UNIVERSIDADE DE SÃO PAULO

Bayesian analysis in M/M/1 queues using sampling methods

JOSEMAR RODRIGUES

SILVIA PRADO CHITTA

Nº 16

N O T A S

Instituto de Ciências Matemáticas de São Carlos

Instituto de Ciências Matemáticas de São Carlos

ISSN - 0103-2577

Bayesian analysis in M/M/1 queues using sampling methods

JOSEMAR RODRIGUES

SILVIA PRADO CHITTA

Nº 16

NOTAS DO ICMSC
Série Estatística

São Carlos
Fev./1995

Bayesian analysis in M/M/1 queues using sampling methods

Josemar Rodrigues and Silvia Prado Chitta
ICMSC-USP-São Carlos-SP- 13560-Brasil

Abstract

The purpose of this paper is to study the Bayesian prediction in M/M/1 queues using the Rubin's Sampling-Importance-Resampling(SIR) to simulate predictive distributions of usual measures of effectiveness, such as the number of customers and the waiting time in the system. The histogram approach as described in Albert (1993) is used via Minitab to assess the prior information of the intensity of traffic in M/M/1 queue in equilibrium. Minitab macros based on SIR algorithm are also developed and illustrated with examples. Some comparisons with the conjugate Bayesian analysis introduced by Armero and Bayarri (1992) are also provided. To apply the sampling procedure formulated in this paper, we only need the prior information of the intensity of traffic of the M/M/1 queue without worrying about the arrival rate or the service rate.

Key words: Bayesian inference; measure of performance; SIR-algorithm; histogram approach.

1 Introduction

The M/M/1 queue can be described as follows: customers arrive to the system where the interarrival times follow an exponential distribution with the arrival rate λ; they may have to wait in line if the single server is busy; service times follow an exponential distribution with the service rate μ; the service times are assumed to be independent of the arrivals. For an M/M/1

queue to be in equilibrium, the parameter $\rho = \lambda/\mu$ (the intensity of traffic) has to be strictly less than one. A good book on the subject is Gross and Harris (1995). Our interest in this paper is the Bayesian analysis of M/M/1 queue and references for this subject are Armero and Bayarri (1992a,b). In those papers the conjugate Bayesian for (μ, λ), ρ and the prediction of various measures of performance of the queue in equilibrium are treated in depth. In this paper, we explore the use of one particular simulation technique, the Sampling-Importance -Resampling (SIR) algorithm (Gelfand and Smith 1992) and the histogram approach as described in Albert (1993), in studying the predictive distribution of usual measures of effectiveness in a M/M/1 queue system such as the number of customers in the system and the waiting time in the system. The SIR algorithm and the histogram approach for the intensity of traffic ρ is described in Section 2 and can be performed somewhat automatically for the Bayesian analysis in a M/M/1 queue. This method has various advantages which were discussed in details by Albert (1993) and in queueing problem we need only the prior information about ρ without worrying about the nuisance parameters μ and λ. Finally, since the these methods are relatively simple, they can be performed using the MINITAB macros which are presented in the Appendix. In Section 3, using these sampling methods, we derive the prior and the posterior distribution of ρ and some comparisons with Armero and Bayarri's results are considered. Section 3 derives the predictive distributions for the steady-state number of customers in the system. Also, in Section 3 , using sampling methods, we deal with prediction about the steady -state waiting time of a customer in the system.

2 The SIR algorithm and the histogram approach in a M/M/1 queue

The Bayesian inferences in a M/M/1 queue can be described as follows: Let us assume that we observe n_a interarrival times and n_s service times for fixed n_a and n_s. Let X_i denote the service time of the i-th customer, $i = 1, 2, \ldots, n_s$ and Y_j denote the time elapsed between the arrivals of customers $j-1$ and j, $j = 1, 2, \ldots, n_a$, where Y_1 denotes the arrival time of the first customer. Following the assumptions of a M/M/1 model, $Y_1, \ldots, Y_{n_a}$ is a random sam-

