

RT-MAE-9221

VERIFICAÇÃO E VALIDAÇÃO DE MODELOS DE SIMULAÇÃO

by

Carlos Alberto Barbosa Dantas

Palavras Chaves: Simulação, Validação, Modelos probabilísticos,
(Key words) Sistemas de atendimentos de emergência.

Classificação AMS: 68J10, 90B22
(AMS Classification)

Verificação e Validação de Modelos de Simulação

Carlos A.B.Dantas

Departamento de Estatística

Instituto de Matemática e Estatística da USP

1 Introdução : Modelos Matemáticos e de Simulação

Modelos matemáticos de sistemas foram introduzidos inicialmente no domínio da astronomia e da física por Galileu, Kepler, Newton entre outros.

Os modelos propostos até meados do século passado eram modelos determinísticos. Nessa época surgem modelos estocásticos através da formulação da Mecânica Estatística, da Termodinâmica, do modelo de Galton para descrever a extinção dos nomes de família, entre outros.

A partir da década de 50 observou-se uma crescente utilização de modelos matemáticos, tanto determinísticos quanto estocásticos, para descrever o comportamento de sistemas de produção e de processos de tomada de decisões nas firmas e no governo.

Muitas vezes a complexidade dos sistemas estudados não permite a construção de um modelo matemático, sendo possível apenas esboçá-lo. Outras vezes o modelo não nos fornece soluções analíticas. Uma ferramenta que tem se mostrado muito útil nessas circunstâncias é a simulação do sistema.

A simulação consiste em construir-se um modelo que reproduza os eventos que ocorrem no sistema e a partir da elaboração de um programa fazer experimentos no computador, produzindo assim, no modelo, observações da evolução do sistema ao longo do tempo.

O advento dos computadores com grande capacidade de processamento deu um grande impulso à simulação de sistemas. Nesse artigo nós consideraremos em especial o problema de julgar se o modelo de simulação é adequado para representar o sistema. Seguindo Glass [4], nós utilizaremos para abordar essa questão os conceitos de verificação e de validação de um modelo de simulação, que serão expostos na próxima seção. Ilustraremos, a aplicação desses conceitos a um modelo de simulação de serviço de atendimentos de guinchos em uma rodovia.

2 Noções gerais : Verificação e Validação

Os conceitos de verificação e de validação não são conceitos restritos à área de simulação. Aplicam-se, ao problema de julgar se um modelo matemático é adequado para descrever(representar) um sistema físico .

A verificação é o procedimento adotado para determinar se o programa de computação roda como pretendido. Se isto não ocorrer procura-se detectar os erros e depurar o programa tanto dos erros de consistência lógica, quanto daqueles que produzem resultados numéricos.

Quando se simula um sistema complexo, é praticamente impossível decidir que o programa esta verificado. É necessário em geral restringir a verificação aos aspectos mais relevantes do problema em estudo. Uma outra alternativa consiste em examinar se a simulação produz em casos particulares, para os quais exista uma solução analítica, os mesmos resultados que essa.

A validação de um modelo de simulação visa testar a concordância do comportamento do modelo com o comportamento do sistema real sendo modelado.

Para tornar mais precisas estas afirmações nós precisaríamos formular testes de hipóteses estatísticas que afirmassem que as observações da simulação e do sistema físico podem ser considerados como produzidos pelo mesmo mecanismo. Para sistemas complexos, como é o caso da maioria dos sistemas simulados esta é uma tarefa praticamente impossível. Uma alternativa consiste, como no caso da verificação ,de restringir o escopo da validação aos aspectos mais relevantes do problema em estudo. Uma outra forma de proceder-se consiste em fazer a validação por etapas, testando-se por exemplo as hipóteses em que se baseia o modelo e quando possível hipóteses referentes a subsistemas no modelo. Pode-se ainda ,como no caso da verificação, considerar a validação de casos especiais do sistema para os quais existam modelos analíticos.

