**Análise de Comprimento de Simulação
Através do Procedimento Sequencial
de Law e Carson**

CARLOS APARECIDO DOS SANTOS

ORIENTADORA: DRA. MARIA CREUSA B. SALLES

Dissertação apresentada ao Instituto de Ciências Matemáticas de São Carlos - USP, como parte dos requisitos para obtenção do Título de Mestre em Ciências - Área: *Ciências de Computação e Matemática Computacional*.

São Carlos

Dezembro/1994

A razão humana é a base de toda atividade. Antes de qualquer empreendimento é necessário que a pessoa reflita muito, não se deixando levar só pelo sentimento. A sabedoria que é aplicada só para si de nada vale, é inútil. A verdadeira pessoa sábia partilha seu saber com seu povo. Essa atitude é a melhor, por ela se obtém muitos frutos. Quando uma pessoa partilha, sua sabedoria será repleta de bênçãos de Deus.

Eclesiástico 37 - 16 ao 26.

*À meus pais Mario e Maria,
por tudo que sou;
e irmãos Sueli, Marcos e
Luis Carlos (in memorian).*

Agradecimentos ¹

A conclusão desta dissertação não seria possível sem o auxílio de várias pessoas, que contribuíram de alguma forma com este projeto, às quais expresso meus mais sinceros agradecimentos.

Principalmente agradeço a Deus, autor e princípio de todas as coisas, sem o qual, nem minha própria existência seria possível.

À Maria Creusa pela orientação, dedicação e paciência durante todo o desenvolver deste trabalho.

Aos professores Josemar Rodriguês e Jorge Achcar, com os quais tive o prazer de trabalhar e aprender mais sobre Estatística.

A meus mestres e amigos, Julia Maria e Alonso Soller, principais incentivadores para que esta dissertação um dia pudesse se tornar realidade, a eles minha eterna gratidão.

Aos professores da UNESP - Presidente Prudente Messias Meneguette e Antonio Assiz, pelo constante apoio e incentivo.

A todos meus amigos adquiridos durante o mestrado, Rosana Satie, Claudinha, Marcelo (Natal), Vera Damasceno, João Tomazella, Carlos Fogo, Luiza, Margareth, Wania, Junior, Jandira, Marta, Sadao, Adriana Favoretti, Mazé, Gaby, Fábio, Rosemeire, Omar; entre tantos outros, dos quais sentirei saudades, obrigado pela felicidade de tê-los conhecido e tornado meu convívio agradável nesta cidade. Também aqueles que colaboraram diretamente na elaboração deste trabalho, Renata, Roberta, Joãozinho, Elisa, Antonio Carlos e Silvia Chitta; meu muito obrigado.

À minha família por entender e apoiar meus objetivos.

Aos funcionários deste instituto por estarem sempre à disposição, não medindo esforços em prestar seus serviços.

A todos os amigos de Presidente Prudente e de Piracicaba, pelo constante acolhimento, companheirismo, carinho e amizade dedicados a mim durante todo o transcorrer de mais esta etapa de minha vida.

¹Este trabalho teve suporte financeiro do CNPq

Resumo

Usualmente, em simulação de sistemas, fixa-se a priori o tamanho da rodada n , e a partir dos resultados calcula-se a média final da variável aleatória de interesse e seu intervalo de confiança.

Law e Carson propõem um procedimento sequencial para determinação do tamanho da rodada, a partir da meia amplitude do intervalo de confiança fixado e da autocorrelação de *lag* 1 para as médias dos batches.

Com o procedimento sequencial de Law e Carson determinamos o comprimento médio das rodadas de simulações da fila M/M/1, com $\rho = 0,8$ e $0,5$; e γ , a meia amplitude relativa do intervalo de confiança com valores ∞ ; $0,15$ e $0,075$; a partir de 100 a 180 rodadas distintas.

O desempenho da simulação foi então analisado através da porcentagem de cobertura do valor real do tempo médio de espera na fila, para três modelos. Os resultados obtidos levou-nos a uma cobertura muito próxima de 90%, que é a confiança do intervalo de confiança construído, com correlação entre os batches muito pequena indicando uma quase independência entre eles.

Abstract

In system simulation, usually we have a run of length n fixed, and then we calculate the final mean of the random variable analysed, and its confidence interval.

Law and Carson's sequential procedure fix the half width of the confidence interval and the autocorrelation *lag* 1 for batch means, and determine the run length.

With Law and Carson's procedure we determine the run length for several M/M/1 queues, with occupation rate $\rho = 0.8$ and 0.5 , and γ relative half width of the confidence interval, as ∞ , 0.15 and 0.075 , from 100 to 180 independent runs.

The simulation results of 200 runs of some models was analysed by the percentagem of times that the real value of the mean waiting time belong to the 90% confidence interval. The results were very close to 90%, and the correlations between the batch means as small as to be considered quasi independence.

Conteúdo

1	Introdução	1
2	Teoria de Filas	6
2.1	Considerações Iniciais	6
2.2	Caracterização de um Sistema de Filas	8
2.2.1	Processo de Chegada	9
2.2.2	Processo de Serviço	9
2.2.3	Disciplina da Fila	11
2.2.4	Capacidade do Sistema	12
2.2.5	Número de Servidores	12
2.3	Notação	12
2.4	A Fila M/M/1.	13
2.4.1	Distribuição de Equilíbrio	14
2.4.2	Medidas de Performance	16
3	Simulação Estocástica	19
3.1	Considerações Iniciais	19

3.2	Tipos de sistemas e modelos	20
3.3	Vantagens e Desvantagens da Simulação	21
3.3.1	Etapas em um Estudo de Simulação	22
3.3.2	Linguagens de Simulação	24
3.4	Geração de Números Pseudo-aleatórios	25
3.5	Métodos de Coleta de Dados	27
3.5.1	Método de Replicações	28
3.5.2	Método do Batch Means	30
3.5.3	Método Regenerativo	33
3.6	Quantidade de Batches	34
3.7	Período Transiente	35
4	Procedimento Sequencial	38
4.1	Considerações Iniciais	38
4.2	Tipos de Tamanho de Amostra	39
4.3	Procedimento de Law e Carson	42
4.4	Resultados Apresentados por Law e Carson	44
5	Descrição e Análise dos Resultados da Simulação	46
5.1	Considerações Iniciais	46
5.2	Descrição do Algoritmo Sequencial de Law e Carson	47
5.3	Um Exemplo de como Funciona o Algoritmo	49
5.4	Simulação da Fila M/M/1 Através do Procedimento Sequencial	51
5.5	Análise dos Resultados	56

<i>Conteúdo</i>	iii
5.6 Análise da Porcentagem de Cobertura	59
5.6.1 Resultados Encontrados	60
5.6.2 Porcentagem de Cobertura Obtida	62
6 Conclusões Finais e Propostas Futuras	65
6.1 Conclusão	65
6.2 Propostas Futuras	67
Bibliografia	68
Apêndices	
A Estimador Jackknife	71
B Programa de Simulação	74
C Tamanhos de Amostras	77

Lista de Figuras

2.1	Sistemas de Fila	8
2.2	A Fila M/M/1	14
3.1	Método de Replicações	28
3.2	Método de Batch Means	31
3.3	Método Regenerativo	34
4.1	Comportamento Padrão Simples	43

Lista de Tabelas

5.1	Obtenção de um Particular Tamanho de Amostra, considerando $\rho = 0,8$ e $\gamma = \infty$; 0,15 e 0,075	53
5.2	Obtenção de um Particular Tamanho de Amostra, considerando $\rho = 0,5$ e $\gamma = \infty$; 0,15 e 0,075	54
5.3	Resultados obtidos considerando $\rho = 0,8$	55
5.4	Resultados obtidos considerando $\rho = 0,5$	56
5.5	Resultados adicionais considerando $\rho = 0,8$	57
5.6	Resultados adicionais considerando $\rho = 0,5$	58
5.7	Resultados para a fila M/M/1 com $\rho = 0,8$ (modelo I e II) . . .	60
5.8	Autocorrelações para a fila M/M/1 com $\rho = 0,8$ e $\gamma = 0,15$ (modelo I)	61
5.9	Autocorrelações para a fila M/M/1 com $\rho = 0,8$ e $\gamma = 0,075$ (modelo II)	61
5.10	Resultados para a fila M/M/1 com $\rho = 0,5$ (modelo III)	62
5.11	Autocorrelações para a fila M/M/1 com $\rho = 0,5$ e $\gamma = 0,075$ (modelo III)	62
5.12	Porcentagem de Cobertura obtida	63

Capítulo 1

Introdução

A teoria de filas estuda sistemas com clientes requerendo algum tipo de serviço, sendo que muitos desses sistemas, por terem capacidade limitada de atendimento, acabam levando à contenção de recursos e consequente formação de filas. Esta é uma área de muita utilidade para vários profissionais; tendo importantes aplicações documentadas na literatura de probabilidade, pesquisa operacional e ciências de engenharia. Também podemos encontra-la em atividades corriqueiras, como ir a um banco ou um supermercado; ou ainda em outras menos usuais, como na utilização de um computador ou no controle de um tráfego aéreo. Em computação encontramos, principalmente na área de redes de computadores e sistemas distribuídos, importantes aplicações devido ao seu grande avanço e dos problemas atuais cada vez mais complexos que surgem nestas áreas.

Quando modelamos um sistema através da teoria de filas, devemos fazê-lo com o máximo de detalhes possível, pois os resultados devem apresentar as características essenciais do sistema analisado. Essas características seriam, por exemplo, o número de servidores disponíveis, os processos de chegada e serviço, se existe ou não prioridade entre os clientes; e se ao terminar um serviço o cliente vai embora ou retorna ao sistema.

Vários tipos de sistemas de filas são apresentados na literatura, entre eles, sistemas com um ou mais servidores, redes de Jackson, filas clássicas com tempos de serviços gerais, filas com férias do servidor, etc (Cooper [COO 90] e Disney e König [DEK 85]).

Como trabalhamos com a fila em equilíbrio, calculamos a distribuição de equilíbrio para o sistema de fila M/M/1, que é utilizado nesta pesquisa. Através da distribuição de equilíbrio encontrada, apresentamos as medidas de performance mais utilizadas. Enfatizamos que não só as médias, mas também a distribuição de probabilidade das diversas medidas de performance, são de interesse do pesquisador. Isto é apresentado no capítulo 2 desta dissertação de mestrado.

Segundo Soares [SOA 90] a técnica analítica é o método geralmente mais rápido e eficiente, portanto o preferido quando aplicável na resolução de problemas, neste caso, aqueles que envolvem filas de espera. Diversos sistemas do mundo real são muito complexos para serem avaliados analiticamente, exigindo que muitas hipóteses de simplificação sejam feitas para que possamos resolver o modelo. Portanto a simulação surge como uma alternativa.

A solução analítica nos fornece um resultado exato para o modelo em estudo, o que não acontece com a simulação. Mas devido as simplificações, o modelo analítico pode não ser uma representação precisa do sistema real, e o resultado exato obtido ser desprovido de qualquer utilidade, pois é bem provável que para resolvermos um modelo de redes de filas mais complexo analiticamente, várias restrições deverão ser feitas.

A simulação estocástica é um método de solução que tenta imitar o comportamento do sistema, constituindo-se em um experimento estocástico que observa o comportamento do modelo no decorrer do tempo. O modelo pode ser exercitado com dados obtidos de medidas do sistema ou mais usualmente, com valores aleatórios gerados a partir de distribuições de probabilidades, com a finalidade de representar o tempo de chegada dos clientes, tempos de serviço e as probabilidades de transição dentro do sistema. Os valores aleatórios gerados não são de fato aleatórios mas, por serem produzidos por um algoritmo, são designados como números pseudo-aleatórios.

Por outro lado, a simulação estocástica é atualmente uma poderosa ferramenta na análise de modelos estocásticos, e mais particularmente, processos de filas. Ela é mais geral e pode ser aplicada em situações mais complexas e novas (aquelas que estão na idéia ou na prancheta do projetista). Para Macdougall [MAC 87], o preço a ser pago pela generalidade é que um tempo maior é exigido na obtenção das medidas de performance.

Contudo, mesmo quando decidimos construir um modelo detalhado de

simulação, é uma boa idéia ter um modelo analítico mais simples, que possa ser resolvido, e que seja uma aproximação do modelo simulado. O modelo analítico além de fornecer um meio de constatação parcial de que o modelo de simulação está funcionando corretamente, poderá ser usado para a obtenção de estimativas das medidas de performance.

Nosso interesse é em simulação de sistemas estocásticos em equilíbrio, não significando que o sistema permanecerá o mesmo a partir de um determinado instante, mas simplesmente que as distribuições de probabilidades associadas com o sistema convergirão para uma distribuição limite.

Em sistemas reais, uma condição transiente vai sempre existir inicialmente, quando o sistema é ativado e os clientes começam a chegar para serem servidos. Se o padrão de chegada não se altera e o sistema tem a capacidade de fornecer o serviço requerido, a condição de equilíbrio pode existir.

Tem-se feito análises comparativas de simulações utilizando, ou não, o período transiente para se verificar a influência do mesmo na porcentagem de cobertura, para alguns métodos de coletas de dados (por exemplo, Bernasconi [BER 93] e Law [LAW 77]). Law [LAW 83] mostrou para uma fila M/M/1 como o tempo de espera na fila W_q , é afetado pelo número de clientes no sistema no tempo 0, depois de digamos n fins de serviços. A mais rápida convergência do sistema só foi possível com um número inicial de clientes muito maior que a média. O problema fica mais difícil para sistemas com múltiplas filas, já que seus comportamentos conjuntos são considerados. O capítulo 3 aborda assuntos relevantes à simulação estocástica.

Neste trabalho, nós analisamos o desempenho da simulação utilizando um modelo de fila M/M/1.

Nosso objetivo principal é determinar tamanhos de amostra, utilizando um procedimento sequencial, com uma *precisão relativa* do intervalo de confiança (meia amplitude do intervalo de confiança dividido pela sua média) fixada.

Um problema importante, mas de difícil solução encontrado em estudos de simulação do mundo real, é o de construir um intervalo de confiança para a média μ em equilíbrio, de um processo estocástico. Este problema tem requerido considerável atenção na literatura, pois, sabemos que o resultado da simulação é um valor estimado, e o intervalo de confiança construído em torno dele, nem sempre contém o valor real da variável de interesse.

O que tem sido observado, na prática, é que em muitos estudos de simulação de sistemas uma grande quantidade de tempo e dinheiro é gasto no desenvolvimento do modelo e em sua programação, mas pouco esforço tem sido feito para analisar os dados de saída da simulação de maneira apropriada. Operacionalmente é comum fazer uma rodada única de comprimento arbitrário e então tratar as estimativas resultantes da simulação como sendo as "reais" medidas de performance para o sistema que está sendo analisado.

Estas estimativas são resultados de variáveis aleatórias, que podem ter variâncias muito grandes, por isso o resultado de uma particular rodada de simulação pode diferir grandemente da medida real. A consequência disto é óbvia, pois podemos ter uma chance significativa de fazer inferências inadequadas sobre o sistema estudado, já que os resultados da simulação refletem o mesmo, mas não seu resultado exato.

Uma maneira de construímos estes intervalos de confiança é fixarmos o comprimento da rodada antes de começarmos a simulação. Esta metodologia não é muito segura, pois se o comprimento da rodada for muito pequeno podemos ter resultados não adequados ao sistema que será simulado. A razão para isso é clara, que diferentes sistemas estocásticos podem funcionar de modo radicalmente distintos, e então requererem comprimentos de rodada muito diferentes para construir adequados intervalos de confiança.

Este problema pode ser resolvido se aumentarmos iterativamente o comprimento da rodada, que é chamado na literatura de *procedimento sequencial*. Law e Kelton [LEK 82] apresentam os principais procedimentos sequenciais encontrados na literatura, abordando com detalhes quatro destes procedimentos, classificando-os em *regenerativos* e *não regenerativos*. No capítulo 4 apresentamos estes procedimentos, dando ênfase ao procedimento de Law e Carson [LEC 79], que é trabalhado nesta pesquisa.

Pawlikowski [PAW 90] propõe dois procedimentos sequenciais para a parada de experimentos de simulação quando uma precisão relativa do intervalo de confiança é atingida. Um dos procedimentos é baseado no *método de análise espectral*, e o outro aplica o *método de batch means nonoverlapping*; sendo que ambos requerem estacionaridade da sequência de observações analisadas, onde cada um desses procedimentos apresentam um mecanismo para a obtenção do período transiente. Porém não foram feito estudos exaustivos sobre a eficiência destes métodos, mas resultados preliminares indicam que o método espectral é mais eficiente.

No procedimento de Law e Carson, inicialmente fixamos um tamanho mínimo de amostra e sequencialmente a aumentamos até que um intervalo de confiança possa ser construído satisfazendo a um critério proposto pelo procedimento. Isto é feito para os dados de saída referentes ao tempo médio de espera na fila W_q para uma fila M/M/1, considerando duas taxas de ocupação, uma de 80% e outra de 50%.

Abordamos mais detalhadamente procedimentos sequenciais no capítulo 4, e o desenvolvimento do algoritmo sequencial de Law e Carson, juntamente com os resultados obtidos deste procedimento, tanto por nós como pelos autores, são apresentados no capítulo 5.

