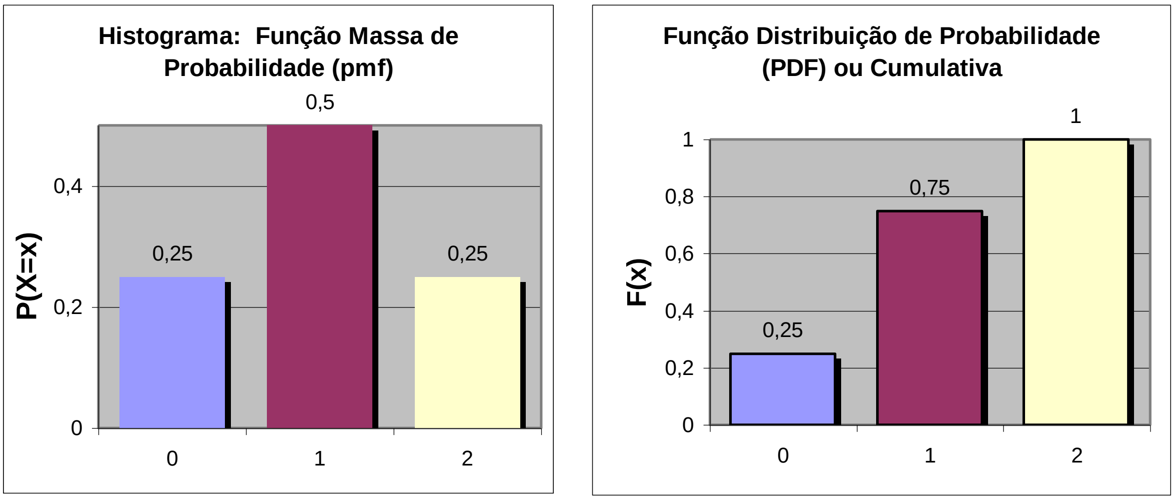


Figura 5: Gráficos da pmf e CDF.

Formalização

- Levantados os resultados possíveis da VA, os valores numéricos da VA são denotados por letras minúsculas: $x_1, x_2, x_3, ..., x_n$.
- Desse modo, para uma variável aleatória X, que assume os valores $x_1, x_2, x_3, ..., x_n$, define-se uma *função de probabilidade* $\mathbf{p(x_i)}$ que tem as seguintes propriedades:
  - $p(x_i) \geq 0$, para todo i, onde $p(x_i)$ é a probabilidade associada a $X=x_i$, ou seja, a probabilidade de ocorrência de um determinado resultado da VA.
  - $\sum_i p(x_i) = 1$
- Se a distribuição de probabilidades de uma VA é explicitamente conhecida, então todo o resumo estatístico (média, desvio padrão, etc.) também será conhecido.

Questões

1. É possível que o próprio resultado do experimento já possa ser expresso como uma VA? Sim. Exemplo, VA resultado da jogada de um dado.
2. Pode-se a um experimento associar-se mais de uma VA? Sim. Por exemplo, no caso das moedas, seja Y uma VA que representa o número de coroas.

Tipos de Variáveis Aleatórias

Diz-se que uma variável aleatória é **discreta** se todos os seus valores podem ser listados, e estes valores pertencem a um conjunto finito ou infinito, numerável. Exemplo: número de chegadas a uma fila, número de caras em uma jogada de duas moedas, resultado da jogada de um dado.

Uma VA é **contínua** se os seus valores não podem ser listados, mas podem assumir um número infinito de valores em um intervalo finito ou infinito. Exemplo: intervalo de tempo entre chegadas, altura de pessoas em uma sala.

O exercício com teste de celulares a seguir exemplifica o levantamento de resultados elementares, definição de uma variável aleatória e a construção das funções pmf e PDF.

Formalização da PDF:

$$\begin{aligned} F(t) &= P(-\infty \leq X \leq t) \\ F(t) &= P(X \leq t) \\ F(t) &= \sum_{i \leq t}^{i} p(x_i) \end{aligned} \tag{3}$$

Exercício

Seja um sistema de teste de celulares. Cada celular tem 80% de chance de ser reprovado em um teste. Em um experimento, três equipamentos são testados. Supondo que cada equipamento é independente do outro, estabeleça a distribuição de probabilidade do número X de equipamentos que são reprovados, e também a função distribuição de probabilidade acumulada (PDF).

<u>Solução</u>

Resultado Elementar	X
000	0
001	1
010	1
100	1
110	2
101	2
011	2
111	3

0 → passar no teste
1 → reprovar no teste

Resultado Elementar	X	Probabilidade
000	0	0,2 x 0,2 x 0,2 = 0,008
001	1	0,2 x 0,2 x 0,8 = 0,032
010	1	0,2 x 0,8 x 0,2 = 0,032
100	1	0,8 x 0,2 x 0,2 = 0,032
110	2	0,8 x 0,8 x 0,2 = 0,128
101	2	0,8 x 0,2 x 0,8 = 0,128
011	2	0,2 x 0,8 x 0,8 = 0,128
111	3	0,8 x 0,8 x 0,8 = 0,512