Muitos desses procedimentos dependem da análise estatística das saidas da simulação. Essa é uma área que tem recebido grande atenção, pois, os métodos classicos de inferência baseados em váriaveis aleatórias independentes e identicamente distribuidas não se aplicam. De fato, as observações ao longo de uma rodada da simulação são em geral dependentes. Além disso, como na maioria das vezes, estamos interessados em medidas de desempenho, de sistemas em equilibrio , temos que decidir como fazer para lidar com as observações da fase transitória. Sugerimos ao leitor interessado que consulte, por exemplo [7].

3 Serviços de emergência : modelo analítico e de simulação.

Os serviços de atendimentos de emergências em zonas urbanas ou em rodovias caracterizam-se pelos seguintes fatos : os instantes e os locais dos chamados são aleatórios ; as unidades de atendimento estão distribuídas na região ; os chamados são dirigidos a uma central que envia uma unidade para atendê-los.

Para o planejamento e operação desses sistemas é necessário responder às seguintes questões : a) Quantas unidades devem ser alocadas ao serviço ? b) Se cada unidade atende prioritariamente um trecho da região, como devem ser formados os trechos? c) Qual deve ser a política de envio de unidades para atender os chamados ? Deseja-se responder essas questões de modo que o nível do atendimento seja satisfatório.

3.1 Modelo hipercúbico

Larson [5] formulou um modelo matemático , denominado modelo hipercúbico para serviços de atendimentos de emergências. Propôs que se avalie o sistema através de um conjunto de medidas de desempenho associadas ao sistema como um todo , aos trechos, às unidades e aos átomos (definidos a seguir).

O modelo hipercúbico é um modelo de filas onde as unidades de serviço estão distribuídas espacialmente e atendem chamados gerados nessa região. A região é dividida em um certo número de subregiões mínimas denominadas átomos. Na construção do modelo os átomos são as menores subregiões onde os chamados são registrados. A reunião de um certo número de átomos será denominada trecho. Em cada trecho há uma unidade de serviço. As unidades de serviço quando não estão atendendo chamados podem ficar estacionadas em um átomo do trecho ou podem permanecer em movimento nos átomos do mesmo. Nesse último caso, a localização de uma unidade pode ser dada estatisticamente através da fração de tempo que ela gasta em cada átomo do trecho.

O modelo hipercúbico foi construído com base nas seguintes hipóteses :

1) O processo de chegadas, isto é, as chamadas para serviço, na região, seguem um Processo de Poisson com taxa λ . 2) Os tempos de serviço de todas as unidades do sistema tem distribuição exponencial com média $\frac{1}{\mu}$ independente do local da chamada e da unidade enviada. 3) Somente uma unidade é enviada para atender um chamado. Para um sistema com N unidades, se não levarmos em conta a identidade das unidades livres e ocupadas temos sob as hipóteses 1), 2) e 3) um modelo de filas $M|M|N$.