ple from an exponential distribution with parameter λ and $X_1, \ldots, X_{n_s}$ is a random sample from an exponential dsitribution with parameter μ, independent of Y's. We represent the observed data by $z = (y_1, \ldots, y_{n_a}, x_1, \ldots, x_{n_s})$ and the joint density is given by:

$$g_\theta(z) = \lambda^{n_a} e^{-\lambda t_a} \mu^{n_s} e^{-\mu t_s}, \tag{1}$$

where $\theta = (\lambda, \mu)$, $t_a = \sum y_j$ and $t_s = \sum x_i$. The quantities t_a and t_s are the total observed arrival time and the total observed service time, respectively. Suppose that ρ, the intensity of traffic, is the parameter of interest, so, we can reparametrize the joint density (1) in terms of (ρ, μ) obtaining the following joint likelihood:

$$L(\rho, \mu) \propto \rho^{n_a} \mu^{n_t} e^{-\mu(t_s + \rho t_a)}, \tag{2}$$

where $n_t = n_s + n_a$ is the total number of observations. Since we are only interested in the parameter ρ, we eliminate the nuisance parameter μ in (2) using the profile likelihood given by

$$\begin{aligned} L_p(\rho \mid z) &= sup_{\{\theta:\lambda/\mu=\rho\}} g_\theta(z) \\ &= \rho^{n_a} \left(\frac{n}{\rho t_a + t_s}\right)^n e^{-n}. \end{aligned} \tag{3}$$

Following Efron's suggestions (1993), we will use the profile likelihood, $L_p(\rho \mid z)$, as an approximated likelihood function for ρ in order to develop a Bayesian analysis without worring about the parameter θ. At this moment, we assume that the M/M/1 queue is in equilibrium, that is, $0 < \rho < 1$. From the Bayesian point of view ρ is considered as a random variable whose distribution gets updated, via Bayes theorem, as prior information is obtained. This prior information about ρ is quantified by prior distribution $\pi(\rho)$ which is updated by the data z and quantified by the posterior distribution

$$\pi(\rho \mid z) = K \pi(\rho) L_p(\rho \mid z) \tag{4}$$

where K is a proportionality constant. In the next subsections, we introduce the histogram approach (Berger, 1985) to take a 'prior sample' from the prior distribution $\pi(\rho)$, and the SIR algorithm to obtain a 'posterior sample' from the posterior distribution $\pi(\rho \mid z)$ using the 'prior sample'.

2.1 The histogram approach (Berger, 1985)

One interesting method to construct the prior distribution for ρ is the histogram approach described by Berger (1985) as follows: The interval [0,1] of possible values for ρ is broken into subintervals and then one subjectively specifies likelihoods of the different intervals. Suppose in this setting that the M/M/1 queue is in equilibrium and that the interval [0,1] is broken into 10 equal subintervals $I_i, i = 1, 2, \ldots, 10$. After some thought, based on prior belief about the intensity of traffic ρ, the likelihoods $L_i, i = 1, 2, \ldots, 10$ of the different subintervals are specified, obtaining the prior distribution histogram in Table 1.

Table 1: Prior histogram for the traffic intensity ρ

I_i	L_i
I_1	L_1
I_2	L_2
$\vdots$	$\vdots$
I_{10}	L_{10}

In order to simulate the posterior distribution of ρ using SIR algorithm which will be described in Section 2.2, we first generate an approximate prior sample $\rho_1, \ldots . \rho_m$ from the prior histogram showed by Table 1. This prior sample is easily generated using MINITAB commands. The 'discrete' subcommand of 'random' chooses an interval I_i with weight proportional to L_i and the 'uniform' subcommand of 'random' randomly chooses a point ρ_i inside the chosed interval I_i. We suggest the reader to see Albert's paper (1993) to see interesting applications of this sampling procedure.