O outro objetivo deste trabalho é, com os tamanhos de amostras encontrados, utilizando o procedimento sequencial de Law e Carson, fazer várias rodadas de simulação visando construir intervalos de confiança em torno do valor estimado da variável analisada W_q . Isto é feito para verificar a *porcentagem de cobertura*, ou seja, a proporção de vezes que o valor real da variável analisada pertence aos intervalos de confiança construídos.

Além da porcentagem de cobertura, verificamos também as autocorrelações de *lag* l ; $l = 1, 2, 3$; considerando a estimação usual e a Jackknife (no apêndice A detalhamos este estimador, que é utilizado por ser menos viciado). Os resultados referentes a porcentagem de cobertura e as autocorrelações também são apresentados no capítulo 5.

As simulações foram feitas utilizando o programa GPSS/H (*General Purpose Simulation System, version H*), sendo que o programa de simulação referente ao sistema estudado pode ser encontrado no apêndice B; e a análise estatística dos resultados de saída foram feitas por intermédio do software estatístico MINITAB. Apresentamos as conclusões finais deste trabalho, bem como algumas propostas futuras para continuidade desta pesquisa no capítulo 6.

□

Capítulo 2

Teoria de Filas

Este capítulo aborda tópicos que apresentam e caracterizam um sistema de filas. O enfoque é dado à fila M/M/1, apresentando sua distribuição de equilíbrio, e algumas de suas medidas de performance que são comumente mais utilizadas. Como uma bibliografia básica para este capítulo temos Gross e Harris [GEH 85], Allen [ALL 78] e White, Schmidt e Bennet [WSB 75].

2.1 Considerações Iniciais

A teoria de filas foi desenvolvida para proporcionar modelos predizendo o comportamento de sistemas que procuram propiciar serviços para clientes¹ que chegam aleatoriamente ou deterministicamente. Inicialmente esta teoria esteve ligada a problemas de congestionamento de tráfego em telefone. Em 1909, A. K. Erlang publicou o trabalho intitulado *The Theory of Probabilities and Telephone Conversations* [ERL 09], onde observou um sistema de telefone.

Telefonia ficou sendo a principal aplicação da teoria de filas durante os anos de 50's. A teoria de filas teve um grande impulso com a Segunda Guerra Mundial, com problemas de estratégia militar, bem como com o desenvolvi-

¹O termo *cliente* é usado em um sentido geral e não implica necessariamente um cliente humano. Por exemplo, um cliente pode ser uma máquina, um livro, uma requisição ou um programa computacional esperando para ser rodado.

mento/uso de computadores que acarretaram no surgimento de vários problemas de filas, que a cada dia se tornam mais complexos. Deste então, numerosas outras aplicações tem sido encontradas, muitas das quais tem sido documentadas nas literaturas de probabilidade, pesquisa operacional e ciências de engenharia (Gross e Harris [GEH 85]).

Algumas das diversas aplicações da teoria de filas são em reparos de máquinas, cabines de pedágio, ponto de táxi, controle de estoques, e embarque e desembarque de navios. Como filas também são comuns em sistemas de computação, temos por exemplo, fila de pessoas esperando para usar um terminal computacional, fila de espera de mensagens para serem processadas por um sistema de computação central, fila de terminais de computadores com entrada/saída, fila de processamento em rede, entre muitas outras situações. Uma área nova particularmente fértil de aplicações é a modelagem de fluxo de mensagens em comunicação digital através de redes complexas.

Atualmente, profissionais das mais variadas áreas, tais como, engenheiros, profissionais em computação (principalmente os que trabalham com redes de computação e sistemas distribuídos), físicos, entre outros, utilizam teoria de filas para resolverem seus problemas.

Um sistema de fila pode ser descrito como clientes chegando aleatoriamente para um serviço, aguardando para serem atendidos numa sala de espera se o serviço não acontece imediatamente, e caso tenham esperado pelo serviço, vão embora depois de serem atendidos; como podemos ver na figura 2.1. O sistema de fila determinístico é um caso particular do modelo aleatório.

Embora alguns sistemas de fila possam ser esquematizados desta maneira, devemos deixar claro que uma representação com razoável precisão de um dado sistema poderá requerer uma detalhada caracterização do processo de interesse.

Sempre que analisamos filas de espera, simples ou complexas, temos que considerar que este é um estudo de sistemas de filas, que tem suas dificuldades, pois se houver fornecimento de serviço além do necessário, resultará em gastos e prejuízos com o servidor e se, por outro lado, o sistema não proporcionar um atendimento mínimo, poderá haver perda de clientes e/ou grandes filas de espera.

Os resultados apresentados a seguir são de caráter geral, podendo ser encontrado na literatura básica de Teoria de Filas.

Figura 2.1: Sistemas de Fila

2.2 Caracterização de um Sistema de Filas

A caracterização de um sistema de filas é feita principalmente de acordo com os processos e características que descreveremos a seguir; o que nos proporciona uma descrição adequada do sistema que está sendo estudado.

2.2.1 Processo de Chegada

O processo de chegada para um sistema de fila é determinado pela distribuição de probabilidade, através da função de distribuição acumulada, $A(t)$, dos tempos entre-chegadas de clientes para o sistema, associado a este processo aleatório.

Então, se $\tau_n = t_n - t_{n-1}$, $n = 1, 2, \dots$, é o tempo entre chegadas (t_n é o instante de chegada do n -ésimo cliente), temos:

$$A(t) = P[\tau_n \leq t]$$

Podemos ter também a possibilidade que as chegadas ocorram em lotes (batches), em lugar de um cliente de cada vez. Nesta situação, não estamos interessados apenas na aleatoriedade dos tempos entre sucessivas chegadas, mas também no tamanho dos batches.

É também necessário conhecer a reação de um cliente após entrar no sistema. Se um cliente decide não entrar na fila após chegar ao sistema, dizemos que ele é um *cliente desistente*. Por outro lado, um cliente pode entrar na fila, mas depois de um certo tempo perde a paciência e decide abandoná-la; neste caso ele é um *cliente impaciente*.

Um processo de chegada que não muda com o tempo, isto é, a forma e os valores dos parâmetros da distribuição de probabilidade descrevendo o processo de chegada são independentes do tempo, é chamado de processo de chegada *estacionário*. Quando não há independência do tempo, neste sentido, dizemos que o processo de chegada é *não estacionário*.

De modo geral, para sistemas estacionários, adota-se sempre que os tempos entre chegadas são variáveis aleatórias independentes e identicamente distribuídas.

2.2.2 Processo de Serviço

Muito da discussão sobre o processo de chegada é apropriado para o processo de serviço. O processo de serviço está relacionado com o tempo necessário para que um servidor execute uma tarefa. Também é descrito através da distribuição de probabilidade dos tempos de serviço.

De qualquer modo existe uma diferença importante entre os processos de chegada e de serviço. Quando falamos em taxa de serviço ou tempo de serviço, estes termos estão condicionados ao servidor, ou seja, estes elementos são características próprias do servidor.

O processo de serviço é especificado pela função distribuição acumulada, $B(t)$, dos tempos de serviço para cada um dos clientes, isto é,

$$B(t) = P[\text{tempo de serviço} \leq t]$$

que geralmente também são considerados independentes e identicamente distribuídos.

Normalmente, assume-se que o processo de serviço é independente do processo de chegada.

Como no processo de chegada, o processo de serviço determinístico é um caso particular do modelo aleatório.

Existem sistemas onde a função de distribuição acumulada da variável aleatória do processo de serviço, pode depender do número de clientes esperando para serem atendidos. Um servidor pode trabalhar mais rápido se vê que a fila está acumulando, ou inversamente, ele pode ficar atrapalhado e se tornar menos eficiente. Esta situação, é denominada como *serviço dependente do estado*. A dependência de estado não depende do instante do sistema, mas somente do estado do mesmo em um dado momento, ou seja, da quantidade de clientes que estão atualmente no sistema.

O serviço pode ser feito um a um ou em batches. Geralmente associa-se um único cliente sendo servido em um dado instante pelo servidor, mas há muitas outras situações onde mais de um cliente são servidos simultaneamente por um mesmo servidor. Exemplos de tal situação seria um computador com processamento em paralelo, turistas guiados em uma excursão, ou pessoas embarcando em um trem.

Em geral, clientes chegam e partem em intervalos aleatórios, então o comprimento da fila, o número de clientes no sistema, etc; também são aleatórios.

Além do que foi apresentado, podemos ter, por exemplo, o *período do tempo de ocupação*, que é uma distribuição do comprimento do tempo no qual

o servidor permanece continuamente atarefado; e o *processo de partida*, que é o processo imerso nos instantes de saídas dos clientes do sistema. O processo de partida é um processo derivado do processo de chegada e do processo de serviço e não é dependente da disciplina da fila, como é o processo do tempo de espera. Não faremos uma abordagem mais detalhada desses processos por não serem de nosso interesse, que são encontrados em Disney [DIS 75].

2.2.3 Disciplina da Fila

A disciplina da fila refere-se ao modo pelo qual os clientes são selecionados para serem servidos quando uma fila está formada. As disciplinas mais comuns são:

- FCFS (*first-come-first-served*): atende aos clientes pela ordem de chegada, sendo também a mais comum.
- LCFS (*last-come-first-served*): atende aos clientes pela ordem inversa de chegada.
- SIRO (*service in random order*) ou RSS (*random-selection-for-service*): seleciona os clientes em ordem aleatória, independente do instante de chegada na fila.
- PRI (*priority service*): associa prioridade aos clientes que chegam no sistema.
- RR (*round robin*): atende cada cliente por um pequeno intervalo de tempo, denominado "quantum". Caso não tenha sido suficiente para completar o serviço requerido pelo cliente, este volta para o final da fila de espera, e o cliente seguinte entra para receber seu "quantum" de serviço. O processo se repete até que o serviço requisitado pelo cliente se complete.

Há duas situações gerais para as disciplinas com prioridade. Na primeira, que é chamado *com preempção*, o cliente com maior prioridade é alocado para entrar em serviço imediatamente, mesmo se um outro com menor prioridade esteja sendo atendido quando ele entra no sistema. Na segunda, chamado *sem preempção*, o cliente de maior prioridade vai encabeçar a fila, mas não pode entrar em serviço até que aquele que está sendo atendido termine, embora este tenha menor prioridade.

2.2.4 Capacidade do Sistema

Em alguns sistemas de fila há uma limitação física do número de clientes que cabem na sala de espera. Quando a fila atinge um comprimento determinado, nenhum cliente pode adentrar a mesma até que um espaço livre seja criado, como por exemplo, através de um serviço completado. Em situações deste tipo temos o tamanho máximo da fila finito. Vale ressaltar que há sistemas limitados com capacidade zero, sendo chamado *sistema sem sala de espera*.

Outros tipos de sistema assumem que a capacidade da fila é infinita, ou seja, todos os clientes que chegam são direcionados a esperar na fila até que possam ser atendidos.

2.2.5 Número de Servidores

O sistema de fila simples, neste sentido, é o *sistema de servidor único* que pode servir somente um cliente de cada vez. Um *sistema multi-servidor* tem usualmente c servidores, podendo portanto proporcionar serviço para c clientes simultaneamente.

Em um *sistema de infinitos servidores*, para cada cliente que chega é imediatamente providenciado um servidor. Embora usualmente não podemos ter sistemas reais com um número infinito de servidores, há sistemas que tem servidores em quantidade suficiente para atender todos os clientes que chegam.

2.3 Notação

Uma notação especial, chamada *notação de Kendall*, foi desenvolvida para descrever sistema de fila. Ela tem a seguinte forma:

$$A/B/c/K/m/Z$$

onde A descreve a distribuição do tempo entre-chegadas, B a distribuição do tempo de serviço, c o número de servidores, K a capacidade do sistema (número

máximo de clientes que o sistema comporta), m o tamanho da população, e Z a disciplina da fila.

Atualmente estuda-se filas mais complexas e a notação utilizada têm se adequado de maneira diferente às filas, onde usamos os três ou quatro primeiros elementos da notação proposta por Kendall e explicamos a disciplina, a prioridade, etc., de modo literal.

Na notação abreviada $A/B/c$, assume-se que não há limite no comprimento da fila e a disciplina é a que atende aos clientes pela ordem de chegada.

É comum usarmos para A e B

M : distribuição do tempo entre-chegadas ou serviço exponencial

GI : tempo entre-chegadas ou de serviço geral independente

G : tempo entre-chegadas ou de serviço geral

H_k : distribuição do tempo entre-chegadas ou serviço hiperexponencial de k -estágios

E_k : distribuição do tempo entre-chegadas ou serviço Erlang- k

D : distribuição do tempo entre-chegadas ou serviço determinística (constante)

U : distribuição do tempo entre-chegadas ou serviço uniforme

2.4 A Fila M/M/1.

Devido ao modelo desenvolvido nas simulações deste trabalho, consideremos uma fila M/M/1, onde os tempos entre-chegadas dos clientes tem distribuição exponencial de parâmetro λ , ou seja, o processo de chegada forma um Processo de Poisson; os tempos de serviço são variáveis aleatórias independentes com distribuição exponencial de parâmetro μ e temos apenas um único servidor, que atende aos clientes pela sua ordem de chegada. A sala de espera para os clientes tem capacidade infinita. A figura 2.2 descreve a fila M/M/1.

Neste caso, a função distribuição acumulada do tempo entre-chegadas bem como a função distribuição acumulada do tempo de serviço são ambas

Figura 2.2: A Fila M/M/1

exponenciais, ou seja,

$$A(t) = 1 - e^{-\lambda t}, \quad t \geq 0$$

onde λ é a taxa de chegada no sistema e,

$$B(t) = 1 - e^{-\mu t}, \quad t \geq 0$$

onde μ é a taxa de atendimento.

Todos os resultados que serão apresentados estão amplamente desenvolvidos na literatura.

2.4.1 Distribuição de Equilíbrio

Para obtermos as medidas de desempenho para este sistema em equilíbrio, que serão distribuições de probabilidades ou no mínimo valores médios, inicialmente devemos encontrar a distribuição de equilíbrio do sistema. Faremos isto porque estamos interessados em analisar a fila M/M/1 em equilíbrio.

Seja o processo estocástico $\{N_t : t \geq 0\}$ que conta o número de clientes no sistema num instante $t \geq 0$. Este é um Processo Markoviano devido ao fato de que os tempos entre-chegadas e de serviço seguem a distribuição exponencial e portanto tem a propriedade de "falta de memória". O espaço de estados é dado por $E = \{0, 1, 2, \dots\}$

Para obtermos a distribuição de equilíbrio para $\{N_t : t \geq 0\}$ precisamos resolver o sistema de equações:

$$\Pi Q = 0 \quad (2.1)$$

sujeito às condições

1. $0 \leq \Pi_n \leq 1; \forall n$
2. $\sum_{n=0}^{\infty} \Pi_n = 1$

onde Π é o vetor distribuição de equilíbrio, ou seja, $\Pi = (\Pi_0, \Pi_1, \dots)$ e $\Pi_n = \lim_{t \rightarrow \infty} Pr[N_t = n] = Pr[N = n]$ é a probabilidade de termos n pessoas no sistema quando o mesmo atingir o equilíbrio; e $Q = (q_{ij}); (i, j) \in E$ que é a matriz de taxas infinitesimais do processo $\{N_t : t \geq 0\}$, e para este caso é dada por

$$Q = \begin{bmatrix} -\lambda & \lambda & 0 & 0 & 0 & \dots \\ \mu & -(\lambda + \mu) & \lambda & 0 & 0 & \dots \\ 0 & \mu & -(\lambda + \mu) & \lambda & 0 & \dots \\ 0 & 0 & \mu & -(\lambda + \mu) & \lambda & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}$$

O sistema de equações lineares 2.1, satisfazendo a condição de equilíbrio

$$\rho = \frac{\lambda}{\mu} < 1 \quad (2.2)$$

terá solução única; caso contrário terá solução nula. Resolvendo o sistema 2.1, obtemos a solução da distribuição de equilíbrio do processo dada por

$$\Pi_n = (1 - \rho)\rho^n; \quad n = 0, 1, 2, \dots$$

onde, ρ é chamado de taxa de ocupação do sistema.

Obtida a distribuição de equilíbrio para o processo estocástico $\{N_t : t \geq 0\}$, podemos encontrar algumas medidas de performance para este processo no equilíbrio.

2.4.2 Medidas de Performance

Com relação a estas medidas, geralmente estamos interessados em três respostas sobre o sistema, que são:

1. Algumas medidas do tempo de espera a que um cliente pode se sujeitar;
2. Uma indicação da maneira pela qual os clientes podem se acumular;
3. Uma medida do tempo de ociosidade dos servidores.

Já que sistemas de filas são aleatórios, estas medidas são frequentemente variáveis aleatórias, e suas distribuições de probabilidade ou no mínimo seus valores esperados, precisam ser determinados.

Para um pesquisador de teoria de filas, a tarefa deve ser a de determinar os valores das medidas de eficiência apropriados a um dado processo, ao planejar um sistema "ótimo" (de acordo com alguns critérios). Ele deve relacionar o tempo da espera e o comprimento da fila de acordo com as propriedades do fluxo de entrada e os procedimentos de serviço. Se o tempo de espera e do período ocioso puderem ser obtidos diretamente, eles serão usados para determinar o número ótimo de canais de serviço, juntamente com suas taxas de operação. Também para planejar a espera é necessário informações a respeito do possível tamanho da fila, afim de projetar a sala de espera.

Em alguns casos o pesquisador resolverá esses problemas de forma analítica, mas quando não for possível, ele deverá recorrer à simulação (Cohen [COH 69]).

