ENSINO DE SIMULAÇÃO: AVALIAÇÃO DO IMPACTO DA ESCOLHA DAS DISTRIBUIÇÕES DE PROBABILIDADE NA AUSÊNCIA DE DADOS

Luana Machado dos Santos – luanavip25@hotmail.com
Maria José Pereira Dantas - mjpdantas@gmail.com
Mestrado em Engenharia de Produção e Sistemas (MEPROS) da Pontifícia Universidade Católica de Goiás (PUC-GO)
Av. Universitária, 1440, Área III, Bloco D, Sala 308.
CEP 74.605-010, Goiânia-GO, Brasil

Resumo: *A simulação é um método numérico que pode ser aplicado para investigar praticamente qualquer tipo de sistema estocástico. A identificação incorreta das distribuições de probabilidade pode levar à obtenção de estimativas das saídas do sistema real que não correspondem à realidade. Entre as muitas aplicações da simulação estão os estudos relacionados aos sistemas de filas. Neste trabalho experimentou-se fazer alterações nas distribuições de probabilidade que descrevem o tempo de atendimento de um sistema de filas simples, com um único canal de atendimento, que atende por ordem de chegada. Supôs-se que o tempo de atendimento fosse corretamente descrito por uma distribuição log-normal, e fez-se estimativas de distribuições alternativas como a triangular e beta, considerando parâmetros correspondentes obtidos por procedimentos heurísticos. Avaliou-se, em seguida, o impacto no tempo de espera na fila. Os experimentos foram conduzidos na versão livre do Software de Simulação FLEXSIM, que pode ser facilmente obtido pelos estudantes e usado em aulas de laboratório. Observou-se nos resultados erros de até 87,25%. Uma etapa importante no ensino de simulação é mostrar ao estudante e futuro analista de modelos a importância da representação correta do comportamento das variáveis aleatórias presentes na descrição de um sistema real, na situação em que não existem dados disponíveis.*

Palavras-chave: *Ensino de Simulação, Teoria das Filas, Escolha de Distribuição de Probabilidade, Software Flexsim.*

1. INTRODUÇÃO

Banks & Chwif (2010) apresentam algumas questões importantes para que se tenha êxito na construção de um modelo de simulação. Entre as questões apresentadas os autores iniciam uma discussão sobre a escolha incorreta das distribuições de probabilidade e avaliam o impacto para algumas situações experimentais. Conway & McClain (2003), abordam pontos importantes para se desenvolver bons projetos de simulação. Um dos pontos enfatizados pelos autores é a melhor forma de se representar os dados em função da dificuldade de não se dispor deles.

A dificuldade de não se ter dados é um problema real. Em algumas situações é necessário definir uma distribuição de probabilidade para representar a aleatoriedade na ausência de dados. Isto pode ocorrer se o sistema ainda não existe; o sistema existe, mas não é possível a coleta dos dados por uma série de motivos (tempo, por exemplo); ou o sistema existe, mas os dados foram coletados automaticamente e estão em um formato diferente do formato adequado. Nestes casos são desenvolvidos procedimentos heurísticos para a escolha da distribuição (LAW, 2007).

Neste trabalho avalia-se um sistema de fila de espera simples com uma taxa de chegada exponencial conhecida e um tempo de atendimento desconhecido. O tempo de atendimento é uma quantidade aleatória e pode ser representada no modelo de simulação por uma variável aleatória contínua. O tempo de atendimento pode ser, por exemplo, o tempo gasto por um técnico para se reparar uma peça de um equipamento após o mesmo ter falhado. O objetivo é considerar alguns modelos para se escolher uma distribuição de probabilidade para o tempo de atendimento e avaliar os erros obtidos, caso eles existam. Os experimentos apresentados têm a finalidade de demonstrar ao estudante de simulação um dos pontos críticos da modelagem, que ocorre na fase de concepção do modelo.

Este artigo está estruturado em 5 seções. A seção 2 apresenta um estudo da Teoria de Filas e a definição das distribuições matemáticas. A seção 3 expõe a metodologia do trabalho. A seção 4 apresenta o desenvolvimento da simulação e seus principais resultados e discussão. A seção 5 contém as considerações finais do estudo e as referências bibliográficas.

2. REFERENCIAL TEÓRICO

Os experimentos foram realizados em sistema de fila de espera. As filas estão presentes em uma série de sistemas reais: aeroportos, praças de pedágio, restaurantes, bancos, supermercados, entre outros (BATEMAN et al., 2013). Assim, a seção 2.1 descreve brevemente a teoria das filas. Em seguida, na seção 2.2, apresentam-se as características das distribuições de probabilidade utilizadas neste trabalho.