X	P(X=x)	F(x)
0	0,008	0,008
1	0,096	0,104
2	0,384	0,488
3	0,512	1
Total	1	

Classificação das Distribuições de Probabilidades

A **<u>Função Distribuição Acumulada</u>** ou **<u>Função Distribuição de Probabilidade</u>** (PDF – *Probability Distribution Function* ou CDF – *Cumulative Probability Function*), também chamada *função de distribuição*, é a probabilidade da VA X assumir valores menores ou iguais a x. É representada por F(x), de modo que:

$$F(x) = P(X \leq x) \tag{4}$$

Para uma variável aleatória *discreta*, a função distribuição de probabilidade F(x) é:

$$F(x) = P(X \leq x) \tag{5}$$
$$F(x) = \sum_{x_i \leq x}^{i} p(x_i) = p(x_1) + p(x_2) + p(x_3) + \cdots + p(x)$$

Para uma variável aleatória *contínua*, a função distribuição de probabilidade F(x) é:

$$F(x) = P(X \leq x) \tag{6}$$
$$F(x) = \int_{-\infty}^{x} f(x)dx$$

onde f(x) é chamada **<u>função densidade de probabilidade</u>** ou pdf. A função densidade de probabilidade associa os valores de X com a probabilidade de cada um deles ocorrer. A **<u>função massa de probabilidade</u>**, ou pmf, é a mesma função, porém definida para variáveis aleatórias discretas, somente.

A F(x) tem as seguintes propriedades:

a) F(x) é uma função não-decrescente.

b) $\lim_{x \to +\infty} F(x) = 1$

c) $\lim_{x \to -\infty} F(x) = 0$

Modelos Matemáticos que Representam Distribuições de Probabilidades

Conhecidos os principais conceitos referentes ao cálculo das probabilidades, pode-se agora estudar distribuições e funções densidade de probabilidade que, pela sua importância, merecem um estudo especial. Tais distribuições partem do pressuposto da existência de certas hipóteses bem definidas. Como diversas situações na vida real se aproximam destas premissas, os modelos descritos a seguir são úteis para o estudo destas situações.

<u>Conceito introdutório</u>

Existem experimentos aleatórios cujos resultados, refletidos em uma VA, seguem um comportamento previsível em relação às suas probabilidades de ocorrência, e portanto podem ser modelados por uma equação específica. Seja por exemplo o número de chamadas telefônicas que chegaram a uma central telefônica, em um determinado tempo *t* (Figura 6).

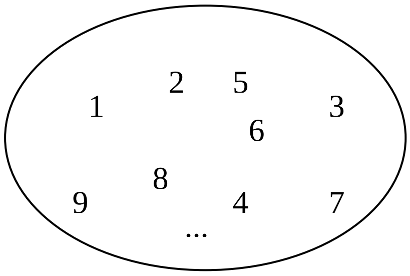


Figura 6: Espaço amostral "número de chamadas que chegaram a uma central telefônica num tempo *t*".

Já que a representação dos resultados já é numérica, uma variável aleatória X, "número de chamadas telefônicas chegadas", pode associar diretamente os resultados com sua probabilidade de ocorrência.

O estudo das observações históricas deste experimento concluiu que a expressão de probabilidade de Poisson pode ser usada para modelá-lo matematicamente.

O Modelo de Poisson

Basicamente, este modelo representa a probabilidade de ocorrência de um certo *número de chegadas* em um determinado tempo. Modela, por exemplo:

- Número de requisições para um servidor em um intervalo de tempo *t*
- Número de falhas em componentes por unidade de tempo
- Número de requisições para um sistema de banco de dados em *t* segundos
- Número de erros de datilografia por formulário
- Número de chegadas telefônicas em uma central em um intervalo de tempo *t*

As seguintes condições permitem o uso do modelo de Poisson para modelar um experimento:

a) O número de chegadas durante qualquer intervalo de tempo depende somente da duração do intervalo de tempo; quanto maior o intervalo, maior tende a ser o número de chegadas.

b) As chegadas ocorrem independentemente, isto é, um excesso ou falta de chegadas em algum intervalo de tempo não exerce efeito sobre o número de chegadas ocorridas durante qualquer outro intervalo. Por exemplo: se num experimento, observou-se que a média de chegadas é 5 (por intervalo de tempo), e observou-se que num determinado intervalo houve nenhuma chegada, isso não significa que, no próximo intervalo, haverá um excesso de chegadas.

c) A possibilidade de duas ou mais chegadas ocorrerem durante um pequeno intervalo de tempo *t* é muito pequena comparada à de uma única chegada.

A distribuição de probabilidades de Poisson é definida como: seja uma variável aleatória X que pode assumir os valores x=0, 1, 2, ..., n, ... A probabilidade de X assumir um determinado valor x é dada pela seguinte expressão:

$$f(x) = P(X = x) = \frac{e^{-\lambda} \cdot \lambda^{x}}{x!} \text{, para } x = 0, 1, 2, 3, 4, ..., n, ... \tag{7}$$

Ou seja, P(X=x) representa a probabilidade de ocorrerem x chegadas na unidade de tempo trabalhada, sendo que λ é a média de chegadas na unidade de tempo. A distribuição de Poisson é discreta, definida apenas para valores inteiros de x.

O resumo das características principais da distribuição de Poisson está na Tabela 5, e a Figura 7 traz a aparência da distribuição, para diversos valores de média (observar que as linhas dos gráficos servem apenas como indicação visual da distribuição; sendo a distribuição de Poisson uma função discreta, estas linhas de fato não existem).

