Modelagem e Avaliação de Desempenho

Pós Graduação em Engenharia Elétrica - PPGEE
Prof. Carlos Marcelo Pedroso
2011

Análise de desempenho

- São disponíveis duas abordagens para realizar a análise de desempenho:
  - *Estudo analítico* através de *modelos* matemáticos;
  - *Modelagem* e *simulação* do sistema em estudo;

Exemplo: Servidor Web

- Suponha um servidor Web que recebe requisições e transmite arquivos em resposta;
- A equação $b=r/t$ descreve a relação entre a quantidade de bits transmitidos *b* e a taxa de transmissão *r (bits por segundo, bps ou b/s)* de um enlace ao longo de um período de tempo *t (segundos)*. A equação é um modelo determinístico;
- Que tipo de estatística pode ser medida em um sistema com relação a *r*?
  - *Média*
  - *Desvio padrão / variância*

Exemplo: Servidor Web

- As estimativas de média e variância são suficientes neste caso?
  - Anote os motivos
- Considere o número de bits transmitidos por unidade de tempo como uma variável discreta *(r=0, 1, 2, ..., n)*.
  - O que mais pode ser feito para descrever o comportamento de *r?*

Estocástico

- A palavra estocástico deriva do grego (*στοχάζεθαι* ou “*adivinhação*”), significando “*aleatório*”, “*chance*”. Seu antônimo é “*determinístico*” , “*certo*”.
- Um modelo estocástico prevê um conjunto de possibilidades de acontecimentos. Exemplo: um lançamento de moeda, com possibilidade ½ para cada um dos resultados.

Caracterização: Freqüência relativa

- Suponha que repetimos n vezes o experimento ε, e sejam A e B dois eventos associados a ε. Admita que sejam n_A e n_B o número de vezes que o evento A e o evento B ocorrem.
  - $f_A = n_A / n$ é denominada freqüência relativa do evento A nas n repetiçoes de ε;
  - É claro que $0 \leq f_A \leq 1$;
  - … e $n = n_A + n_B$.

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 s seguintes propriedades:
  - $0 \leq P(A) \leq 1$;
  - $P(S)=1$;

Variável aleatória

- Sejam ε um experimento e S um espaço amostral associado ao experimento. Uma função X, que associe a cada elemento de $s \in S$ um número real, $X(s)$, é denominada variável aleatória.

Exemplo

- Suponha que foi medido o tráfego em um servidor de páginas. São exemplos de variáveis aleatórias:
  - X=número de bits transmitidos por segundo
  - Y= tamanho do pacote;
  - Z= intervalo entre pacotes;
  - W= tempo de atendimento da requisição.

...anotar o espaço amostral para cada caso...

Variáveis aleatórias - Notação

- Letras maiúsculas, como X, Y ou Z, denotam variáveis aleatórias;
- Letras minúsculas, como x, y ou z, denotam números reais;
- A expressão $\{X=x\}$ é o evento que a variável aleatória assume um valor igual a x.
- A expressão $\{X\leq x\}$ é o evento que a variável aleatória assume um valor menor ou igual a x.

Função densidade de probabilidade

- $f(x)$ é a função que indica a probabilidade de obter-se exatamente o valor x em um experimento, ou seja, $P(X=x)$;
- Normalmente, a função densidade de probabilidade é denotada por letras minúsculas:

$$f(x) = P(X=x);$$

Anotar o exemplo para o X=número de bits transmitidos por unidade de tempo;

Anotar: qual a relação necessária para existência da distribuição de probabilidade?

Função densidade de probabilidade

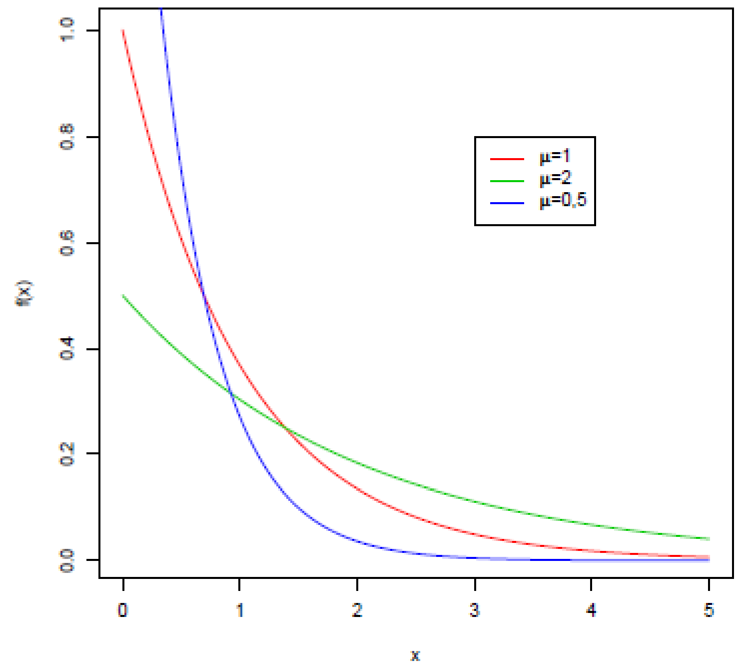
- Desta forma, pode-se escrever:

$$P\{a < X \leq b\} = \int_a^b f(x)\,dx$$

Função densidade de probabilidade

- Exemplo: distribuição exponencial.

Distribuição Exponencial	
Parâmetros	μ $\mu =$ média, $\mu > 0$
Limites	$0 \leq x < \infty$
Densidade de Probabilidade	$f(x) = \frac{1}{\mu} e^{-x/\mu}$
Distribuição Acumulada	$F(x) = 1 - e^{-x/\mu}$
Esperança ($E[X]$)	μ
Variança ($Var[X]$)	μ^2



Função distribuição acumulada

- *A função de distribuição acumulada de probabilidade é definida da seguinte forma:*
  - $F(x) = P(X \leq x)$
  - Esta é a distribuição de probabilidade acumuladada variável aleatória X, normalmente denotada por letras capitais: *F(x)*.
- *F(x)* e *f(x)* são intimamente relacionados, onde:

$$F(x) = \int_{-\infty}^{x} f(\xi)\, d\xi, \quad -\infty < x < \infty$$

Função densidade de probabilidade

- Se F(x) é diferenciável em x, então pode-se escrever que X tem uma função densidade de probabilidade dada por:

$$f(x) = \frac{d}{dx} F(x) = F'(x)$$

Distribuição de probabilidade

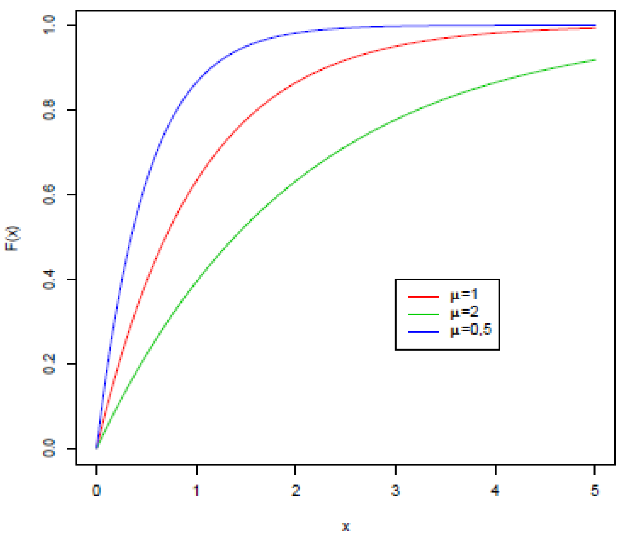
- A probabiliade de um evento ocorrer é escrito $P(X \leq x)$. Permitindo a variação de x, define-se a função:
  - $F(x) = P(X \leq x), \infty \leq x \leq \infty;$
  - Esta é a distribuição de probabilidade acumuladada variável aleatória *X*, normalmente denotada por letras capitais: F(x).

Distribuições de probabilidade

- Uma distribuição de probabilidade muita informação disponível sobre uma V.A.
- Pode ser calculado, por exemplo,
  - $P\{X>a\}=1-F(a)$
  - $P\{a \leq X \leq b)=F(b)-F(a)$
  - *Momentos da V.A.*

Distribuição exponencial

- A função de distribuição acumulada de probabilidade exponencial é dada por ***F(x)=1-e$^{-\lambda x}$, x>0,*** onde λ é 1/média.
  - *Prove como exercício*
- Graficamente:



Exercícios

1. Considere a distribuição:

$$F(x) = \begin{cases} 0 & x \leq 0 \\ x^3 & 0 < x < 1 \\ 1 & x \geq 0 \end{cases}$$

Determine:

A. Esta realmente uma distribuição de probabilidade? Prove matematicamente.

B. Calcule a função densidade de probabilidade *P(X=x)*

C. P(X>0,7)

D. P(0,4<X≤0,9)

E. Calcule E[X] e V[X]

Exercícios

2. Suponha a distribuição de probabilidade uniforme, dada por $f(x)=0,01$ *para* $0<x<100$.
   A. Calcule a expressão para F(x).
   B. Calcule P(X<20) utilizando f(x) e F(x).
   C. Calcule E[X] e V[X].