Uma política de atendimento é definida quando é dada uma partição da região em trechos, a localização das unidades nos mesmos e uma regra que associa a cada átomo em ordem decrescente de prioridade a unidade que atendem chamados aí gerados. Quando todas as unidades estão ocupadas há duas possibilidades: a chamada não é atendida pelo sistema ou, se o for, entra em fila e é atendida de acordo com a ordem de chegada, pela primeira unidade que ficar livre. Se as unidades forem numeradas de 1 a N e, além disso, a cada unidade for atribuído o número 0 ou 1 conforme ela esteja livre ou ocupada, então o espaço de estados do sistema, no caso de capacidade do compartimento de espera nula, será o conjunto das seqüências de zeros e uns de comprimento N , isto é, o hipercubo N -dimensional. Com esse espaço de estados, sob as hipóteses adotadas o processo é uma Cadeia de Markov a parâmetro contínuo. Como os tempos de serviço tem uma distribuição exponencial com parâmetro μ , segue-se que a taxa de transição de um estado com k unidades ocupadas para um estado com $k - 1$ unidades ocupadas é igual a μ , para $k = 1 \dots N$. A taxa de transição de um estado com k para um estado com $k+1$ unidades ocupadas é difícil de determinar, porque ela depende do estado e da regra de prioridade usada para enviar as unidades para atender os chamados. Larson [5] construiu um algoritmo para determinar essas taxas. Tendo as taxas de transição obtêm-se as equações de balanço detalhado cuja solução dá a distribuição estacionária para o sistema com 2^N estados. Com a distribuição estacionária obtêm-se a taxa de ocupação ρ_i , que representa a fração do tempo que a unidade i esta ocupada atendendo chamados, para $i = 1 \dots N$. Usando as taxas de ocupação nós obtemos as medidas de desempenho associadas ao sistema como um todo, aos trechos e às unidades. Larson introduziu um modelo aproximado que reduz a N o número de equações a serem resolvidas para determinar a distribuição estacionária. Como pode ser visto em [6] os resultados da aproximação são muito bons.

3.2 O Modelo de Simulação

No modelo hipercúbico dados de entrada como as taxas de chamados e as probabilidades dos chamados ocorrerem nos átomos devem ser estimados. Além disso, o tempo total de serviço das unidades também deve ser estimado. Este depende dos tempos de atendimento no local e dos tempos de deslocamento, que por sua vez dependem da configuração do sistema. No modelo hipercúbico não é possível formar fila em uma unidade quando outras unidades estão desocupadas. Quando as distâncias são grandes essa possibilidade pode ser de utilidade. Estes fatos e principalmente a constatação que os tempos de serviço não tem distribuição exponencial levaram-nos a construir o modelo de simulação. Ver [3].

No modelo de simulação como no modelo hipercúbico a região é dividida em átomos, reuniões de subconjuntos dos mesmos formando os trechos. Supomos ainda que : a) O processo de chegadas é um Processo de Poisson com taxa λ b) O tempo de serviço no local tem distribuição exponencial, e é independente do local do chamado e da unidade enviada. c) A velocidade de deslocamento dos veículos quando se dirigindo para o local da ocorrência ou quando realizando uma tarefa é constante. d) As unidades quando não estão atendendo um chamado permanecem em seu trecho. Observemos que se os tempos de serviço no local não forem exponenciais basta substituir a rotina que gera as exponenciais por uma rotina que gere a distribuição apropriada.

Política de atendimento

Uma política de atendimento é definida no modelo de simulação por meio de tres tabelas : a tabela 1 dá os átomos que formam os trechos e o átomo onde se localiza a unidade no trecho ; a tabela 2 , associada a um chamado , dá para cada trecho as unidades que o atendem em ordem decrescente de prioridade ; a tabela 3 , associada a um fim de serviço , dá para cada unidade os trechos que ela atende em ordem decrescente de prioridade.