Tabela 5: Resumo das características da Distribuição de Poisson.

Distribuição de Poisson (taxa)**:**
Parâmetro: λ (evento/UT)
pmf: $f(x) = P(X = x) = \lambda^{x} \frac{e^{-\lambda}}{x!}$
CDF: $F(x) = \sum_{i} p_i$, para $f(x) = P(X = x) = p_i$
Média: λ
Variância: λ

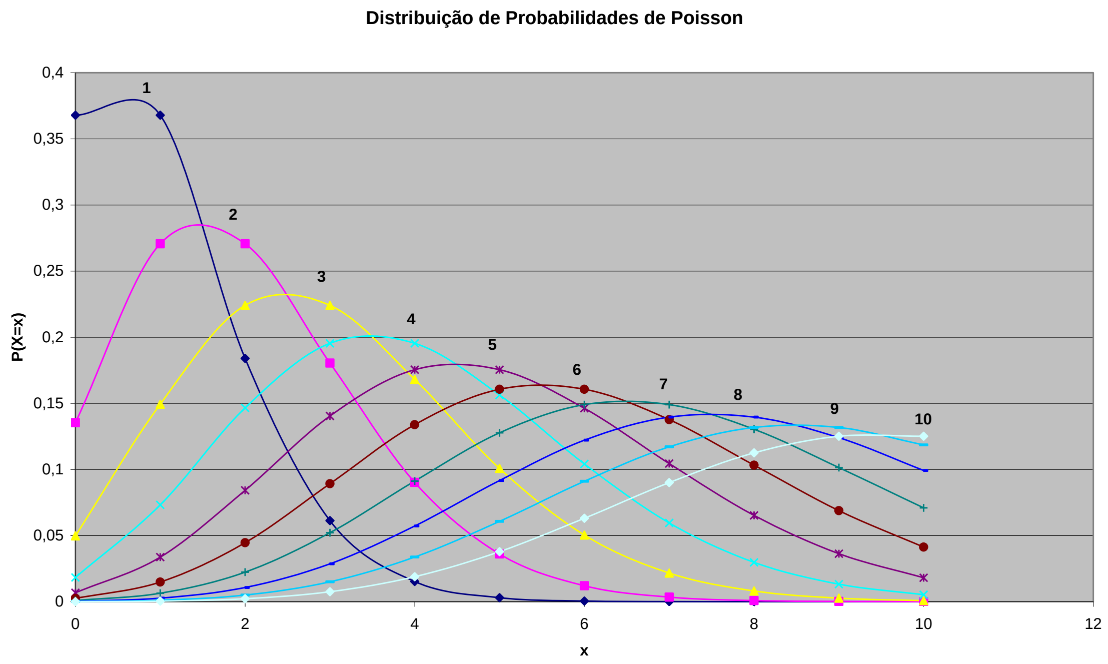


Figura 7: Gráficos para as Distribuições de Poisson, com várias médias diferentes.

Exercício

1. Um escritório de advocacia recebe, em média, cinco telefonemas por hora. Determine a probabilidade de que em determinada hora, selecionada aleatoriamente, sejam recebidas exatamente três chamadas.

2. Segundo dados históricos em determinada empresa, 3 é o número médio de chamadas em 20 minutos:

    a. determine a distribuição de probabilidade para esse exemplo.

    b. determine a probabilidade de haver, no máximo, 2 chamadas em 40 minutos, em um intervalo escolhido aleatoriamente.

Solução

1. $\lambda = 5 \text{ tel/h} \quad \therefore P(X = 3) = \lambda^x \dfrac{e^{-\lambda}}{x!} = 5^3 \dfrac{e^{-5}}{3!} = 0{,}14$

2.

    a. Não é preciso trabalhar com uma unidade de tempo unitária. A unidade de tempo pode ser 20 minutos. Isso facilita o cálculo para a letra (b). Portanto, $\lambda = 3 \text{ cha/20 min}$

    b. Mudar a média. Se chegam 3 a cada 20 minutos, então, em 40 minutos, a média será 6 chamadas. Fazer então a distribuição e calcular a Probabilidade Acumudada (PDF).

Questão 2: Distribuição de Poisson

	média=	3
Chamadas	P(X=x)	
0	0,049787	
1	0,149361	
2	0,224042	
3	0,224042	
4	0,168031	
5	0,100819	
6	0,050409	
7	0,021604	
8	0,008102	
Total	0,996197	

Questão 2: Distribuição de Poisson

	média=	6
Chamadas	P(X=x)	
0	0,002479	
1	0,014873	
2	0,044618	
Total	0,061969	

O Modelo Exponencial

Conforme visto anteriormente, a Distribuição de Poisson está relacionada com ritmo ou taxa de chegadas. A **Distribuição Exponencial** (também chamada em algumas literaturas como Distribuição Exponencial Negativa) é a correspondente da Distribuição de Poisson para a intervalos entre chegadas, ou *tempos de interchegada*. Quando um fenômeno, portanto, segue Poisson em sua taxa de chegada, ele também comporta-se segundo a Distribuição Exponencial em termos de tempo entre chegadas (Figura 8).

