

COPPEAD/UFRJ

RELATÓRIO COPPEAD Nº 187

ESTUDO DA EFICIÊNCIA DA AMOSTRAGEM  
DESCRITIVA NA SIMULAÇÃO DE UM  
PROBLEMA DE FILAS

Eduardo Saliby \*  
Gilberto Funchal \*\*

Maio de 1987

\* Professor Adjunto da área de Métodos Quantitativos e pesquisador da COPPEAD - Instituto de Pós-Graduação e Pesquisa em Administração/UFRJ.

\*\* Mestrando em Engenharia de Sistemas - Pesquisa Operacional pelo Instituto Militar de Engenharia (IME).

## RESUMO

Uma das idéias básicas da simulação por Monte Carlo consiste no uso da amostragem aleatória simples, reproduzindo assim o comportamento probabilístico de uma variável aleatória. Recentemente, no entanto, foi sugerido um método de amostragem que se considera mais adequado à simulação: a amostragem descritiva. Em essência, a amostragem descritiva baseia-se numa seleção totalmente determinística dos valores amostrais e na sua permutação aleatória.

Este trabalho tem por objetivo mostrar, através de uma comparação empírica dos dois métodos, a maior eficiência estatística da amostragem descritiva. Para isso, foi estudada a simulação de um problema de filas relativo à descarga de navios em um terminal portuário.

Além da comprovação de que a amostragem descritiva conduz a estimativas mais precisas verificou-se também uma substancial redução do tempo de processamento.

## ABSTRACT

The use of simple random sampling, in order to reproduce the probabilistic behavior of a random variable, represents a basic idea in simulation. Recently, however, it was suggested a new sampling method that seems to be more appropriate: descriptive sampling. Descriptive sampling is based on a fully deterministic selection of the sample values and into their random permutation.

The purpose of the present work is to show, by means of an empirical test, that descriptive sampling is statistically more efficient. It was studied a queueing problem concerning the unloading operation of iron ore at a port.

The results do confirm that descriptive sampling produces more precise simulation estimates. Apart from it, a substantial reduction on the processing time is also reported.

## 1. INTRODUÇÃO

Recentemente Saliby (1980 e 1982b) propôs uma nova abordagem em simulação por Monte Carlo. Tal proposta se refere à utilização da Amostragem Descritiva (AD) em lugar da Amostragem Aleatória Simples (AAS), até então o método usual. A justificativa da utilização da amostragem descritiva, conforme descreve o seu autor, se prende ao fato de que ela produz resultados mais precisos que os obtidos com a amostragem aleatória simples, sendo portanto estatisticamente mais eficiente.

Este trabalho se propõe a comparar resultados da simulação de um problema de filas, obtidos segundo as duas abordagens (amostragem aleatória simples e amostragem descritiva). No caso da amostragem aleatória simples, procedeu-se também à aplicação do Modelo Linear de Resposta (MLR) (Saliby, 1980) onde parte da variabilidade das estimativas é explicada pelos desvios observados entre as médias das amostras de entrada e seus respectivos valores teóricos.

O problema estudado, descrito na seção seguinte, refere-se ao caso Great Western (Vatter, P.A. et al., 1978) relativo à operação de um porto de descarga de minério.

Este trabalho também representa uma contribuição para o desenvolvimento de procedimentos para a geração de amostras descritivas que, por se tratar de uma teoria recente, ainda carece de métodos mais adequados para esse fim.

Embora não se tivesse preocupado com uma avaliação do seu desempenho computacional, a amostragem descritiva, além de levar

a estimativas mais precisas, resultou ainda em menores tempos de processamento.

Antes de passar à descrição do problema, algumas considerações sobre a amostragem em simulação são necessárias (Saliby, 1986).

São duas as fontes de variabilidade das estimativas obtidas com a amostragem aleatória simples :

- a variação de conjunto e
- a variação de seqüência.

A variação de conjunto, como o próprio nome indica, está associada ao conjunto de valores das amostras de entrada, independentemente de sua ordem de ocorrência.

Já a variação de seqüência está associada, justamente, à ordem em que os valores das amostras são utilizados.

A primeira fonte de variabilidade pode ser eliminada com a utilização da amostragem descritiva.

A segunda fonte, associada à seqüência, é na verdade a única a continuar presente numa simulação.

A amostragem descritiva caracteriza-se por uma seleção determinística dos valores amostrais que são permutados aleatoriamente. Dessa forma, utilizando sempre os mesmos valores amostrais em diferentes corridas, a variação de conjunto pode ser praticamente eliminada.

