

MOMENT ESTIMATION OF CUSTOMER LOSS RATES FROM TRANSACTIONAL DATA

D.J. DALEY¹

*Australian National University
School of Mathematical Sciences
Canberra ACT 0200 Australia*

L.D. SERVI

*GTE Laboratories Incorporated
40 Sylvan Road
Waltham, MA 02254 USA*

(Received October, 1997)

Moment estimators are proposed for the arrival and customer loss rates of a many-server queueing system with a Poisson arrival process with customer loss via balking or reneging. These estimators are based on the lengths $\{S_{j1}\}$ of the initial inter-departure intervals of the busy periods $j = 1, \dots, M$ observed in a dataset consisting of service starting and finishing times and encompassing both busy and idle periods of the process, and whether those busy periods are of length 1 or > 1 . The estimators are compared with maximum likelihood and parametric model-based estimators found previously.

Key words: Balking, Reneging, Customer Loss, Moment Estimation, Transactional Data.

AMS subject classifications: 60K25, 62M05.

1. Motivation

Under all but the simplest modeling assumptions, quantifying customer service quality from transactional records of, for example, automatic teller machines (cf. Larson [11]) or wireless communication systems (cf. Daley and Servi [4]) necessitates first inferring the customer arrival rate and behavior in the face of congestion, and only then estimating derivative information such as concerns the transient queue length process.

This paper discusses deductions from records of the epochs where services start and end for each customer served in a c -server queueing system having Poisson arri-

¹Work done with partial support from GTE while visiting GTE Laboratories Incorporated.

vals with unknown rate parameter λ under one of the following two distinct assumptions for customer behavior.

Assumption B (Balking): An arriving customer who finds all servers busy, independently of all other such customers, either balks with probability ϖ , never to return again, or else stays in the system and waits for service.

Assumption R (Reneging): An arriving customer who finds all servers busy, independently of all other such customers, stays in the system and then either reneges after an exponentially distributed time (and this random variable has mean $1/\eta$), or enters service after waiting some time, whichever of these two time periods is the shorter, i.e., each waiting customer has a constant hazard rate η of reneging.

A queueing system with mean service time b and satisfying one of those assumptions can exist in a stationary state for $\lambda(1 - \varpi)b < c$ under the balking assumption, for all $\eta > 0$ under the reneging assumption, and for $\lambda b < c$ if $\varpi = 0$ or $\eta = 0$.

In previous work (Daley & Servi [5, 7]), we have shown how to estimate λ and either ϖ or η . In the latter, we used an additional assumption of exponentially distributed service times (the queueing model is then of M/M/c type). In the former, we developed maximum likelihood (ML) procedures without this unnecessary additional assumption.

This paper uses less information than the likelihood approach, but provides both reader access to the parameter estimates and their properties. It also has computational advantages over the ML method, and its results can be used as a starting point for exploring the likelihood surface. None of these approaches gives uniformly better estimates.

Every statistical procedure is based on modeling assumptions. Relative to standard model-building in queueing theory, our data do not provide observations of the commonly used key quantities like queue size or waiting time, i.e., direct measures of congestion. Nevertheless, we can make deductions about their behavior (for waiting time, cf. Bertsimas & Servi [1], Hall [8], Jones & Larson [9], Daley & Servi [6]), because we assume that the arrival process is Poisson and that customers balk or renege independently of each other; we then exploit the mathematical consequences of these probabilistic assumptions.

2. The Data and Notation

Let $\mathfrak{D} \equiv \{d_n\}$ and $\mathfrak{T} \equiv \{t_n\}$, $n = 1, 2, \dots$, represent respectively the times of starting and finishing service on customer n , the label n being applied only to the customers who enter service. At any epoch in time, either all servers are busy or at least one server is idle, so the dataset $(\mathfrak{D}, \mathfrak{T})$ partitions the time interval \mathfrak{J} from which the data are drawn into busy periods and idle periods, characterized by $t \in \mathfrak{J} \setminus (\mathfrak{D} \cup \mathfrak{T})$ being either a busy epoch or idle epoch respectively. The dataset also suffices to determine the counting function

$$N_S(t) = \text{number of servers occupied at } t; \quad (2.1)$$

its points of increase, i.e., $\{t: N_S(t-) < N_S(t+)\}$, are the service commencement epochs that coincide with arrivals, and so this set is the set of *observed arrival*

epochs; while we know that other arrivals occur, we do not know the epochs where they occur.

