Transactional Data Inference
For Telecommunications Models

Kavitha Chandra * and Lee K. Jones **
*Center for Advanced Computation and Telecommunications
**Department of Mathematical Sciences
University of Massachusetts, Lowell, MA 01854

Abstract

Bayesian methods are presented for inferring customer behavior from transactional data in telecommunications systems. Transactional data is defined as the set of recorded times of state changes of a continuous time stochastic process. We consider two telecommunications examples: one is a K-channel radio system. When all K channels are in use new customers continuously monitor channel use (form an invisible queue) until they either gain access to an available channel or renege. The transactional data in this case is a sequence of beginning and ending times of each channel use. Customer behavior is inferred by estimating the arrival rate and renege parameter. The second example is from multimedia service systems. Suppose there are many (100 or more) satellite channels available. Customers make active inquiries by telephone in which they request a certain video. After negotiation the customer either reneges or agrees to purchase the service. In this case a start time for the video now becomes public. New customers make passive inquiries and either balk or request the service. The transactional data are the times of active inquiries leading to purchase, reneging active inquiries and purchase resulting from passive inquiries. Customer behavior is inferred by estimating an arrival rate and a combined balking-reneging parameter. The methods developed use extensions and variations of the balking Queue Inference Engine.

Keywords: transactional data, parameter estimation

1. Introduction and Preliminaries

Transactional data consists of a particular subset of the sample of times at which a discrete valued continuous time stochastic process changed state. The particular subset usually contains only the (conveniently) recorded times of certain types of state changes and is hence not a random subsample but rather a structured subsample. For instance,
consider a single automated teller machine (ATM). Arriving customers queue and then ultimately insert their cards, receive service and withdraw their cards. The associated stochastic process has values which are the number of customers in queue. The jumps occur when new customers arrive and at service completions (which are virtually simultaneously followed by service initiations when customers are in queue). The recorded data however is limited to the times of service completion and initiation and does not include all of the arrival times. The problem is then to infer properties of the arrival process with this incomplete information.

Looking at a sequence of transaction times we see certain periods when the system is congested. These begin right after a recorded service initiation that does not immediately follow a service completion. Then there is a sequence of completions immediately followed by initiations and then the congestion period ends with a completion not immediately followed by an initiation. It is in these congestion periods when queuing occurs. Between congestion periods there is no queuing so that the service initiation just before the start of a congestion period coincides with a customer arrival. However, during congestion the arrival times are not recorded. The same construction of congestion periods holds with K ATM's (but there may be more than one recorded arrival between congestion periods).

Larson [1] analyzed such transactional data and gave an algorithm (The Queue Inference Engine) for estimating expected queue length in a given congestion period. He assumed a homogeneous Poisson arrival process but his estimates did not require knowledge of the arrival parameter. Servi and Daley [2] outline methods for estimating queue lengths with homogeneous Poisson arrivals but where the customers were allowed to (a) renge with a constant hazard rate or (b) balk with a constant probability when the queue length exceeds a prescribed threshold. Their estimates depended on the arrival and balking / reneging parameters. Jones [3] then gave an estimation algorithm for more general than Poisson arrival processes and arbitrary balking process. The balking process was quantified by a sequence of probabilities of abandoning a queue as a function of queue length at arrival. Also a mixed likelihood method was developed for estimating the arrival rate and balking parameter for the homogeneous Poisson case with a parametric balking process. This likelihood function was a product of two likelihood functions. The first was a function of only the recorded arrivals during non-congestion and the second was the probability of congestion persisting in the given congestion periods given the transaction times with their periods.

These likelihood methods readily apply to any other balking system like fast food restaurants or car washes etc. In this paper we apply the method to the transactional data problems from telecommunications which involve reneging as a function of time spent in queue or balking as a function of the announced time until service commences.
2.0 A K-Channel Radio System with Reneging

Whenever all K channels are busy potential users continuously monitor channel use and attempt to acquire a channel as soon as one becomes free. Hence an invisible queue is formed and as soon as one of the K communications is terminated a (randomly chosen) queued customer is able to "lock in" the available channel. This is Example 2 of [1].

Let us assume that the arriving customers have unknown constant Poisson rate λ but that each waiting customer has an unknown constant hazard rate η of leaving the system. (The probability of leaving during an interval of length δt is \(1 - e^{-\eta \delta t}\).) Now, by monitoring the channels, the telecommunications company is able to record the beginning and end of customer communications. As in the ATM case, congestion periods are identified by an initial channel use followed by a sequence of service completions immediately followed by service initiations and ending with a service completion not immediately followed by a service initiation. Using the algorithms and notation of [2] we obtain: The number of Poisson arrivals in an interval of length \(S_r\) with a mean of \(\lambda S_r\).

