

Trabalho de Simulação de MAB-515
Primeiro Semestre de 2019 - Prof. Paulo Aguiar
Grupo de 3 alunos máximo
V.1

A data limite para a entrega do trabalho é 08/07/2019, como definido no site.

Os grupos devem estar definidos até **22/04**, segunda-feira, devendo ser definido em sala de aula ou enviado um e-mail para aguiar@dcc.ufrj.br com o assunto Grupo AD 2019-1.

Alunos que estiverem sem grupos serão considerados como grupos individuais.
Após o dia **22/04**, alteração de grupo só com a autorização do professor.

Objetivo

O foco deste trabalho de simulação será a implementação de uma simulação orientada a eventos discretos e que permita a obtenção de intervalos de confiança para algumas métricas de uma fila M/M/1, variando a disciplina de atendimento entre FCFS e LCFS.

Como a simulação de uma fila M/M/1 é essencialmente o mais fácil que podemos pedir, não será permitido o uso de quaisquer rotinas ou facilidade de geração de amostras da distribuição exponencial. O único recurso que poderá ser utilizado é o gerador de números aleatórios da linguagem de programação. Qualquer linguagem poderá ser utilizada. O tempo de simulação deverá ser uma variável real e necessariamente deverá ser implementada a lista de eventos.

O tempo médio de serviço será igual a 1s. Para as disciplinas FCFS e LCFS, a fila será analisada para as seguintes utilizações: 0,2; 0,4; 0,6; 0,8; 0,9.

Deverão ser obtidos ICs de 95% e com precisão de 5%, ou seja, o tamanho do intervalo deverá ser no máximo 10% do centro do intervalo. Para a obtenção dos ICs será padronizado o método batch para todos os grupos, de forma que teremos apenas uma semente inicial. Cada grupo irá definir um método para o término da fase transiente para cada utilização e cada disciplina, e esta metodologia deverá estar explicada e documentada em seção própria do relatório.

Como para obter a precisão de 5% com a Chi quadrado são necessárias 3200 rodadas de simulação (independentemente do número k de coletas por rodada), será definido que deverá ser usado $n=3200$ também na obtenção do IC com a t-Student. Cada grupo deverá determinar o número k de coletas por rodada mais adequado e suficiente, fazendo uso do próprio simulador. Lembrar que rodadas não podem ser muito pequenas devido aos requisitos teóricos de obtenção dos ICs (vide apostila/slides).

Para cada fator de utilização e para cada disciplina, deveremos obter o IC para as seguintes métricas:

- a) Tempo médio de espera em fila, usando o IC para a média pela t-Student, usando 3200 rodadas. Para garantir a precisão de 5% em todos os níveis de utilização, o número mínimo de coletas por rodada para cada disciplina, designados k_{min} (FCFS) e k_{min} (LCFS), deverão ser obtidos. Estes tamanhos de rodada deverão garantir a precisão de 5% para todos os níveis de utilização de forma independente da semente inicial. ***Esta independência da semente deverá ser comprovada no trabalho através de comprovação mostrada com os resultados comparados de rodadas do simulador. Somente relatar e indicar o que foi usado não é suficiente. Use o menor k_{min} que satisfizer a precisão. Isso tem que estar documentado no seu relatório.***
- b) Variância do tempo de espera em fila para cada disciplina. Serão obtidos os ICs tanto com a Chi-quadrado como com a t-Student, como especificado abaixo.

Para o IC pela Chi-quadrado, deve ser usada a expressão da pág. 131 da apostila, que mostra o cálculo do IC da variância para o tempo de espera em fila, em função do tamanho da rodada e do número de rodadas, a partir das médias do tempo de espera por rodada. Não será usada a fórmula da pág. 130. Vide aula 7. O tamanho da rodada tem que ser tal que $k_{min} \gg 1$ (vide teoria).

Para o IC pela t-Student, será usado o estimador da variância por rodada. Espera-se que estas duas metodologias conduzam a ICs convergentes.

Use inicialmente o valor k_{min} determinado no item (a), como o tamanho da rodada. Após as 3200 rodadas, calcule os ICs pela t-Student e pela Chi-quadrado. Se não houver sobreposição total dos dois intervalos (há sobreposição completa quando o centro do IC da t-Student estiver dentro do IC da Chi quadrado e vice-versa), o tamanho da rodada k_{min} deverá ser incrementado de 100 até que a sobreposição completa ocorra.