Amongst the elements of the set $\mathfrak{D} \cup \mathfrak{T}$ are the epochs where simultaneously one service finishes and another starts, namely $t' \in \mathfrak{D} \cap \mathfrak{T}$. During an idle period, $N_S(t)$ increases on the set of epochs $\mathfrak{D} \setminus (\mathfrak{D} \cap \mathfrak{T})$ and decreases on the set of epochs $\mathfrak{T} \setminus (\mathfrak{D} \cap \mathfrak{T})$. A busy period starts at the end of an idle period where $N_S(\cdot)$ increases to c , and ends at an epoch where $N_S(\cdot)$ decreases from c . The integrals

$$N_I \equiv \int_{a+}^{b-} (dN_S(u))_+ \quad \text{and} \quad T_I \equiv \int_a^b I_{\{N_S(u) < c\}} du \tag{2.2}$$

represent respectively the number of observed arrival epochs on (a, b) and the total length of intervals on (a, b) spent ‘waiting for an observed arrival’ as known from the dataset; mnemonically, N_I is the Number of observed arrivals occurring in Time intervals of total length T_I when at least one server is Idle. Consequently,

$$\frac{1}{\widehat{\lambda}_I} \equiv \frac{T_I}{N_I} \tag{2.3}$$

is a moment estimator of the average time waiting for an arrival; under the Poisson process assumption, $\widehat{\lambda}_I$ is also the MLE of λ based on these waiting times (see Daley and Servi [5]).

Suppose there are M busy periods in all, with the i th consisting of N_i customers and inter-departure time intervals $\{S_{ik}: k = 1, \dots, N_i\}$ between successive departure epochs. (There is a minor abuse of terminology here: the ‘inter-departure’ intervals $\{S_{i1}: i = 1, \dots, M\}$ start at epochs where busy periods start, and end with the first departure epochs in the respective busy periods). In Daley and Servi [5] we constructed likelihood functions using all of these data. Here we show how they can be used quite differently, and different from Daley and Servi [7] as well, in constructing moment estimates of the balking or reneging parameters. The approach described below arose as an alternative one to the use of an ML estimator of η based on a likelihood function that is quite flat in the region of its maximum.

3. Balking and Reneging: Moment Estimates

As in Daley and Servi [5, 7], we estimate λ by $\bar{\lambda} \equiv \widehat{\lambda}_I$ defined in (2.3).

We propose moment estimators for the parameters ϖ and η corresponding to the two modes of loss behavior described in Section 1. These estimators are based on the number n_1^{BP} of busy periods of length 1 in an observation interval $(0, T)$, from which the transactional dataset of this paper is assumed drawn. On $(0, T)$, we suppose, consistent with our earlier notation, that the M busy periods have ‘initial inter-departure intervals’ $\{S_{i1}: i = 1, \dots, M\}$ (see note below (2.3)).

Suppose the initial inter-departure interval of some busy period has length S_1 . There is just one customer in this busy period when there is balking if there is no arrival during S_1 that stays or when there is reneging if any arrival during S_1 has left by the end of that period. Simple computation (cf. (b) and (d) of Example 2 of Daley and Servi [4]) gives

$\Pr\{\text{busy period has length } 1 \mid S_1\} \equiv$

$$\begin{cases} p_1^B(\varpi \mid S_1) = e^{-\lambda(1-\varpi)S_1} & \text{(balking model),} \\ p_1^R(\eta \mid S_1) = \exp\left[-\frac{\lambda(1-e^{-\eta S_1})}{\eta}\right] & \text{(renewing model),} \end{cases} \tag{3.1}$$

and these two probabilities agree over a range of values of S_1 if and only if $(\varpi, \eta) = (0, 0)$ or $(1, \infty)$. Then, given the data $\{S_{i1}: i = 1, \dots, M\}$, moment estimators of ϖ or η as appropriate follow by equating n_1^{BP} to the respective expectations that come from (3.1), namely

$$E(n_1^{BP}) = \begin{cases} \sum_{i=1}^M p_1^B(\varpi \mid S_{i1}) \equiv g_B(\varpi) & \text{(balking),} \\ \sum_{i=1}^M p_1^R(\eta \mid S_{i1}) \equiv g_R(\eta) & \text{(renewing),} \end{cases} \tag{3.2}$$