Assim, seja um fenômeno qualquer. Seu processo de chegadas é baseado em Poisson, de modo o número de chegadas em um intervalo de tempo *t* é uma V.A. *discreta*, e a média de chegadas no intervalo *t* é λ (chegadas/unidade de tempo-UT). O tempo entre as ocorrências destas chegadas é definido segundo a **Distribuição Exponencial.**

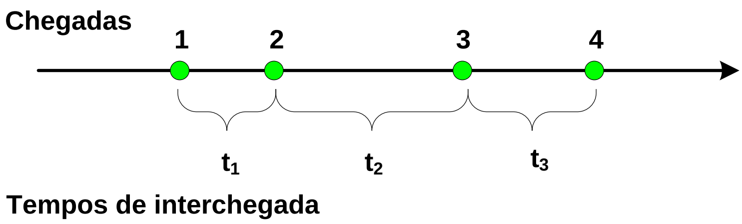


Figura 8: Relação entre chegadas e tempos de interchegada.

O período de tempo T entre contagens sucessivas de um processo de Poisson, com média λ>0, é uma *Variável Aleatória contínua*, cuja **função densidade de probabilidade** (pdf) é dada por:

$$f(x) = P(X = x) = \lambda e^{-x\lambda}, \text{ para } x \geq 0 \qquad (8)$$

Nesta equação, *x* representa o *tempo* e λ é a *taxa de chegada* ou taxa de ocorrência de eventos por unidade de tempo, a mesma utilizada como parâmetro em Poisson. Novamente, *x* é uma V.A. <u>contínua</u>.

A distribuição cumulativa de probabilidade (PDF ou CDF) da expressão de probabilidade do modelo exponencial é calculada a seguir:

$$F(x) = \int_0^x f(x)dx = \int_0^x \lambda \cdot e^{-x\lambda} dx \qquad (9)$$
$$\therefore F(x) = P(X \leq x) = 1 - e^{-x\lambda}$$

O resumo das características principais da distribuição exponencial segue na Tabela 6.

Tabela 6: Resumo das características da Distribuição Exponencial com parâmetro UT/evento.

Distribuição Exponencial (período ou intervalo)**:**

Parâmetro: a (UT/evento)

pdf: $f(x) = \frac{1}{a} e^{-x/a}$

CDF: $F(x) = 1 - e^{-x/a}$

Média: a

Variância: a^2

Observar que, no resumo acima, o parâmetro utilizado **não é** a taxa média de ocorrência do evento por unidade de tempo, mas sim seu **inverso**, o tempo médio de interocorrência do evento (unidade de tempo por evento). A relação entre o parâmetro a e a taxa média λ, para fins práticos, é:

$$\lambda = \frac{1}{a} \qquad (10)$$

E o resumo das características da distribuição exponencial pode ser reescrito conforme a Tabela 7. A Figura 9 ilustra os gráficos da pdf e CDF da distribuição exponencial, que é contínua.

Tabela 7: Resumo das características da Distribuição Exponencial com parâmetro evento/UT.

Distribuição Exponencial (período ou intervalo)**:**

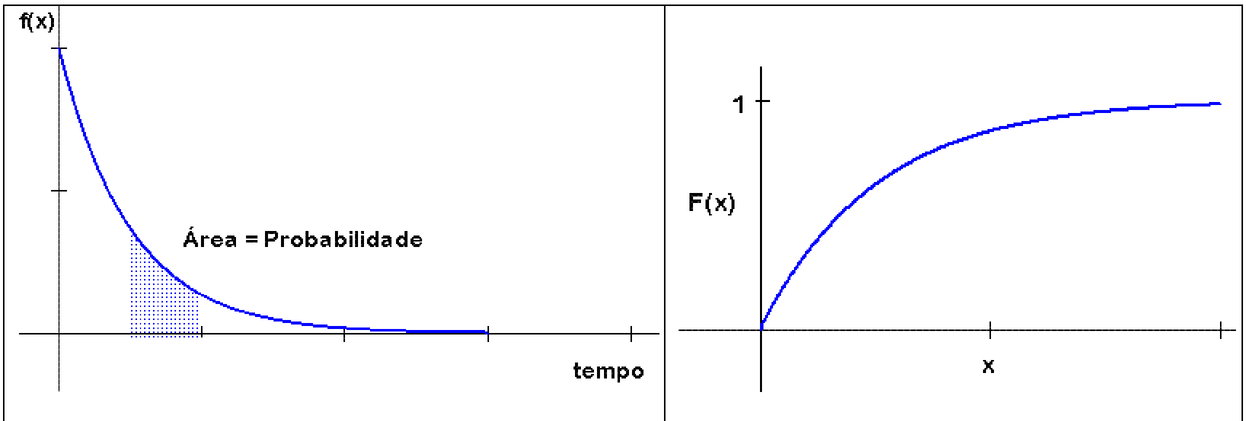
Parâmetro: λ (evento/UT)

pdf: $f(x) = \lambda e^{-x\lambda}$

CDF: $F(x) = 1 - e^{-x\lambda}$

Média: $1/\lambda$

Variância: $(1/\lambda)^2$



Função Densidade de Probabilidade (pdf) | **Função Cumulativa de Probabilidade (CDF)**

Figura 9: Gráficos da pdf e CDF para distribuição exponencial.