2.2 The SIR algorithm

Suppose that ρ is the parameter of interest and that a convenient prior sample $(\rho_1 \ldots . , \rho_m)$ from the prior distribution $\pi(\rho)$ is taken via histogram approach. In this section we formulate the SIR algorithm (Rubin, 1987, 1988) to use this " prior sample " to obtain an approximate posterior sample from $\pi(\rho \mid z)$. Rubin's algorithm is as follows: Take a prior sample $(\rho_1, \ldots, \rho_m)$ from $\pi(\theta)$ and compute the sample weights

$$w(\rho_i) = \frac{\pi(\rho_i \mid z)}{\pi(\rho_i)} = K L_p(\rho_i \mid z). \qquad i = 1, \ldots, m.$$

Then one obtains a new approximate sample $(\rho_1^*, \ldots, \rho_m^*)$ with replacement of $(\rho_1, \ldots, \rho_m)$ with unequal probabilities proportional to

$$(L_p(\rho_1 \mid z), \ldots, L_p(\rho_m \mid z)).$$

The sample $\{\rho_i^*\}$ is approximately distributed from the posterior distribution $\pi(\rho \mid z)$. Furthermore, Albert (1993) showed that this method can be implemented using the MINITAB 'random' command with the 'discrete' subcommand. It is interesting to note that the prior sample $\{\rho_i\}$ reflects the beliefs of the user about the intensity of traffic ρ and the second sample $\{\rho_i^*\}$ reflects the beliefs after observing the data z . For the examples described in Section 3, we the take prior and the posterior sample size m equal to 500. This choice of m was suggested by Albert (1993) because it appears to be sufficiently large in many problems to provide an accurate description of $\pi(\rho \mid z)$. Since the SIR algorithm is an approximate method, we suggest the user to read Albert's paper to know which cautions should be taken in a real situation. One important problem in the queue theory is to predict the number N of customers in the system or the total time T that a customer stays in the system. Then, for a M/M/1 queue in steady-state, the distribution of N, given ρ, is geometric with parameter $(1 - \rho)$, that is (see, i.e. Gross and Harris, 1985)

$$p(N = n \mid \rho) = (1 - \rho)\rho^n, \quad n = 1, 2, \ldots \tag{5}$$

for $\rho < 1$. If the M/M/1 queue has N future customers in the system, then the predictive density of N is given by

$$\pi(N = n \mid z) = \int (1 - \rho)\rho^n \pi(\rho \mid z) d\rho. \tag{6}$$

From the posterior sample $\{\rho_1^*, \ldots, \rho_m^*\}$, $\pi(N = n \mid z)$ is approximated by

$$\frac{1}{m} \sum_{i=1}^{m} (1 - \rho_i^*)(\rho_i^*)^n. \tag{7}$$

Also, for any given values of μ and ρ , for $\rho < 1$, the conditional distribution of T is exponential with parameter $\mu(1-\rho)$ (see, i.e. Gross and Harris, 1985). Then, its density is given by

$$w(t \mid \mu, \rho) = \mu(1-\rho)e^{-\mu(1-\rho)t}, \quad t > 0. \tag{8}$$

The predictive density of T is given by

$$\pi(t \mid z) = E^{(\mu,\rho)|z}[w(t \mid \mu, \rho)]. \tag{9}$$

To obtain an approximate predictive density for T using the SIR algorithm, we suggest the following sampling procedure:

- Generate an approximate prior sample $(\rho_1, \ldots, \rho_m)$ using the histogram approach as in previous sections.
- Generate an approximate posterior sample $(\rho_1^*, \ldots, \rho_m^*)$ from the prior sample using the SIR algorithm.
- Using the maximum likelihood estimator of μ, given ρ, that is

$$\hat{\mu}(\rho) = \frac{n}{\rho t_a + t_s}, \tag{10}$$

  we generate an approximate joint posterior sample for (μ, ρ) which is given by

$$\{(\rho_1^*, \hat{\mu}(\rho_1^*)), \ldots, (\rho_m^*, \hat{\mu}(\rho_m^*))\}. \tag{11}$$

- From the approximate joint posterior sample given in the third step, $\pi(t \mid z)$ is approximated by

$$\frac{1}{m}\sum_{i=1}^{m} w(t \mid \hat{\mu}(\rho_i^*), \rho_i^*) \tag{12}$$

3 Some illustrative examples

In this section, we apply our sampling procedure described in previous sections to a real-life example in Vogel (1979) and compare with the results obtained by Armero and Bayarri (1992) which used a conjugate Bayesian analysis and a non trivial non-informative prior. Also, some simulated data via GPSS/H is obtained to illustrate the sampling procedure along with informative priors.