subject to $0 < \varpi < 1, 0 < \eta < \infty$. Such equations, conditional on λ , yield unique solutions $\bar{\varpi}$ and $\bar{\eta}$ say. First we note that the extreme cases below yield extreme values of $\bar{\varpi}$ or $\bar{\eta}$:

$$\begin{aligned} n_1^{BP} = M & \quad \text{implies } \bar{\varpi} = 1 \text{ or } \bar{\eta} = \infty; \\ n_1^{BP} \leq g_B(0) = g_R(0) \equiv g(0) = \sum_{i=1}^M e^{-\lambda S_{i1}} & \quad \text{implies } \bar{\varpi} = 0 \text{ or } \bar{\eta} = 0. \end{aligned} \tag{3.3}$$

We discuss these extreme cases and their complement in Section 4 below.

An attractive feature of these estimators $\bar{\varpi}$ and $\bar{\eta}$ is that approximate estimates of $\text{var}(\bar{\varpi})$ and $\text{var}(\bar{\eta})$ are available as follows. For given λ and $\{S_{i1}\}$, the function $g_B(\varpi)$ at (3.2) has

$$\frac{dg_B}{d\varpi} = \sum_{i=1}^M \lambda S_{i1} e^{-\lambda(1-\varpi)S_{i1}}. \tag{3.4}$$

Then, considering $\bar{\varpi} = \bar{\varpi}(n_1^{BP})$ as a function of the random variable n_1^{BP} whose variance equals $\sum_{i=1}^M p_1^B(\varpi \mid S_{i1})[1 - p_1^B(\varpi \mid S_{i1})]$, an elementary Taylor series argument (see e.g., Kendall and Stuart [10] 1958, §§10.6-7) yields

$$\text{var}(\bar{\varpi}) \approx \left(\frac{d\bar{\varpi}}{dn_1^{BP}}\right)^2 \text{var}(n_1^{BP}) = \frac{\sum_{i=1}^M e^{-\lambda(1-\bar{\varpi})S_{i1}}(1 - e^{-\lambda(1-\bar{\varpi})S_{i1}})}{\left(\sum_{i=1}^M \lambda S_{i1} e^{-\lambda(1-\bar{\varpi})S_{i1}}\right)^2} \tag{3.5}$$

$$\approx \frac{1}{\left(\sum_{i=1}^M \lambda S_{i1} / M\right)^2} \left(\frac{1}{E(n_1^{BP})} - \frac{1}{M}\right), \tag{3.6}$$

with equality at (3.6) if $S_{i1} = S$ (all i), so that then

$$\text{var}\bar{\varpi} \approx \frac{1}{(\lambda S)^2} \left(\frac{1}{E(n_1^{BP})} - \frac{1}{M}\right) = \frac{e^{\lambda S(1-\bar{\varpi})} - 1}{M(\lambda S)^2}. \tag{3.7}$$

Replacing g_B in this argument by $g_R(\eta)$ and supposing, as before, that λ and

$\{S_{i1}\}$ are known, we have

$$\frac{dg_R}{d\eta} = \sum_{i=1}^M \exp(-\lambda(1 - e^{-\eta S_{i1}})/\eta) \frac{\lambda[1 - e^{-\eta S_{i1}}(1 + \eta S_{i1})]}{\eta^2}, \tag{3.8}$$

and hence

$$\text{var}(\bar{\eta}) \approx \left(\frac{d\eta}{dg_R}\right)^2 \sum_{i=1}^M \exp(-\lambda(1 - e^{-\eta S_{i1}})/\eta) [1 - \exp((1 - e^{-\eta S_{i1}})/\eta)] \tag{3.9}$$

$$\approx \frac{1}{\left(\sum_{i=1}^M \lambda[1 - e^{-\eta S_{i1}}(1 + \eta S_{i1})]/M\eta^2\right)^2} \left(\frac{1}{E(n_1^{BP})} - \frac{1}{M}\right), \tag{3.10}$$

with equality at (3.10) if $S_{i1} = S$ (all i), in which case (cf. (3.7))

$$\text{var}(\bar{\eta} S) \approx \left(\frac{(\eta S)^2}{\lambda S[1 - e^{-\eta S}(1 + \eta S)]}\right)^2 \left(\frac{1}{E(n_1^{BP})} - \frac{1}{M}\right). \tag{3.11}$$