Valor Esperado e Momentos

O enésimo momento de uma variável aleatória X é dado por

$$E[X^n] = \int_{-\infty}^{+\infty} x^n f(x)dx \tag{1.7}$$

O primeiro momento, correspondendo a $n = 1$ é chamado de valor esperado (ou média) de X, e normalmente denotado por $E[X]$ (e muitas vezes por μ_x). O primeiro momento central é zero, enquanto o segundo momento central é chamado de *variância* de X, escrito normalmente como $Var[X]$ ou σ^2 e dado por $Var[X] = \sigma^2 = E[(X - \mu_x)^2] = E[X] - \mu^2$. Normalmente σ denota o desvio padrão de X, dado por $\sigma = \sqrt{Var[X]}$.

A *mediana* de uma variável aleatória X é o valor v tal que $P[X \geq v] \geq 1/2$ ou $P[X \leq v] \geq 1/2$.

Valor Esperado e Momentos

- Se X é uma V.A. discreta, então o n-ésimo momento é dado por

$$E[X^n] = \sum_i x_i^n P\{X = x_i\}$$

- Quando a soma não converge, o momento não existe.
- Se X é contínuo com função densidade de probabilidade dada por f(x), então o n-ésimo momento é dado por

$$E[X^n] = \int_{-\infty}^{+\infty} x^m f(x)\,dx$$

Valor Esperado e Momentos

- A variância é dada por σ^2 ou Var[X].
- Formas equivalentes para variância:

$Var[X]=E[(X-\mu)^2]$

$Var[X]=E[X^2]-\mu^2$

- A raiz quadrada da variância é chamada de desvio padrão, e denotada por σ.

Estimadores

- A média *amostral* é obtida a partir da repetição do experimento, e normalmente é denotada por:

$$\bar{x} = \sum_{i=1}^{n} \frac{x_i}{n} = \frac{1}{n} \sum_{i=1}^{n} x_i$$

- A variância *amostral* é dada por:

$$s^2 = \sum_{i=1}^{n} \frac{\left(\bar{x} - x_i\right)^2}{n}$$

Exemplo

- Suponha que foi realizada a medição das seguintes variáveis em um servidor de páginas:
  - Intervalo entre requisições;
  - Tamanho do pacote;
  - Quantidade de pacotes transmitidos por requisição;