A evolução do IC da Chi-quadrado a cada incremento deverá ser documentada e constar do relatório o caminho para a convergência. Caso o k_{min} inicial já atinja a sobreposição completa, o objetivo já estará conseguido. O k_{min} que permitiu a convergência dos ICs para a variância será usado para o número na fila de espera.

Espera-se que um aumento do tamanho da rodada garanta a sobreposição completa dos intervalos de confiança e o relatório deverá mostrar isso. Faça os comentários pertinentes na seção de resultados.

Os grupos devem observar que valores pequenos de k_{min} , embora permitam a obtenção de ICs com precisão de 5%, podem estar fornecendo intervalos irreais, onde os valores estimados não estão contidos no IC, devido a violar a aproximação pela Normal.

- c) Número médio da fila de espera em fila, usando o IC da t-Student, sendo o número médio por rodada calculado pela área. O IC para o número médio será aquele obtido com $n = 3200$ e o tamanho de rodada k_{min} obtido em (a).
- d) Variância do número de pessoas na fila de espera.

Calcule o IC pela Chi quadrado, com $n = 3200$ rodadas e o k_{min} (número de coletas por rodada) que garantiu a sobreposição completa. Usa-se o conjunto de amostras formado pelo número médio de pessoas no sistema por rodada, como amostras da normal $N(E[N_q], V(N_q))$ e, em n rodadas, usa-se a fórmula da página 130 da apostila.

Como tarefa complementar, calcule o IC pela t-Student a partir do estimador da variância do número no sistema por rodada, calculado pela pmf de N_q . A pmf é obtida do histograma de N_q ao longo de uma rodada da simulação. Calcula-se $p_j = P(N_q = j)$, como sendo a relação entre a soma de todos os períodos de tempo em que $N_q = j$ ao longo da rodada pelo tempo virtual da rodada. Este p_j é a probabilidade de encontrar a fila de espera com j pessoas numa amostragem aleatória. A partir da pmf da i -ésima rodada, calcula-se $E[N_{qi}^2] = \text{somatório de } j^2 p_j$, o segundo momento de N_{qi} . Isso pode ser feito de forma incremental, a cada alteração do número de pessoas na fila de espera, somando j^2 vezes cada intervalo de tempo transcorrido com j pessoas e apenas no final dividindo pelo tempo virtual da rodada. A média também pode ser recalculada e checada com o valor obtido pela área, pois deve dar a mesma resposta e pode ser uma forma adicional para garantir a correteza da obtenção da área. Para a i -ésima rodada, $V(N_{qi}) = E[N_{qi}^2] - E[N_{qi}]^2$.

Os ICs obtidos pelas duas metodologias devem ser comparados. Faça os comentários pertinentes. Espera-se que estes ICs tenham também sobreposição completa. Se não houver, verifique o que acontece com aumento de k_{min} e comente.

Os resultados deverão ser fornecidos na forma de uma tabela ou forma equivalente, indicando a precisão do intervalo de confiança ou os seus limites, além do valor central do intervalo, e, adicionalmente, o valor analítico de cada métrica. Este valor analítico deve estar dentro do IC.

Os resultados poderão também ser apresentados na forma de gráficos, com o intervalo de confiança indicado. Procure a forma mais simples e conveniente.

Apesar do número médio de pessoas na fila de espera estar relacionado analiticamente com o tempo médio na fila de espera, estas métricas serão computadas diretamente pela simulação e a relação que existe entre elas será uma consequência e poderá ser usada como uma possível indicação de correção do programa de simulação.

O intuito do trabalho é ter um simulador correto e, a partir disso, obter as métricas pedidas. Todavia, antes de mais nada, é preciso que a correção do simulador esteja garantida e esta tarefa será um item específico do relatório. A prova da operação correta do simulador deverá ser feita acompanhando um número selecionado de eventos em situações particulares (chegadas e serviços determinísticos para demonstração de correção pode ser interessante), onde valores das métricas possam ser obtidos facilmente. Lembre-se que você deve mostrar que as estruturas estão sendo tratadas corretamente e que as estatísticas estão também sendo coletadas sem erro. Você tem que anexar resultados dos experimentos e mostrar que as estatísticas batem com os resultados analíticos esperados. Você deverá documentar esta depuração e anexá-la ao relatório. Não adianta apenas relatar. Tem que apresentar dados.