3.1 Example 1: Becton Dickinson example

In this section, we illustrate the sampling procedures obtained in previous sections with a real-life example reported in Vogel (1979) and discussed in details by Armero and Bayarri (1992) using queueing theory. The problem was the follows: the manufacturing machines seem to suffer jams and the major function of attendants was clearing these jams. In the optimum assignment, a machine operator was responsible for 5 machines. The situation was modelled as an M/M/1 queue in which the jammings were considered the customers and the machine attendant, the server. We suggest the reader to see Armero and Bayarri's paper to known how the assumptions of an M/M/1 queue were checked. Furthermore, based on the data, it was estimated that each machine is called for service at a mean rate of 60 times per hour;hence there are 5 machines, the estimated mean arrival rate (expected number of jams in hour) is $\hat{\lambda} = 300$. Also, it was found that the estimated mean service rate (expected number of jams fixed by an operator per hour) is $\hat{\mu} = 449.82$, so that, it took on operate an average of time of 8 seconds approximately to fix a jam. I was assumed that $n_a = n_s = 500$. Motivated by the conjugate Bayesian analysis, Armero and Bayarri used the following non-informative prior:

$$\pi_{AB}(\rho, \mu) \propto \frac{1}{\rho(1-\rho)^2\mu} \tag{13}$$

To apply our sampling procedure discussed in previous sections and compare with the Armero and Bayarri's results, we adopt the following non-informative prior distribution histogram for ρ:

Table 2: Non-informative prior histogram for ρ

I_i	L_i
I_1	1.0
I_2	1.0
$\vdots$	$\vdots$
I_{10}	1.0

The MINITAB macros that were used in this example are listed in the Appendix. The MINITAB 'dotplot' command is used to give parallel graphs

the two simulated samples for ρ which are presented in Figure 1.

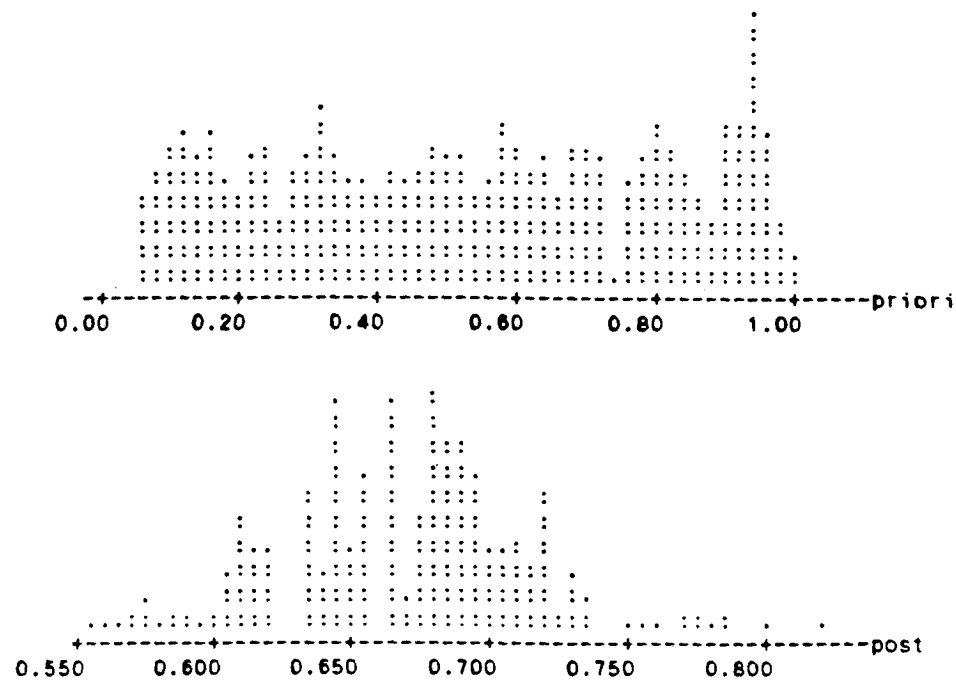

Figure 1:Parallel Dotplots of prior and posterior samples for ρ.