Exemplo: Intervalo entre reqs

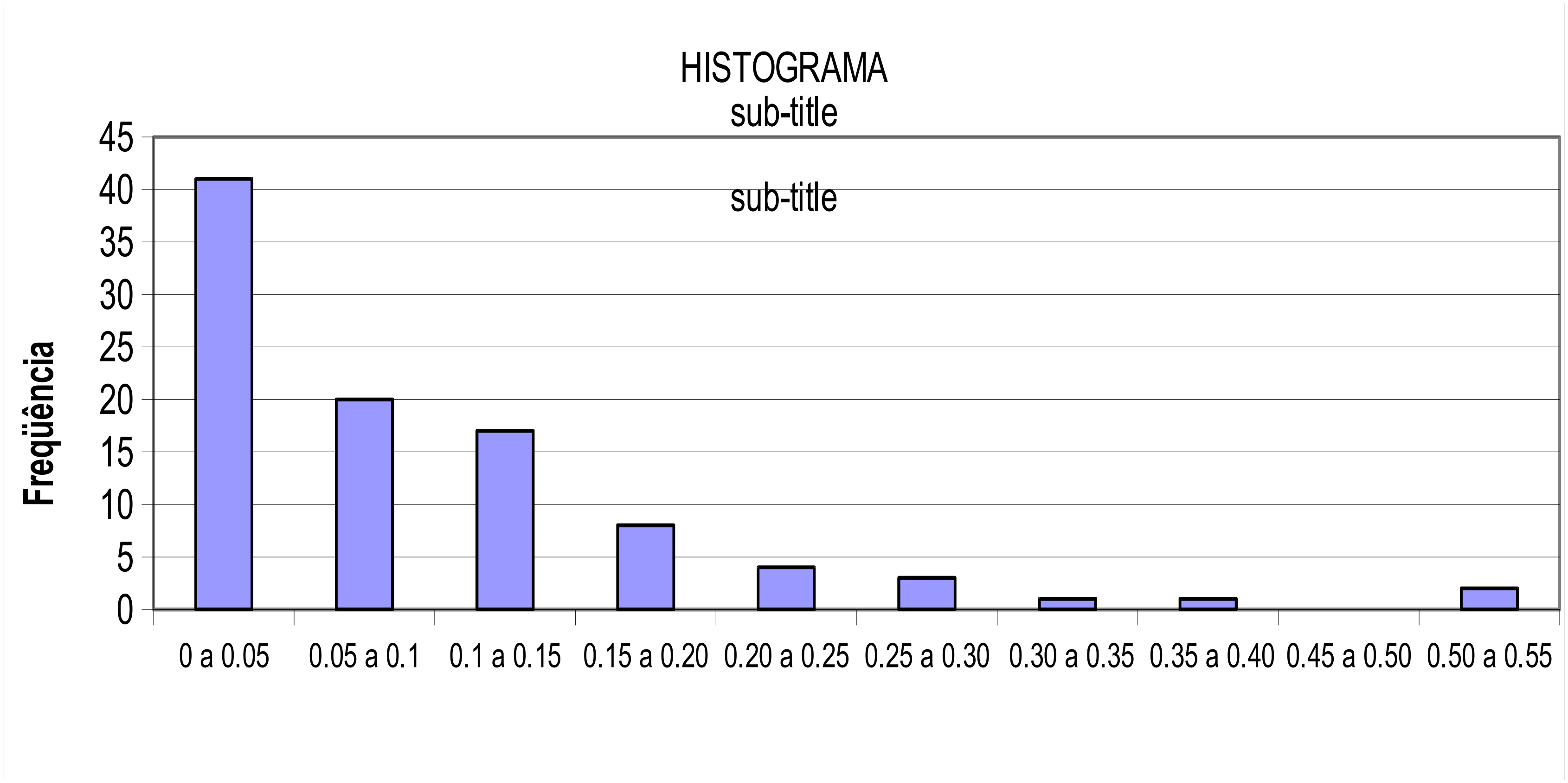
0.1493	0.0357	0.0273	0.2188	0.2188	0.0788	0.1557	0.1605	0.0337	0.2607
0.2009	0.0366	0.2935	0.0237	0.4973	0.0306	0.0059	0.0403	0.0024	0.0187
0.1448	0.0479	0.1043	0.3216	0.0102	0.0109	0.0089	0.1096	0.0579	0.0102
0.1031	0.0613	0.0358	0.0579	0.4676	0.1101	0.0259	0.0112	0.0532	0.0018
0.1628	0.1193	0.0595	0.0141	0.1058	0.0277	0.1169	0.1221	0.1622	0.0345
0.0020	0.0823	0.0105	0.0622	0.0454	0.1983	0.0673	0.1023	0.1090	0.1952
0.0066	0.0281	0.0308	0.0286	0.1379	0.0624	0.1028	0.1088	0.1308	0.1700
0.0956	0.1246	0.0285	0.0166	0.1226	0.0325	0.0434	0.0699	0.2957	0.0754
0.0719	0.0473	0.1027	0.0327	0.0074	0.0603	0.0092	0.1291	0.0518	0.3773
0.0138	0.0095	0.0584	0.1092	0.0396	0.0716	0.0271	0.1024	0.1219	0.1712

Exemplo: intervalo entre reqs

- A média pode ser estimada em *0.1139*
- O desvio padrão pode ser estimado em *0.1524*
- O histograma nos leva a:

Intervalo	Fo
0 a 0.05	41
0.05 a 0.1	20
0.1 a 0.15	17
0.15 a 0.20	8
0.20 a 0.25	4
0.25 a 0.30	3
0.30 a 0.35	1
0.35 a 0.40	1
0.45 a 0.50	0
0.50 a 0.55	2

Exemplo: Intervalo entre reqs



Teste de aderência de Chi-Quadrado

- O histograma mostrado possui semelhanças com a distribuição exponencial;
- Para verificar a qualidade da aproximação oferecida pela distribuição exponencial negativa existe o teste de Chi-quadrado.
- Calcula-se:
  - $D=(fo-fe)^2/fe$;
  - f_e é a freqüência esperada (obtida a partir da distribuição teórica e f_0 é a freqüência observada;

Teste de aderência de Chi-Quadrado

- No exemplo, obtemos:

Intervalo	Fo	Fe	D
0 a 0.05	41	35.51965	0.845568
0.05 a 0.1	20	22.90319	0.368007
0.1 a 0.15	17	14.76806	0.33732
0.15 a 0.20	8	9.522498	0.243423
0.20 a 0.25	4	6.14014	
0.25 a 0.30	3	3.959184	
0.30 a 0.35	1	2.552896	
0.35 a 0.40	1	1.646116	
0.45 a 0.50	0	1.061422	
0.50 a 0.55	2	0.684408	1.585849
			3.380167

Agrupar sempre que o número de ocorrências for pequeno

Teste de aderência de Chi-Quadrado

- O teste de aderência de Chi-quadrado compara o valor de *D* com o valor tabelado da distribuição de Chi-quadrado:
  - $\chi_{1-\alpha,\ k-r-1}$
  - onde α é o nível de significância e k é o número de graus de liberdade (é o número de classes) e r é o número de estimadores da distribuição em estudo.
  - Para que a hipótese seja aceita, *D* deve ser menor que $\chi_{1-\alpha,\ k-1}$

Teste de aderência

- Ao nível de significância de $\alpha=0.10$,
  - Da tabela: $\chi_{0.9,3}=6.251$
  - Como D<6.25, a hipótese não pode ser rejeitada.

	p										
n	0.005	0.010	0.050	0.100	0.200	0.500	0.800	0.900	0.950	0.990	0.995
1	0.000	0.000	0.004	0.016	0.064	0.455	1.642	2.706	3.841	6.635	7.879
2	0.010	0.020	0.103	0.211	0.446	1.386	3.219	4.605	5.991	9.210	10.597
3	0.072	0.115	0.352	0.584	1.005	2.366	4.642	6.251	7.815	11.345	12.838
4	0.207	0.297	0.711	1.064	1.649	3.357	5.989	7.779	9.488	13.277	14.860
5	0.412	0.554	1.145	1.610	2.343	4.351	7.289	9.236	11.070	15.086	16.750
6	0.676	0.872	1.635	2.204	3.070	5.348	8.558	10.645	12.592	16.812	18.548
7	0.989	1.239	2.167	2.833	3.822	6.346	9.803	12.017	14.067	18.475	20.278
8	1.344	1.647	2.733	3.490	4.594	7.344	11.030	13.362	15.507	20.090	21.955
9	1.735	2.088	3.325	4.168	5.380	8.343	12.242	14.684	16.919	21.666	23.589
10	2.156	2.558	3.940	4.865	6.179	9.342	13.442	15.987	18.307	23.209	25.188

Exemplo

- No exemplo anterior, suponha que os estimadores para os valores médios foram:
  - Intervalo entre requisições=*0.1139s*
  - Tamanho do pacote=503 bytes
  - Quantidade de pacotes transmitidos por requisição=12;
- Qual a capacidade mínima do enlace para este servidor?
- Quais fatores podem afetar este dimensionamento (para pior)?

Exercício

- Realize o teste de aderência para os dados abaixo, que representam o tamanho do pacote transmitido pelo servidor.

512	448	498	587	435	583	489	373	342	470
630	582	458	517	565	531	673	541	502	639
449	346	473	453	507	456	431	429	550	410
469	531	505	537	357	486	612	529	370	451
591	438	612	423	488	558	601	444	628	525
461	471	441	392	690	384	566	630	365	654
571	617	584	530	448	533	569	724	370	433
408	551	452	468	387	447	324	543	479	489
435	555	542	249	452	485	586	422	528	502
506	487	474	352	352	577	554	341	273	669

Testes de Aderência disponíveis

- Teste de Chi-Quadrado: realizado com *f(x)*
- Kolgomorov-Smirnov: realizado com *F(x)*
- QQPlot: método gráfico – plotar quantiles de *F(x)* da distribuição teórica e empírica.

Trabalho

- Fazer uma coleta de dados em um sistema:
  - Tamanho de arquivos no Unix;
  - Tamanho de arquivos no Windows;
  - Tamanho de arquivos em um servidor de páginas;
  - Outras ideias são bem vindas.
- Realizar a modelagem do sistema utilizando distribuições de probabilidade:
  - Identificar a distribuição teórica que melhor se adapta a VA observada;
  - Realizar os testes de aderência de Chi-Quadrado, Kolgomorov-Smirnov e também o

Trabalho

Referências:

1. R. Jain. The art of computer systems performance analysis: techniques for experimental design, measurement, simulation and modeling. John Wiley & Sons, 1991.
2. V. Ricci, Fitting Distributions with R, disponível em http://cran.r-project.org/doc/contrib/Ricci-distributions-en.pdf2005.