O Programa

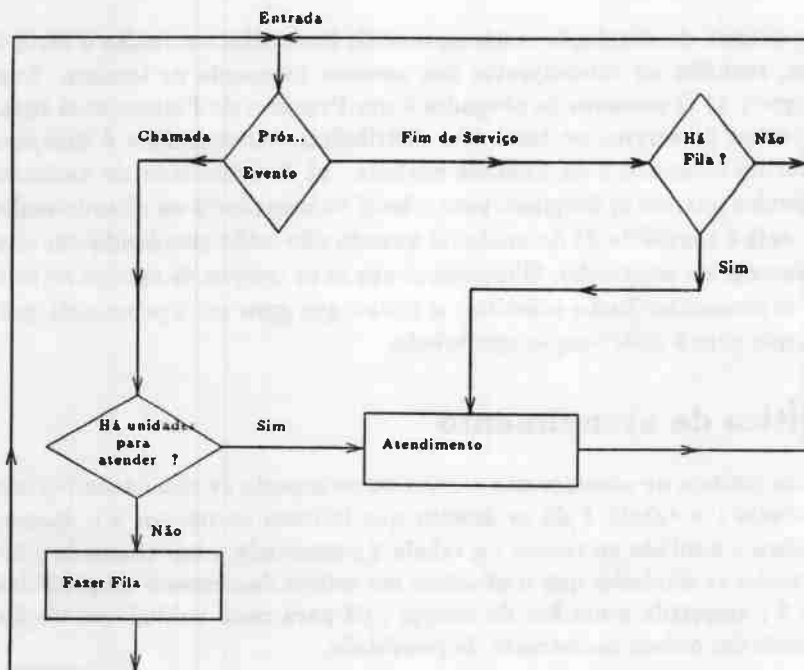
O controle do programa é feito através do escalonamento de eventos. Um evento será uma chamada ou um fim de serviço. Um diagrama de fluxo simplificado, para o programa, é o seguinte:

Próximo evento : Nesse bloco são gerados os instantes das chamadas, e é decidido se o próximo evento é uma chamada ou um fim de serviço. Ai ,atualiza-se o relógio e os parâmetros que dependem dele.

Há unidade para atender : Este bloco localiza o chamado e verifica se há alguma unidade para atendê-lo.

Fazer fila : Atualiza a fila da unidade do trecho onde ocorreu a chamada.

Há fila ? : a unidade que termina um serviço verifica se há fila nos trechos que ela atende em ordem decrescente de prioridade.



Atendimento : Este bloco gera o tipo de serviço, o tempo de serviço no local e para serviços realizados em duas etapas gera o local da segunda etapa. Calcula ainda os vários tempos de deslocamento utilizando-se da matriz das distâncias médias entre os átomos e a velocidade de deslocamento das unidades

4 Aplicação : verificação do modelo de simulação de serviços de emergência.

Antes de tratar da verificação do modelo vamos dar um exemplo de suas saídas. Nós consideraremos um sistema de guinchos em uma rodovia. Os átomos são formados pela parte da rodovia delimitada por dois trevos, isto é, por pontos onde há comunicação entre as pistas num e noutro sentido. A rodovia foi dividida em 21 átomos e formaram-se 5 trechos, com um carro guincho em cada trecho. Cada chamado pode dar origem à remoção de um veículo para o acostamento, portanto havendo somente um atendimento no local, ou, pode ser necessário guinchar o veículo para determinados locais da rodovia. Nesse caso os locais são dados

por um vetor $x = (x_i, i = 1 \dots 21)$, onde x_i designa o destino de um guincho para um chamado no átomo i .

O tempo de serviço total (TST) é decomposto, conforme a chamada necessite ou não de guinchamento, de uma das seguintes maneiras :

$$TST = TD + TSL_1 + TR$$

$$TST = TD + TSL_2 + TG + TRG$$

onde TD é o tempo de deslocamento ao átomo do chamado; TSL_1 é o tempo de serviço no local, sem guinchamento; TSL_2 é o tempo de serviço no local quando há guinchamento; TR é o tempo de retorno do local do chamado ao átomo onde o guincho está localizado; TG é o tempo de guinchamento e TRG é o tempo de retorno do destino do guincho ao átomo de origem. A política adotada é especificada pelas tabelas 1, 2, e 3. Ver pag.3.