Table 1 gives values of the approximations to $\sqrt{\text{var}\bar{\omega}}$ and $\sqrt{\text{var}(\bar{\eta} S)}$, together with the exact values of the former and the exact value of $\sqrt{\text{var}(\bar{\eta} S \mid \bar{\eta} < \infty)}$ for a range of values of ϖ and λS , when $S_i = S$ (all i) and $M = 25$. The model value for ηS is determined as at (5.2) below. For the middle (third) line of each set, see around (6.3) below.

3.1 Remarks Concerning Moment Estimators

It should be evident from (2.3), (3.6), (3.10) and Table 1, that the moment estimators described around (3.2) are potentially useful provided both N_I is large enough (to ensure a good estimator of λ), and that M is large enough. What is ‘large enough’ here depends on the balance between all three of the arrival rate λ , the mean service time b and the customer loss rate ϖ or η as applicable. This dependence is illustrated in Table 2 of the next section.

4. When Can Uncertain Call Loss be Detected?

The extreme case $n_1 \equiv n_1^{BP} = M$ occurs with probability one when $\varpi = 1$ or $\eta = \infty$; see (3.3). Otherwise, whenever $M \geq 2$ and $\sum_{i=1}^M e^{-\lambda S_i} = g(0) < M - 1$, each of the possibilities below,

- (a) that $\{g(0) < n_1 \leq M - 1\}$ and the estimate of ϖ or η is positive;
- (b) that $\{n_1 \leq g(0)\}$ and $\bar{\omega}$ or $\bar{\eta}$ is zero; and
- (c) that $\{n_1 = M\}$ and $\bar{\omega} = 1$ or $\bar{\eta} = \infty$;

has positive probability, so any may occur. To simplify discussion, we consider the balking case for the rest of this section; the renegeing case can be treated similarly at the cost of greater algebraic (and computational) complexity.

Recall that the random variable n_1 has a generalized binomial distribution when the S_i are different. If $S_i = S$ (all i), then n_i is binomially distributed with parameters $(M, e^{-\lambda S(1 - \varpi)}) \equiv (M, P)$ say, and again, for the sake of simplicity, we treat only this case below. We discuss the two cases $\varpi = 0$ and $0 < \varpi < 1$ in turn.

Table 1

Approximate and exact standard deviations of the moment estimators $\bar{\omega}$ (top two lines) and $\bar{\eta} S$ (bottom two lines), and of the MLE $\hat{\omega}$ (middle line), for $M = 25$.