Exercício

1. Ao observar-se a duração das baterias de *videogames* do tipo *Gameboy*, conclui-se que esta vida nada mais é do que o intervalo entre falhas sucessivas das baterias; para essas falhas, pode-se aplicar o processo de Poisson. Desse modo, o tempo médio entre falhas vem a ser a vida média da bateria.

   Considere que inúmeras baterias foram usadas e anotou-se (algo raro de ocorrer no dia-a-dia, somente as fábricas o fazem) que a cada sete dias havia necessidade de trocá-las (ou seja, a vida média da bateria é de uma semana). As falhas das baterias são aleatórias e independentes e atendem às condições da distribuição de Poisson; então, para o tempo de vida da bateria, pode-se utilizar a distribuição exponencial.

   a. determine a probabilidade de a bateria durar pelo menos 2 semanas;
   b. determine a probabilidade de uma bateria falhar dentro de 3 dias;
   c. determine a probabilidade de uma bateria durar de 3 a 4 semanas;
   d. determine o desvio padrão do tempo de vida de uma bateria.

<u>Solução</u>

a = 7 dias/ocorrência ou a = 1 semana/ocorrência

a) em semanas:

$$P(X \geq x) = 1 - P(X \leq x) = 1 - \left(1 - e^{-x/a}\right) \therefore P(X \geq 2) = 1 - \left(1 - e^{-2/1}\right) = 1 - 0{,}864 = 0{,}135$$

b) em dias: $P(X \leq 3) = 1 - e^{-3/7} = 0{,}349$

c) em semanas:

$$P(3 \leq X \leq 4) = P(X \leq 4) - P(X \leq 3) = 1 - e^{-4/1} - \left(1 - e^{-3/1}\right) = 0{,}981 - 0{,}95 = 0{,}0314$$

d) $\sigma = \sqrt{a^2} = a = 1\ \text{semana}$

Geração de Números Aleatórios

Um dos passos-chave no desenvolvimento de uma simulação é a geração de números aleatórios ou randômicos, que nortearão os valores assumidos pelas variáveis aleatórias do modelo simulado. Por exemplo, num modelo de fila com um servidor e uma fila, os tempos de interchegada, bem como os tempos de serviço, deverão ser gerados aleatoriamente, dentro de uma média estipulada pelo modelo.

A rotina de geração de números aleatórios para simulação envolve duas etapas. Primeiro, é obtida uma seqüência de números aleatórios distribuída uniformemente entre 0 e 1 (ou seja, a probabilidade de se obter cada número na seqüência é exatamente igual para todos os números na seqüência). Então, esta seqüência é transformada de modo a produzir números aleatórios que satisfaçam a distribuição correta desejada (por exemplo, exponencial ou normal). O primeiro processo é chamado *geração aleatória de números*, e o segundo, *geração aleatória de variáveis aleatórias*.

Gerador de Números Aleatórios

O método mais comum para a geração de números aleatórios é o uso de uma relação recursiva, onde o próximo número na seqüência é função do último ou dois últimos números.

$$x_n = f(x_{n-1}, x_{n-2}, \ldots) \tag{11}$$

Uma destas funções pode ser vista a seguir.

$$x_n = (5x_{n-1} + 1) \bmod 16 \tag{12}$$

Começando com $x_0=5$, obtém-se x_1:

$$x_1 = (5(5) + 1) \bmod 16 = 26 \bmod 16 = 10 \tag{13}$$

Os primeiros 32 números obtidos através deste gerador 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.

Os números são inteiros entre 0 e 15. A fim de se obter uma seqüência aleatória entre 0 e 1, basta dividir os números por 16, ou seja:

- 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.

Semente

Se a função *f* geradora é conhecida, pode-se obter a mesma seqüência de números a qualquer tempo, desde que o valor inicial x_0 seja conhecido. Este valor x_0, usado para iniciar a seqüência, é chamado de ***semente*** ou ***seed***. Mudando-se a semente, outra seqüência pode ser gerada.

Números Pseudo-Aleatórios

Uma observação importante que se pode fazer acerca do exemplo anterior é que a função *f* é *determinística*. Ou seja, dada uma semente, a seqüência de números gerada pode ser prevista com certeza absoluta. Os números são considerados aleatórios, entretanto, porque satisfazem os testes estatísticos para aleatoriedade. Estes números são portanto chamados de ***pseudo-aleatórios***, pois são apenas "parcialmente" aleatórios. Os geradores de números aleatórios incluídos na maioria dos compiladores ou interpretadores de linguagem de programação são geradores pseudo-aleatórios, que usam algoritmos matemáticos conhecidos, como por exemplo:

- Geradores lineares-congruentes
- Geradores Tausworthe
- Geradores Fibonacci Extendidos
- Geradores Combinados

Uma seqüência pseudo-aleatória de números são geralmente preferíveis no lugar de números totalmente aleatórios em se tratando de aplicações de simulação, pois freqüentemente, na depuração e teste de modelos, é desejável repetir-se o mesmo experimento de simulação, exatamente como feito anteriormente (usando, pois, a mesma seqüência de números aleatórios). Quando resultados diferentes forem requeridos, basta alterar o valor da semente, culminando assim num maior controle sobre a simulação e a obtenção de resultados.