2.1. Teoria das filas

A teoria de filas é um método da Pesquisa Operacional que utiliza fórmulas matemáticas e aplica cálculo de probabilidades para solucionar problemas de filas (ALVES et al., 2013; PRAIA & GOMES, 2013).

Um sistema de filas de espera é caracterizado por um processo de chegada que descreve como os clientes chegam ao sistema, o processo de serviço apresenta como os clientes são atendidos e uma disciplina da fila representa de que forma os clientes que estão chegando serão atendidos. As principais regras são Primeiro que Chega e o Primeiro que sai - First In, First Out (FIFO); Último a Chegar e o Primeiro a Sair - Last In First Out (LIFO).

Um modelo de filas é representado pelos clientes que chegam (fonte de entrada) ao longo do tempo em busca de um determinado atendimento que é representado pela fonte de entrada, que resultam em um sistema de fila. Em um determinado momento, um cliente será chamado para atendimento de acordo com a disciplina da fila que foram criadas as regras para chamada de clientes. Com isso o cliente é atendido pelo mecanismo de atendimento e assim que finalizar o serviço, o cliente deixara o sistema (HILLIER E LIEBERMAN, 2010).

Para descrever um sistema de filas, é utilizado a Notação de Kendall que é uma forma simplificada a partir da representação de cinco características que apresentam diante das possibilidades de filas (FOGLIATTI E MATOS, 2007).

A notação utilizada por Kendall (1953) é utilizada para representar as filas, sendo A/B/C/D/E, em que: A, indica o processo de chegada; B, refere–se ao processo de tempo de atendimento; C, representa a quantidade de canais de serviços; D, representa pela capacidade de usuários no sistema; e, E representa o tamanho da população (SINAY, 2004).

O modelo analítico M/M/1 é o caso mais simples e clássico no ensino de teoria de filas e de simulação. Pereira & Dantas (2015) confrontam os resultados dos dois modelos para se mostrar o regime estável e analisar os resultados da simulação frente aos resultados teóricos. Ocorre que em sistemas reais o tempo entre chegadas e tempo de atendimento com distribuições exponenciais não é um modelo muito realista.

As medidas de desempenho de uma fila M/M/1 são descritas na Tabela 1. A fila M/M/1 não foi utilizada neste trabalho na parte experimental, mas esclarece ao estudante diferenças marcantes entre os modelos analíticos e um modelo de simulação. No modelo de simulação os alunos podem ver cada cliente chegando no sistema e o que ocorre à medida que o tempo passa, podendo-se variar as distribuições de probabilidade e ver o impacto na saída da simulação.

Tabela 1 – Fórmulas para as medidas de desempenho do modelo com um único servidor (M/M/1)

Parâmetros	Símbolo	Fórmula	Eq.
Número de clientes Esperado Sistema	L	$\frac{\lambda}{\mu - \lambda}$	(1)
Número de clientes Esperado na Fila	Lq	$\frac{\lambda^2}{\mu(\mu - \lambda)}$	(2)
Tempo Previsto de Espera no sistema (Incluindo o tempo do serviço)	W	$\frac{1}{\mu - \lambda}$	(3)
Tempo médio de espera na Fila	Wq	$\frac{\lambda}{\mu(\mu - \lambda)}$	(4)
Probabilidade de que o sistema esteja vazio	Po	$1 - \frac{\lambda}{\mu}$	(5)
Utilização do Sistema	ρ	$\frac{\lambda}{s\mu}$	(6)

Fonte: adaptado de Moore & Wheatherford (2005) apud Florêncio

Neste trabalho aborda-se o problema em que o tempo de atendimento segue uma distribuição desconhecida. O modelo analítico mais adequado é o M/G/1/∞/FIFO. Neste caso os tempos entre as chegadas sucessivas são exponencialmente distribuídos com média $\frac{1}{\lambda}$, os tempos de atendimento seguem uma distribuição desconhecida com média igual a $\frac{1}{\mu}$ e desvio padrão σ, existe um único canal de serviço que atende por ordem de chegada e não há limitação física para a fila (FOGLIATTI E MATOS, 2007). Para o modelo M/G/1 o tempo de espera na fila, w_q é dado pela Eq. (7).

$$w_q = \frac{\rho^2 + \lambda^2 \sigma_s{}^2}{2\lambda(1 - \rho)} \tag{7}$$

2.2. Tipos de distribuições de probabilidade

A distribuição de probabilidade **exponencial** é utilizada na teoria das filas para modelar tempos aleatórios, tais como a chegada de clientes (FERNANDES, 2005). Esta distribuição é a mais fácil de usar, porque tem apenas um parâmetro e tem boas propriedades matemáticas que levam as equações de estado estacionário, bastante simples para calcular as medidas de desempenho em Filas M/M/1 (BANKS & CHWIF, 2010).