Tabela 1

Trecho	Local (atomo)	Atomos do trecho					
1	3	1	2	3	4	5	
2	8	6	7	8	13		
3	9	9	10	11	12	20	21
4	15	14	15	16	17	18	
5	19	19					

Tabela 2

Trecho	Unidades		
1	1	4	2
2	2	3	1
3	3	5	
4	4	1	2
5	5	3	

Tabela 3

Unidade	Trecho			
1	1	4	2	
2	2	3	1	4
3	3	5	2	
4	4	1		
5	5	3		

Dados de entrada e saída

Foram simulados 400 períodos de operação de 18 horas cada um, com os seguintes dados de entrada : taxa de chamados dos guinchos $\lambda = 3.24$ chamados por hora; velocidade de deslocamento dos veículos 60 km por hora; tempos de serviço no local quando não há guinchamento (TSL_1) e quando há guinchamento (TSL_2) com distribuições exponenciais com médias 30 e 15 minutos respectivamente. As tabelas que dão os locais para onde são guinchados os veículos a partir do átomo da ocorrência e que dão as probabilidades de ocorrência de chamados nos átomos são omitidas por economia de espaço. As tabelas 4 e 5 dão as medidas de desempenho associadas às unidades e aos trechos. Os dados da tabela são as médias das variáveis indicadas no topo.

Tabela 4 : Medidas de desempenho associadas às unidades

Un	TO	FAFT	TST	TD	TSL	TG	TR	F_m
1	.50	.51	48.72	14.17	17.57	12.00	14.89	.038
2	.45	.49	46.15	13.26	17.77	12.11	12.97	.027
3	.55	.37	43.53	11.02	17.65	8.46	13.35	.109
4	.54	.18	49.23	11.64	17.12	16.19	17.64	.077
5	.54	.33	52.29	9.02	17.63	7.97	24.30	.106
Média			47.81	11.77	17.55	11.26	16.52	

TO : é a taxa de ocupação de cada unidade, definida como a fração do tempo em que a unidade esta ocupada atendendo chamados. $FAFT$: é a fração de atendimentos de uma unidade fora de seu trecho. TG : é o tempo que o guincho gasta para levar o veículo guinchado para o seu destino; TR é o tempo que o guincho leva para retornar para o átomo sede em seu trecho ; F_m é, para cada unidade o comprimento médio de sua fila.

Tabela 5 : Medidas de desempenho associadas aos trechos

Trecho	FAUF	TDT_r	TOT_r	TEF	TEFF
1	.41	12.22	22.03	4.36	19.61
2	.35	9.54	18.88	3.56	19.07
3	.34	10.06	37.39	8.72	22.99
4	.43	15.07	20.03	4.98	23.31
5	.30	10.11	37.39	10.72	28.89

FAUF : é a fração de chamados em um trecho atendidos por uma unidade de outro trecho; **TOT_r** : é a fração do tempo em que todas as unidades que atendem um dado trecho estão ocupadas atendendo chamados ; **TEF** : é o tempo de espera na fila ; **TEFF** : é o tempo de espera na fila dado que há fila.

As medidas de desempenho gerais ,isto é ,associadas à região como um todo são as médias gerais das medidas de desempenho associadas às unidades e aos trechos. A fração geral de despachos, que são despachos entre trechos é igual a .37 e o desequilíbrio entre as taxas de ocupação das unidades,que é a diferença máxima entre as taxas de ocupação das unidades é igual a .105

Verificação.

Das hipóteses do modelo de simulação é fácil concluir que para um sistema com uma unidade de guincho temos uma fila $M|G|1$. De fato, o processo de chamados é um processo de Poisson com taxa λ . Os tempos de serviço totais são independentes e identicamente distribuídos porque, dado o instante de um chamado ,sua localização é obtida de uma distribuição discreta independente do local do chamado ; os tempos de serviço no local são i.i.d. com distribuição exponencial ;as distâncias percorridas só dependem do átomo selecionado e dos locais para onde o veículo é guinchado.