Tamanho do Ciclo, Cauda (*Tail*) e Período

Novamente, observando o exemplo anterior, somente os primeiros 16 números são únicos. O 17º é igual ao primeiro número e, subseqüentemente, a seqüência se repete ciclicamente com os primeiros 16 valores. Em outras palavras, este gerador de números aleatórios tem um ***tamanho de ciclo*** de 16.

Alguns geradores não repetem uma certa parte inicial da seqüência. Esta parte inicial é chamada de ***cauda*** ou ***tail***. Nestes casos, o ***período*** do gerador é a soma do tamanho da cauda e o tamanho do ciclo. A Figura 10 ilustra estes conceitos.

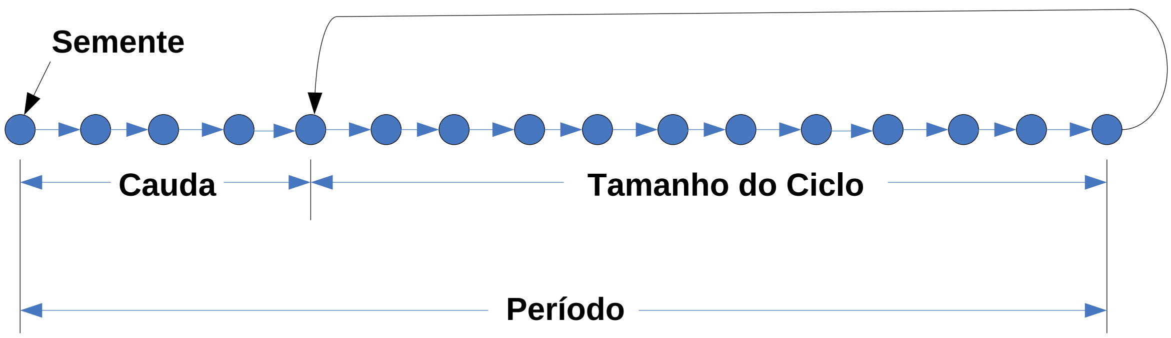


Figura 10: Tamanho do ciclo, tamanho da cauda e período de um gerador de números aleatórios.

Propriedades Desejadas de uma Função Geradora de Números Aleatórios

Uma função geradora de números aleatórios, para fins de simulação, deve ter as seguintes características:

1. *Deve ser computacionalmente eficiente*. Como as simulações tipicamente requerem a geração de milhares ou milhões de números aleatórios a cada execução, o tempo de processamento requerido para a geração dos números deve ser pequeno.

2. *O período deve ser grande*. Um período pequeno pode resultar numa seqüência que reinicia muito cedo, causando uma seqüência de eventos repetida. Os tempos de simulação poderão então ficar limitados.

3. *Os valores sucessivos devem ser independentes e distribuídos uniformemente* (IID – *Independent and Identically Distributed*). A correlação entre números sucessivos deve ser pequena. A correlação, se for significante, indica que há dependência entre os números sucessivos.

Geração de Variáveis Aleatórias Randomicamente

Há uma série de métodos usados para gerar valores para variáveis aleatórias não-uniformes. Cada método é aplicável somente para uma parte da distribuição em questão. Ainda, para uma distribuição particular, um determinado método pode ser mais eficiente do que outros. Descrever-se-á aqui apenas um método, usado para gerar valores para uma variável aleatória que segue uma distribuição exponencial.

Transformação Inversa

Este método baseia-se na observação de que, para uma dada uma variável aleatória x, com uma CDF $F(x)$, a variável $u = F(x)$ é uniformemente distribuída entre 0 e 1. Assim, os valores de x podem ser obtidos gerando-se números aleatórios uniformemente distribuídos e computando-se $x = F^{-1}(u)$, conforme pode ser visto na Figura 11. A prova desta observação está demonstrada em (JAIN, 1991).

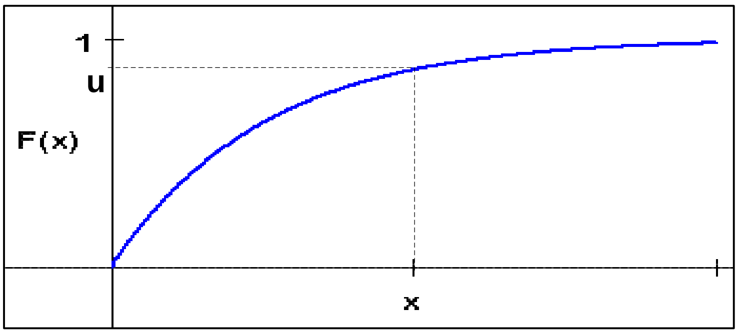


Figura 11: Transformação inversa da CDF.

<u>Geração de Valores Exponencialmente Distribuídos</u>

A função de probabilidade exponencialmente distribuída já foi vista. Dado um determinado valor x, e uma taxa média de chegada λ, calcula-se a probabilidade de ocorrência de x ou ainda a probabilidade acumulada $P(X \leq x)$.