A distribuição de probabilidade **log-normal** é utilizada para modelar tempos de processamentos manuais (FERNANDES, 2005; BANKS E CWHIF, 2010). Esta distribuição possui dois parâmetros, a média e o desvio padrão. O desvio padrão representa a forma da distribuição. Quanto menor for o desvio padrão mais a curva se aproxima da vertical. O segundo parâmetro é a média, que não afeta na curva e sim na escala (LAW, 2007).

A distribuição da probabilidade **triangular** é caracterizada por 3 parâmetros: *a* (valor mínimo), *b* (valor máximo) e *m* (valor mais provável, a moda), sendo representada por um gráfico em formato de triângulo (FERNANDES, 2005; MACHADO & FERREIRA, 2012). É importante mencionar que está distribuição é utilizada em situações em que não se conhece a forma exata da distribuição de probabilidade. É como um modelo na ausência de dados, ou seja, quando não há dados disponíveis utiliza-se esta distribuição (LAW, 2007; VILCAPOMA et al., 2013). O principal problema é ser uma distribuição truncada. É utilizada normalmente no início da simulação para modelar o tempo de chegada e o tempo de atendimento (BANKS, CHWIF, 2010).

A última distribuição de probabilidade utilizada neste artigo é a distribuição **beta** que é aplicada para modelar o tempo de serviço (FERNANDES, 2005). Devido à sua flexibilidade, é muito utilizada quando não se dispõe de dados.

3. METODOLOIA

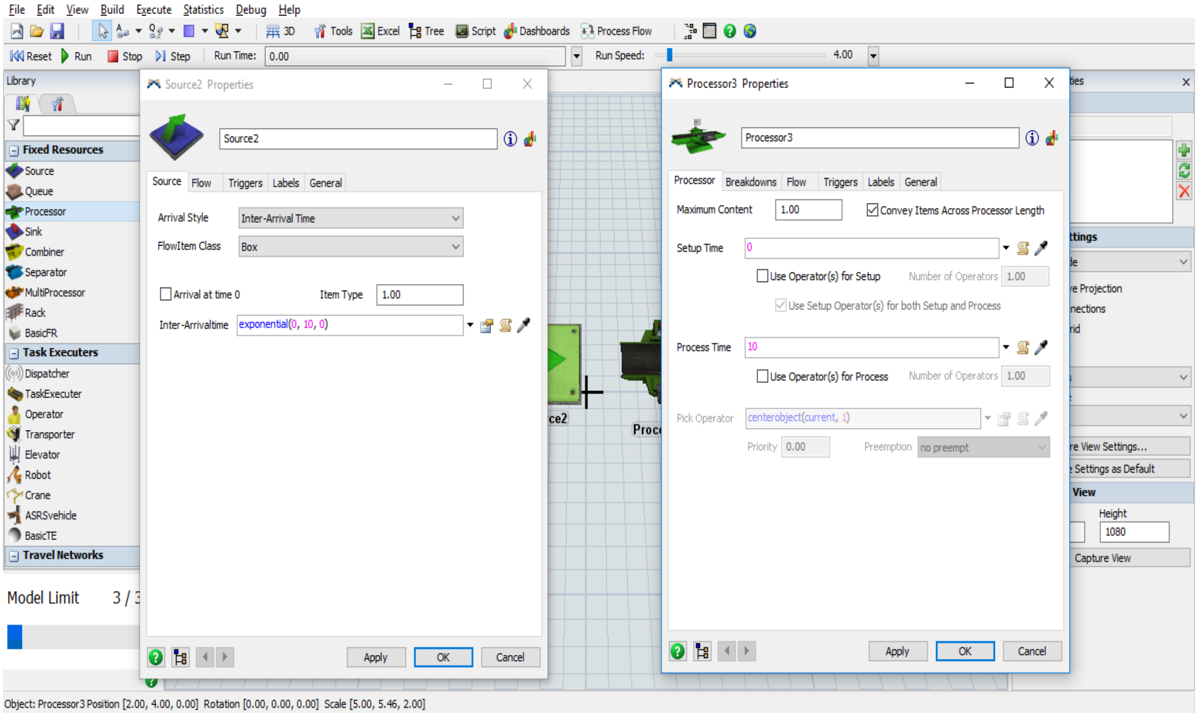
Para demonstrar qual o impacto da escolha da distribuição de probabilidade quando não se tem dados escolheu-se o *Software Flexsim*, que possui versão livre disponível no site www.flexsim.com/pt.