O Modelo Linear de Resposta (Saliby, 1980) foi o ponto de partida para a teoria da amostragem descritiva. Segundo ele, as estimativas de uma simulação tendem a se relacionar com os momentos amostrais das variáveis de entrada, segundo um modelo de regressão linear. Dessa forma, o MLR não só explica o efeito de conjunto como também avalia, através do coeficiente de

determinação ( $R^2$ ), sua influência nas variâncias das estimativas.

Assim, uma importante comparação surge entre a parcela da variação explicada pelo MLR e a redução de variância obtida com a amostragem descritiva.

## 2. DESCRIÇÃO DO PROBLEMA SIMULADO

O problema "Great Western Steel Corporation" (Vatter, P.A. et al, 1978) pode ser resumidamente descrito da seguinte forma :

A Great Western Steel Corporation operava duas docas em um porto da costa oeste dos Estados Unidos. Nelas eram descarregados em média 250 navios de minério de ferro por ano, vindos da Venezuela. Os navios eram aproximadamente do mesmo tipo e tamanho. Cada doca podia descarregar somente um navio por vez e o tempo requerido para essa descarga variava de 24 a 36 horas, conforme mostra o quadro 1.

Quadro 1 - Distribuição de probabilidade dos tempos de descarga das docas

Tempo de Descarga ( horas )	Porcentagem de navios
24	4
25	9
26	18
27	13
28	10
29	5
30	4
31	4
32	6
33	8
34	11
35	6
36	2

A jornada de trabalho era de 24 horas por dia, 7 dias por semana. Algumas vezes ocorria que um navio a caminho do porto encontrava as duas vagas ocupadas, tendo de esperar para ser

descarregado; dificilmente, no entanto, isto demorava mais do que umas poucas horas.

Entretanto, uma nova usina siderúrgica estava sendo concluída, o que aumentaria a necessidade de minério de ferro e, conseqüentemente, a chegada de navios. Cerca de 500 carregamentos anuais seriam agora necessários e a gerência estava preocupada com a possibilidade de os navios, pelos quais a companhia pagava um aluguel de \$1400 por dia, terem de esperar um longo período de tempo para serem descarregados.

Elaborou-se um estudo acerca da possibilidade de tornar mais regulares as chegadas dos navios, mas uma variedade de condições encontradas durante as viagens tornava isso impossível. Um estudo dos dados passados mostrava que a chegada dos navios era completamente imprevisível, sendo o tempo entre duas chegadas consecutivas, para o novo tráfego previsto, dado pela distribuição exponencial com média de  $365 / 500 = 0.73$  dias.

Outra opção seria a construção de mais uma doca nas proximidades, semelhante às existentes, que custaria cerca de \$1.3 milhão.

Para um estudo da viabilidade econômica desta alternativa necessitava-se determinar o tempo médio de espera dos navios considerando-se os dois casos : 2 e 3 docas.

Este problema foi resolvido por simulação.

### 3. METODOLOGIA UTILIZADA

Simulação utilizando amostragem aleatória simples

Para se comparar a operação do porto com 2 e 3 docas, adotou-se o seguinte procedimento :

- Um experimento de simulação foi definido por  $M = 100$  corridas independentes.
- Para cada caso (2 e 3 docas) foram testadas corridas com duração de 5 e 10 anos.
- Em cada corrida, foram levantadas as seguintes estatísticas :
  - \* Tempo médio entre chegadas ( $\overline{TC}$ ).
  - \* Tempo médio de descarga de cada doca ( $\overline{TD}$ ).
  - \* Tempo médio de espera dos navios ( $\overline{W}$ ).

As duas primeiras médias referem-se às amostras de entrada; elas foram levantadas para serem utilizadas como dados de entrada do MLR. A terceira corresponde à estimativa em estudo, que também é utilizada no MLR.

- Sintetizando os resultados do experimento, foram calculadas as seguintes estatísticas globais :
  - \* Média dos  $M = 100$  tempos médios de espera dos navios ( $\overline{W}$ ).
  - \* Variância dos  $M = 100$  tempos médios de espera dos navios ( $\text{Var}(\overline{W})$ ) que serve como medida de precisão das estimativas.



\* Coeficiente de determinação ( $R^2$ ) calculado através da regressão múltipla do MLR, tendo como variáveis independentes o tempo médio entre chegadas  $\overline{TC}$  e os tempos médios de descarga  $\overline{TD}$  de cada doca, e como variável dependente o tempo médio de espera  $\overline{W}$ . Cada corrida originou uma observação para esta regressão.