Table 3 shows a comparison between the results obtained from our sampling procedure and the conjugate Bayesian analysis introduced by Armero and Bayarri (1992).

Table 3: Prior mean and posterior mean for ρ

prior mean	posterior mean
$\hat{\rho_H} = 0.531$	$\hat{\rho_{SIR}} = 0.669$
	$E_{AB}[\rho \mid z] = 0.668$

The prior mean $\hat{\rho}_H$, the posterior mean $\hat{\rho}_{SIR}$ and the dotplots suggest that the performance of the M/M/1 queue is greater than the performance reflected by the prior. Furthermore, from Table 3, the posterior mean of ρ using SIR algorithm is approximately equal to the exact posterior mean obtained by Armero and Bayarri (1992) using a non trivial conjugate Bayesian

analysis. It is important to emphasize that we get only information of ρ without worrying about μ as in Armero and Bayarri's approach.

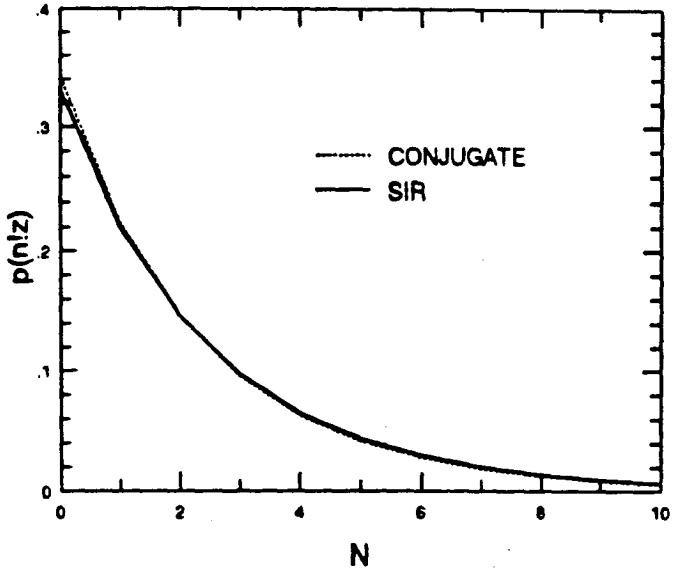

Figure 2: Predictive distributions of N using the SIR algorithm and the conjugate Bayesian analysis

Figure 2 presents the predictive distributions of N, number of machines in the system, using the sampling method discussed in (8) and the conjugate Bayesian analysis (Armero and Bayarri, 1992). In this example, a machine is in the queueing system when it is inactive because either it is waiting for the attendant to clean up the jam, or because the attendant is already working in it. From Figure 2, we observe that these predictive densities are quite closed showing a very nice performance of our sampling procedure by using only an approximate uniform prior information for the intensity of traffic ρ. Also, the probability that the system is empty (that is, all machines are working) can be easily computed from this distribution and is summarized as follows:

$$\begin{aligned} Pr(N = 0 \mid z) &= 0.342 \quad \text{(Armero and Bayarri)} \\ \hat{Pr}(N = 0 \mid z) &= 0.330 \quad \text{(SIR algorithm)} \end{aligned}$$

Figure 3 shows the predictive distributions of T, the time that a machine spends in the system to be fixed, using the sampling procedure introduced in