Para uma fila $M|G|1$ com taxa de chegada λ e tempo médio de serviço $\frac{1}{\mu}$, o comprimento médio da fila F_m e o tempo médio de espera na fila $E(TEF)$ são dados por :

$$F_m = \frac{\rho^2 + \lambda^2(\text{var}(TST))}{(1 - \rho)} \quad E(TEF) = \frac{F_m}{\lambda}$$

onde $\rho = \frac{\lambda}{\mu}$ e $\text{var}(TST)$ é a variância do tempo de serviço total.

$$\begin{aligned} E(TST) &= E(TT + TSL_1 + TR|X = 1)P(X = 1) + \\ &+ E(TT + TSL_2 + TG + TRG|X = 2)P(X = 2) \end{aligned}$$

$$\text{Var}(TST) = E(\text{var}(TST|X)) + \text{Var}(E(TST|X))$$

Rodamos 400 períodos de 18 horas da simulação com taxas de serviço λ iguais a 0.25, 0.50 e 0.75 chamados por hora respectivamente e com uma unidade de guincho localizada em um átomo central. Todos os demais dados são iguais aos do exemplo acima. Na tabela abaixo apresentamos os resultados obtidos calculando-se ρ , a taxa de ocupação, F_m o comprimento médio da fila e o tempo médio de

espera na fila (na tabela representado por TEF) a partir da simulação e das fórmulas do modelo $M|G|1$.

	SIM	$M G 1$	SIM	$M G 1$	SIM	$M G 1$
ρ	0.2462	0.2416	0.4759	0.4619	0.7011	0.7064
F_m	0.0479	0.0463	0.2370	0.2420	0.894	.85000
TEF	11.3	11.1	28.15	29	72	68

Validação

Essa etapa nos moldes propostos pode ser realizada para sistemas em funcionamento . Observa-se o sistema com uma dada configuração e obtém-se as medidas de desempenho empiricamente. Roda-se a simulação com a mesma configuração e compara-se os resultados .

Para sistemas projetados para funcionamento futuro outros métodos devem ser utilizados para tentar-se obter evidências na direção da validação do modelo de simulação. Uma possibilidade consiste em reunir-se pessoal técnico com experiência de operação de sistemas análogos e submeter a eles resultados de rodadas da simulação para várias configurações do sistema . A avaliação da plausibilidade dos resultados por essa equipe nos dá indicações da validade do modelo.

5 Conclusão

A validação de um modelo de simulação envolve procedimentos que não podem se restringir a testes de hipóteses estatísticas . A análise estatística dá um referencial importante para o estudo do problema ,porem não dispensa outros procedimentos que não podem ser formulados de modo sistemático, ou melhor, fechado. Aconselhamos o leitor interessado a consultar Naylor e Finger [8] que formulam a questão de modo conceitual inspirando-se em estudos de validação de modelos econômicos. O artigo de Gass [4] aborda principalmente questões de validação de modelos de "análise de políticas " , isto é , a análise sistemática de questões relacionadas com o planejamento governamental ou com processos

de tomada de decisões que afetam esses mesmos processos. O artigo de Carson [2] aborda os aspectos gerais de validação .A bibliografia mais atualizada que encontramos esta no livro de Law e Kelton [7].

Referências

- [1] Balci,O. and Sargent,R. G. *A Bibliography on the Credibility, Assessment and Validation of Simulation and Mathematical Models*. Simmuletter 15,15-27. (1984b)
- [2] Carson, J. S.*Convincing Users of Model's Validity is Challenging Aspect of Modeler's Job*. Ind.Eng. 18,74-85 (june 1986)
- [3] Dantas, C. A. B. e Collucci, E.*Serviços de Emergência : Comparação dos Resultados de Simulação e do Modelo Hipercúbico*. Simpósio Nacional de Probabilidade e Estatística ,4,Rio de Janeiro,1980.Atas 1982,p.251-262.
- [4] Gass, S. I.*Decision Aiding Models: Validation,Assessment and Related Issues in Policy Analysis*. Operat. Res. 31,603-631.(1983)
- [5] Larson, R. C.*A Hipercubic Queucing Model for Facility Location and Redistricting in Urban Emergency Services*. Comput. and Operat. Res. 1,67-95 (1974)
- [6] Larson, R. C.*Approximating the Performance of Urban Emergency Service Systems*. Operat. Res.23,5,845-868.(1975)
- [7] Law, A. M. and Kelton, W. D..*Simulation Modeling and Analysis*. McGraw-Hill Inc.2nd. ed. (1991)
- [8] Naylor,T. II. and Finger J. M..*Verification of Computer Simulation Models*.Manag. Sci.14,92-101.(1967)
- [9] Van Horn, R. L..*Validation of Simulation Results*. Manag. Sci. 17,247-258.(1971)