Esta versão possui diversos recursos básicos para modelar um pequeno sistema, permitindo salvar o modelo que foi criado. É uma ótima alternativa para realizar o experimento com um *software* que desenvolve relatórios de análise estatísticas e animação gráfica para facilitar o entendimento da simulação, permitindo que os usuários possam identificar os problemas e avaliar as soluções em um curto tempo (GUIMARÃES, 2015).

Passo 1: No ambiente de trabalho do *Software Flexsim* existem vários objetos utilizados na construção de um modelo de simulação. Foram utilizados para a chegada (*Source*), para a fila (*Queue*), para o processo (*Processor*) e a saída (*Sink*). Esses objetos podem ser visualizados no painel da biblioteca que está localizado no lado esquerdo da tela. Todos os objetos do *Software Flexsim* necessitam ser ligados para se comunicarem uns com os outros e para definir o fluxo dos processos. Neste caso, para fazer estas conexões, utiliza-se a letra A do teclado pressionada para ligar uma porta de saída a uma porta de entrada de cada objeto. Se houver necessidade de desconectar algum objeto utiliza a letra Q do teclado pressionada.

Passo 2: Nesta fase ocorreu a escolha das distribuições e definição de parâmetros. Com um duplo click em cima do objeto abriu-se uma tela para a definição de propriedades (*properties*). A Figura 2 mostra a tela para os objetos do tipo *Source* e *Processor* que são os objetos que foram utilizados para definição dos parâmetros no experimento a ser realizado.

Figura 2 – Telas de entrada para definição dos parâmetros dos objetos *Source* (chegada) e *Processor* (atendimento)



Fonte – Flexsim (2016)

O Quadro 1 apresenta as heurísticas para as estimativas das distribuições do tempo de atendimento e o Quadro 2 apresenta um resumo dos parâmetros.

Quadro 1- Algoritmo para estimação das distribuições.

1: **o especialista estima** os valores de *a* (estimativa pessimista) e *b* (estimativa otimista) do tempo para executar uma tarefa, tal que os valores de *a* e *b* sejam números reais, $a < b$, e que a probabilidade $P\ (a \leq X \leq b) \approx 1$. Em seguida o especialista atribui uma distribuição de probabilidade para representar a variável *X*.
2: Se o especialista associa a distribuição **triangular**, basta agora definir o valor *m* que significa o valor mais provável, ou seja, a moda. Assim a Triangular está estimada T(*a*, *m*, *b*). A distribuição Triangular é truncada. Ela difere substancialmente de uma log-normal que tem uma longa cauda pela direita.
3: Se o especialista opta por uma **beta**, no mesmo intervalo deve estimar os parâmetros de *shape*, α_1 e α_2. Esta opção é mais flexível pois existe uma variedade de parâmetros de *shape* que a distribuição pode assumir. Por outro lado, não está claro como se deve proceder. Uma primeira ideia é assumir que a variável pode tomar valores uniformemente distribuídos entre *a* e *b*, neste caso $\alpha_1=1$ e $\alpha_2=1$. Isto resulta que *X* terá uma distribuição uniforme U(*a*,*b*). Este modelo pode ser usado quando se conhece pouco sobre o comportamento da variável na faixa [*a*, *b*]. Uma ideia mais realista é assumir que a distribuição é assimétrica a direita, ou seja $\alpha_2 > \alpha_1 > 1$. Assim, pode-se estimar os parâmetros de *shape*. A distribuição beta possui média μ e moda *m*, dadas pelas Equações (8) e (9).

$$\mu = a + \frac{\alpha 1(b-a)}{\alpha_1 + \alpha_2} \tag{8}$$

$$m = a + \frac{(\alpha_1 - 1)(b - a)}{\alpha_1 + \alpha_2 - 2} \tag{9}$$

Estas equações podem ser resolvidas para se obter as estimativas de $\tilde{\alpha}_1$e $\tilde{\alpha}_2$ por meio das Eq. (10) e (11).

$$\tilde{\alpha}_1 = \frac{(\mu - a)(2m - a - b)}{(m - \mu)(b - a)} \tag{10}$$

$$\tilde{\alpha}_2 = \frac{(b - \mu)\tilde{\alpha}_1}{(\mu - a)} \tag{11}$$

Não se deve esquecer que $\mu < m$, para garantir a assimetria pela direita.