| Model | values | λS | 0.10 | 0.20 | 0.30 | 0.50 | 0.75 | 1.00 | 1.25 | 1.50 |
|-------|------------------|-------------|-------|-------|-------|-------|-------|-------|-------|-------|
| 0.464 | ηS 0.20 | | 0.577 | 0.417 | 0.347 | 0.281 | 0.242 | 0.221 | * | * |
| | | | 0.378 | 0.343 | 0.341 | 0.339 | 0.324 | 0.335 | * | * |
| | | | 0.167 | 0.175 | 0.184 | 0.207 | 0.253 | 0.358 | * | * |
| | | | 1.56 | 1.13 | 0.94 | 0.76 | 0.65 | 0.60 | 0.57 | 0.55 |
| | | | 0.95 | 1.41 | 1.31 | 0.88 | 0.65 | 0.54 | 0.49 | 0.47 |
| 1.126 | 0.40 | | 0.497 | 0.357 | 0.296 | 0.237 | 0.201 | 0.181 | 0.169 | 0.161 |
| | | | 0.354 | 0.302 | 0.274 | 0.243 | 0.210 | 0.200 | 0.191 | 0.179 |
| | | | 0.124 | 0.128 | 0.133 | 0.143 | 0.162 | 0.190 | 0.240 | 0.379 |
| | | | 2.03 | 1.46 | 1.21 | 0.97 | 0.82 | 0.74 | 0.69 | 0.66 |
| | | | 0.98 | 1.61 | 1.74 | 1.39 | 1.00 | 0.81 | 0.73 | 0.68 |
| 2.232 | 0.60 | | 0.404 | 0.289 | 0.238 | 0.188 | 0.158 | 0.140 | 0.129 | 0.121 |
| | | | 0.322 | 0.292 | 0.245 | 0.190 | 0.162 | 0.144 | 0.134 | 0.126 |
| | | | 0.082 | 0.083 | 0.085 | 0.089 | 0.096 | 0.103 | 0.113 | 0.126 |
| | | | 3.08 | 2.20 | 1.82 | 1.44 | 1.20 | 1.07 | 0.98 | 0.92 |
| | | | 1.01 | 1.83 | 2.26 | 2.38 | 1.92 | 1.50 | 1.24 | 1.10 |
| 3.197 | 0.70 | | 0.349 | 0.249 | 0.205 | 0.161 | 0.134 | 0.118 | 0.108 | 0.101 |
| | | | 0.304 | 0.315 | 0.268 | 0.188 | 0.143 | 0.123 | 0.112 | 0.105 |
| | | | 0.061 | 0.062 | 0.063 | 0.065 | 0.068 | 0.072 | 0.076 | 0.081 |
| | | | 4.31 | 3.07 | 2.52 | 1.99 | 1.65 | 1.46 | 1.33 | 1.24 |
| | | | 1.02 | 1.97 | 2.54 | 3.10 | 2.96 | 2.46 | 2.02 | 1.70 |
| 4.965 | 0.80 | | 0.284 | 0.202 | 0.166 | 0.130 | 0.107 | 0.094 | 0.085 | 0.079 |
| | | | 0.282 | 0.353 | 0.333 | 0.244 | 0.160 | 0.115 | 0.094 | 0.083 |
| | | | 0.040 | 0.041 | 0.041 | 0.042 | 0.043 | 0.045 | 0.046 | 0.048 |
| | | | 7.31 | 5.20 | 4.26 | 3.34 | 2.76 | 2.42 | 2.19 | 2.03 |
| | | | 0.99 | 2.14 | 2.89 | 3.90 | 4.52 | 4.50 | 4.10 | 3.60 |
| 10.00 | 0.90 | | 0.201 | 0.142 | 0.116 | 0.091 | 0.074 | 0.065 | 0.058 | 0.054 |
| | | | 0.237 | 0.369 | 0.409 | 0.393 | 0.323 | 0.251 | 0.191 | 0.146 |
| | | | 0.020 | 0.020 | 0.020 | 0.021 | 0.021 | 0.021 | 0.021 | 0.022 |
| | | | 20.06 | 14.22 | 11.64 | 9.06 | 7.45 | 6.49 | 5.84 | 5.37 |
| | | | 0.86 | 2.20 | 3.33 | 5.04 | 6.55 | 7.69 | 8.50 | 8.95 |
| 20.00 | 0.95 | | 0.142 | 0.100 | 0.082 | 0.064 | 0.052 | 0.045 | 0.041 | 0.037 |
| | | | 0.186 | 0.321 | 0.390 | 0.445 | 0.448 | 0.421 | 0.382 | 0.340 |
| | | | 0.010 | 0.010 | 0.010 | 0.010 | 0.010 | 0.010 | 0.010 | 0.010 |
| | | | 56.64 | 40.10 | 32.78 | 25.46 | 20.85 | 18.11 | 16.25 | 14.88 |
| | | | 0.68 | 1.93 | 3.20 | 5.59 | 8.08 | 10.06 | 11.69 | 13.08 |

* System unstable

First we assert that when MP is not small, $\Pr\{\bar{\omega} = 0 \mid \omega = 0\} \approx 0.5$, i.e., in hypothesis testing terms, the Type I error rate is about 50%. This assertion follows from the relations

$$\Pr\{\bar{\omega} = 0 \mid \omega = 0\} = \Pr\{n_1 \leq g(0) \mid \omega = 0\} = \Pr\{n_1 \leq E(n_1) \mid \omega = 0\} \approx 0.5,$$

where the approximation follows from the central limit theorem.