A partir da distribuição de equilíbrio do sistema de fila M/M/1, podemos calcular suas medidas de performance. As medidas de interesse imediato para nosso trabalho são:

- *Número esperado de clientes no sistema*

$$L = E[N] = \sum_{n=0}^{\infty} n \Pi_n = \frac{\rho}{1 - \rho}$$

Equivalentemente

$$L = \frac{\lambda}{\mu - \lambda}$$

- *Número esperado de clientes na fila*

$$L_q = \mathbf{E}[N_q] = 0\Pi_0 + \sum_{n=1}^{\infty} (n-1)\Pi_n = \frac{\rho^2}{1-\rho}$$

onde N_q é o número de clientes que estão na fila, isto é

$$N_q = \begin{cases} N-1 & \text{se } N \geq 1 \\ 0 & \text{se } N = 0 \end{cases}$$

Equivalentemente

$$L_q = \frac{\lambda^2}{\mu(\mu - \lambda)}$$

As medidas de performance referentes ao *tempo médio de permanência do cliente no sistema* e *tempo de espera do cliente*, para a fila M/M/1, também podem ser determinadas através da distribuição do tempo de espera na fila.

- *Tempo médio de permanência do cliente no sistema*

$$W = \frac{1}{\mu - \lambda}$$

- *Tempo médio de espera do cliente na fila*

$$W_q = \frac{\lambda}{\mu(\mu - \lambda)} \quad (2.3)$$

Estas medidas de performance *tempo médio de permanência do cliente no sistema* e *tempo médio de espera do cliente na fila*, citadas anteriormente, também podem ser obtidas mais facilmente através da fórmula de Little, que é dada por:

$$L = \lambda W$$

onde W é *tempo médio de permanência do cliente no sistema*.

Também é importante constatar que o *tempo médio de permanência do cliente no sistema* é igual ao seu *tempo médio de espera na fila* mais o seu *tempo médio de serviço*, ou seja:

$$W = W_q + 1/\mu$$

Dos resultados mencionados anteriormente, o que utilizaremos no contexto deste trabalho é o *tempo médio de espera na fila* (W_q), que será o valor comparado com os resultados obtidos através das simulações.

Nosso interesse é em simulação de modelos matemáticos de sistemas de filas, em que o modelo seja uma representação lógica, não física, do sistema estudado. Como trabalhamos a fila M/M/1, visando obter melhores resultados utilizando simulação, apresentamos o capítulo seguinte abordando esta alternativa de resolução de problemas.

□

Capítulo 3

Simulação Estocástica

Este capítulo discute tópicos relevantes sobre modelagem e simulação de sistemas discretos. Abordamos as vantagens e desvantagens no uso da simulação; e descrevemos as etapas envolvidas em uma simulação. Também enfocamos alguns métodos de coleta de dados. Os resultados e definições apresentados neste capítulo podem ser encontrados em Ripley [RIP 87], Law e Kelton [LEK 91], e Cooper [COO 90].

3.1 Considerações Iniciais

Como abordado anteriormente, os problemas atuais enfrentados pela indústria, centros de pesquisas, comércio, órgãos públicos, entidades de prestação de serviços e a sociedade em geral são cada vez maiores e mais complexos, exigindo, cada vez mais, técnicas e procedimentos para resolvê-los. Uma técnica que é usada na resolução e análise desses problemas é a simulação.

Simulação pode ser definida como uma metodologia onde um sistema é estudado pela observação de um modelo através de dados gerados aleatoriamente. Embora o conceito de simulação não dependa da existência de um computador, foi com o advento do computador digital de alta velocidade que tem feito da simulação uma ferramenta muito útil na análise de sistemas. Falaremos de sistemas e modelos na seção 3.2.

Sabemos que no sentido técnico, simulação envolve o uso de um modelo computacional que reflita um sistema real que existe. Quando testes e experimentos com o sistema real seriam caros demais e/ou destrutivos; ou o sistema ainda será implementado, usamos esta técnica para obter os resultados desejados. Por exemplo, ela é usada para verificar o tempo de utilização de terminais em uma rede conectada a um computador central. Se o sistema tem um elemento estocástico (tempo, volume, entre outros), usamos a Simulação Estocástica, de modo que possamos analisar este sistema no decorrer de um intervalo de tempo e não apenas num dado instante.

Em algumas aplicações, a simulação estocástica é o primeiro passo para se ter uma idéia intuitiva das generalidades do problema, antes de tentar resolvê-lo com outras ferramentas.

Existe uma *regra de ouro* para a simulação estocástica, que é enunciada da seguinte maneira (Bustos e Frery [BEF 92]):

Nenhum problema para o qual existam soluções teóricas alcançáveis no tempo requerido e a custos toleráveis deve ser resolvido usando Simulação Estocástica.

Como trabalhamos com simulação estocástica visando principalmente validá-la, enfatizamos esta regra com o objetivo de chamar a atenção dos usuários da mesma.

A simulação é uma ferramenta poderosa e muito vantajosa se bem utilizada. Como trabalhamos com modelos que procuram representar o sistema real de interesse, é imprescindível não esquecermos as simplificações feitas no modelo e resistir à tentação de generalizar as conclusões obtidas na simulação.

3.2 Tipos de sistemas e modelos

Podemos caracterizar os sistemas como sendo discretos e contínuos. Um *sistema discreto* é aquele para o qual a variável observada muda somente em uma quantidade enumerável (ou finita) de pontos, dentro de um intervalo de

tempo. Um *sistema contínuo* é aquele no qual a variável de estado muda continuamente com respeito ao tempo. Poucos sistemas, na prática, são completamente discretos ou contínuos, mas normalmente há uma predominância nos mesmos, que os tornam possíveis de serem classificados como discretos ou contínuos.

Modelo é uma descrição do sistema que elimina detalhes não essenciais, ou seja, modelo é uma abstração do sistema real.

O tipo de modelo escolhido para descrever algum sistema real dependerá do conhecimento e bom senso do analista. Vale salientar que um modelo construído deve refletir um equilíbrio entre exatidão e tratabilidade.

Um sistema modelado em um instante qualquer é o que chamamos *modelo de simulação estático*. O método de Monte Carlo de simulação é tipicamente deste tipo, já que ele pode ser definido como uma técnica empregando números pseudo-aleatórios (seção 3.4), que são usados principalmente na resolução de problemas numéricos.

O modelo de fila M/M/1 abordado neste trabalho é do tipo discreto e estocástico, e por ter essas características é denominado *modelo de simulação de evento discreto*.

3.3 Vantagens e Desvantagens da Simulação

Apresentamos a seguir algumas das possíveis razões para a popularidade da simulação:

1. Sistemas do mundo real muito complexos, com elementos estocásticos, dificilmente podem ser descritos por um modelo matemático que possa ser resolvido analiticamente. Então a simulação é muitas vezes o único meio de investigação possível;
2. Simulação permite estimar a performance de um sistema já existente sob novas condições de operação;
3. Planejamento de sistemas propondo alternativas (ou política de operação alternativa para um sistema único), que podem ser comparadas via simulação

para ver qual melhor atende a uma necessidade específica;

4. Em uma simulação podemos manter um controle maior sobre as condições experimentais, do que geralmente seria possível quando fazemos a experiência com o próprio sistema;
5. Simulação permite estudar um sistema por um grande período de tempo, como por exemplo, um sistema econômico ou o estudo detalhado do desenvolvimento de um sistema em tempo expandido.

Simulação não está sem seus inconvenientes. Em particular, algumas de suas desvantagens são:

1. Cada rodada de um modelo de simulação estocástica produz somente *estimativas* das verdadeiras características do modelo, para um conjunto particular de parâmetros de entrada. Portanto, várias rodadas independentes do modelo provavelmente serão requeridas para cada conjunto de parâmetros de entrada a ser estudado;
2. O grande volume de dados produzidos por uma simulação, muitas vezes cria uma tendência a se ter maior confiança nos resultados obtidos do que é justificado. Se o modelo não é uma representação válida do sistema em estudo, os resultados da simulação, não interessam quão impressionantes pareçam, não irão fornecer nenhuma informação útil sobre o sistema.
3. Modelos de simulação de sistemas complexos são geralmente caros e demorados para serem desenvolvidos e analisados. Alguns pesquisadores consideram isto como uma vantagem, pois se o sistema é complexo, normalmente a simulação é a única forma de investigação do mesmo.

Para decidirmos se um estudo de simulação é adequado ou não em uma dada situação, temos que pensar sobre as vantagens e as desvantagens da simulação, e sempre usá-la de maneira correta visando obter resultados adequados do sistema em estudo.

3.3.1 Etapas em um Estudo de Simulação

Embora as etapas do processo de simulação sejam descritas de forma sequencial, elas de fato não o são, pois geralmente existe interação entre as diver-

sas etapas. É importante também destacar que algumas das etapas que compõem o desenvolvimento de uma simulação são comuns a uma resolução analítica.

Estas etapas podem ser esquematizadas da seguinte maneira:

1. *Formulação do problema e planejamento do estudo:* Cada estudo deverá começar com a definição clara dos objetivos a serem estudados, planejando também o projeto global;
2. *Coleta de dados e definição do modelo:* Os dados serão coletados do sistema de interesse e usados para estimar os parâmetros de entrada e obter distribuições de probabilidades das variáveis aleatórias usadas no modelo;
3. *Validação do modelo:* Devemos validar o modelo de simulação definido anteriormente, para verificarmos se o mesmo está correspondendo ao modelo proposto para o sistema real. Além disso, devemos verificar a adequabilidade das funções de distribuições de probabilidades sobre os dados observados, através de testes estatísticos de aderência, tais como *Qui-quadrado* e *Kolmogorov-Smirnov*;
4. *Construção do programa computacional e verificação:* O usuário de simulação deve decidir se programa o modelo em uma linguagem de propósito geral, como C ou FORTRAN; ou se ele usa um software específico (linguagem de simulação), como no caso de simulação de eventos discretos, o GASP, GPSS, ou SIMSCRIPT, entre outros;
5. *Execução de rodadas piloto:* Rodadas piloto do programa de simulação são feitas com o propósito de validação;
6. *Validação do programa de simulação:* As rodadas piloto são usadas para testar a sensibilidade dos resultados do modelo de simulação para pequenas mudanças nos parâmetros de entrada. É aconselhável fazer também uma comparação dos resultados da simulação piloto com os resultados obtidos na coleta de dados;
7. *Planejamento do experimento da simulação:* Para cada sistema a ser simulado, decisões devem ser tomadas com relação às condições iniciais para a rodada de simulação, o comprimento (tamanho) das rodadas de simulação e o número de rodadas a serem feitas;
8. *Execução das rodadas de simulação:* São executadas as rodadas de simulação para a obtenção dos dados de interesse do sistema em estudo.

9. *Análise dos resultados obtidos:* As entradas da simulação são obtidas através da geração de variáveis aleatórias, e as saídas também são funções dessas variáveis, e portanto a análise de saída é um problema estatístico. Técnicas estatísticas são usadas para analisar os dados obtidos através das rodadas de simulação. O objetivo usual é calcular média, desvio padrão e construir intervalos de confiança para alguma medida de performance do sistema em estudo;
10. *Documentação e implementação dos resultados:* Pelo fato dos modelos de simulação serem muitas vezes usados para mais que uma aplicação, é importante documentar as suposições que foram feitas sobre o modelo, bem como seu correspondente programa computacional. Finalmente é importante verificar se os resultados atingem os objetivos propostos em (1).

Vale ressaltar que esses passos, utilizados na construção de uma simulação, devem ser repetidos quantas vezes forem necessárias, até a obtenção do resultado final.

3.3.2 Linguagens de Simulação

O uso de uma linguagem adequada para implementar a simulação torna-se importante, uma vez que a construção do programa é frequentemente, uma fase bastante complexa do processo de simulação.

Vale ressaltar que a princípio, toda linguagem de programação é uma candidata a linguagem de simulação. Porém algumas como FORTRAN, PASCAL, C, MODULA-2, que são projetadas para uso geral e sendo também flexíveis, relativamente fáceis de entender e disponíveis em quase todos os computadores atuais, podem ser usadas para escrever programas de simulação.

Já que linguagens de simulação tem muitos elementos em comum, por exemplo, o conceito de um marcador de tempo (relógio), atualização de posição e sequenciamento de eventos; linguagens com propósitos específicos tem sido desenvolvidas. Entre as mais populares, temos GPSS, SIMSCRIPT, GASP e SLAM. (Para maiores detalhes, ver Breuer [BRE 75] e Schriber [SCH 91]).

As maiores vantagens do uso de linguagens de simulação (ou software específicos de simulação) são:

1. podemos ter uma grande redução no tempo de programação, pois a elaboração do programa fica mais simples;
2. elas são apropriadas para trabalhar com filas (em caso de estarmos utilizando software de filas), sendo pré-estruturada para tratar problemas de linha de espera de modo natural;
3. é geralmente muito fácil alterar os códigos quando o programa necessita ser reexecutado por alguma razão;
4. tais programas tem poucas linhas de códigos, sendo bastantes eficientes para detectar erros.

Neste trabalho simulamos uma fila M/M/1 e utilizamos o pacote de simulação GPSS/H (*General Purpose Simulation System, version H*) [BCN 89], que é de manuseio bastante simples, mas também muito poderoso. Além de ser muito bom em simulação de sistemas de filas, este pacote serve também para simulação de outros tipos de sistemas.

3.4 Geração de Números Pseudo-aleatórios

Como esses números alimentam o simulador, é importante que eles, de fato sejam provenientes de uma variável aleatória com distribuição uniforme $[0,1]$. Eles são usados para vários propósitos, tais como determinando os tempos entre-chegadas, os tempos de serviço, ou a rota (percurso) dos clientes; podem ser usados ainda para testar uma condição probabilística.

Os primeiros métodos de geração de números aleatórios foram essencialmente realizados a mão, utilizando dados ou bolas numeradas. Por volta de 1940 foram desenvolvidos os primeiros geradores numéricos (ou aritméticos) de números pseudo-aleatórios, através de uma função matemática. Uma vez que a sequência de números é produzida deterministicamente por uma equação, os números não são verdadeiramente aleatórios, e por isso são usualmente chamados de *números pseudo-aleatórios*.

Devemos ter bem claro que na maioria das vezes, a performance de uma simulação estará fortemente relacionada com a do gerador de números

pseudo-aleatórios utilizado.

As propriedades desejáveis da função de números pseudo-aleatórios são:

- os números devem ser uniformemente distribuídos no intervalo $[0,1]$;
- os números devem ser independentes e, portanto, nenhuma correlação deve existir na sequência de números pseudo-aleatórios;
- O gerador deve produzir uma sequência distinta de números pseudo-aleatórios bastante grande (também chamado de *período do gerador*);
- a sequência de números pseudo-aleatórios deve ser reprodutível.

E do computador são:

- o gerador deve ser rápido;
- o espaço de armazenamento requerido para a geração deve ser o menor possível.

Um método popular, e que melhor atende as propriedades citadas anteriormente, é o chamado *método congruencial linear*, introduzido por Lehmer em 1951, que diz o seguinte:

Resultado 3.1 *Seja M inteiro positivo ≥ 2 . Uma sequência $(y_1, y_2, \dots)$ em $0, 1, \dots, M-1$ é dita "gerada" pelo gerador congruencial linear com parâmetros a , c e M e semente y_0 se*

$$y_{i+1} = (ay_i + c) \bmod M; \quad i \geq 1$$

onde a , c e y_0 são inteiros do conjunto $0, 1, \dots, M-1$. Devemos observar se $a < M$, $c < M$ e $y_0 < M$, além de condições especiais que tornem a sequência a maior possível.

Os números pseudo-aleatórios desejados U_i em $[0,1]$ (para $i = 1, 2, \dots$) são obtidos por:

$$U_i = \frac{y_{i+1}}{M}$$

Os valores a , c e M dependem do gerador da linguagem e da máquina que se está utilizando, sendo recomendado o uso de geradores de números pseudo-aleatórios especificamente projetados na linguagem que executará o modelo de simulação.

O gerador de números pseudo-aleatórios do GPSS/H é o congruencial linear de Lehmer, e tem um período correspondente de $2^{31} - 2$ números. Isto significa que mais de dois bilhões de números podem ser gerados antes que a sequência comece a se repetir.

3.5 Métodos de Coleta de Dados

Durante os últimos 31 anos de discussão da metodologia de análise estatística dos dados de saída de uma simulação em equilíbrio, que foi iniciada por Conway (apud [PAW 90]), com o trabalho intitulado *Some Tactical Problems in Digital Simulation*, uma variedade de métodos para a coleta e análise de dados de saída de uma simulação tem sido propostos.

Estes métodos tentam diminuir ou remover as dependências entre as observações; ou ainda procuram explorar a natureza correlacionada das observações na análise de variância, necessária para determinar os intervalos de confiança dos parâmetros estimados.

Há diversos métodos para coleta dos dados de saída de uma simulação, tais como: *Replicações*, *Batch Means*, *Regenerativo*, *baseado em Análise Espectral*, *Batch Means Sobrepostos*, *Representação Autoregressiva*, entre outros. Na elaboração desse trabalho abordaremos apenas os três primeiros métodos, sendo que maiores detalhes sobre os outros, podem ser encontrados, por exemplo, em Pawlikowski [PAW 90].

Estes métodos são aplicados para estimar as características e/ou parâmetros do processo simulado.