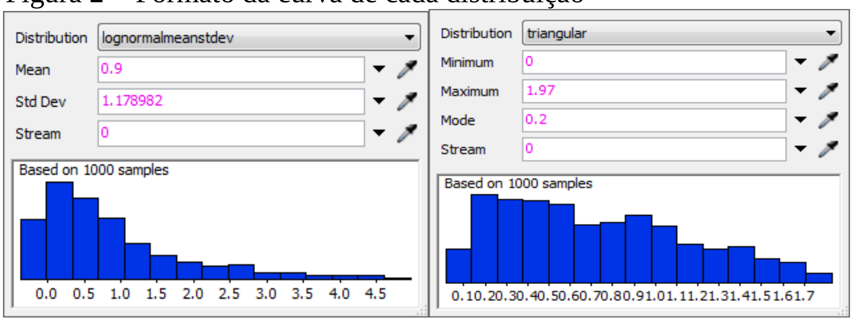
Quadro 2 – Tempo entre as chegadas e tempo de atendimento

Tempo de Chegada	Tempo de atendimento, X	
Distribuição	Distribuição	Justificativa
Exponencial (λ=1)	Log-normal Com média 0,9 e variância 1,39	Original, desconhecida
	Triangular 1 (*a*=0, *m*=0,2, *b*=1,97)	*b*= quantil 0,9 da distribuição log-normal
	Triangular 2 (*a*=0, *m*=0,2, *b*=2,5	*b* =Média da lognormal, (a+b+m)/3=0,9
	Beta (*a*=0, *shape* 1=1,08, *shape* 2=1,92, *m*=0,2)	μ=0,9

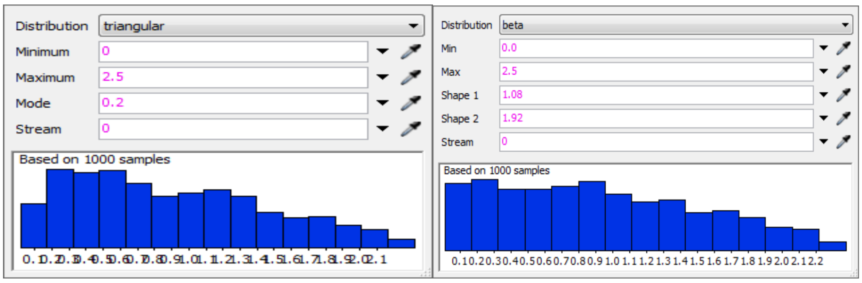
Fonte: Adaptado de Law (2007)

Ao implantar os parâmetros no *Processor* de acordo com cada distribuição estimada, é possível observar o formato da curva. A Figura 2 (a)-(d) apresenta o modelo do gráfico que representa cada distribuição aproximada.