Alternatively, when $0 < \varpi < 1$, we seek conditions under which $0 < \bar{\varpi} < 1$. Appealing again to the central limit theorem for n_1 ,

$$\begin{aligned} \Pr\{0 < \bar{\varpi} < 1 \mid \varpi\} &= \Pr\{g(0) - MP < n_1 - MP < M(1 - P) \mid \varpi\} \\ &\approx \Phi\left(\sqrt{M(e^{\lambda S(1-\varpi)} - 1)}\right) - \Phi\left(-\frac{\sqrt{M}(1 - e^{-\lambda S\varpi})}{\sqrt{e^{\lambda S(1-\varpi)} - 1}}\right) \equiv \alpha(M, \lambda S, \varpi), \\ &\rightarrow 1 \quad (M \rightarrow \infty \text{ for fixed } P), \end{aligned}$$

where Φ denotes the standard Gaussian distribution function with zero mean and unit variance. Table 2 gives the least values of M required to ensure that $\alpha(M, \lambda S, \varpi) \geq 0.95$ for a representative set of values of λS and ϖ (ϖ and ηS are related as at (5.2)).

Table 2
Critical values of M for $\alpha(M, \lambda S, \varpi) = 0.95$

| ηS | ϖ | λS | 0.10 | 0.20 | 0.30 | 0.50 | 0.75 | 1.00 | 1.25 | 1.50 |
|----------|----------|-------------|--------|--------|-------|-------|-------|-------|-------|-------|
| 0.215 | 0.10 | | 2571.0 | 1361.8 | 960.7 | 646.7 | 499.8 | 436.2 | 407.8 | 398.6 |
| 0.464 | 0.20 | | 574.4 | 305.4 | 216.5 | 147.0 | 114.7 | 100.9 | 95.0 | 93.5 |
| 0.761 | 0.30 | | 224.6 | 119.9 | 85.4 | 58.4 | 46.0 | 40.8 | 38.7 | 38.3 |
| 1.126 | 0.40 | | 114.1 | 60.3 | 42.7 | 29.1 | 22.9 | 20.5 | 19.5 | 19.4 |
| 1.594 | 0.50 | | 78.9 | 40.7 | 28.0 | 18.2 | 13.6 | 11.7 | 11.0 | 10.9 |
| 2.232 | 0.60 | | 72.3 | 36.4 | 24.5 | 15.1 | 10.6 | 8.6 | 7.5 | 7.0 |
| 3.197 | 0.70 | | 89.0 | 43.9 | 28.9 | 17.0 | 11.2 | 8.4 | 6.9 | 6.0 |
| 4.965 | 0.80 | | 133.9 | 66.3 | 43.8 | 25.7 | 16.7 | 12.2 | 9.5 | 7.8 |
| 10.00 | 0.90 | | 269.1 | 134.0 | 88.9 | 52.8 | 34.7 | 25.7 | 20.3 | 16.7 |
| 20.00 | 0.95 | | 539.5 | 269.3 | 179.1 | 106.9 | 70.8 | 52.8 | 42.0 | 34.7 |

5. Distinguishing Between Balking and Reneging

Depending on whether the balking or reneging model is assumed, the moment estimates $\bar{\varpi}$ and $\bar{\eta}$ are determined either by one of the extremes at (3.3), or else by solving either

$$g_B(\bar{\varpi}) = n_1^{BP} \quad \text{or} \quad g_R(\bar{\eta}) = n_1^{BP} \tag{5.1}$$

as appropriate. This equation begs the question as to which model is appropriate while showing that these estimates are functionally related: in the simplest case, when $S_i = S$ (all i) as in an M/D/1 queueing system, the equations at (5.1) yield

$$\exp(-\lambda(1 - \bar{\varpi})S) = \exp\left(-\frac{\lambda S(1 - e^{-\bar{\eta} S})}{\bar{\eta} S}\right), \tag{5.2}$$

so the relationship is independent of λ , and equivalent to

$$\bar{\varpi} = 1 - \frac{1 - e^{-\bar{\eta} S}}{\bar{\eta} S} = \frac{\bar{\eta} S}{2} - \frac{(\bar{\eta} S)^2}{3!} + \dots \tag{5.3}$$

This relation is not the same one as that existing between equivalent ϖ and η in $M/M/c$ systems with the same arrival rate, service rate, and customer rate loss as discussed around Figure 1 and equation (3.7) in Daley and Servi [7].

We turn to the more substantive question of distinguishing between the two models, i.e., is one model to be preferred because it explains the data better than the other? We show that in principle, we can make the distinction provided that both the set $\{S_{i1}\}$ is varied enough and M is large enough.