3.5.1 Método de Replicações

Replicação é, ao menos conceitualmente, um método fácil de coleta de dados. Neste método replicamos ou repetimos um certo número k de vezes um mesmo experimento de simulação, cada uma gerando m valores amostrais de uma variável aleatória de interesse. Para cada uma das k rodadas de simulação usamos uma sequência independente e diferente de números pseudo-aleatórios, e a média das observações coletadas durante cada rodada é calculada. Estes valores médios coletados a cada rodada são computados e usados na análise estatística do processo de interesse $\{N_t : t \geq 0\}$. É de se esperar que os dados de saída (os resultados de cada simulação) sejam independentes e identicamente distribuídos. Na figura 3.1, mostramos esquematicamente o método de replicação.

Figura 3.1: Método de Replicações

Em cada rodada, m observações são coletadas, gerando um total de n observações ($n = km$). Esta sequência de observações pode ser escrita como

$$(x_{11}, x_{12}, \dots, x_{1m}), (x_{21}, x_{22}, \dots, x_{2m}), \dots, (x_{k1}, x_{k2}, \dots, x_{km})$$

cujas médias são:

$$\bar{X}_1(m), \bar{X}_2(m), \dots, \bar{X}_k(m)$$

onde

$$\bar{X}_i(m) = \frac{1}{m} \sum_{j=1}^m x_{ij} \quad i = 1, 2, \dots, k$$

que são usados para obter o estimador pontual da média do processo estocástico $\{N_t : t \geq 0\}$ na i -ésima replicação. O estimador média das médias das k rodadas é:

$$\bar{\bar{X}}(k, m) = \frac{1}{k} \sum_{i=1}^k \bar{X}_i(m)$$

que é numericamente equivalente à *média global* dos dados, $\bar{X}(n)$. Este é o resultado geralmente apresentado como resultado da simulação.

As médias são independentes, já que sequências diferentes de números pseudo-aleatórios são usadas para cada rodada, e para m suficientemente grande, o *Teorema Central do Limite* garantirá que as médias das replicações, $\bar{X}_i(m)$, $i=1, 2, \dots, k$; poderão ser consideradas com distribuição $N(\nu, \sigma^2)$.

Desta forma, um intervalo de confiança com $100(1 - \alpha)\%$ de confiança, a partir de uma distribuição t-Student com $k_* - 1$ graus de liberdade para a média das médias das k replicações, pode ser construído, ou seja

$$\bar{\bar{X}}(k, m) \pm t_{k-1, 1-\alpha/2} \hat{\sigma}[\bar{\bar{X}}(k, m)] \quad (3.1)$$

onde

$$\hat{\sigma}^2[\bar{\bar{X}}(k, m)] = \sum_{i=1}^k \frac{(\bar{X}_i(m) - \bar{\bar{X}}(k, m))^2}{k(k-1)} \quad (3.2)$$

é o estimador da variância de $\bar{\bar{X}}(k, m)$ e $t_{k-1, 1-\alpha/2}$ é o limite superior do ponto crítico $(1 - \alpha/2)$ da distribuição t-Student para $0 < \alpha < 1$.

O método de replicações parece ser mais sensível a não estacionaridade das observações coletadas durante o período transiente (seção 3.7) do que os métodos baseados numa única rodada de simulação, já que a cada rodada o sistema começa em desequilíbrio. Há diferentes opiniões sobre a aplicabilidade deste método, comparado com outros métodos de coletas de dados baseados em uma única rodada de simulação longa.

Este método de replicações serve também para reduzir a variância dos dados. Observe que quando consideramos as médias em cada rodada, estamos diminuindo a variância (3.2), pois

$$\text{Var}[\bar{X}_i(m)] = \frac{\sigma^2}{m} \quad i = 1, 2, \dots, k$$

3.5.2 Método do Batch Means

Suponha agora obtermos uma sequência de n ($n = km$) observações originais $(x_1, x_2, \dots, x_n)$ do processo observado $\{N_t : t \geq 0\}$ que é dividida em uma série de k batches (lotes) não sobrepostos

$$(x_{11}, x_{12}, \dots, x_{1m}), (x_{21}, x_{22}, \dots, x_{2m}), \dots, (x_{k1}, x_{k2}, \dots, x_{km})$$

de tamanho m , com as médias destes batches dadas por

$$\bar{X}_1(m), \bar{X}_2(m), \dots, \bar{X}_k(m)$$

obtidas da mesma maneira que no método das replicações.

O método de *batch means* controla o problema transiente através desta divisão da rodada em batches. Assumindo que apenas uma deleção é usada para encontrar as condições iniciais de equilíbrio para o primeiro batch, cada batch subsequente começa com o sistema em equilíbrio.

Analogamente ao método das replicações calculamos o estimador pontual e a média das médias. Se o tamanho m do batch é suficientemente longo, as médias dos batches serão aproximadamente independentes e com distribuição $N(\nu, \sigma^2)$; então um intervalo de confiança, do tipo da equação 3.1, pode ser construído, como no caso de replicações, com as médias dos batches substituindo as médias das rodadas (figura 3.2).

Há três fontes de erros potenciais quando usamos batch means para construir um intervalo de confiança para a média em equilíbrio, que são:

1. A correlação entre os $\bar{X}_i(m)$'s, e o consequente estimador da variância viado; se m não for grande o bastante;

Figura 3.2: Método de Batch Means

2. O fato que os $\bar{X}_i(m)$'s não são identicamente distribuídos com média igual à teórica do processo de interesse, requerendo um aumento do tamanho da amostra para melhor estimar esse parâmetro;
3. A não normalidade dos $\bar{X}_i(m)$'s, ocorrendo também quando m não for grande o bastante.

O método de batch means tem como suposição que observações mais separadas no tempo são menos correlacionadas, então para batches suficientemente longos de observações temos que os batches serão quase não correlacionados. A dificuldade é encontrar tamanho de batches que assegurem a não correlação entre as médias dos batches.

Este método é frequentemente usado na prática por ser mais interessante (com uma única rodada podemos obter, por exemplo 30 dados médios; enquanto que na replicação necessitamos fazer 30 rodadas diferentes) e requerer menor tempo de computador, sendo com isto menos oneroso.

Correlação entre os dados

Um fato que ocorre em sistemas de fila é a dependência dos eventos, gerando portanto, correlação entre as observações coletadas. Para diminuir a correlação dos dados de saída da simulação, as observações individuais são substituídas pelas médias dos batches, por serem menos correlacionadas. Então o problema da análise dos intervalos de confiança das observações correlacionadas é substituído pelo problema da determinação do tamanho do batch, tais que os batches sejam (quase) não correlacionados.

A correlação entre as médias dos batches de tamanho m podem ser calculadas pelo estimador do coeficiente de autocorrelação

$$\hat{\rho}_*(k, m) = \frac{\hat{R}(l, m)}{\hat{R}(0, m)} \quad (3.3)$$

onde

k = número de batches;

l = estimador de autocorrelação de lag l ;

m = tamanho do batch; e

$$R(l, m) = \frac{1}{k-l} \sum_{i=l+1}^k [\bar{X}_i(m) - \bar{\bar{X}}(k, m)][\bar{X}_{i-k_l}(m) - \bar{\bar{X}}(k, m)]$$

é o estimador de autocovariância de lag l , $l = 1, 2, 3, \dots$, da sequência das médias obtidas $\bar{X}_1(m), \bar{X}_2(m), \dots, \bar{X}_k(m)$.

A sequência das médias pode ser considerada como não correlacionada quando todos os $\hat{\rho}_*(k, m)$, $k=1, 2, 3, \dots$, assumem pequenas magnitudes, digamos se eles são menores que 0,05. Segundo Mood, Graybill e Boes [MGB 74], o usual é considerarmos uma autocorrelação desprezível quando esta é menor, em módulo, que $1/\sqrt{y}$, onde y representa o número de elementos considerados no cálculo da autocorrelação.

Geralmente, em uma sequência de dados correlacionados, os coeficientes de autocorrelação de lag l , $l = 1, 2, 3, \dots$, não são necessariamente decrescentes quando o lag aumenta.

Neste trabalho, um procedimento sequencial é proposto justamente

com o objetivo de encontrar tamanho de batch, supondo que sejam distribuídos normalmente e com pequena correlação.

3.5.3 Método Regenerativo

No método regenerativo as n observações do processo $\{N_t : t \geq 0\}$, são também agrupadas em batches, mas os batches são de comprimento aleatório, determinado por sucessivos instantes de tempo em que o processo simulado probabilisticamente recomeça, isto é, em que a transição para o estado futuro não depende do estado passado. Tais instantes de tempo são chamados de *pontos de regeneração*, e os estados do processo nesses pontos são chamados de *estado de regeneração*.

A natureza especial deste método reside no fato de que a cada dois pontos de regeneração temos batches de observações, que são coletados durante diferentes *ciclos regenerativos*. Um *ciclo* começa quando o estado de regeneração do sistema é atingido, e neste estado, o comportamento futuro do sistema é independente do passado. Quando um retorno é feito a tal estado, o ciclo termina.

O comprimento do ciclo regenerativo não é determinístico, mas sim uma variável aleatória, que determina o tamanho do ciclo. Esses ciclos regenerativos são sempre independentes e identicamente distribuídos. Por exemplo, podemos admitir que o ponto de regeneração de um sistema de fila M/M/1 seja o instante em que um cliente recém chegado encontra o sistema vazio, embora saibamos que para esta fila todos os estados são regenerativos.

Logo, o estado de regeneração tem as seguintes propriedades:

- o modelo volta periodicamente ao estado de regeneração (com probabilidade 1), levando a ocorrência de ciclos;
- quando o sistema entra no estado de regeneração, o comportamento futuro do sistema depende apenas deste estado.

A principal consideração prática que pode ser tirada dessas propriedades, é a necessidade do estado de regeneração ocorrer frequentemente, em uma simulação de tamanho razoável, isto é, que tenhamos um número mínimo de ciclos de regeneração na simulação (digamos uns 30). Se não tivermos essa propriedade, não poderemos aplicar o método com segurança.

EXEMPLO. Uma realização do processo $\{N_t : t \geq 0\}$ que conta o número de clientes no sistema para uma fila M/M/1, como mostra a figura 3.3.

Figura 3.3: Método Regenerativo

Por causa do comprimento aleatório dos batches, este método requer estimador especial, usualmente na forma de uma razão de duas variáveis aleatórias. Os resultados podem ser encontrados em Pawlikowski [PAW 90].

Sabemos que muitos sistemas do mundo real não possuem pontos de regeneração, ou alguns tem ciclos de comprimento tão grandes que somente uns poucos ciclos podem ser simulados, de tal modo que restringem a aplicação deste método. Em alguns casos relatados, especialmente se um pequeno número de ciclos regenerativos são considerados, a performance do método regenerativo demonstrou ser inferior que o método de batch means (Law e Kelton [LEK 84]).

3.6 Quantidade de Batches

Quando um pesquisador executa um experimento de simulação, ele requer a rodada de simulação longa o suficiente para:

1. proporcionar uma precisão desejada do parâmetro analisado;
2. obter um intervalo de confiança válido para o parâmetro de interesse.

Outro fator a ser analisado é o número de batches necessários para estimarmos um intervalo de confiança. Para determinar o número de batches, Schmeiser [SCH 82] utilizando amostras de tamanho fixado, estudou o efeito do número de batches, k , relacionando com as propriedades da meia amplitude dos intervalos de confiança, sobre a média do parâmetro de interesse. Os resultados mostraram que usualmente o número de batches usados na análise dos intervalos de confiança não devem ser menos que 10, e não mais que 30, se a rodada de simulação é grande o bastante para podermos assumir normalidade e independência dos batches. Isto significa que tendo determinado um tamanho de batch que propicia uma correlação desprezível e as médias dos batches com distribuição aproximadamente normal, não há necessidade de usar mais que $k = 30$ batches para obter intervalos de confiança com boa cobertura.

Conclui ainda que $k < 10$ batches deve ser evitado porque produzirão alta variabilidade, e $k > 30$ batches é válido somente se a estabilidade do intervalo de confiança é importante.

Um outro problema relacionado ao método de batch means, que se estende também a todos os outros métodos de coletas de dados, é a escolha de um tamanho apropriado para o período transiente, que descreveremos a seguir.

3.7 Período Transiente

Todos os tópicos de simulação tratados até agora, são referentes à simulação de qualquer tipo de sistema. Neste trabalho, por simularmos a fila M/M/1 em equilíbrio, nos preocupamos também com um outro problema que pode ser fonte de muitos erros na análise dos resultados obtidos através das simulações, que é o período transiente.

Em teoria de filas é importante saber, no mínimo aproximadamente, quão longa deve ser a quantidade de tempo para que o sistema efetivamente encontre a estacionaridade. Isto se deve ao fato de que sistemas dinâmicos, tais

como sistemas de fila ou redes, inicialmente se encontram em um período transiente (também chamado de *desequilíbrio*). Depois desse período, onde suas características variam com o tempo, o sistema se aproxima do equilíbrio se ele é estável, ou fica permanentemente em *desequilíbrio* se o sistema não for estável. A duração desse período varia conforme as características do sistema.

Notamos que na prática, por exemplo, a fila M/M/c com uma população infinita de clientes e capacidade ilimitada de fila de espera, pode nunca entrar em equilíbrio, se a taxa de chegada for maior ou igual a taxa de atendimento, ou seja

$$\lambda \geq c\mu$$

onde λ é a taxa de chegada, μ é a taxa de serviço e c é o número de servidores. Neste caso a fila sempre aumentará sua extensão com o tempo, e o sistema ficará permanentemente congestionado. Por outro lado, sistemas de fila deste mesmo tipo com capacidade de fila de espera limitada, sempre atingirá o equilíbrio, mesmo que tenhamos uma taxa de ocupação do sistema maior que 1, ou seja,

$$\rho = \frac{\lambda}{c\mu} > 1.$$

A determinação do período transiente é um fator importante no planejamento da simulação, pois normalmente estudamos o sistema em equilíbrio sendo, portanto, não recomendável a coleta dos resultados nesse período inicial. Os dados coletados durante este período podem causar desvios significativos nos resultados finais, em relação aos verdadeiros parâmetros que estamos estimando.

Uma série de regras para a determinação do tamanho do período transiente são apresentadas em Pawlikowski [PAW 90].

Uma delas é a regra de Odoni e Roth, que propõem a seguinte aproximação para o comprimento do período de *desequilíbrio*:

$$\tau = \frac{C_a^2 + C_s^2}{2,8\mu(1 - \sqrt{\rho})^2} \quad (3.4)$$

onde C_a^2 e C_s^2 são os coeficientes de variação para o *tempo entre-chegadas* e o *tempo de serviço*, respectivamente, $1/\mu$ é o *tempo médio de serviço*, e ρ é a *taxa de ocupação do sistema*.

Para chegar a equação 3.4, Odoni e Roth estudaram sistema de fila Markovianos com capacidade infinita, fila única e apenas um servidor (Odoni e Roth [OER 83]).

Por ser o período de estabilização um fator de grande importância, utilizamos esta regra no contexto do nosso trabalho, quando trabalhamos com o procedimento sequencial, pois conhecemos as taxas referentes aos tempos entre chegadas e de serviços; e também a taxa de ocupação do sistema.

Na segunda etapa do trabalho, fazendo o cálculo da porcentagem de cobertura para a fila M/M/1, utilizamos uma outra regra, que também está proposta em Pawlikowski, afirmando que o efeito do período transiente termina logo após o comprimento inicial correspondendo a 2% do total da rodada de simulação. Para termos mais garantia com respeito aos resultados da simulação, utilizamos o equivalente a 3% do tamanho da rodada fixado.

No capítulo seguinte falamos de procedimentos sequenciais, onde um deles é a base deste trabalho.

□

Capítulo 4

Procedimento Sequencial

Este capítulo aborda alguns procedimentos sequenciais utilizados para construir intervalos de confiança sobre o parâmetro de interesse, enfatizando o procedimento sequencial proposto por Law e Carson.

4.1 Considerações Iniciais

Como mencionado anteriormente, em estudos de simulação tem sido observado que é comum se fazer uma rodada única de comprimento arbitrário do modelo de simulação, para estimar os resultados de saída de um processo de interesse. Estes resultados são aceitos como sendo as "reais" medidas de performance do modelo, mas eles podem diferir grandemente do valor teórico da variável analisada. A motivação deste trabalho é analisar a melhor maneira de utilizar a simulação, para que o modelo simulado represente adequadamente o sistema que está sendo estudado.

Os dados de saída da simulação podem ser obtidos através dos vários métodos para a coleta de dados existentes, tais como replicações, batch means e ciclos de regeneração.

Para se obter os dados de saída da simulação, utilizando qualquer um dos métodos de coleta de dados, podemos executar um modelo de simulação com

um tamanho de amostra fixado inicialmente, ou ainda determinar o tamanho da amostra, através de procedimentos (ditos) sequenciais. A seguir abordaremos melhor esse assunto.

4.2 Tipos de Tamanho de Amostra

Uma maneira de avaliar os dados de saída de uma simulação, é construirmos um intervalo de confiança em torno da média observada, fixando inicialmente o tamanho da amostra.

Law e Kelton [LEK 84], utilizando os cinco mais usados métodos de coleta de dados (replicação, batch means, representação auto-regressiva, análise espectral e ciclos regenerativos), apresentam resultados comparativos sobre o desempenho desses métodos, quando o comprimento da rodada de simulação é fixado inicialmente.