Figura 2 - Formato da curva de cada distribuição



(a) Log-normal (b) Triangular 1

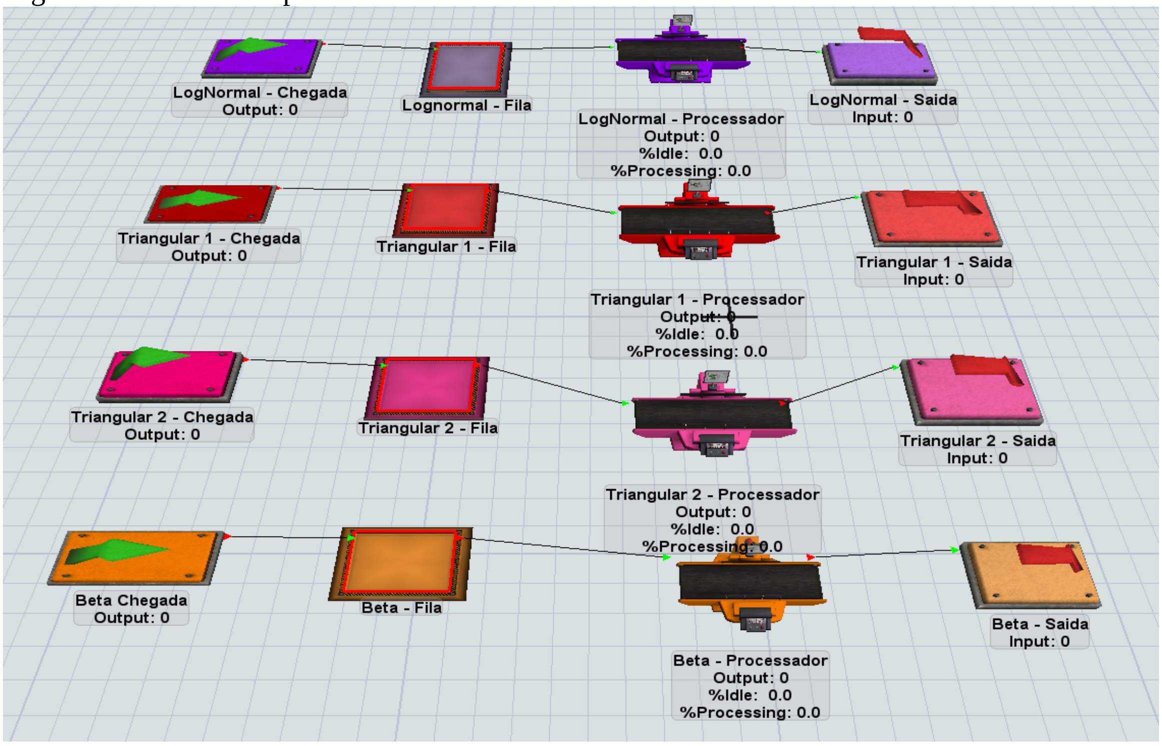


(c) Triangular 2 (d) Beta

Fonte: Flexsim (2016)

Passo 3: Na Figura 3 é apresentado o modelo computacional do sistema após a definição dos modelos no *source* e *processor*. Pode-se agora desenvolver os experimentos.

Figura 3 – Modelo computacional – Cenário Inicial Flexsim

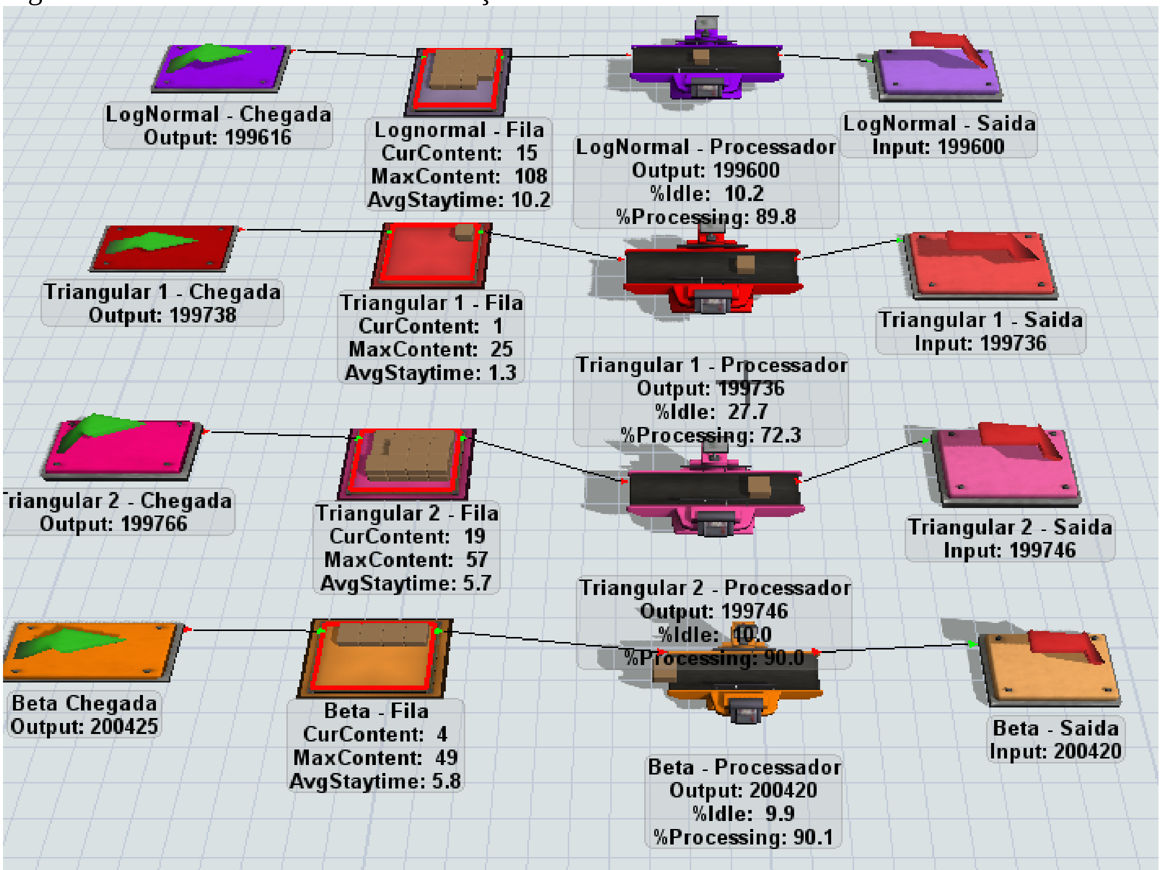


Fonte – Flexsim (2016)