Os valores analíticos para a média e a variância do número de pessoas na fila de espera, bem como a média e variância para o tempo de espera, podem ser determinados e serão utilizados na aferição dos resultados. Se você conhece estes resultados analíticos, eles podem servir de rumo para a compreensão da simulação, mas não podem ser utilizados para demonstrar a correção do simulador. Numa simulação mais complexa, valores analíticos não são conhecidos e a indicação de implementação correta deve ser encontrada através de cenários específicos.

A dedução dos valores analíticos é obrigatória.

Tanto para FCFS como para LCFS, a transformada do tempo de espera é conhecida. A transformada do número de pessoas na fila de espera também é conhecida ($Nq(z)$) pode ser obtida a partir de $N(z)$, sendo igual para ambas as disciplinas.

Para este trabalho, como taxa de serviço $= \mu = 1$, então taxa de chegada $= \lambda =$ utilização $= \rho$. A partir destas simplificações, obtenha as expressões apenas em função ρ .

Procure simplificar as transformadas a partir destas condições dadas e obtenha de maneira mais fácil as expressões para as métricas. Cuidado com a obtenção dos momentos a partir das transformadas. Estas deduções devem ser apresentadas no relatório feito.

Lembre-se que o simulador deve gerar ICs onde os valores analíticos estão contidos. Não erre a dedução, pois é essencial, neste trabalho, que ela seja uma indicação de que algo está errado com o simulador caso os ICs obtidos não contenham o valor analítico. Se isso ocorrer, tem que ir nos testes de correção e reverter.

RELATÓRIO

O relatório feito deverá incluir necessariamente as seções especificadas abaixo. Notas serão dadas a cada sessão individualmente. A soma das notas das sessões constituirá a nota final do relatório.

Na página de rosto do seu trabalho descreva a participação de cada integrante do grupo na produção do trabalho. Esta indicação de participação é obrigatória e deve ser assinada por todos os integrantes. Sem esta indicação o trabalho não será corrigido.

1) Introdução (10 pontos)

Descreva com detalhes a implementação do simulador, explicando:

- Funcionamento geral do simulador;
- Os eventos escolhidos e explicado como cada um deles é tratado;
- Estruturas internas utilizadas e sua função;
- A linguagem utilizada e a forma de geração das variáveis aleatórias envolvidas (facilidades de geração de número aleatório da linguagem utilizada); mostre como a semente inicial é escolhida ou determinada; você terá que mostrar independência da semente e tem que ser mostrado como garante que os intervalos gerados pelas sementes não se sobrepõem.
- Indicar como implementou o conceito de cores ou equivalência e explicar como o término de uma rodada acontece. Seja preciso. Lembre-se que as estatísticas dos fregueses de qualquer rodada que não seja a rodada atual devem ser desprezadas. Apenas no cálculo do número na fila de espera tudo é levado em conta, obviamente.
- Na determinação da precisão do intervalo de confiança, indicar como se determinou o tamanho da rodada (k_{min}) e documente o procedimento para parar o incremento de k_{min} frente a convergência dos ICs da variância pela Chi quadrado e pela T-Student.
- Para a coleta das estatísticas, especialmente no caso do número de pessoas na fila de espera, indique como o cálculo das médias e da variância foi feito.
- Indique a máquina utilizada e o tempo total para rodar a simulação e apresentar todos os resultados, para todas as utilizações e para as duas disciplinas de atendimento.
- O relatório deve abordar o que foi implementado e não entrar em considerações desnecessárias sobre a teoria. O trabalho é essencialmente prático e usar o simulador para comprovar as alternativas válidas para o IC da variância.
- Outras informações pertinentes

2) Testes de Correção (10 pontos)

Nesta seção você descreverá e anexará os resultados dos testes de correção que foram efetuados para garantir o pleno funcionamento do simulador. Você deve demonstrar que o seu programa está simulando exatamente e com correção o esquema proposto. Fórmulas analíticas não podem ser utilizadas para garantir a correção. Servem apenas de orientação, pois na maioria das vezes partimos para a simulação exatamente por não termos os resultados analíticos.

Procure rodar o simulador com cenários determinísticos com estatística conhecida, demonstrando que a geração dos eventos e a coleta de estatística estão corretos.

Tem que anexar resultados dos experimentos e não apenas relatar.