Todos estes métodos tem a inconveniência de que se o tamanho amostral total escolhido for demasiadamente pequeno para o sistema simulado, a análise dos intervalos de confiança resultantes dos dados de saída podem proporcionar uma porcentagem de cobertura, que é a proporção de vezes que os intervalos de confiança contém o valor teórico da variável de interesse, consideravelmente menor que a desejada. Por outro lado, se o tamanho amostral total é suficientemente grande, a porcentagem de cobertura proporcionada pelos intervalos de confiança para os dados de saída será adequada apenas para o sistema considerado. De qualquer forma, o que é suficientemente grande para um processo simulado pode não ser para outro, pois sistemas podem funcionar de modo radicalmente diferentes, exigindo comprimentos de rodadas também muito diferentes.

Outro modo é fixar um intervalo de confiança, e procurarmos o tamanho da amostra necessário para a construção deste intervalo de confiança. Daí começamos com um comprimento de rodada mínima e sequencialmente a aumentamos, durante o curso da simulação, até que um intervalo de confiança que satisfaça esta condição possa ser construído.

Vários procedimentos sequenciais são apresentados por Law e Kelton [LEK 82].

Os procedimentos propostos por Fishman [FIS 77] e por Lavenberg e Sauer [LES 77] assumem que o processo analisado é regenerativo, e então exploram o fato de que os ciclos de regeneração são independentes e identicamente distribuídos, facilitando a aplicação dos métodos da estatística clássica. Portanto é importante que o sistema de interesse possua pontos de regeneração autênticos, uma propriedade não compartilhada por muitos sistemas do mundo real, como por exemplo, um sistema de filas com dois ou mais fluxos de chegada não Poisson. Isto limita a aplicação destes métodos.

Os procedimentos propostos por Mechanic e McKay [MEM 66] e por Law e Carson [LEC 79] não fazem a suposição regenerativa, possibilitando a simulação de modelos mais complexos onde os pontos de regeneração não ocorrem ou ocorrem em intervalos muito grandes de difícil investigação no modelo. Estes procedimentos se baseiam no método de batch means que, como já foi dito, é um método que tem suas vantagens de aplicação, mas possui dependência devido a autocorrelação entre os batches.

Cada um desses procedimentos procura construir um intervalo de confiança com $100(1 - \alpha)\%$, para a média em equilíbrio, dada por

$$v = \lim_{n \rightarrow \infty} \sum_{i=1}^n \frac{X_i}{n}$$

significando que este intervalo de confiança conterá v com probabilidade perto de $1 - \alpha$.

Suscintamente descreveremos três desses procedimentos, e somente enfatizaremos o procedimento de Law e Carson, por ser de nosso interesse.

- *O Procedimento de Fishman*

O procedimento de Fishman, utilizando o método regenerativo, visa construir um intervalo de confiança baseado na distribuição t-Student à $100(1 - \alpha)\%$ para v , com meia largura do intervalo de confiança fixada $\gamma > 0$. Os dados de saída da simulação são agrupados em k blocos de j ciclos regenerativos cada um, e um estimador pontual, $\tilde{h}_i$, para a média observada da variável de interesse, é obtido dos j ciclos do i -ésimo bloco, $i = 1, 2, 3, \dots, k$.

Este procedimento apresenta um método de estimação sequencial que quando usado com a propriedade regenerativa, permite que os resultados sejam

obtidos com uma exatidão estatística especificada. Também descreve um teste para testar a normalidade dos dados obtidos da simulação, e apresenta um método de redução de variância para este procedimento sequencial. O algoritmo detalhado deste procedimento é apresentado em Fishman [FIS 77].

O procedimento de Fishman procura obter variáveis aleatórias independentes e identicamente distribuídas que possam ser consideradas como não viciadas para a média observada e com distribuição normal, e então usa-se os métodos de estatística clássica em suas análises.

- *O Procedimento de Lavenberg e Sauer*

No procedimento regenerativo de Lavenberg e Sauer constroem-se um intervalo de confiança à $100(1 - \alpha)\%$ para a média v , tal que a razão entre a meia amplitude do intervalo de confiança e o ponto médio, não exceda uma constante fixada $\delta > 0$, onde δ é a largura relativa desejada. A partir de um número qualquer de ciclos regenerativos, constroem-se um intervalo de confiança para a média observada, usando a aproximação clássica.

Os autores proporcionaram uma justificativa teórica para este procedimento; em particular provaram que quando δ se aproxima de zero, o intervalo de confiança conterá a média observada com uma probabilidade que se aproxima do nível desejado $1 - \alpha$, além de uma porcentagem de cobertura perto deste nível. Porém não é dito o que significa exatamente "pequeno δ ", e por outro lado, observaram que o comprimento de rodada ficará grande quando δ é especificado pequeno.

O procedimento de Lavenberg e Sauer é operacionalmente bastante simples (ver seu algoritmo em Lavenberg e Sauer [LES 77]).

- *O Procedimento de Mechanic e McKay*

Na implementação do procedimento de Mechanic e McKay, utilizando o método de batch means, as n observações amostrais são divididas em J , $J > 2$ batches de um dado tamanho e é calculado um estimador de autocorrelação das correspondentes médias dos batches. Então o tamanho do batch é aumentado sucessivamente (com o tamanho total da amostra n permanecendo constante, e conseqüentemente diminuindo o número de batches), resultando uma sequência

de estimadores de autocorrelação, uma para cada correspondente tamanho de batch. Se esta sequência de estimadores de auto-correlação satisfizer um certo conjunto de condições, então está determinado um tamanho de batch, tal que as correspondentes médias dos batches são (supostamente), não correlacionadas; e essas médias dos batches são usadas para construir um intervalo de confiança para o parâmetro de interesse. Se essas condições não são satisfeitas, deve-se aumentar n e nova tentativa é feita para identificar o tamanho de batch apropriado.

Diferentemente dos outros procedimentos sequenciais descritos neste trabalho, este procedimento não impõem nenhuma condição sobre a metade do comprimento do intervalo de confiança que é construído.

Na próxima seção exporemos detalhadamente o procedimento sequencial de Law e Carson, que será aplicado no nosso trabalho.

4.3 Procedimento de Law e Carson

Suponha uma rodada de simulação de comprimento n e então dividimos as observações resultantes $x_1, x_2, \dots, x_n$ que são assumidas serem de um processo de covariância estacionária, em k batches, k par, de comprimento m ($n = km$). Seja $\bar{X}_j(m)$, $j = 1, 2, \dots, k$ a média amostral das m observações no j -ésimo batch e seja $\bar{\bar{X}}(k, m)$ a média amostral global, ou seja, a média das médias. Estimamos a variância de $\bar{\bar{X}}(k, m)$ pela equação 3.2.

Os autores constroem um intervalo de confiança para a média resultante $\bar{\bar{X}}(k, m)$ após cada rodada de simulação, do acordo com a equação 3.1.

O processo de interesse, tanto no caso do trabalho de Law e Carson como no nosso, é o tempo médio de espera na fila W_q .

Sabemos que ao aumentarmos o tamanho do batch, pela *Lei Fraca dos Grandes Números*, a média obtida em cada batch convergirá para a média teórica. Em Law e Carson [LEC 79], temos que quando aumentamos o tamanho do batch é de se esperar que o estimador de autocorrelação correspondente ao mesmo, tenderá a zero. Isto implica dizer que ao aumentarmos o tamanho dos batches, devemos esperar que as médias correspondentes a cada batch se aproximará da média teorica e as correlações entre esses batches se aproximarão de zero.

Neste procedimento os autores sugeriram usar lk batches de comprimento m para encontrar os k batches de comprimento lm , obtidos no final da execução do algoritmo do procedimento, aproximadamente não correlacionados.

As relações entre $\rho_*(m)$ e $\rho_*(lm)$, que são as autocorrelações de *lag* 1, que ocorrem nos batches de tamanho m e lm , respectivamente, também podem ocorrer na prática. Eles estudaram alguns processos de filas para os quais $\rho_*(m)$ e $b(k, m)$, onde $b(k, m)$ é o vício do parâmetro da variável de interesse, podem ser analiticamente calculados, tais como

- $\{W_{q,t} : t \geq 0\}$ para a fila M/M/1 com $\rho = 0,5; 0,8; 0,9$, onde $\{W_{q,t} : t \geq 0\}$ corresponde ao processo que mede o tempo de espera na fila dos clientes.

Law e Carson encontraram um comportamento para a autocorrelação de *lag* 1 como uma função de m , resultando daí o *valor de parada* c do algoritmo. As autocorrelações dos elementos obtidos da simulação da fila M/M/1, exibem o comportamento *padrão simples*, que é apresentado na figura 4.1.

Figura 4.1: Comportamento Padrão Simples

Como na subseção 3.5.2, o estimador de autocorrelação de *lag* 1, de acordo com a equação (3.3) pode ser escrito como

$$\hat{\rho}_*(k, m) = \frac{\sum_{j=1}^{k-1} [\bar{X}_j(m) - \bar{\bar{X}}(k, m)] [\bar{X}_{j+1}(m) - \bar{\bar{X}}(k, m)]}{\sum_{j=1}^k [\bar{X}_j(m) - \bar{\bar{X}}(k, m)]^2}$$

onde $\hat{\rho}_*(k, m)$ é o estimador usual de autocorrelação de lag 1.

Para estimarem a autocorrelação no procedimento os autores utilizaram o estimador Jackknife (ver apêndice A), que usualmente é menos viciado que o estimador usual. Um estimador Jackknife de coeficiente de autocorrelação de lag 1 para uma sequência de batch means de tamanho m é calculado pela seguinte fórmula:

$$\tilde{\rho}_*(k, m) = 2\hat{\rho}_*(k, m) - [\hat{\rho}_{*1}(k/2, m) + \hat{\rho}_{*2}(k/2, m)]/2$$

onde $\hat{\rho}_*(k, m)$ é o estimador usual de lag 1 sobre todos os k batches, ao passo que $\hat{\rho}_{*1}(k/2, m)$ e $\hat{\rho}_{*2}(k/2, m)$ são, respectivamente, os estimadores usuais de lag 1 baseados na primeira e segunda metade da sequência dos k batches.

O algoritmo deste procedimento sequencial é apresentado e desenvolvido no capítulo seguinte.

Através de resultados experimentais, Law e Kelton [LEK 82] mostraram que o procedimento sequencial de Law e Carson é preferível no caso não regenerativo. Por isto optamos em trabalhar este procedimento, pois além de utilizar o método de batch means, que pode ser aplicado para sistemas mais complexos, produz resultados melhores que o procedimento de Mechanic e McKay, que também utiliza este mesmo método.

4.4 Resultados Apresentados por Law e Carson

Os autores simularam vários sistemas, dentre eles uma fila M/M/1 de $\rho = 0,8$, com $\lambda = 1$ e $1/\mu = 0,8$; e também simularam esta mesma fila considerando $\rho = 0,5$, com $\lambda = 1$ e $1/\mu = 0,5$.

De acordo com o processo de interesse, temos que o valor teórico para o tempo médio de espera na fila é $W_q = 3,2$, calculado através da equação (2.3).

Law e Carson [LEC 79] fixaram $l = 10$ e $k = 40$, embora estes não sejam os únicos possíveis valores, e escolheram $n_0 = 600$ e $n_1 = 800$. Fizeram 100 rodadas independentes de simulação usando o procedimento para valores fixados inicialmente de $\gamma = \infty; 0,15; 0,075$ e $c = 0,4$, onde c foi fixado a partir do comportamento *padrão simples* da figura 4.1.

Para cada rodada eles construíram um intervalo com 90% de confiança em torno da média das médias, usando uma distribuição t-Student, para verificar se o intervalo continha o valor teórico correspondente ao tempo médio de espera na fila. A partir destes intervalos de confiança, eles construíram um intervalo para a *cobertura obtida*, $\hat{p}$, cuja expressão é apresentada no capítulo seguinte.

Os resultados obtidos por Law e Carson para esses dois sistemas de fila, são apresentados no próximo capítulo; juntamente com os resultados que nós encontramos utilizando este procedimento sequencial.

□

Capítulo 5

Descrição e Análise dos Resultados da Simulação

Neste capítulo apresentamos os resultados, obtidos através do procedimento sequencial de Law e Carson, para o sistema de fila M/M/1; posteriormente, em posse desses resultados, apresentamos a porcentagem de cobertura para alguns tamanhos de amostras encontrados.

5.1 Considerações Iniciais

Inicialmente construímos um programa de simulação para o sistema de fila M/M/1 utilizando o pacote de simulação GPSS/H (*General Purpose Simulation System, version H*) que, como já foi mencionado anteriormente, é uma linguagem de simulação bem ajustada para sistemas de filas. O programa de simulação correspondente ao sistema estudado está apresentado no Apêndice B.

A primeira etapa deste trabalho foi determinar tamanhos de amostra para a fila M/M/1 com uma precisão relativa fixada, de acordo com o procedimento sequencial proposto por Law e Carson. Depois disso, a partir de tamanhos de amostras médios encontrados, fizemos os cálculos da porcentagem de cobertura.

A regra de parada utilizada para terminar as simulações, nos casos apresentados nesta dissertação, foi o número de clientes que abandonam o sistema após serem atendidos, tanto na determinação sequencial do tamanho da amostra, como na análise. A medida de performance avaliada neste trabalho é o *tempo médio de espera dos clientes na fila* (W_q). Para fazermos análises estatísticas dos resultados fornecidos pela simulação, correspondentes aos tempos de espera dos clientes que realmente chegaram a sair do sistema, utilizamos o software estatístico MINITAB.

Em virtude da grande importância do gerador de números pseudo-aleatórios, que já foi mencionada, no caso do GPSS/H essa sequência está pronta no interior do pacote, e quando a utilizamos escolhemos apenas a sua posição. O gerador do GPSS/H foi testado quanto a uniformidade por Bernasconi [BER 93] e verificou-se sua eficiência.

5.2 Descrição do Algoritmo Sequencial de Law e Carson

A seguir apresentamos o algoritmo, juntamente com um exemplo, de como funciona este procedimento sequencial.

Neste algoritmo, observamos que as **regras de parada** nos passos do mesmo, estão relacionadas com as autocorrelações de *lag* 1 entre as médias dos batches, e com a precisão relativa dos intervalos de confiança.

- *Algoritmo de Law e Carson*

Passo 0: Fixar inteiros positivos l, k, n_0, n_1 (lk par, $lk/2$ par, $n_0 < n_1 < 2n_0$, n_1 e $2n_0$ ambos divisíveis por lk), o valor de parada $c > 0$ e a precisão relativa $\gamma > 0$.

Seja $i = 1$; coletar n_1 observações.

Passo 1: a) Dividir as n_i observações em lk batches de tamanho $m = n_i/lk$ e calcular $\hat{\rho}_*(lk, m)$ de $\bar{X}_j(m)$, $j = 1, 2, \dots, lk$.

Se $\tilde{\rho}_*(lk, m) \geq c$, vá para o **Passo 2** (veja NOTA 1).

Se $\tilde{\rho}_*(lk, m) < 0$, vá para o **Passo 1c**.

Caso contrário, vá para o **Passo 1b**.

b) Dividir n_i em $lk/2$ batches de tamanho $2m$ e calcular $\tilde{\rho}_*(lk/2, 2m)$ de $\bar{X}_j(2m)$, $j = 1, 2, \dots, lk/2$ (veja NOTA 2).

Se $\tilde{\rho}_*(lk/2, 2m) < \tilde{\rho}_*(lk, m)$, vá para o **Passo 1c** (veja NOTA 3).

Caso contrário, vá para o **Passo 2**.

c) Dividir n_i em k batches de tamanho lm . Calcular $\bar{X}(k, lm)$ e $\delta = t_{k-1, 1-\alpha/2} \sqrt{\hat{\sigma}^2[\bar{X}(k, lm)]}$ a partir de $\bar{X}_j(lm)$, $j = 1, 2, \dots, k$ (veja NOTA 2).

Se $\delta/\bar{X}(k, lm) < \gamma$, então usar $\bar{X}(k, lm)$ e δ em 3.1, para construir um intervalo de confiança para a média das médias e_i (veja NOTA 4)

PARE (regra de parada do algoritmo).

Caso contrário, vá para o **Passo 2**.

Passo 2: Substituir i por $i+1$, colocar $n_i = 2n_{i-2}$, coletar as adicionais $(n_i - n_{i-1})$ observações requeridas, e ir ao **Passo 1a** (veja Nota 5).

As observações citadas no algoritmo são:

- NOTA 1: Este controle foi sugerido pelo comportamento *padrão simples*.
- NOTA 2: Um número apropriado de batches, de tal modo que os $\bar{X}_j(m)$'s obtidos, possam ser suficientes para calcular os $\bar{X}_j(2m)$'s e os $\bar{X}_j(lm)$'s.
- NOTA 3: Testar se $\rho_*(m')$ é decrescente para $m' \geq m$, verificando se $\rho_*(2m) < \rho_*(m)$.
- NOTA 4: Este controle adicional é para exigir que a precisão relativa dos intervalos de confiança, ou seja, a meia amplitude dividida pelo estimador pontual (média), seja menor que um número especificado γ (precisão relativa).
- NOTA 5: Os sucessivos tamanhos amostrais são $n_1, 2n_0, 2n_1, 4n_0, 4n_1, \dots$. Então o tamanho amostral total é dobrado a cada duas iterações.

5.3 Um Exemplo de como Funciona o Algoritmo

A seguir para melhor entendimento, exemplificamos com dados da simulação da fila M/M/1 com $\rho = 0,8$, um dos possíveis desenvolvimentos do algoritmo.