4. EXPERIMENTOS E DISCUSSÕES

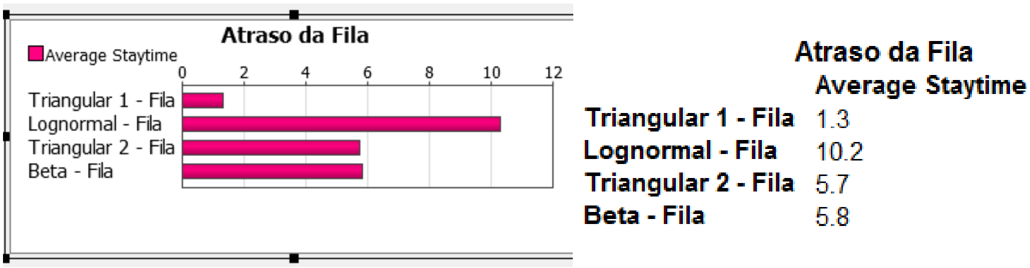
O experimento foi replicado por duzentas mil vezes para a obtenção do regime estável, e os resultados obtidos de cada distribuição foram analisados considerando o tempo de espera na fila conforme resultados apresentados na Figura 4 e Figuras 5(a)-(b). E a Tabela 2 apresenta o percentual de erro para cada distribuição.

Figura 4 - Resultados de saída da simulação



Fonte – Flexsim (2016)

Figura 5 - Tempo de espera na fila, W_q



(a) Gráfico (b) Tabela

Fonte – Flexsim (2016)

Tabela 2- Percentual de erro das distribuições alternativas

Distribuição	Média de espera na fila	Percentual de erro
Log-normal	10,2	0
Triangular 1	1,3	87,25
Triangular 2	5,9	42,16
Beta	5,8	43,14

Fonte: Autores (2017)

A distribuição log-normal original com (μ=0,9, σ^2=1,39) apresentou o tempo de espera na fila na fila de 10,2 min. Entre as distribuições alternativas avaliadas, no caso de não se ter dados e fazer estimativas por procedimentos heurísticos: a distribuição triangular com os parâmetros (*a*=0, *b*=2,5, *m*=0,2) teve seu tempo de espera de 5,7 min; a distribuição triangular com os parâmetros (*a*=0, *b*=1,97, *m*=0,2) com a média do tempo de espera de 1,3 min (pior caso); a distribuição Beta com parâmetros (*a*=0, *b*=2.5, $\tilde{\alpha}_1$=1,08, $\tilde{\alpha}_2$=1,92, *m*=0,2) com média de tempo de espera de 5,8 min.

O experimento mostra que os erros são de fato muito grandes. Extraiu-se da Tabela 2 que se a triangular é utilizada e tendo em vista suas características os erros variaram de 42,16% a 87,25%. A Beta, por sua vez, teve erro de 43,14%.

5. CONSIDERAÇÕES FINAIS

Uma etapa importante no ensino de simulação é mostrar ao estudante e futuro analista de modelos a importância da representação correta do comportamento das variáveis aleatórias presentes na descrição de um sistema real, na situação em que não existem dados disponíveis. O *software Flexsim* permitiu a obtenção de resultados de uma forma simples e de fácil visualização para a comprovação do impacto da escolha das distribuições de probabilidades no tempo de espera na fila.

Supondo-se que a log-normal (μ=0,9, σ2=1,39) fosse o modelo adequado para os dados, na medida de desempenho tempo de espera na fila, ocorreram erros de até, aproximadamente, 87,25% para o pior caso, que seria atribuir o modelo Triangular (*a*=0, *b*=1,97, *m*=0,2) para os dados. Em gerenciamento de prazos em projeto a triangular e beta são bastante utilizadas quando não se tem dados, por exemplo.

Conclui-se então que a escolha da distribuição triangular ou beta para representar aleatoriedade quando não se tem dados, ou mesmo quando o analista não entende a importância da escolha correta da distribuição pode gerar saídas que não representam o sistema, como erros muito grandes. Isto pode comprometer as decisões acerca do sistema real.

Agradecimentos

Agradeço a CAPES e a FAPEG pelo incentivo, reconhecimento e apoio ao desenvolvimento de projetos de pesquisa científica

REFERÊNCIAS BIBLIOGRÁFICAS

ALVES, L. F. P.; REZENDE, F. A.; ALVES, T. F. P.; BOIKO, T. J. P.; MORAIS, M. F. Teoria das Filas: Conceitos e Aplicações. *In:* VII Encontro de Engenharia de Produção Agroindustiral, 2013.