To do so, we recast our estimation procedure above into a likelihood setting that is simpler than that in Daley and Servi [5]. We use the lengths $\{S_{i1}\}$ of the initial inter-departure intervals of the busy periods, and indicator r.v.s I_i defined by

$$I_i = \begin{cases} 1 & \text{if } N_i = 1, \\ 0 & \text{if } N_i \geq 2 \end{cases} \quad (i = 1, \dots, M).$$

Then the likelihood of the dataset $\{(S_i, I_i): i = 1, \dots, M\}$ equals

$$\prod_{i=1}^M p(\varpi | S_i)^{I_i} [1 - p(\varpi | S_i)]^{1 - I_i} \tag{5.4}$$

for the balking case and a similar expression in the reneging case (cf. Cox [2] for binary data regression approaches to such datasets). The MLE of ϖ based on this likelihood function, $\hat{\varpi}'$ say, satisfies the equation

$$\sum_{i=1}^M \frac{\lambda S_i [I_i - p(\hat{\varpi}' | S_i)]}{1 - p(\hat{\varpi}' | S_i)} = 0.$$

Now $1 - p(\hat{\varpi}' | S_i) < \lambda S_i (1 - \hat{\varpi}')$ and $\sum_i I_i = n_1 = \sum_{i=1}^M p(\bar{\varpi} | S_i)$, so it follows that

$$0 \geq \sum_{i=1}^M [I_i - p(\hat{\varpi}' | S_i)] = \sum_{i=1}^M [p(\bar{\varpi} | S_i) - p(\hat{\varpi}' | S_i)].$$

Since $p(\varpi | S)$ is monotonically increasing in ϖ , we have that $\hat{\varpi}' < \bar{\varpi}$, whenever they lie in the interior of $(0, 1)$.

In the reneging case, the MLE $\hat{\eta}'$ based on the analogue of (5.4) satisfies the equation

$$\sum_{i=1}^M \frac{\lambda S_i^2 [I_i - p(\hat{\eta}' | S_i)]}{1 - p(\hat{\eta}' | S_i)} \cdot \frac{1 - (1 + \hat{\eta}' S_i) e^{-\hat{\eta}' S_i}}{(\hat{\eta}' S_i)^2} = 0.$$

We simulated initial inter-departure intervals S_i of $M = 100$ busy periods of an $M/M/1$ system with $\lambda = 1 = \mu$, and simulated 100 random variables uniformly distributed on $(0, 1)$ so as to determine two sets of indicator variables I_i , one set determined by probabilities $p^R(\varpi | S_i)$ with balking probability $\varpi = 0.5$, and the other by probabilities $p^B(\eta | S_i)$ with reneging rate $\eta = 1$. Each of the two resulting datasets $\{(S_i, I_i): i = 1, 100\}$ was fitted to both the balking and reneging models, both by moment estimation and by maximum likelihood, and the log likelihoods evaluated. This experiment suggested two conclusions, neither of them surprising:

- 1) the moment and likelihood estimators of the loss parameters are closer when the correct loss model is used; and
- 2) the log likelihood of the correct model is generally larger than that of the incorrect model, but the margin is not necessarily large enough to discriminate between the models.

We repeated this experiment 19 times. For the log likelihood difference (True – False), we obtained values in the range (–1.08, 4.82), with median 2.02, for data generated by the balking model, and in the range (–1.10, 5.35), with median 3.20, for data from the renegeing model. For 19 such experiments with $M = 400$, the corresponding ranges and medians were (2.77, 20.16) and 8.19 (balking model), and (–1.72, 20.14) and 11.10 (renegeing model). These results suggest that the models are not distinguishable, on likelihood alone, unless M is around 100 or more.

A proper question to ask of a real dataset is whether arrivals balk or renege or both. Analysis in Daley and Servi [7] and above suggests that observation would be better than indirect inference!

6. Comparison of $\bar{\omega}$ with ML Estimator $\hat{\omega}$

We attempt in this section to compare our estimator $\bar{\omega}$ described below (3.2) with the ML estimator $\hat{\omega}$ given in Daley and Servi [5], namely, for an interval of length $T = T_I + \sum_{j=1}^M Z_j$ where there is a total idle time T_I and M busy periods of durations Z_j ($j = 1, \dots, M$) determined by N_j inter-departure intervals,

$$\hat{\omega} = 1 - \frac{\sum_j (N_j - 1)}{\sum_j Z_j} \cdot \frac{T_I}{N_I} \tag{6.1}$$

We base our comparison on the asymptotic standard deviations of the estimators.