P.0

Fixar inteiros: $l = 6$; $k = 30$; $n_0 = 630$; $n_1 = 900$; $c = 0,4$ e $\gamma = 0,075$

P.1(a)

$$\left. \begin{array}{l} n_1 = 900 \\ lk = 180 \\ m = 5 \end{array} \right\} \Rightarrow \tilde{\rho}_*(lk, m) = 0,41 \geq c \rightarrow \text{P.2}$$

Considere agora,

P.2

$$n_2 = 2n_0$$

P.1(a)

$$\left. \begin{array}{l} n_2 = 1.260 \\ lk = 180 \\ m = 7 \end{array} \right\} \Rightarrow \tilde{\rho}_*(lk, m) = 0,33 < c \rightarrow \text{P.1(b)}$$

P.1(b)

$$\left. \begin{array}{l} n_2 = 1.260 \\ lk/2 = 90 \\ 2m = 14 \end{array} \right\} \Rightarrow \tilde{\rho}_*(lk/2, 2m) = 0,38 \geq \tilde{\rho}_*(lk, m) \rightarrow \text{P.2}$$

Considere agora,

P.2

$$n_3 = 2n_1$$

P.1(a)

$$\left. \begin{array}{l} n_3 = 1.800 \\ lk = 180 \\ m = 10 \end{array} \right\} \Rightarrow \tilde{\rho}_*(lk, m) = 0,25 < c \rightarrow \text{P.1(b)}$$

P.1(b)

$$\left. \begin{array}{l} n_3 = 1.800 \\ lk/2 = 90 \\ 2m = 20 \end{array} \right\} \Rightarrow \tilde{\rho}_*(lk/2, 2m) = 0,19 < \hat{\rho}_*(lk, m) \rightarrow \mathbf{P.1(c)}$$

P.1(c)

$$\left. \begin{array}{l} n_3 = 1.800 \\ k = 30 \\ lm = 60 \end{array} \right\} \Rightarrow \bar{\bar{X}}(k, lm) = 3,04; \quad \delta = 0,31$$

Então

$$\delta/\bar{\bar{X}}(k, lm) = 0,10 \geq \gamma \rightarrow \mathbf{P.1(a)}$$

P.1(a)

$$\left. \begin{array}{l} n_4 = 2.520 \\ lk = 180 \\ m = 14 \end{array} \right\} \Rightarrow \tilde{\rho}_*(lk, m) = 0,19 < c \rightarrow \mathbf{P.1(b)}$$

P.1(b)

$$\left. \begin{array}{l} n_4 = 2.520 \\ lk/2 = 90 \\ 2m = 28 \end{array} \right\} \Rightarrow \tilde{\rho}_*(lk/2, 2m) = -0,09 < \tilde{\rho}_*(lk, m) \rightarrow \mathbf{P.1(c)}$$

P.1(c)

$$\left. \begin{array}{l} n_4 = 2.520 \\ k = 30 \\ lm = 84 \end{array} \right\} \Rightarrow \bar{\bar{X}}(k, lm) = 3,23; \quad \delta = 0,24$$

Então

$$\delta/\bar{\bar{X}}(k, lm) = 0,074 < \gamma$$

Portanto

$$i.c. = (2, 80; 3, 41)$$

5.4 Simulação da Fila M/M/1 Através do Procedimento Sequencial

Nosso objetivo inicial, na simulação, é determinar tamanho de amostra n com γ fixado; utilizando o procedimento sequencial proposto por Law e Carson.

Simulamos uma fila M/M/1, considerando duas taxas de ocupação, que descrevemos a seguir:

1. com $\rho = 0,8$ e as seguintes suposições iniciais:

- (a) a média dos tempos entre-chegadas dos clientes $1/\lambda = 1 \text{ ut}$
- (b) a média dos tempos de serviço $1/\mu = 0,8 \text{ ut}$
- (c) a taxa de ocupação do sistema, conforme a equação 2.2 é

$$\rho = \frac{1}{1/0,8} = 0,8$$

(d) o tempo médio de espera de um cliente, conforme a equação 2.3 é

$$W_q = \frac{1}{1/0,8(1/0,8 - 1)} = 3,2$$

2. com $\rho = 0,5$ as suposições iniciais foram as seguintes:

- (a) a mesma média dos tempos entre-chegadas dos clientes, considerada anteriormente
- (b) a média dos tempos de serviço $1/\mu = 0,5 \text{ ut}$
- (c) a taxa de ocupação do sistema, é

$$\rho = \frac{1}{1/0,5} = 0,5$$

(d) o tempo médio de espera de um cliente, é

$$W_q = \frac{1}{1/0,5(1/0,5 - 1)} = 0,5$$

Desenvolvemos as simulações, tanto para $\rho = 0,8$ como $\rho = 0,5$; utilizando $k = 30$ batches e fixamos $l = 6$ de tal forma que lk e $lk/2$ sejam pares. O valor de k foi escolhido considerando o resultado de Schemeiser (seção 3.3). Consideramos também o valor $c = 0,4$, que é comparado com as autocorrelações de lag 1 do algoritmo, e valores de precisão relativa $\gamma = \infty; 0,15$ e $0,075$.

Calculamos o período transiente para os nossos sistemas utilizando à aproximação proposta por Odoni e Roth, pois inicialmente conhecemos as taxas de chegada e de atendimento dos clientes. De acordo com a equação (3.4) da seção 3.7, temos como resultado para o período transiente

$$\tau = \frac{C_a^2 + C_s^2}{2,8\mu(1 - \sqrt{\rho})^2} = 51,269 \text{ ut}$$

A utilização desta aproximação proposta por Odoni e Roth exige conhecidas as taxas de ocupação do sistema e a de chegada, que são dadas para o sistema de fila M/M/1 desta pesquisa. Como os coeficientes de variação para os tempos de chegada e de serviço são iguais a 1 para fila do tipo M/M/1, pudemos utilizar esta aproximação.

O valor obtido para o período transiente equivale aproximadamente ao número de clientes entrando no sistema, pois estamos considerando que em média, chega um cliente por unidade de tempo ($\lambda = 1$). Entretanto utilizamos, como período transiente, um valor aproximadamente doze vezes maior que este, de tal forma que o mesmo não afetasse os resultados de nossas simulações, desenvolvidas para o procedimento sequencial.

Fixamos um tamanho de amostra inicial, $n_0 = 630$, igual ao período transiente; e tomamos $n_1 = 900$ satisfazendo as exigências do algoritmo, ou seja, $n_0 < n_1 < 2n_0$, com n_1 e $2n_0$ divisíveis por lk .

Fizemos 180 rodadas independentes de simulação para a fila M/M/1 com $\rho = 0,8$, considerando os valores mencionados acima, utilizando o algoritmo sequencial de Law e Carson. O número de vezes de execução do procedimento foi aumentada para verificar a estabilidade dos tamanhos de amostra encontrados, e

notamos que os tamanhos de amostra médios obtidos tanto para 100, como para 180 rodadas, são muito próximos.

Apresentamos, como exemplo, a tabela 5.1 que mostra o número de iterações necessárias, bem como de que modo varia o tamanho das amostras e dos batches; e os valores das autocorrelações *lag* 1 que são comparados com o valor c do algoritmo. Esta tabela ilustra apenas uma possível obtenção de tamanho de amostra através do procedimento, sendo portanto um caso particular.

número da iteração	tamanho da amostra	autocorrelação de lag 1	tamanho do batch
1	900	0,43	5
2	1.260	0,54	7
3	1.800	0,53	10
4	2.520	0,54	14
5	3.600	0,44	20
6	5.040	0,40	28
7	7.200	0,48	40
8	10.080	0,42	56
9 *	14.400	0,34	80
10	20.160	0,19	112
11	28.800	0,12	160
12	40.320	0,13	224
13	57.600	0,09	320

Tabela 5.1: Obtenção de um Particular Tamanho de Amostra, considerando $\rho = 0,8$ e $\gamma = \infty$; 0,15 e 0,075

Observação da tabela 5.1: Na linha correspondendo a 9 iterações, o asterístico indica que para os casos de $\gamma = \infty$ e 0,15, os tamanhos de amostras obtidos até então, são adequados, de acordo com o procedimento proposto. O último valor da tabela corresponde ao tamanho da amostra final para $\gamma = 0,075$.

Apresentamos no Apêndice C os resultados correspondentes aos 100 tamanhos de amostras obtidos, para $\gamma = \infty$; 0,15 e 0,075.

Fizemos 100 rodadas independentes de simulação para a fila M/M/1 com $\rho = 0,5$ com as mesmas considerações acima, mudando a semente antes

de iniciar cada rodada de simulação do procedimento. A tabela 5.2, traz as iterações necessárias para obter um tamanho particular de amostra, juntamente com as autocorrelações de lag 1 e os tamanhos dos batches.

número da iteração	tamanho da amostra	autocorrelação de lag 1	tamanho do batch
1	900	0,52	5
2	1.260	0,40	7
3 *	1.800	0,30	10
4	2.520	0,61	14
5	3.600	0,57	20
6	5.040	0,80	28
7	7.200	0,45	40
8	10.080	0,35	56
9 **	14.400	0,25	80
10	20.160	0,17	112
11	28.800	-0,05	160

Tabela 5.2: Obtenção de um Particular Tamanho de Amostra, considerando $\rho = 0,5$ e $\gamma = \infty$; 0,15 e 0,075

Observação da tabela 5.2: Na linha correspondendo a 3 iterações, o (*) indica que para $\gamma = \infty$ este tamanho de amostra é apropriado; assim como (**) na nona linha indica o tamanho final da amostra no caso de $\gamma = 0,15$. A última linha da tabela traz o tamanho da amostra para $\gamma = 0,075$.

O apêndice C apresenta também os resultados para os 100 tamanhos de amostra encontrados para $\rho = 0,5$.

Em ambos os casos, para cada rodada de simulação foi construído um intervalo com 90% de confiança em torno da média amostral, baseado numa distribuição t-Student, para W_q . Através destes intervalos de confiança construídos para cada particular tamanho de amostra, foi verificado a porcentagem de cobertura do tempo médio de espera na fila. Do resultado encontrado foi calculado um intervalo de confiança, para o que se convencionou chamar de *cobertura obtida*, $\hat{p}$, da seguinte maneira:

$$\hat{p} \pm 1,645\sqrt{\hat{p}(1-\hat{p})/100} \quad (5.1)$$

onde $\hat{p}$ é a proporção de cobertura obtida nas 100 rodadas.

Calcular este intervalo de confiança para a cobertura obtida foi possível porque para obtermos $\hat{p}$, consideramos um *teste binomial*, onde "sucesso" significa o intervalo de confiança construído contém o valor teórico W_q , e "fracasso" é a não ocorrência deste evento.

Law e Carson, simularam o sistema com $k = 40$ batches, $l = 10$, $c = 0,4$, $n_0 = 600$ e $n_1 = 800$; e na tabela 5.3 apresentamos os resultados obtidos por eles e por nós, para a fila M/M/1 com $\rho = 0,8$. As siglas da tabela tem os seguintes significados

i.c para $\hat{p}$: intervalo de confiança para a cobertura obtida;

$\bar{n}$: tamanho médio das 100 rodadas;

p.r.: precisão relativa;

m: tamanho do batch.

	Resultados de Law e Carson		
γ	∞	0,15	0,075
i.c. para $\hat{p}$	$0,85 \pm 0,06$	$0,85 \pm 0,06$	$0,87 \pm 0,06$
$\bar{n}$	32.960	32.960	75.648
<i>p.r.</i>	0,096	0,096	0,065
<i>m</i>	824	824	1.891
	Nossos Resultados		
i.c. para $\hat{p}$	$0,76 \pm 0,07$	$0,76 \pm 0,07$	$0,86 \pm 0,06$
$\bar{n}$	15.905	17.076	70.976
<i>p.r.</i>	0,121	0,117	0,044
<i>m</i>	522	569	2.366

Tabela 5.3: Resultados obtidos considerando $\rho = 0,8$

Os resultados obtidos, tanto por Law e Carson, ainda com os mesmos k , l , c , n_0 e n_1 , como por nós, para a fila M/M/1 com $\rho = 0,5$ são apresentados na tabela 5.4.

Resultados de Law e Carson			
γ	∞	0,15	0,075
i.c. para $\hat{p}$	$0,91 \pm 0,05$	$0,90 \pm 0,05$	$0,87 \pm 0,06$
$\bar{n}$	3.768	4.008	19.472
<i>p.r.</i>	0,136	0,131	0,064
m	94	100	487
Nossos Resultados			
i.c. para $\hat{p}$	$0,75 \pm 0,07$	$0,85 \pm 0,06$	$0,88 \pm 0,05$
$\bar{n}$	1.320	4.901	18.562
<i>p.r.</i>	0,027	0,035	0,019
m	44	163	619

Tabela 5.4: Resultados obtidos considerando $\rho = 0,5$

Se estamos interessados em intervalos de confiança com pequena amplitude, e não há uma preocupação particular com o tamanho da amostra resultante, então uma escolha de $\gamma = 0,075$ parece ser apropriada, onde também temos uma menor precisão relativa.

Estas tabelas foram feitas com o objetivo de comparar nossos resultados com os de Law e Carson.

5.5 Análise dos Resultados

Notamos que em ambos os casos houveram diferenças significativas entre os nossos resultados e os apresentados por Law e Carson. Pelos resultados apresentados nas tabelas 5.3 e 5.4 verificamos que os intervalos de confiança para a *cobertura obtida*, $\hat{p}$, são diferentes, menos para o valor fixado $\gamma = 0,075$; para os outros valores de γ , essas diferenças foram bastante acentuadas. Porém, com relação as meias amplitude dos intervalos de confiança, que é um valor comparativo do procedimento, estes valores foram muito próximos.

Considerando a fila M/M/1 com $\rho = 0,8$, notamos que houve uma diminuição nos tamanhos das amostras encontradas por nós, para $\gamma = \infty$ foi em torno de 51,744%, para $\gamma = 0,15$ foi de 48,192%, e finalmente para $\gamma = 0,075$ foi de 6,176%.

Para esse mesmo sistema com $\rho = 0,5$, nos resultados apresentados notamos que houve diminuição nos tamanhos das amostras encontradas, a não ser para $\gamma = 0,15$ em que houve um aumento da mesma, na ordem de 22,280%. Para $\gamma = \infty$ a diminuição do tamanho da amostra foi de 64,968%, e quando consideramos $\gamma = 0,075$ esta diminuição foi de 4,673%.

Para o sistema com $\rho = 0,8$, observamos que no caso do tamanho dos batches houve uma diminuição dos nossos resultados em torno de mais ou menos 36,650% para $\gamma = \infty$ e 30,946% para $\gamma = 0,15$, mas em compensação houve um aumento de 20,076% para $\gamma = 0,075$. Considerando $\rho = 0,5$, notamos que houve uma diminuição apenas para $\gamma = \infty$, na ordem de 53,191%. Com relação aos outros dois parâmetros analisados, existiu um aumento de 63,144% quando consideramos $\gamma = 0,15$, e de 21,325% quando o valor considerado foi $\gamma = 0,075$.

Outros resultados, relativos ao tempo médio de espera na fila W_q , foram obtidos para o sistema com $\rho = 0,8$ e são apresentados na tabela 5.5. Nesta tabela as siglas tem o seguinte significado:

$\bar{h}$: número médio de iterações para obtenção do tamanho da amostra;

$d.p.$: desvio padrão médio;

δ : meia amplitude média;

$lag\ 1$: autocorrelação de lag 1

γ	∞	0,15	0.075
$\bar{h}$	8,78	9,07	13,43
$d.p$	1,328	1,233	0,673
δ	0,411	0,398	0,209
$lag\ 1$	0,105	0,096	-0,002

Tabela 5.5: Resultados adicionais considerando $\rho = 0,8$

Notamos, através da tabela 5.5 que o desvio padrão médio decresce quando diminuimos γ , e a autocorrelação de *lag* 1 também decresce, em módulo, com o aumento da amostra (passa de 0,105 para -0,002). A meia amplitude média δ , dos intervalos de confiança melhora em aproximadamente 43,049% quando passamos de $\gamma = \infty$ para $\gamma = 0,075$ (passa, respectivamente, de $\delta = 0,411$ para $\delta = 0,209$).

A tabela 5.6 traz resultados relativos a W_q na fila M/M/1 com $\rho = 0,5$, com as siglas da mesma tendo o mesmo significado que os da tabela 5.5.

γ	∞	0,15	0,075
h	1,87	5,50	9,50
$d.p$	0,325	0,197	0,101
δ	0,103	0,062	0,031
<i>lag</i> 1	0,091	0,004	-0,037

Tabela 5.6: Resultados adicionais considerando $\rho = 0,5$

Pela tabela 5.6 verificamos que o desvio padrão médio decresce quando γ diminui, e a autocorrelação de *lag* 1 também decresce, passando de 0,091 no começo da simulação, para -0,037 no final da mesma. A meia amplitude dos intervalos de confiança obtidos no início da execução do algoritmo para $\gamma = \infty$ é 0,103, e passa para 0,031, quando consideramos $\gamma = 0,075$. Notamos que neste caso, há uma melhora na meia amplitude dos intervalos de confiança em 69,903%, que é bastante considerável.

Observamos que em ambos os casos estudados, a precisão relativa fixada do intervalo de confiança (γ), foi fundamental na obtenção dos resultados estatísticos, pois quanto menor for estes valores fixados, melhores estatísticas obtemos dos sistemas analisados.