ULTIMOS RELATORIOS TECNICOS PUBLICADOS

1992

- 9201 - BOLFARINE, H., NASCIMENTO, J.A. & RODRIGUES, J. Comparing Several Regression Models with Measurement Errors. A Bayesian Approach, 16p.
- 9202 - BOLFARINE, H. & SANDOVAL, M.C. Empirical Bayesian Prediction in the Location Error in Variables Superpopulation Model, 26p.
- 9203 - BUSSAB, W.O. & BARROSO, L.P. Painei Multivariado - Análise Através do Modelo de Componentes de Variância, 07p.
- 9204 - LEITE, J.G. & PEREIRA, C.A.B. Urn Scheme to Obtain Properties of Stirling Numbers of Second Kind, 09p.
- 9205 - BELITSKY, V. A Stochastic Model of Deposition Processes with Nucleation, 21p.
- 9206 - BOLFARINE, H. & NASCIMENTO, J.A. Bartlett Correction Factors for the Structural Regression Model with Known Reliability Ratio, 11p.
- 9207 - FERRARI, P.A. Growth Processes on a Strip, 23p.
- 9208 - FERRARI, P.A., GALVES, J.A. & LANDIM, C. Exponential Waiting Time for a Big Gap in a One Dimensional Zero Range Process, 8p.
- 9209 - LOSCHI, R.H. Coerência e Probabilidade, 17p.
- 9210 - CRIBARI-NETO, F. & FERRARI, S.L.P. An Improved Lagrange Multiplier Statistic for the Test of Heteroskedasticity, 22p.
- 9211 - LEITE, J.G. & BOLFARINE, H. Bayesian Estimation of the Number of Equally Likely Classes in a Population, 10p.
- 9212 - BOLFARINE, H. & SANDOVAL, M.C. On Predicting the Finite Population Distribution Function, 9p.
- 9213 - FERRARI, P.A. & FONTES, L.R.G. Fluctuations in the Asymmetric Simple Exclusion Process, 5p.

- 9214 - FERRARI, P.A. & FONTES, L.R.G. Current fluctuations for the Asymmetric Simple Exclusion Process, 14p.
- 9215 - IRONY, T.Z. & PEREIRA, C.A.B. Motivation for the Correct Use of Discrete Distributions in Quality Assurance, 12p.
- 9216 - IRONY, T.Z. & PEREIRA, C.A.B. Bayesian Hypotesis Test: Using Surface Integrals to Distribute Prior Information Among the Hypotheses, 25p.
- 9217 - FERRARI, P.A. & MAURO, E.S.R. Ergodicity and Invariance Principle for the One Dimensional S.O.S. Stochastic Model, 10p.
- 9218 - PEREIRA, C.A.B. & TIWARI, R.C. A Nonparametric Bayesian Analysis of Competing Risks Models, 20p.
- 9219 - PEIXOTO, C. Tempos Exponenciais e Aproximação do Equilíbrio para um Passeio Aleatório no Hipercub, 24p.
- 9220 - FERRARI, S.L.P. & CRIBARI-NETO, F. On the Corrections to the Wald Test of Nonlinear Restrictions, 8p.

The complete list of Relatórios do Departamento de Estatística, IME-USP, will be sent upon request.

- Departamento de Estatística
- IME-USP
- Caixa Postal 20.570
- 01498-970 - São Paulo, Brasil