The probability of \(n\) arrivals in the interval is given by

\[
K(\mu, n) = e^{-\mu} \frac{\mu^n}{n!}
\]

(1)

where \(\mu = \lambda S_r\). If one considers reneging with a hazard rate of η, the probability that exactly \(n\) of the arrivals remain at the end of the interval is shown in [2] to be Eqn. (1) with a mean value

\[
\mu = \lambda \frac{1 - e^{-\eta S_r}}{\eta}
\]

(2)

\(S_r = t_{r+1} - t_r\)

Then the one step transition probabilities* can be shown [2] to be

\[
[Q(r)]_{i,j} = \min(i-1,j) \sum_{l=0}^{\min(i-1,j)} K(\mu, j-l) \left(1 - e^{-\eta S_r} \right)^l \left[1 - e^{-\eta S_r} \right]^{i-l-1}
\]

(3)

The congestion starts at transition time \(t_0\) and ends at time \(t_n\). Here, the state of the system at time \(t_r\) represents the number of users trying to gain access just prior to time \(t_r\). The probability (given state \(i\) at time \(t_{r_1}\)) of ending in state \(j\) at time \(t_{r_2}\) is given by \((i, j\text{-th})\) element of the matrix,

* From \(t_r\) to \(t_{r+1}\)
\[ P_{r_1r_2} = \prod_{l=r_1}^{r_2-1} Q(l) \]  

The probability of congestion is obtained by taking \( i=1, j=0, r_1=0 \) and \( r_2=n \).

From the aforementioned equations we can compute a probability of congestion persisting for each congestion period. The product of these probabilities over all congestion periods is our congestion likelihood and is a function of the unknown \( \lambda \) and \( \eta \). The likelihood of the non-congestion data is just the product of exponential densities evaluated at \( N \) inter-arrival times \( \delta t_i \) between congestion periods. It has the form,

\[ L_{\text{non-congestion}} = \lambda^N e^{-\lambda \sum_{i=1}^{N} \delta t_i} \]  

As a numerical example, with \( K=1 \) we took simulated data involving 10 congestion periods with \( \lambda=150 \) and \( \eta=25 \) and calculated the likelihoods for varying \( \lambda \) and \( \eta \). Figure (1) shows the maximum likelihood results obtained for \( \lambda \) when \( \eta \) was assumed known. These likelihoods may be interpreted as follows. First put a constant prior distribution on the unknown parameter \( \lambda \). This corresponds to admitting that all \( \lambda \) values are equally likely prior to data acquisition. Now the transactional data are processed and the product of the congestion probabilities as well as the product of the non-congestion probabilities are plotted for each \( \lambda \). The results shown are the posterior distributions using non-congestion data, congestion data and both. Note the reduction in variance in the parameter value when comparing non-congestion posteriors to with-congestion posteriors. In Fig. (2), the likelihood function for estimating \( \eta \) is shown, assuming \( \lambda \) to be known. The posterior for \( \eta \) using the congestion probabilities is shown. The likelihood function is seen to be a maximum at the value of \( \eta \) used in the simulation. Again the graph may be interpreted as a posterior density for \( \eta \) given a uniform prior on \( \eta \). Inferring the joint posterior for both \( \lambda \) and \( \eta \) will be a subject of future research.

3.0 Video on Demand

We now give a simple example of video on demand transactional data where the likelihoods may be easily derived without needing sophisticated algorithms as in [2] and [3]. We anticipate that the techniques may be expanded to include more complicated systems.

Let us consider requests for one particular video. The sequence of inquiries and requests can be explained with respect to Figure (3). The time \( t_0 \) represents the start of records. At \( t_1 \) a potential customer who arrives at a Poisson rate \( \lambda \) inquires via telephone if they
will schedule this video and how long he must wait. Suppose the delay in scheduling the video is $\Delta$. Assume customers balk at waiting the time $\Delta$ with a probability $1-e^{-\alpha \Delta}$. So this first customer either purchases the service, with probability $e^{-\alpha \Delta}$ or balks. Assume he balks. An arriving customer at $t_2$ purchases the service. The video would then be scheduled at $p_1=t_2+\Delta$. The video offering at $p_1$ is now made public (until $p_1$). So new customers arriving at $t_2 < t \leq p_1$ would make passive inquiries balking with probability $1-e^{-\alpha (p_1-t)}$ or purchase otherwise. However because the queries in this period are passive only a decision to purchase will be recorded. Thus we call the period from $t_2$ until $p_1$ a congestion period. Times of purchase within this period are now inhomogeneous Poisson with rate

$$\hat{\lambda}(t) = \lambda e^{-\alpha (p_1-t)}$$  \hspace{1cm} (6)