O valor de  $R^2$  do MLR define a parcela da variância da estimativa que é explicada pela variação das médias  $\overline{TC}$  e  $\overline{TD}$ . Trata-se, portanto, de um erro associado ao conjunto de valores de entrada. Utilizando os desvios das médias como controle, chega-se à variância reduzida da estimativa, dada pela relação :

$$\text{Var}_R(\overline{W}) = (1 - R^2) \times \text{Var}(\overline{W}).$$

Assim, pode-se prever a variância da estimativa, uma vez descontado o efeito de conjunto.

#### Simulação utilizando a amostragem descritiva

Foram feitos os mesmos experimentos que no caso anterior, exceção feita à análise do MLR que não se aplica à amostragem descritiva. Isto se dá porque as médias das amostras de entrada não mais variam.

- As amostras descritivas para os tempos de descarga (TD), cujos valores são agora deterministicamente selecionados, foram definidas da seguinte maneira :

\* A amostra de entrada de cada corrida foi dividida em

subamostras descritivas, cada uma delas composta de 500 observações. Como se tem 500 chegadas anuais, foram usadas 5 subamostras numa corrida de 5 anos e 10 numa de 10 anos. Esta divisão de uma amostra descritiva em subamostras, a que se denominou blocagem, levou a uma simplificação no processo de geração. No entanto, o seu efeito foi previamente estudado, para se certificar de que não resultaria em distorções nas estimativas.

- \* Os valores TD, para as 500 observações de uma subamostra descritiva constam do quadro 2. Note-se que as frequências relativas de cada valor são iguais às respectivas probabilidades.

Quadro 2 - Valores que compõem uma subamostra descritiva dos tempos de descarga das docas

TD (horas)	Quantidade de valores (TD) na subamostra descritiva
24	20
25	45
26	90
27	65
28	50
29	25
30	20
31	20
32	30
33	40
34	55
35	30
36	10
Total	500

- \* Os tempos de descarga para todas as docas foram obtidos a partir de uma única amostra descritiva.
- As amostras descritivas para os tempos de chegada (TC), também com valores deterministicamente selecionados, foram definidas da seguinte forma :

- \* O tamanho de cada subamostra descritiva foi de 2500 observações.

- \* Os valores TC, referentes às 2500 observações da distribuição de chegada exponencial com média 0.73 dias, foram definidos segundo a fórmula :

$$TC_i = - 0.73 \times \ln[(i - 0.5) / 2500], i = 1, \dots, 2500,$$
que corresponde à aplicação do método da transformada inversa para a geração de amostras descritivas (Saliby, 1980 e 1982b).

A simulação do problema, para os dois métodos amostrais, foi feita em um microcomputador de 16 bits, utilizando-se a linguagem TURBO-PASCAL.

## 4. RESULTADOS OBTIDOS COM A AMOSTRAGEM ALEATÓRIA SIMPLES

Um resumo dos resultados globais obtidos, utilizando-se a amostragem aleatória simples, consta do quadro 3.

Destes resultados verifica-se que :

- os valores de  $R^2$  são significativamente altos ( 36 a 49 % ), especialmente se for considerado que representam uma parcela de erro que poderia ser evitada.
- as médias globais ( $\bar{W}$ ) são aproximadamente iguais, independentemente da duração ser de 5 ou 10 anos. Por outro lado, as variâncias ( $\text{Var}(\bar{W})$  e  $\text{Var}_R(\bar{W})$ ) mostraram-se inversamente proporcionais ao tamanho das amostras ( lei do inverso de n ).

Quadro 3 - Resumo dos resultados obtidos com AAS para M = 100 corridas independentes

Número de Docas	Duração (anos)	$\bar{W}$ (dias)	$\text{Var}(\bar{W})$	$R^2$	$\text{Var}_R(\bar{W})$
2	5	1.37667	0.06555	0.493	0.03322
	10	1.40023	0.02773	0.403	0.01654
3	5	0.15007	0.00039	0.360	0.00025
	10	0.15025	0.00016	0.404	0.00010

Como informação complementar, o tempo de processamento para cada experimento ( $M = 100$  corridas) foi cerca de 100 minutos no caso de 5 anos e, portanto, 200 minutos no caso de 10 anos.

## 5. RESULTADOS OBTIDOS COM A AMOSTRAGEM DESCRITIVA

Os resultados globais, utilizando-se amostragem descritiva, constam do quadro 4. Como distinção, o tempo médio de espera é agora representado por  $\overline{WD}$ .