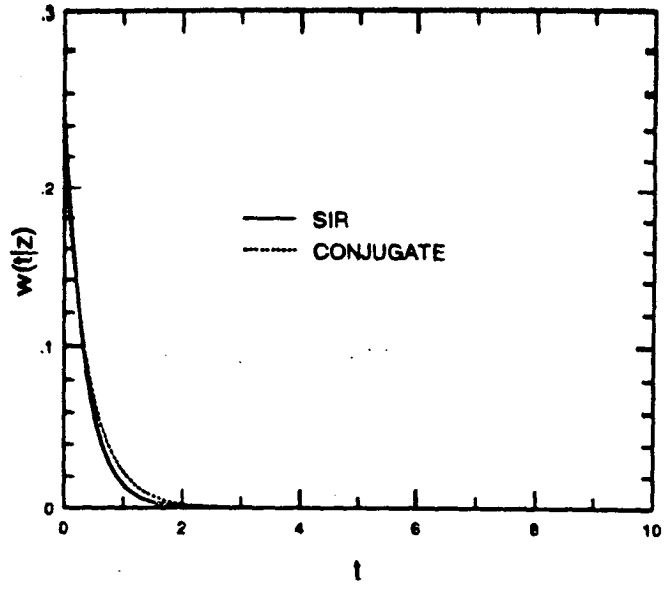

Figure 3: Predictive distributions of T, time a machine spends in the system, using the SIR algorithm and the conjugate Bayesian analysis

the previous section and the Armero and Bayarri's method. The distribution appearing is not actually that of T, but the one of $U = T/\bar{t}_s$, where $\bar{t}_s = 8$ seconds, is the average time needed to fix the 500 jams. Also, the predictive distributions are quite closed and the means and variances are :

$$\begin{aligned} E[T \mid z] &= 2.960\bar{t}_s \quad \text{(Armero and Bayarri)} \\ Var[T \mid z] &= 9.229\bar{t}_s^2 \quad \text{(Armero and Bayarri)} \\ \hat{E}[T \mid z] &= 3.105\bar{t}_s \quad \text{(SIR algorithm)} \\ \hat{Var}[T \mid z] &= 9.937\bar{t}_s^2 \quad \text{(SIR algorithm)} \end{aligned}$$

3.2 Example 2: Predictive distributions for the number of customers and the waiting time in the system: Informative prior

In this example, we undertake a Bayesian analysis of a simulated experiment in which $n_a = 28$, $n_s = 25$, $\rho = 0.8$, $t_s = 20.1558$ and $t_a = 27.4949$ along with an informative prior. The sample prior for ρ was generated from the prior histogram given by

Table 4: Prior histogram for ρ

I_i	L_i
I_1	10
I_2	20
I_2	30
I_3	40
⋮	⋮
I_9	90
I_{10}	95

Note that the subinterval of highest likelihood is (0.9, 1.0), indicating that the Bayesian statistician expects a strong intensity of traffic in the system.

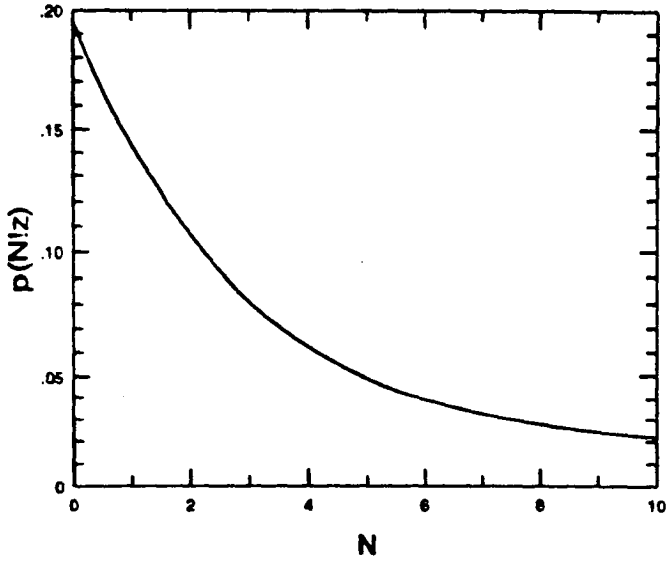

Figure 4: Predictive distributions of N using SIR-algorithm

Figure 4 shows the predictive distribution of N, the number of customers in the system obtained from the SIR- algorithm. To see how the statistician's prior opinions about ρ can be modified by the data, some quantiles of the prior and the posterior distribution of ρ are given in Table 4.