After $p_1$ we have non-congestion until some customer makes an active inquiry and decides to purchase. The associated likelihood functions for the illustrated sequence in Figure (3) is given below. The likelihood function for the non-congestion periods are given by products of terms of the form,

$$L_{\text{non-congestion}} = \begin{cases} \lambda e^{-\lambda(t_{1-t_0})}(1-e^{-\alpha \Delta}) \lambda e^{-\lambda(t_{2-t_1})} e^{-\alpha \Delta} & \\
\lambda e^{-\lambda(t_{3-t_2})}(1-e^{-\alpha \Delta}) \lambda e^{-\lambda(t_{6-t_5})} e^{-\alpha \Delta} & \\
\end{cases}$$  \hspace{1cm} (7)

The corresponding function for the congestion periods is given by products of the form,

$$L_{\text{congestion}} = \begin{cases} \lambda e^{-\alpha (p_1-t_3)} e^{-\lambda \alpha}[e^{-\alpha (p_1-t_3)} e^{-\alpha (p_1-t_4)}] & \\
\lambda e^{-\alpha (p_1-t_4)} e^{-\lambda \alpha}[e^{-\alpha (p_1-t_3)} e^{-\alpha (p_1-t_4)}] & \\
\lambda e^{-\alpha (p_2-t_5)} e^{-\lambda \alpha}[e^{-\alpha (p_2-t_5)} e^{-\alpha (p_2-t_6)}] & \\
\lambda e^{-\alpha (p_2-t_6)} e^{-\lambda \alpha}[e^{-\alpha (p_2-t_5)} e^{-\alpha (p_2-t_6)}] & \\
\end{cases}$$  \hspace{1cm} (8)

and the product likelihood function

$$L = L_{\text{congestion}} \cdot L_{\text{non-congestion}}$$  \hspace{1cm} (9)

The simulation was carried out using $\lambda=50$ and balking parameter $\alpha=10$. The value of $\Delta=0.1$ time units was used. A total of 30 congestion and non-congestion periods was considered in the calculation of the likelihood functions. Fig. (4) demonstrates the behavior of the likelihood functions obtained by assuming $\alpha$ and varying $\lambda$. The
likelihood functions denoted by $L_{uc}(\lambda)$, $L_{c}(\lambda)$ and $L(\lambda)$ have a mean value and variance of ($L_{uc}$: $\{49.1, 31.7\}$; $L_{c}$: $\{44.8, 23.6\}$; $L$: $\{46.5, 13.5\}$) respectively. It can be seen that the variance in the parameter estimate substantially decreases when one considers the congestion probabilities. This feature is also evident in Fig. (5) which demonstrates the maximum likelihood estimation of $\alpha$ assuming $\lambda=50$. The corresponding values of mean and variance are ($L_{uc}$: $\{9.4, 2.1\}$; $L_{c}$: $\{13.3, 5.04\}$; $L$: $\{10.6, 1.6\}$) respectively.

Acknowledgements: This research was supported in part by NSF Grant No. DMS-9202161 (LKJ) and by the Department of Energy (KC).

References


Fig. 1: The posterior distributions for $\lambda$ assuming the reneging parameter $\eta$ to be known. $L_{uc}$: likelihood function using non-congestion probabilities; $L_c$: congestion probabilities; $L = L_{uc} \cdot L_c$.

Fig. 2: The likelihood function for the reneging parameter using congestion probabilities.
Figure 3

t0: Start of Records;
t1, t5: Active inquiry (Balks);
t2, t6: Active inquiry (requests service);
t3, t4, t7: Purchase from public scheduling;
p1, p2: Video begin times.

Likelihood Functions for Estimating Arrival Rate

Fig. 4: The posterior distributions for $\lambda$ assuming the balking parameter $\alpha$ to be known.

Likelihood Functions for Estimating Balking Parameter

Fig. 5: The posterior distributions for $\alpha$ assuming the arrival rate $\lambda$ to be known.