5.6 Análise da Porcentagem de Cobertura

Numa segunda etapa desse trabalho, agora com n fixado tivemos como objetivo analisar a porcentagem de cobertura da variável observada, que também foi W_q , o tempo médio de espera dos clientes, para o sistema de fila M/M/1 estudado. Dentre os tamanhos de amostras obtidos, tanto para $\rho = 0,8$ (tabela 5.3) e $\rho = 0,5$ (tabela 5.4), consideramos três modelos para verificar a porcentagem de cobertura.

A partir dos resultados da tabela 5.3 da fila M/M/1 com $\rho = 0,8$, determinados pelo procedimento sequencial de Law e Carson, usamos como **modelo I** o tamanho de amostra médio $\bar{n} = 17.076$; que é o tamanho de amostra obtido quando fixamos $\gamma = 0,15$ no procedimento.

Fizemos então 200 rodadas de simulação considerando este modelo, utilizando o método de batch means, e cada rodada foi realizada contendo um período transiente de 513 partidas, o que equivale a 3% do total da amostra; e $k = 30$ batches, cada batch tendo 569 partidas.

Em seguida, considerando ainda como taxa de ocupação $\rho = 0,8$, tomamos como **modelo II** o tamanho de amostra médio $\bar{n} = 70.976$, obtido com $\gamma = 0,075$ e o fixamos. Também fizemos 200 rodadas de simulação para este sistema, onde cada rodada foi realizada com um período transiente de 2.129 partidas, equivalente a 3% do total da amostra; e mesmo $k = 30$ batches, cada um deles com 2.366 partidas.

E por fim, da fila M/M/1 com $\rho = 0,5$, a partir dos resultados da tabela 5.4, consideramos como **modelo III**, o tamanho de amostra médio $\bar{n} = 18.562$ que foi obtido para $\gamma = 0,075$ fixado. Novamente foram feitas 200 rodadas de simulação para esse sistema, considerando esse n médio fixo. Cada rodada com um período transiente de 557 partidas (3% do total da amostra); $k = 30$ batches, e cada um deles com 619 clientes partindo do sistema.

Para cada um desses três sistemas simulados, determinamos os seguintes valores:

1. tempo médio de espera em cada batch;

2. média das médias dos tempos de espera para os 30 batches, $\overline{\overline{X}}$;
3. desvio padrão;
4. meia amplitude do intervalo de confiança;
5. intervalo com 90% de confiança em torno da média $\overline{\overline{X}}$, utilizando a distribuição t-Student;
6. autocorrelações de *lag* 1, 2 e 3 dos tempos médios entre os batches, considerando o cálculo através do estimador usual e do estimador Jackknife, para comparar os resultados.

Não precisamos calcular autocorrelação para *lags* maiores que 3, pois teoricamente esperamos que para os batches mais distantes, mais eles deverão ser considerados independentes.

5.6.1 Resultados Encontrados

Para a fila M/M/1 com $\rho = 0,8$, os resultados obtidos, tanto para o tamanho de amostra médio, considerando $\gamma = 0,15$ ($\bar{n} = 17.076$), como para $\gamma = 0,075$ ($\bar{n} = 70.976$), apresentados na tabela 5.7. Lembramos que o tempo médio teórico de espera na fila é 3,2.

	Resultados	
	$\gamma = 0,15$	$\gamma = 0,075$
Média	3,173	3,187
Desvio padrão	1,317	0,699
Meia amplitude	0,410	0,217

Tabela 5.7: Resultados para a fila M/M/1 com $\rho = 0,8$ (modelo I e II)

Com relação as autocorrelações obtidas para as amostras correspondentes à fila M/M/1 com $\rho = 0,8$, notamos que elas foram menores que 0,05 e também é inferiores, em módulo, ao valor 0,183 (subseção 2.5.2). Os resultados referentes as autocorrelações encontradas estão nas tabelas 5.8 e 5.9.

	<i>Estimador Usual</i>	<i>Estimador Jackknife</i>
<i>Lag 1</i>	-0,017	0,021
<i>Lag 2</i>	-0,046	-0,020
<i>Lag 3</i>	-0,031	-0,009

Tabela 5.8: Autocorrelações para a fila M/M/1 com $\rho = 0,8$ e $\gamma = 0,15$ (**modelo I**)

	<i>Estimador Usual</i>	<i>Estimador Jackknife</i>
<i>Lag 1</i>	-0,044	-0,009
<i>Lag 2</i>	-0,022	0,018
<i>Lag 3</i>	-0,032	-0,001

Tabela 5.9: Autocorrelações para a fila M/M/1 com $\rho = 0,8$ e $\gamma = 0,075$ (**modelo II**)

Finalizando, os resultados apresentados na tabela 5.10, referem-se ao tamanho de amostra $n = 18.562$, obtido para o sistema de fila M/M/1 com $\rho = 0,5$, correspondendo ao **modelo III**.

	Resultados
Média	0,499
Desvio padrão	0,106
Meia amplitude	0,033

Tabela 5.10: Resultados para a fila M/M/1 com $\rho = 0,5$ (modelo III)

Neste caso, os resultados encontrados para os estimadores de autocorrelação estão na tabela 5.11.

	<i>Estimador Usual</i>	<i>Estimador Jackknife</i>
Lag 1	-0,026	0,012
Lag 2	-0,053	0,000
Lag 3	-0,050	-0,021

Tabela 5.11: Autocorrelações para a fila M/M/1 com $\rho = 0,5$ e $\gamma = 0,075$ (modelo III)

Notamos que neste caso, os resultados obtidos nas autocorrelações são menores, em módulo, quando utilizamos o estimador Jackknife, portanto podemos considerar a independência entre os batches. Algumas considerações a respeito destes resultados serão apresentadas no próximo capítulo, fazendo parte das conclusões deste trabalho.

5.6.2 Porcentagem de Cobertura Obtida

Como resultado final para análise, em cada rodada de simulação, a partir dos intervalos de 90% de confiança determinados, verificamos se o valor

teórico do tempo médio de espera na fila ($W_q = 3,2$ para as filas simuladas com $\rho = 0,8$ e $W_q = 0,5$ para a fila com $\rho = 0,5$), estava contido nesses intervalos, ou seja, obtemos as porcentagens de cobertura. Estes resultados são apresentados na tabela 5.12.

	$\rho = 0,8$		$\rho = 0,5$
$\bar{n}$	17.076	70.976	18.562
cobertura	0,880	0,905	0,905

Tabela 5.12: Porcentagem de Cobertura obtida

Para estes três modelos considerados neste trabalho, obtemos resultados muito interessantes em termos da porcentagem de cobertura. Considerando os tamanhos de amostras da fila M/M/1 quando fixamos $\gamma = 0,075$, tanto para $\rho = 0,8$ e $\rho = 0,5$ (**modelo II e III**), temos que mais de 90% dos intervalos de confiança obtidos contém o valor real da variável W_q , que corresponde aos tempos de espera na fila. Como estes intervalos de confiança foram construídos de acordo com uma distribuição t-Student com 90% de confiança, estes resultados por nós obtidos, em termos estatísticos, são muito bons. Para a fila M/M/1 com $\rho = 0,8$ e $\gamma = 0,15$ (**modelo I**), a cobertura obtida foi de 88% para os intervalos de confiança construídos de igual maneira dos sistemas anteriores, sendo um resultado bastante razoável, pois ficou apenas 2% abaixo dos 90% desejados.

Entretanto, outros resultados interessantes foram encontrados. Comparando os resultados para as duas amostras correspondentes a fila M/M/1 com $\rho = 0,8$, notamos que as médias relativas aos tempos de espera deram bastante parecidas: foi de 3,173 para $\gamma = 0,15$ (**modelo I**), e 3,187 para $\gamma = 0,075$ (**modelo II**), como pode ser visto na tabela 5.7. Verificamos que ambas as médias são menores que a média teórica $W_q = 3,2$, mas estão próximas desse valor, sendo inferior em 0,406% e 0,844% respectivamente.

Notamos também que o desvio padrão e a meia amplitude dos intervalos de confiança se reduzem praticamente a metade quando passamos de $\gamma = 0,15$ ($\bar{n} = 17.076$) para $\gamma = 0,075$ ($\bar{n} = 70.976$), ou seja, com o aumento da amostra obtemos intervalos de confiança com menores amplitude e variância, além de maior cobertura, como foi visto anteriormente.

Observando os resultados para a fila M/M/1 com $\rho = 0,5$, notamos que a média obtida para os tempos de espera na fila foi de 0,499 conforme a tabela 5.10, estando a mesma muito próxima da média teórica, $W_q = 0,5$. Além disso, vemos na tabela 5.10, que o desvio padrão e a meia amplitude dos intervalos de confiança também foi pequena, e a cobertura obtida de 90,5% foi a esperada.

Em relação as autocorrelações de *lag* l ; $l = 1, 2$ e 3 , tanto as calculadas utilizando o estimador usual ou o Jackknife, temos que para os três experimentos realizados podemos aceitar independência entre os batches, devido aos baixos valores encontrados para estas autocorrelações, como podemos ver nas tabelas 5.8, 5.9 e 5.11.

Law [LAW 77] utilizando o método de batch means e considerando tamanho de amostras fixadas, estudou o comportamento da porcentagem de cobertura para a fila M/M/1. Neste trabalho ele apresenta os resultados referentes a esta fila com taxa de ocupação de $\rho = 0,9$, tomando $k = 40$ batches, com um período transiente (ele não apresenta o tamanho desse período), e um tamanho de amostra fixado $n = 12.800$. O resultado obtido como porcentagem de cobertura para esse sistema foi de 75,50%, sendo que para esta mesma fila, mas com $\rho = 0,5$ e $\rho = 0,7$ os resultados foram bem próximos deste, contudo não são analiticamente apresentados.

Em todos os casos, as porcentagens de cobertura encontradas por ele, foram inferiores aquelas obtidas por nós, evidenciando a importancia dos procedimentos sequenciais para obtenção do tamanho de amostras.

No próximo capítulo apresentamos as conclusões que foram retiradas deste trabalho.

□

Capítulo 6

Conclusões Finais e Propostas Futuras

Neste capítulo apresentamos as conclusões que foram extraídas desse trabalho, bem como algumas propostas para futuras pesquisas nesta linha apresentada.

6.1 Conclusão

O objetivo deste trabalho foi determinar o tamanho de amostra considerando um intervalo de confiança fixado anteriormente. Utilizamos para isto o trabalho do Law e Kelton [LEK 82] sobre procedimentos sequenciais, e dentre os que foram apresentados, usamos o procedimento sequencial proposto por Law e Carson para alcançar este objetivo.

Modelamos por simulação uma fila M/M/1 com taxas $\rho = 0,8$ e $\rho = 0,5$ de ocupação do sistema, e coletamos os dados de saída referentes ao tempo médio de espera na fila. Utilizamos o algoritmo do procedimento sequencial de Law e Carson para fazermos esse trabalho.

Vimos que ocorreram diferenças entre os nossos resultados e os obtidos por Law e Carson. Pelos resultados apresentados nas tabelas 5.4 e 5.3 notamos

principalmente, que os intervalos de confiança para a *cobertura obtida*, $\hat{p}$, não são muito parecidos, a não ser para o valor fixado $\gamma = 0,075$. Esta cobertura foi melhor nos resultados apresentados por Law e Carson, mas deve-se notar que destes intervalos de confiança nos interessa devido ao procedimento, as meia amplitudes dos mesmos, que não apresentaram diferenças significativas.

Os tamanhos de amostras obtidos por nós, para a fila M/M/1 com $\rho = 0,8$, foram nos três casos considerados ($\gamma = \infty$, $\gamma = 0,15$ e $\gamma = 0,075$), inferiores aquelas encontradas por Law e Carson. No caso desse sistema com $\rho = 0,5$, houve diminuição nos tamanhos das amostras encontradas, a não ser para $\gamma = 0,15$ em que ocorreu um aumento da mesma na ordem de 22,280%. Vale ressaltar que o nosso tamanho de amostra n , é menos oneroso que o de Law e Carson.

Notamos que a escolha do ρ é fundamental na determinação do tamanho da amostra, ou seja, para $\rho = 0,8$ temos que os tamanhos das amostras são grandes, enquanto que ao considerarmos $\rho = 0,5$ houve uma diminuição bastante acentuada nos tamanhos das amostras finais.

No nosso segundo objetivo, que foi de verificar a porcentagem de cobertura para os modelos,

Modelo I: $\rho = 0,8$; $\gamma = 0,15$ e $\bar{n} = 17.076$

Modelo II: $\rho = 0,8$; $\gamma = 0,075$ e $\bar{n} = 70.976$

Modelo III: $\rho = 0,5$; $\gamma = 0,075$ e $\bar{n} = 18.562$

verificamos que para os modelos II e III esta porcentagem se encontra acima dos 90% desejados, pois quando determinamos intervalos de confiança sabemos que dentre todos os possíveis intervalos construídos, 90% deles devem conter a média teórica. Apenas para o modelo I, não obtemos os intervalos dentro do nível de confiança desejado, sendo o mesmo 2% inferior.

Sabemos que no uso diário da simulação em problemas reais, geralmente não conhecemos o valor real da variável analisada, levando-nos na maioria das vezes a requerermos intervalos de confiança de menor amplitude, que correspondem a estimadores estatísticos de mínima variância. Notamos por este trabalho, que com o aumento do tamanho da amostra, temos intervalos de confiança de menor amplitude, o que era esperado, pois para isto fixamos, de acordo com o algoritmo sequencial de Law e Carson a meia amplitude (que está relacio-

nada com o intervalo de confiança), cada vez menor para obtenção dos tamanhos de amostras.

Como podemos ver pelos resultados anteriores, este procedimento sequencial proporciona resultados muito bons para o sistema estudado nesta pesquisa, tanto para $\rho = 0,8$ como para $\rho = 0,5$. Porém é bom salientar que este método sequencial é bastante oneroso, no ponto de vista de utilização de computador, pois para obter os resultados foi necessário uma demanda muito grande de tempo de máquina. Logo se o pesquisador está interessado em uma solução rápida para seu problema, o que é muito comum na prática, esta resposta não poderá ser dada rapidamente, a não ser que o usuário tenha uma estrutura muito boa para trabalhar com este procedimento. Além disso, não é comum na prática trabalharmos com tamanhos de amostra tão grandes, o que restringe a aplicação deste procedimento.

6.2 Propostas Futuras

Apresentamos a seguir algumas propostas que deverão ser implementadas para dar continuidade a esta nossa pesquisa.

- Aplicar o procedimento sequencial de Law e Carson em outros sistemas, como por exemplo uma fila M/M/2 em série, que é relativamente simples de ser implementada, e se certificar mais da aplicabilidade deste procedimento. Posteriormente fazermos uma análise de cobertura para este procedimento, considerando os tamanhos de amostras encontrados.
- Verificar o que acontece com o sistema estudado nesta dissertação se utilizarmos outros valores para c (ver Law e Carson [LEC 79]), como 0,3 e 0,2, com relação aos tamanhos das amostras e a consequente porcentagem de cobertura.
- Avaliar a porcentagem de cobertura para tamanhos de amostras obtidos através deste procedimento sequencial para a fila M/M/1 com feedback Bernoulli, que foi outro projeto de mestrado.

□

Bibliografia

- [ALL 78] ALLEN, A.O. (1978). *Probability, Statistics and Queueing Theory*. 2ª ed., New York, Academic Press.
- [BCN 89] BANKS, J.; CARSON, J.S.; NGO SY, J. (1989). *Getting Started with GPSS/H*. 2ª ed., Annandale, Wolverine Software Corporation.
- [BEF 92] BUSTOS, O.H.; FRERY, A.C. (1992). *Simulação Estocástica: Teoria e Algoritmos*. Rio de Janeiro, IMPA. (Monografias de Matemática N° 49).
- [BER 93] BERNASCONI, G.C.R. (1993). *Números Aleatórios Antitéticos: Uma Comparação com os Resultados Usuais em Simulação*. Dissertação (Mestrado). Instituto de Ciências Matemáticas de São Carlos - Universidade de São Paulo.
- [BRE 75] BREUER, M.A. (1975). *Digital System Design Automation: Languages, Simulation and Data Base*. California, Computer Science Press.
- [COH 69] COHEN, J.W. (1969). *The Single Server Queue*. New York, Wiley.
- [COO 90] COOPER, R.B. (1990). *Introduction to Queueing Theory*. 3ª ed., Washington, Ceep Press Books.
- [DIS 75] DISNEY, R.L. (1975). Random Flow in Queueing Networks: A Review and Critique. *AIEE Transaction*, **7**(3), 268-288.
- [DEK 85] DISNEY, R.L.; KÖNIG, D. (1985). Queueing Networks: A Survey of Their Random Process. *SIAM Review*, **27**(3), 335-403.