Quantiles of the prior and the posterior distributions of ρ

order	0.25	0.50	0.75	0.95
prior quantiles	0.500	0.724	0.991	0.999
posterior quantiles	0.705	0.826	0.915	0.997

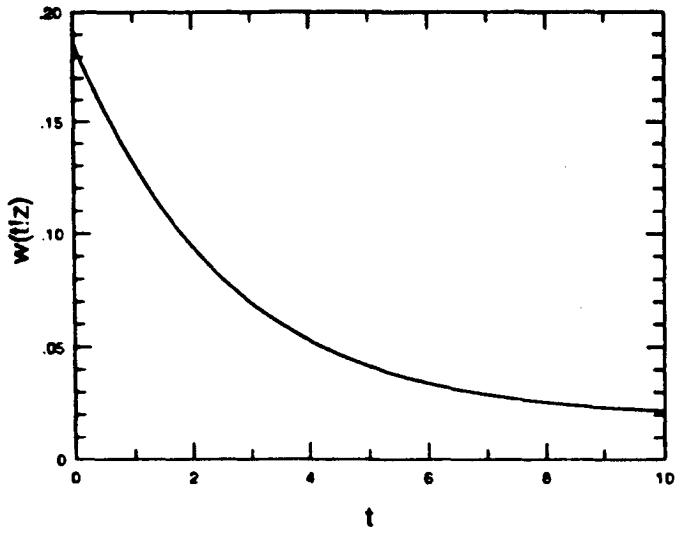

Figure 5: Predictive distributions of T using the SIR algorithm.

References

Armero, C. and Bayarri, M.J. (1992). Prior assessment for prediction in queues, *technical report*, no. 10-92.

Albert, J. (1993). Teaching Bayesian statistics using sampling methods and Minitab, *The American Statistician*, vol.47, no.3, pp. 182-191.

Berger, J.O. (1985). *Statistical Decision Theory and Bayesian Analysis*, New York; Spring -Verlag.

Efron, B. (1993). Bayes and likelihood calculations from confidence intervals, *Biometrika*, 80, 1, pp.3-26

Gelfand, A.E. and Smith, A.F.M. (1990). Sampling-based appoaches to calculating marginal densities, *Journal of the American Statistical Association*, 85, pp. 398-409

Gross, D. and Harris, C.M. (1985). *Fundamentals of queueing theory*, second edition, New York:Wiley.

Rubin, D.B. (1987). Comment to ' The calculation of posterior distributions by data augmentation' by Tammer and Wong, *Journal of the American Statistical Association*, 82, pp. 543-546