Agora, dada a probabilidade $P(X \leq x)=F(x)$, pode-se calcular o valor correspondente de x usando a transformação inversa.

$$\text{pdf:} \quad f(x) = \lambda e^{-\lambda \cdot x} \tag{14}$$

$$\text{CDF:} \quad F(x) = 1 - e^{-\lambda \cdot x} = u \therefore x = -\frac{1}{\lambda}\ln(1-u)$$

Assim, valores para a variável aleatória x_i podem ser produzidos através da geração de uma variável u_i, uniforme, e usando a equação anterior para determinar x_i. Como u é uniformemente distribuída entre 0 e 1, a expressão $1 - u$ também é distribuída uniformemente entre 0 e 1. Desta forma, o algoritmo de geração pode ser simplificado para:

$$x = -\frac{1}{\lambda}\ln(u) \tag{15}$$

Usando, ao invés da taxa média de chegada λ, a média de tempo de interchegada a, a expressão torna-se:

$$x = -a\ln(u) \tag{16}$$

Uso da Técnica em Simulação

Um simulador de eventos discretos necessita gerar uma série de eventos aleatoriamente. Por exemplo, é preciso gerar uma chegada de um cliente numa rede de filas que vai acontecer num determinado tempo aleatório t_1. Este cliente será atendido por um servidor, que demorará um tempo aleatório t_2 para servir o cliente. Finalmente, o próximo cliente a chegar no sistema será gerado pelo simulador após um intervalo de tempo aleatório t_3 depois da chegada do cliente anterior. Cada um destes tempos segue uma distribuição de probabilidades característica, inerente ao modelo simulado.

Por conseguinte, a fim de gerar estes tempos aleatórios, usa-se um gerador de números aleatórios de distribuição uniforme, que produzirá um número uniforme entre 0 e 1. Este número vem a ser a *probabilidade u*, que, conforme visto anteriormente, é distribuída uniformemente. Usando-se a transformação inversa da CDF, calcula-se portanto o valor do tempo correspondente àquela probabilidade u.

Verifique-se que o valor de u corresponde à probabilidade de que o tempo seja *menor ou igual* ao valor calculado pela transformação inversa, já que a função transformada é a CDF. Em outras palavras, gerando-se um número aleatório u, este número será a probabilidade de que o tempo de interchegada esteja compreendido entre 0 e t, onde este t é calculado pela transformação inversa. Assume-se, para fins de simulação, o pior caso, ou seja, o tempo de interchegada obtido pelo cálculo da transformação inversa será t.

Exercício

Seja uma simulação de um sistema de redes de filas M/M/1. O tempo de serviço de cada cliente e os tempos de interchegada de clientes são modelados através de uma distribuição exponencial. Considere que o taxa de chegada é de **0,5** cliente/hora e o tempo médio de serviço é de **1,25** hora/cliente. A tabela abaixo mostra os resultados de uma geração aleatória para as probabilidades $P(X \leq x)$, para obtenção dos tempos de interchegada e tempo de serviço de uma simulação.

	Tempo de Interchegada		**Tempo de Serviço**	
Cliente	**P(X ≤ x)**	**Tempo**	**P(X ≤ x)**	**Tempo**
1	0,10		0,15	
2	0,13		0,28	
3	0,06		0,33	
4	0,22		0,08	
5	0,39		0,14	
6	0,18		0,43	
7	0,15		0,27	
8	0,12		0,33	

a) Calcule os tempos baseados na distribuição exponencial e na geração aleatória e elabore um gráfico de simulação

b) Calcule o **tempo médio em fila** desta simulação.

c) Calcule a **utilização** do servidor nesta simulação.

<u>Solução</u>

(a) Usando-se a equação 14, calcula-se os tempos de interchegada e de serviço. Monta-se a tabela de início e fim de serviço e então se preenche a grade de simulação.