Quadro 4 - Resumo dos resultados obtidos com AD para  $M = 100$  corridas independentes

Número de Docas	Duração (anos)	$\overline{WD}$ (dias)	Var( $\overline{WD}$ )
2	5	1.40018	0.03323
	10	1.41995	0.01930
3	5	0.15245	0.00022
	10	0.15174	0.00009

Observa-se a partir destes resultados que :

- Como no caso anterior, as médias globais  $\overline{WD}$  são aproximadamente iguais para 5 e 10 anos, enquanto as variâncias seguem a lei do inverso de n.
- as médias globais das estimativas são praticamente as mesmas que no caso da AAS. Isto permite concluir que a

AD não resultou em estimativas tendenciosas.

- após constatada a inexistência de viés relevante, comparando-se  $\text{Var}(\overline{WD})$  com os valores de  $\text{Var}(\overline{W})$  do quadro 3, verifica-se que as variâncias são agora sensivelmente menores.
- além do mais, conforme previsto pela teoria do Modelo Linear de Resposta, verifica-se que as variâncias reduzidas  $\text{Var}_R(\overline{W})$  da amostragem aleatória simples são muito próximas das variâncias  $\text{Var}(\overline{WD})$  obtidas com a amostragem descritiva.

Confirma-se, dessa forma, que a redução de variância obtida com a amostragem descritiva, corresponde ao efeito de conjunto que é desnecessariamente introduzido com a amostragem aleatória simples.

Quanto ao tempo de processamento, cada experimento levou cerca de 40 minutos para as corridas de 5 anos e 80 minutos para as de 10 anos. Vê-se portanto que, com a amostragem descritiva, o esforço computacional reduziu-se em 60 %. Isto se explica porque o conjunto de valores descritivos é gerado uma única vez para todas as corridas de um experimento; apenas sua permutação aleatória - um procedimento mais rápido - é gerada novamente a cada corrida. Este ganho torna-se maior à medida em que cresce o número de corridas do experimento.

Considerando-se ainda a redução de variância obtida ( da ordem de 40 % ), o ganho computacional global foi de 76 %. Assim, com apenas 1/4 do tempo de processamento, a amostragem descritiva

produziria estimativas de mesma precisão que a amostragem aleatória simples.



## 6. CONSIDERAÇÕES FINAIS

Tal como ocorreu nos testes anteriores, os resultados deste estudo confirmaram plenamente a validade da amostragem descritiva em simulação.

Neste caso, as estimativas para o tempo médio de espera não mostraram qualquer tendenciosidade que fosse relevante. Além disso, sua variância foi significativamente menor.

Outra constatação importante para a validação da amostragem descritiva é que as previsões de redução de variância dadas pelo MLR foram plenamente confirmadas. Assim, as variâncias da estimativa no caso da amostragem descritiva foram bastante próximas das variâncias reduzidas para a amostragem aleatória simples.

Um resultado igualmente importante refere-se à acentuada redução do tempo de processamento.

Aumentando a eficiência estatística e reduzindo o tempo de processamento, a amostragem descritiva viabiliza novas aplicações da simulação, notadamente quando o tempo de resposta é crítico. Isto ocorre, por exemplo, em aplicações militares.

## REFERENCIAS

SALIBY, E. A reappraisal of some simulation fundamentals.  
Lancaster, Universidade de Lancaster, 1980. Tese de Doutorado.

\_\_\_\_\_ Uma nova comprovação de que o uso de amostragem aleatória simples em simulação é incorreto : estudo do caso particular do problema do jornaleiro. Rio de Janeiro, COPPEAD/UFRJ, 1981. (Relatório de Pesquisa, 29).

\_\_\_\_\_ Sobre o uso incorreto de amostragem aleatória simples em simulação : um estudo da distribuição da amplitude de uma amostra uniforme. Revista de Administração, 17 (3) : 78 - 86, 1982a.

\_\_\_\_\_ Uma revisão dos fundamentos da simulação : o uso incorreto de amostragem aleatória simples. Pesquisa Operacional, 2 (2) : 1 - 16, 1982b.

\_\_\_\_\_ Sources of variation of simulation estimates : an empirical study. Rio de Janeiro, COPPEAD/UFRJ, 1986. (Relatório de Pesquisa, 64).

VATTER, P.A. et al. Quantitative methods in management : text and cases. Homewood, Irwin, 1978.