É preciso garantir um simulador correto antes de mais nada. Lembrar que com os valores analíticos em mãos, os ICs coletados com o simulador devem ter os valores analíticos neles contidos. Comprove e comente o que encontrar.

3) Estimativa da fase transiente (10 pontos)

Nesta seção você descreverá como a fase transiente foi estimada e justificar qual métrica que foi observada para determinar o final da fase transiente para os diversos valores de ρ . Com o simulador operando, você pode testar experimentalmente e inferir uma fase transiente adequada, mas tem que apresentar dados que documentem a conclusão encontrada. O método exato de determinar o final da fase transiente e os valores obtidos para cada disciplina e para cada valor de utilização devem ser documentados, qualquer que seja o método escolhido para determinar o fim da fase transiente.

Ainda que a fase transiente no método batch, que deverá ser usado por todos, tenha menor influência no resultado, a determinação da fase transiente é obrigatória, pois é um exercício para determinar a entrada em equilíbrio do sistema. Você terá que justificar suas escolhas. Este é um processo empírico, mas as dicas na apostila ajudam.

Procure demonstrar a influência da escolha da fase transiente na qualidade das medidas. A determinação da fase transiente deve ser **independente da semente inicial**. Comprove isso no relatório, anexando dados de testes realizados com o simulador.

Com 3200 rodadas de simulação sendo executadas, o tamanho da fase transiente tende a ser relativamente irrelevante, se o tamanho de rodada for escolhido de forma adequada e usado o método batch. Determine a fase transiente com a importância que teria se fosse utilizado o método replicativo. Este é um exercício de aprendizado.

4) Tabelas com os resultados e comentários pertinentes (50 pontos)

Inicialmente, você terá que incluir a tabela de resultados analíticos junto com a dedução. Isso será essencial e tenha certeza de que os valores estão corretos. A avaliação usa esta tabela e fica fácil saber se há problema ou não com o simulador, ainda que os testes de correção devessem ter garantido a correteude.

Apresente todos os ICs como solicitado, não esquecendo de incluir o valor analítico nas tabelas, bem como indicando o tamanho de rodada e número de rodadas.

Para o caso da evolução do IC para a variância, procure mostrar a convergência com o incremento do número de coletas por rodada (incremento de kmin). Mostre isso graficamente ou tabelando. Há que se ter uma visão bem clara da convergência dos ICs.

Lembre-se que o tempo médio de espera para FCFS e LCFS são iguais. Além disso, há a relação esperada entre o número médio na fila de espera e o tempo médio de espera. FCFS e LCFS têm transformada do número do sistema igual. Logo as métricas de número no sistema e na fila de espera para estas disciplinas têm que ser iguais.

5) Conclusões (10 pontos)

Coloque aqui seus comentários finais.

Descreva dificuldades encontradas, as otimizações feitas, e outras conclusões que você tirou do trabalho. Comente o que poderia ser melhorado, como, por exemplo, o tempo de execução do seu programa, caso haja soluções nitidamente mais eficientes. Adicione quaisquer comentários que você julgar relevante.

Comente os erros que foram encontrados ao longo do trabalho e tudo que teve que ser refeito para obter os resultados corretos.

Cada uma das seções terá sua avaliação. Portanto, não deixe de colocar nenhuma seção no seu relatório. Não deixe de ler o capítulo de simulação na apostila.

6) Anexo - Listagem documentada do programa (10 pontos)

A documentação do programa fonte deverá ser feita com rigor, explicando cada sub-rotina ou passo da programação. Código fonte sem comentários não é aceitável. A listagem do código fonte deve ser um anexo ao relatório.

As instruções para rodar o executável (em Linux ou Windows 10) devem estar no relatório.

Se o Grupo usar alguma linguagem específica, ele deve compilar o ambiente e apresentar um executável que rode sem necessidade de instalação especial, em Windows 10 ou Ubuntu. O programa será testado com alguns valores particulares para averiguar sua correção e integridade.

Se solicitado, o grupo deve estar preparado para uma apresentação do trabalho e demonstração de sua operacionalidade. O executável deve ser passível de ser rodado, no mínimo, com uma escolha arbitrária do número de fregueses por rodada, da utilização e da semente inicial.

O relatório completo deve ser entregue impresso (e não em mídia eletrônica).

Em caso de dúvida com o enunciado, procure o professor o mais rápido possível.