BANKS, J; CHWIF, L. Warnings about Simulation. Jounal of Simulation, 2010.

BATEMAN, Robert E.; BOWDEN, Royce G.; GOGG, Thomas J.; HARREL, Charles R.; MOTT, Jack R.A.; MONTEVECHI, José Arnaldo B. Simulação de Sistemas: Aprimorando Processos de Logistica, Serviços e Manufatura. São Paulo: Campus, 2013.

CONWAY, R.W., MCCLAIN, J.O. The Conduct of an Effective Simulation Study. INFORMS Transactions on Education, v. 3, n. 3, p. 13-22, 2003.

FERNANDES, C. A. Gerenciamento de Riscos em Projetos: Como usar o Microsoft Excel para realizar a simulaçao Monte Carlo, 2005. Disponível em: < http://www.bbbrothers.com.br/files/pdfs/artigos/simul_monte_carlo.pdf> Acesso em 01 jun. 2017.

FOGLIATTI, Maria Cristina; MATTOS, Néli Maria Costas. Teoria de filas. 1.ª ed. Rio de Janeiro: Ed. Interciência, 2007.

FLORENCIO, P. H.; DANTAS, M. J. P. Estudos sobre Modelagem e Simulação de Sistemas de Filas M/M/1 e M/M/2. *In:* IX - Simpósio Acadêmico de Engenharia de Produção, 2014.

HILLIER, Frederick S.; LIEBERMAN, Gerald J. Introdução à Pesquisa Operacional. 8ª Edição, São Paulo, 2010.

GUIMARÃES, A. M. C. Pontificia Universidade Catolica do Rio de Janeiro. Simulação Computacional: Um modelo de maturidde e de seleção para uso dos softwares em manufaturas, 2015, 175p, il Tese (Doutorado).

LAW, Averill M.; KELTON, W. David. Simulation: Modeling and Analysis. Boston: Mcgraw – Hill, 2007.

MACHADO, N. R. S.; FERREIRA, A. O. Metodo de simulacão de monte carlo em planilha excel: desenvolvimento de uma ferramenta versátil para análise quantitativa de riscos em gestão de projetos. Revista de Ciências Gerenciais, v. 16, n. 23, p.- 223 – 244, 2012.

PEREIRA, M. M.; DANTAS, M. J. P. Aplicação da Modelagem e Simulação nos Sistemas de Filas M/M/S - Entendendo a Natureza da Simulaçao. *In:* XXXXV - Encontro Nacional de Engenharia da Produção, 2015.

PRAIA, C. R.; GOMES C. F. S. Simulação computacional aplicada à modelagem do processo de recebimento de uniformes na marinha do brasil. Natal. *In*: XLV - Simpósio Brasileiro de Pesquisa Operacional, 2013

SINAY, M. C. F. Modelagem de Filas a Partir de Diagramas de Fluxos. *In*: XXVI - Simpósio Brasileira De Pesquisa Operacional, 2004.

VILCAPOMA, A. A. I.; MOURA, L. M.; SAMPAIO, L. M. D. S. Uso da simulação de monte carlo em projetos de construção de rodovias no Norte Fluminense. In: XVII - Simpósio de Pesquisa Operacional e Logística da Marinha, 2014.

SIMULATION TEACHING: EVALUATING THE IMPACT OF THE CHOICE OF PROBABILITY DISTRIBUTIONS IN THE ABSENCE OF DATA

Abstract: *Simulation is a numerical method that can be applied to investigate virtually any type of stochastic system. Incorrect identification of probability distributions can lead to estimations of real system outputs that do not correspond to reality. Among the many simulation applications are studies related to queuing systems. Experienced in this work, changes in the probability of distributions that describe the service time of a single-server queuing system, which answer in order of arrival. It was assumed that the service time was correctly described by a log-normal distribution, and estimates were made of alternative distributions such as triangular and beta, considering corresponding parameters obtained by heuristic procedures. The impact on waiting time in the queue was then evaluated. The experiments were conducted in the free version of FLEXSIM Simulation Software, which can be easily obtained by students and used in laboratory classes. In the results, errors of up to 87.25% were observed. An important step in teaching simulation is to show the student and future model analyst the importance of the correct representation of the behavior of the random variables present in the description of a real system, in a situation in which no data are available.*

Key-words: *Teaching Simulation, Queuing Theory, Choice of Probability Distribution, Flexsim Software.*