Recall that L , the log likelihood function from which the MLE $\hat{\omega}$ is deduced, is given by

$$L - \lambda T_I + N_I \log(\lambda T_I) - \lambda(1 - \varpi) \sum_j Z_j + \sum_j (N_j - 1) \log(\lambda(1 - \varpi) \sum_j Z_j) + \text{const.},$$

where the constant term is independent of λ and ϖ . The Fisher information matrix is thus the expectation of the second derivatives, i.e., of the matrix

$$\begin{pmatrix} -\frac{N_I + \sum_j (N_j - 1)}{\lambda^2} & \sum_j Z_j \\ \sum_j Z_j & -\frac{\sum_j (N_j - 1)}{(1 - \varpi)^2} \end{pmatrix} \tag{6.2}$$

Here we are faced with a dilemma: how should we compare (asymptotic) variances based on rather contrasting data, namely $\{n_1^{BP}, S_{j1}(j = 1, \dots, M)\}$ for the moment estimator and (N_j, Z_j) ($j = 1, \dots, M$) for the MLE. We do so on the basis of expectations on the assumption that M is not small. Then each busy period, without knowing N_j , has expected duration $ES/[1 - \lambda(1 - \varpi)ES]$ where ES denotes an expected inter-departure interval, while the determinant of the matrix at (6.2) equals $EN_I E \sum_j (N_j - 1)/[\lambda(1 - \varpi)]^2$. Combining this with the expectation of (6.2) gives

$$\begin{aligned} \text{var} \hat{\varphi} &\approx \left(\frac{1}{EN_I} + \frac{1}{E \sum_j (N_j - 1)} \right) (1 - \varpi)^2, \\ &\approx \frac{1}{M} \left(1 + \frac{\lambda ES(1 - \varpi)}{1 - \lambda ES(1 - \varpi)} \right) (1 - \varpi)^2 = \frac{(1 - \varpi)^2}{M[1 - \lambda ES(1 - \varpi)]}. \end{aligned} \quad (6.3)$$

In the third of each set of five lines in Table 1, we evaluate this approximation to facilitate comparison with the approximate and exact standard deviation of $\bar{\omega}$. Neither the moment estimator nor the MLE is uniformly better.

References

- [1] Bertsimas, D.J. and Servi, L.D., Deducing queueing from transactional data: the queue inference engine revisited, *Oper. Res.* **40** (1992), Supplemental Issue 2, S217-S228.
- [2] Cox, D.R., *The Analysis of Binary Data*, Methuen, London 1970.
- [3] Daley, D.J. and Servi, L.D., Exploiting Markov chains to infer queue-length from transactional data, *J. Appl. Probab.* **29** (1992), 713-732.
- [4] Daley, D.J. and Servi, L.D., A two-point Markov chain boundary value problem, *Adv. Appl. Probab.* **25** (1993), 607-630.
- [5] Daley, D.J. and Servi, L.D., Likelihood estimation of balking and renege rates from transactional data (1996), manuscript.
- [6] Daley, D.J. and Servi, L.D., Estimating waiting times from transactional data, *ORSA J. Comp.* **9** (1997), 224-229.
- [7] Daley, D.J. and Servi, L.D., Estimating customer loss rates from transactional data, In: *Applied Probability and Stochastic Processes: A Festschrift for Julian Keilson* (ed. by J.G. Shantikumar and U. Sumita) (1998), to appear.
- [8] Hall, S., New directions in queue inference for management implementation, unpublished Ph.D. thesis, Operations Research Center, Massachusetts Institute of Technology (1992).
- [9] Jones, L. and Larson, R.C., Efficient computation of probabilities of events described by order statistics and applications to queue inference, *ORSA J. Comp.* **7** (1995), 89-100.
- [10] Kendall, M.G. and Stuart, A., *The Advanced Theory of Statistics*, Vol. 1, Charles Griffin, London 1958.
- [11] Larson, R.C., The queue inference engine: Deducing queue statistics from transactional data, *Management Sci.* **36** (1990), 586-601. Addendum **36**, 1062.