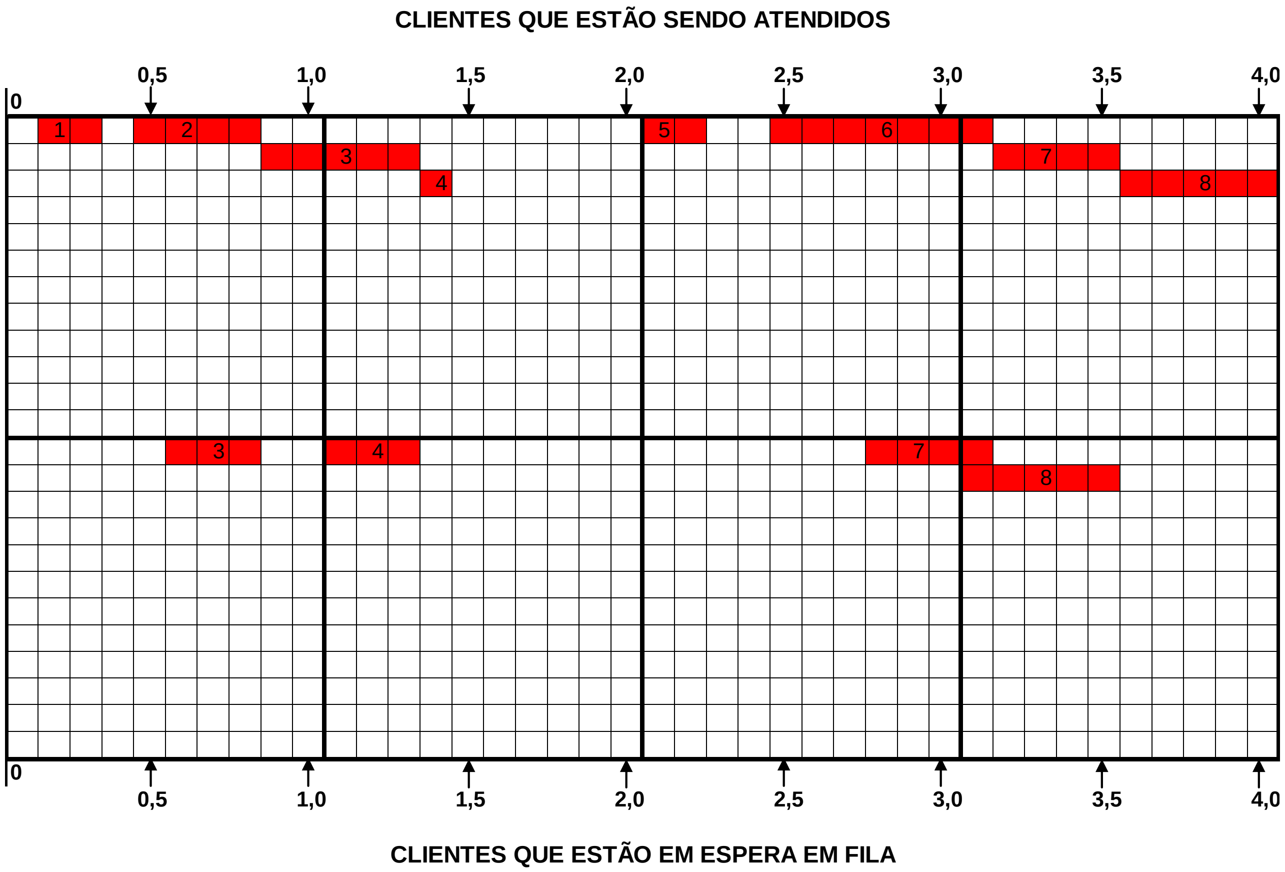


Figura 12: Gráfico representativo da simulação.

Tabela 8: Cálculo de valores para as variáveis aleatórias, usando Transformação Inversa.

Random-Variate Generation

lambda	0,5	job/ut	**tau**	ut/job	2
Mi	0,8	job/ut	**s**	ut/job	1,25

Interchegada							Tempo de Serviço	
P(X<=x)	*Tempo*	*Chegada (atual + anterior)*	*Início Serviço (maior entre Final Serviço e Chegada)*	*Tempo Serviço (random-variate)*	*Final Serviço (início serviço + tempo serviço)*	*Tempo Espera Fila*	*P(X<=x)*	*Tempo*
0,10	0,21	0,21	0,21	0,20	0,41	0,00	**0,15**	0,20
0,13	0,28	0,49	0,49	0,41	0,90	0	**0,28**	0,41
0,06	0,12	0,61	0,90	0,50	1,40	0,286879	**0,33**	0,50
0,22	0,50	1,11	1,40	0,10	1,50	0,290554	**0,08**	0,10
0,39	0,99	2,10	2,10	0,19	2,29	0	**0,14**	0,19
0,18	0,40	2,50	2,50	0,70	3,20	0	**0,43**	0,70
0,15	0,33	2,82	3,20	0,39	3,59	0,377611	**0,27**	0,39
0,12	0,26	3,08	3,59	0,50	4,09	0,515332	**0,33**	0,50

(b): Tempo médio em fila = integral, no tempo, dos tamanhos de fila tomados periodicamente / número de clientes que passaram pela fila (todos, sejam os que ficaram em fila ou não).

$$Qt = \frac{\sum_{i=1}^{n} Q_i \cdot (t_i - t_{i-1})}{T} \cdot \frac{T}{\text{TotCli}} \tag{17}$$

<u>Método 1</u>: somar cada tamanho de fila tomado em cada bloquinho (0,1 hora) e dividir pelo total de clientes. Resultado: 1,5/8 = **0,1875 h**. (0*0,1 + 0*0,1 + ... + 1*0,1 + 1*0,1 + ... + 2*0,1 + 1*0,1 + 2*0,1 + ... = 1,5/8)

<u>Método 2</u>: somar todos os tempos totais em fila de cada clientes que ficou em fila e dividir pelo total de clientes.

Obs.: Ambos os métodos já consideram os clientes que apenas passaram pela fila e nela não ficaram, ou seja, tiveram um tempo de fila igual a zero.

(c): somar todos os tempos ocupados do servidor e dividir pelo tempo total de simulação. Resultado: 30*0,1/4=0,75. (Atenção: o resultado vindo da teoria de filas, onde utilização é lambda/mi não se aplica aqui, pois foi pedido a utilização com os dados da questão.) Para esta questão, nem é preciso montar o gráfico, basta calcular os tempos a partir da geração exponencial, somá-los e dividir pelo tempo total da simulação.

Referências

BRITO, Sérgio de Figueiredo. Material de aulas da disciplina Avaliação de Desempenho de Sistemas do curso de Engenharia Elétrica, Departamento de Engenharia e Arquitetura, Universidade Salvador – UNIFACS, 2001.

JAIN, Raj. **The art of computer systems performance analysis**. John Wiley & Sons, Inc., 1991.

PRADO, Darci. **Teoria das filas e da simulação.** vol. 2. Nova Lima: INDG Tecnologia e Serviços Ltda, 2004.