- [EFR 79] EFRON, B. (1979). Bootstrap Methods: Another Look at the Jackknife. *The Annals of Statistics*, **1**(7), 1-26.
- [EFR 82] EFRON, B. (1982). *The Jackknife, the Bootstrap and Other Resampling Plans*. Philadelphia, SIAM. Cap 1.
- [ERL 09] ERLANG, A.K. (1909). The Theory of Probabilities and Telephone Conversations. *Nyt Tidsskrift Matematik*, **B**(20), 33-39.
- [FIS 77] FISHMAN, G.S. (1977). Achieving Specific Accuracy in Simulation Output Analysis. *Communications of the ACM*, **20**(5), 310-315.
- [GEH 85] GROSS, D.; HARRIS, C.M. (1985). *Fundamentals of Queueing Theory*. 2^a ed., New York, John Wiley.
- [LES 77] LAVENBERG, S.S.; SAUER, C.H. (1977). Sequential Stopping Rules for the Regenerative Method of Simulation. *IBM J. Res. Logist. Quart.*, **24**, 545-558.
- [LAW 77] LAW, A.M. (1977). Confidence Interval in Discrete Event Simulation: A Comparison of Replication and Batch Means. *Naval Res. Logist. Quart.*, **24**, 667-678.
- [LAW 83] LAW, A.M. (1983). Statistical Analysis of Simulation Output Data. *Operations Research*, **31**, 983-1029.
- [LEC 79] LAW, A.M.; CARSON, J.S. (1979). A Sequential Procedure for Determining the Length of a Steady State Simulation. *Operations Research*, **27**, 1011-1025.
- [LEK 82] LAW, A.M.; KELTON, W.D. (1982). Confidence Intervals for Steady-State Simulations, II: A Survey of Sequential Procedures. *Management Science*, **28**, 550-562.
- [LEK 84] LAW, A.M.; KELTON, W.D. (1984). Confidence Intervals for Steady-State Simulations, I: A Survey of Fixed Sample Size Procedures. *Operations Research*, **33**, 1221-1239.
- [LEK 91] LAW, A.M.; KELTON, W.D. (1991). *Simulation Modeling and Analysis*. 2^a ed., New York, McGraw-Hill.
- [MAC 87] MACDOUGALL, M.H. (1987). *Simulating Computer Systems, Techniques and Tools*. Massachusetts, MIT Press.

- [MEM 66] MECHANIC, H.; McKAY, W. (1966). Confidence Intervals for Dependent Data in Simulations II. *Technical Report*, ASDD, 17-22.
- [MIL 74] MILLER, R.G. (1974). The Jackknife - a Review. *Biometrika*, **61**, 1-15.
- [MGB 74] MOOD, A.M.; GRAYBILL, F.A.; BOES, D.C. (1974). *Introduction to the Theory of Statistics*. 3^a ed., New York, McGraw-Hill.
- [OER 83] ODONI, A.R.; ROTH, E. (1983). An Empirical Investigation of the Transient Behaviour of Stationary Queueing Systems. *Operations Research*, **31**, 432-455.
- [PAW 90] PAWLIKOWSKI, K. (1990). Steady-State Simulation of Queueing Processes: A Survey of Problems and Solutions. *ACM Computing Surveys*, **22**(2), 123-170.
- [RIP 87] RIPLEY, B.D. (1987). *Stochastic Simulation*. 2^a ed., New York, Wiley.
- [SCH 82] SCHEMEISER, B. (1982). Batch Size Effects in the Analysis of Simulation Output. *Operations Research*, **30**, 556-568.
- [SCH 91] SCHRIBER, T.J. (1991). *An Introduction to Simulation Using GPSS/H*. New York, John Wiley.
- [SOA 90] SOARES, L.F.G. (1990). *Modelagem e Simulação Discreta de Sistemas*. Escola de Computação, **7**; Minicurso. São Paulo, IME - USP.
- [WSB 75] WHITE, J.A.; SCHMIDT, J.W.; BENNET, G.K. (1975). *Analysis of Queueing Systems*. San Diego, Academic Press.

Apêndice A

Estimador Jackknife

O estimador Jackknife foi apresentado por Quenouille em 1949 (apud [MIL 74]) como um estimador não paramétrico; com o objetivo de reduzir o vício do estimador de correlação através da divisão do tamanho da amostra n . Esta técnica é utilizada por Law e Carson [LEC 79] e por nós, para estimar os coeficientes de autocorrelação que aparecem no algoritmo do procedimento sequencial usado no trabalho.

Suponha termos uma amostra aleatória independente e identicamente distribuída com valores $X_1, X_2, \dots, X_n$. Tendo observado $X_1 = x_1, X_2 = x_2, \dots, X_n = x_n$; calculamos algumas estatísticas de interesse, digamos:

$$\hat{\theta} = \hat{\theta}(x_1, x_2, \dots, x_n)$$

Estamos interessados no vício de $\hat{\theta}$ para estimarmos o verdadeiro valor de θ .

Seja F uma função de distribuição de probabilidade desconhecida, então $\theta(F)$ é uma estatística funcional de algum parâmetro de interesse, tal como uma esperança, um quantil, uma correlação, etc., que estimamos pela estatística

$$\hat{\theta} = \theta(\hat{F}) \tag{A.1}$$

onde $\hat{F}$ é uma *distribuição de probabilidade empírica*, dada sobre os valores observados $x_1, x_2, \dots, x_n$.

A expressão (A.1) garante que $\hat{\theta}(x_1, x_2, \dots, x_n)$ é invariante sobre permutações dos seus argumentos, e mais importante, que o conceito de vício (V) é

bem definido, dado por

$$V = E_F[\theta(\hat{F}) - \theta(F)]$$

onde E_F indica a esperança sobre F , que no caso unidimensional, $\mathcal{R}$, é dado por

$$E_F(X) = \int_{\mathcal{R}} x \, dF(X)$$

O método de Quenouille é baseado em retirar pontos x_i , e recalculando $\hat{\theta}$. Removendo pontos x_i do conjunto de dados, temos uma distribuição de probabilidade empírica diferente $\hat{F}_{(i)}$ sobre os valores observados $x_1, x_2, \dots, x_{i-1}, x_{i+1}, \dots, x_n$.

Em função disso, um correspondente valor recalculado da estatística, passa a ser:

$$\hat{\theta}_{(i)} = \theta(\hat{F}_{(i)}) = \hat{\theta}(x_1, x_2, \dots, x_{i-1}, x_{i+1}, \dots, x_n).$$

Seja

$$\hat{\theta}_{(.)} = \frac{1}{n} \sum_{i=1}^n \hat{\theta}_{(i)}$$

Daí, segue que o estimador do vício de Quenouille é por definição

$$\hat{V} = (n-1)(\hat{\theta}_{(.)} - \hat{\theta})$$

Também destacou o vício corrigido do "estimador Jackknife" de θ , como sendo

$$\tilde{\theta} = \hat{\theta} - \hat{V} = n\hat{\theta} - (n-1)\hat{\theta}_{(.)}$$

Posteriormente, em 1956, o mesmo autor generalizou esta idéia dividindo a amostra em g grupos, onde cada um deles tinha tamanho h . Por exemplo, primeiro remove-se $x_1, x_2, \dots, x_h$ observações, depois remove-se $x_{h+1}, x_{h+2}, \dots, x_{2h}$, etc. Portanto $n = gh$, e explorou esta aplicação geral.

Temos também $\hat{\theta}$ como um estimador do parâmetro θ baseado na amostra de tamanho n , e finalmente seja $\hat{\theta}_{(-i)}$ o correspondente estimador da amostra de tamanho $(g-1)h$, onde o i -ésimo grupo de tamanho h foi ignorado.

Definindo

$$\tilde{\theta}_{(i)} = g\hat{\theta} - (g-1)\hat{\theta}_{(-i)}; \quad i = 1, 2, \dots, g \quad (\text{A.2})$$

O estimador

$$\begin{aligned} \tilde{\theta} &= \frac{1}{g} \sum_{i=1}^g \tilde{\theta}_{(i)} \\ &= g\hat{\theta} - (g-1) \frac{1}{g} \sum_{i=1}^g \hat{\theta}_{(-i)} \end{aligned} \quad (\text{A.3})$$

que tem a propriedade de eliminar o termo de ordem $1/n$ de um vício da forma

$$E(\hat{\theta}) = \theta + \frac{a_1}{n} + \mathcal{O}\left(\frac{1}{n^2}\right).$$

Pouco tempo depois, Tuckey em 1958 (apud [EFR 79]) argumentou que os valores g em (A.2) poderiam ser tratados como variáveis aleatórias, aproximadamente, independentes e identicamente distribuídas em muitas situações. A estatística

$$f(\theta) = \frac{\sqrt{g(\tilde{\theta} - \theta)}}{\left[\frac{1}{g-1} \sum_{i=1}^g (\tilde{\theta}_{(i)} - \tilde{\theta})^2\right]^{1/2}}$$

pode então, aproximadamente, ter uma distribuição t-Student com $(g-1)$ graus de liberdade e constituir-se numa boa estatística para estimar um intervalo robusto. Tuckey posteriormente chamou os valores g em (A.2) de pseudo-valores e denominou (A.3) de *estimador Jackknife* com o propósito de ser uma ferramenta estatística.

Schucary, Gray e Owen em 1971 (apud [MIL 74]) publicaram uma generalização da técnica Jackknife para lidar com as formas mais gerais de vício. Este resultado, assim como muitos outros que estão ligados principalmente com a redução do vício e a estimação do intervalo, que são os dois diferentes aspectos do *jackknife*, estão todos relatados nos artigos do Miller [MIL 74], ou mesmo em Efron [EFR 82]. A abordagem destes resultados não é justificável no contexto deste trabalho, por isto não os fazemos.

□

Apêndice B

Programa de Simulação

Este apêndice apresenta o programa de simulação feito em GPSS/H para simular uma fila M/M/1, utilizando o método de batch means. Ele foi utilizado para gerar os dados que foram utilizados no desenvolvimento do algoritmo sequencial de Law Carson, e também no cálculo da porcentagem de cobertura.

SIMULATE		Executa o modelo.
*		
*Declaracao de Amper-variaveis		
INTEGER	&I	Rodadas de simulacao.
REAL	&A	Variavel auxiliar.
*		
*Controle de declaracao		
RMULT	,50120,2600834	Posicao do ponteiro
*		na sequencia de nume-
*		ros pseudo-aleatori-
*		os.
*		
SAIDA FILE	'LEC1.MIN'	Declara que os resul-
*		tados de saida da fi-
*		la serao colocados no
*		arquivo "LEC1.MIN".
*		

*Secao de Blocos do GPSS/H			
* * * * *	GENERATE	RVEXPO(2,1.0)	Entrada de clientes no sistema obedecendo uma distribuicao exponencial com media 1,0.
* * *	QUEUE	FWAIT	Inicia os calculos estatisticos da fila espera.
* *	SEIZE	SERVER	O cliente e capturado pelo servidor.
* * *	DEPART	FWAIT	Finaliza os calculos estatisticos da fila de espera.
* * * * *	ADVANCE	RVEXPO(3,0.8)	O cliente e atendido pelo servidor, segundo uma distribuicao exponencial de media 0,8.
* *	RELEASE	SERVER	O cliente libera o servidor.
* * *	TERMINATE	1	Decrementa de um cliente a cada rodada de simulacao.
*Controle de declaracao do GPSS/H			
* *	START	630	Periodo transiente n0
* * *	DO	&I=1,30	Executa 30 rodadas de simulacao por batches.

*	RESET	Zera as estatísticas, menos os relógios e as transições.
*		
*		
*	START	1920,NP
*		Simula a entrada de 1920 clientes no sis- tema.
*		
*	BLET	&A=M1
*		Declara que a varia- vel real &A receberá o valor do tempo de espera na fila.
*		
*	BPUTPIC	FILE=SAIDA, LINES=1,(&A)
*		Declara que os tempos de espera na fila se- rão colocados no ar- quivo "LEC1.MIN", na formatação de um va- lor por linha.
*		
*		
*		
*		
****.***		Estabelece o formato numérico de saída que desejamos.
*		
*		
*		
*	ENDDO	Termina um batche.
*	END	Termina a simulação.

□

Apêndice C

Tamanhos de Amostras

Neste apêndice apresentamos os resultados referentes aos tamanhos das amostras obtidos utilizando procedimento sequencial de Law e Carson, para a fila M/M/1 com taxa de ocupação ($\rho = 0,8$) e ($\rho = 0,5$)

rodada	$\rho = 0,8$			$\rho = 0,5$		
	$\gamma = \infty$	$\gamma = 0,15$	$\gamma = 0,075$	$\gamma = \infty$	$\gamma = 0,15$	$\gamma = 0,075$
1	10080	10080	57600	1800	14400	28800
2	7200	10080	40320	900	2520	7200
3	14400	14400	40320	900	2520	14400
4	7200	7200	40320	1800	5040	14400
5	14400	14400	57600	900	5040	14400
6	10080	10080	80640	900	2520	10080
7	14400	14400	57600	900	2520	20160
8	14400	14400	40320	2520	5040	20160
9	14400	14400	40320	1260	3600	14400
10	20160	28800	80640	900	3600	14400

rodada	$\rho = 0,8$			$\rho = 0,5$		
	$\gamma = \infty$	$\gamma = 0,15$	$\gamma = 0,075$	$\gamma = \infty$	$\gamma = 0,15$	$\gamma = 0,075$
11	10080	10080	57600	900	5040	20160
12	20160	20160	115200	1260	14400	28800
13	14400	14400	57600	900	5040	20160
14	10080	10080	80640	1260	3600	14400
15	10080	10080	80640	900	7200	28800
16	40320	40320	80640	1260	3600	10080
17	14400	14400	57600	1800	5040	20160
18	57600	57600	80640	900	2520	10080
19	7200	7200	57600	3600	5040	20160
20	14400	14400	57600	900	2520	14400
21	20160	20160	80640	1800	5040	14400
22	10080	14400	40320	900	3600	40320
23	14400	28800	57600	1800	10080	28800
24	10080	10080	40320	1260	3600	20160
25	14400	14400	57600	1260	5040	14400
26	14400	14400	57600	900	7200	14400
27	10080	20160	40320	900	3600	14400
28	7200	7200	80640	1800	5040	20160
29	40320	40320	115200	900	10080	20160
30	40320	40320	80640	900	1800	14400
31	20160	20160	57600	900	3600	14400
32	14400	14400	57600	900	5040	57600
33	57600	57600	161280	1800	3600	14400
34	14400	14400	80640	900	7200	28800
35	7200	7200	57600	1260	3600	20160
36	5040	5040	80640	1800	2520	20160
37	14400	14400	57600	900	1800	14400
38	5040	5040	28800	1260	3600	14400
39	5040	20160	57600	900	3600	20160
40	7200	7200	57600	900	5040	14400

rodada	$\rho = 0,8$			$\rho = 0,5$		
	$\gamma = \infty$	$\gamma = 0,15$	$\gamma = 0,075$	$\gamma = \infty$	$\gamma = 0,15$	$\gamma = 0,075$
41	10080	10080	80640	1260	2520	14400
42	20160	20160	57600	900	3600	14400
43	5040	5040	80640	900	2520	14400
44	7200	10080	57600	900	7200	20160
45	7200	7200	57600	1260	2520	28800
46	20160	20160	115200	900	2520	10080
47	7200	10080	80640	1260	5040	28800
48	20160	20160	80640	1260	7200	14400
49	20160	20160	40320	1800	5040	14400
50	28800	28800	57600	900	3600	10080
51	5040	5040	57600	2520	5040	28800
52	7200	7200	57600	900	3600	10080
53	10080	10080	40320	1260	3600	14400
54	10080	14400	80640	900	2520	7200
55	28800	28800	57600	900	5040	28800
56	20160	20160	57600	900	5040	20160
57	20160	20160	80640	2520	7200	14400
58	40320	40320	115200	1800	3600	14400
59	28800	28800	115200	1800	3600	20160
60	10080	10080	57600	2520	5040	40320
61	14400	14400	115200	900	3600	40320
62	7200	7200	80640	1260	1260	14400
63	10080	10080	80640	1260	7200	28800
64	28800	28800	57600	900	2520	10080
65	14400	80640	161280	900	5040	28800
66	14400	14400	57600	1260	7200	14400
67	7200	10080	57600	900	10080	20160
68	7200	7200	57600	1800	5040	14400
69	20160	20160	80640	900	3600	20160
70	20160	20160	57600	900	3600	14400

rodada	$\rho = 0,8$			$\rho = 0,5$		
	$\gamma = \infty$	$\gamma = 0,15$	$\gamma = 0,075$	$\gamma = \infty$	$\gamma = 0,15$	$\gamma = 0,075$
71	7200	20160	115200	2520	3600	10080
72	20160	20160	80640	2520	28800	28800
73	7200	7200	40320	1800	3600	14400
74	10080	14400	40320	900	2520	20160
75	14400	14400	115200	900	3600	10080
76	7200	7200	80640	1260	3600	20160
77	20160	20160	57600	1260	5040	14400
78	7200	7200	40320	1260	3600	40320
79	20160	20160	80640	900	7200	20160
80	40320	40320	80640	900	3600	14400
81	10080	10080	115200	1800	5040	20160
82	7200	7200	80640	900	3600	14400
83	14400	14400	40320	900	3600	10080
84	10080	14400	115200	1800	3600	40320
85	7200	14400	80640	1800	3600	20160
86	10080	10080	80640	1260	7200	10080
87	14400	14400	57600	1260	3600	14400
88	10080	14400	115200	1260	2520	10080
89	10080	10080	80640	900	14400	20160
90	10080	10080	40320	1800	2520	10080
91	7200	7200	57600	900	5040	14400
92	10080	10080	57600	900	2520	20160
93	10080	14400	40320	1260	2520	10080
94	40320	40320	115200	1800	3600	14400
95	10080	10080	57600	900	5040	14400
96	14400	14400	40320	3600	10080	20160
97	28800	28800	80640	900	3600	14400
98	40320	40320	80640	900	3600	10080
99	14400	14400	115200	1800	3600	14400
100	28800	28800	57600	900	5040	20160

□