Rubin, D.B. (1988).Using the SIR algorithm to simulate posterior distributions, *In Bayesian Statistics 3*, eds. J. M. Bernardo, M. H. DeGroot, D. V. Lindley, and A.F.M. Smith, New York: Oxford University Press.

```
##############################################################################
# Minitab macro 'data-setup'
# Example 2

# Set up data columns and enter histogram for traffic intensity

name c1 'vat' c2 'hts' c3 'p' c4 'rvat' c5 'unif' c6 'priori'
name c7 'loglike' c8 'post' c9 'probs'

set 'vat'
0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 0.95
end
set 'hts'
10 20 30 40 50 60 70 80 90 95
end
let k1 = 500      # number of simulated values
let k2 = 0.1      # width of each histogram
let k3 = 28       # total the interarrival times
let k4 = 53       # total of the observation
let k5 = 27.4949 # total time of the interarrival
let k6 = 20.1558 # total time of the service

##############################################################################


##############################################################################
# MINITAB macro 'priori.sim'

let 'p' = 'hts'/sum('hts')              #
random k1 'rvat';                       #
discrete 'vat' 'p'.                     #  Simulate from
random k1 'unif';                       #  prior histogram
uniform 0 k2.                           #
let 'prior' = 'rvat' + 'unif' - k2/2  #

##############################################################################


##############################################################################
# MINITAB macro 'post.sim'

# Simulate of posterior

let 'loglike' = k3*log(c6)+k4*log(k3)-k3*log(c6*k5+k6)-k4 # Compute likelihood
let 'probs'= & exp('loglike'-max('loglike'))/             # at prior sample
             sum('loglike'-max('loglike')))               # and normalize
                                                          # likelihood values
random k1 'post';                       # Simulate posterior
discrete 'prior' 'probs'.               # sample
let 'servi'= k4 / ('post'* k5 + k6)     # maximum likelihood estimator
                                        # of service rate

##############################################################################
```

```
##############################################################################

# MINITAB macro 'prednu.sim'

# Compute predictive probability of number the custormers in the system

name c13 'predn'

let k7 = 0                                  # number of custormers in the system
let c12 = (1- 'post') * ('post'** k7)       # distribution geometric
let 'predn = sum(c12)/k1                    # predictive prob
print 'predn'


##############################################################################
# MINITAB macro 'predsi.sim'

# Computes predict the total time T that a customer stays in the system

name c15 'predsi'

let k12 = T                                                    # time
let c14 = 'servi'*(1-'post')*exp(-'servi'*(1-'post')*k12) # density the waiting
                                                               # time in the system
let 'predsi'= sum(c14)/k1
print 'predsi'

##############################################################################
# MINITAB macro 'meanvar.num'

# Computes the mean and variance of the distribution of the
# number custormes in the system

name c17 'mean' c19 'var'

let c16 = 'post'/ (1- 'post')
let 'mean' = sum(c16)/k1
let c18 = ('post'** 2 )/(( 1 - 'post') ** 2)
let 'var' = sum(c18)/k1
print 'mean' 'var'

##############################################################################
# MINITAB macro 'meanvar-syste'

# Computes the mean and variance of the distribuition
# the waitaing time in the system

name c21 'meansy' c223 'varsy'

let c20 =1 / ( 'servi'* (1 - 'post'))
let 'meansy'= sum(c20)/ k1
let c22 = 1 / (('servi'** 2) * (( 1 - 'post')** 2))
let 'varsy'= sum(c22)/k1
print 'meansy' 'varsy'

##############################################################################
```

NOTAS DO ICMSC

SÉRIE ESTATÍSTICA

015/95 ACHCAR, J.A.; PEGORINM M.A. - Laplace's approximations for posterior expectations when the mode is not in the parameter space.

014/94 ACHCAR, J. A.; FOGO, J.C. - Accurate inferences for the reliability function considering accelerated life tests.

013/94 RODRIGUES, J. - Bayesian Solutions to a lass of selections problems using weighted loss functions.

012/94 ACHCAR, J.A.; DAMASCENO, V.L. - Extreme value models: an useful reparametrization for the survival function

011/94 ACHCAR, J.A.- Approximate bayesian analysis for non-normal hierarchical classification model.

010/94 RODRIGUES, J.; RODRIGUES, E.F. - Bayesian estimationin the study of tampered Random variables.

009/94 ACHCAR, J.A.; FOGO, J.C. - An useful reparametrization for the reliability in the Weibull case.

008/94 RODRIGUES, J.; LEITE, J.G. - A note on Bayesian analysis in M/M/1 queues from confidence intervals.

007/94 ACHCAR, J.A.; DAMACENO, V.L. - An useful reparametrization for the survival function considering an exponential regression model.

006/94 RODRIGUES, J. - Bayesian estimation of a normal mean parameter using linex loss function and robustness considerations.

005/93 ACHCAR, J.A.; FOGO, J.C.- Some useful reparametrizations for the reliability function considering an exponential regession model.

004/93 LEME, E.C. - A m.v.u. estimator v.s. a m.l. estimator: a way of comparing their precisions by using the Pareto distribution.

003/93 ACHCAR, J.A.- Some practical aspects of approximate bayesian inference