Multifactor Poisson and Gamma-Poisson models for call center arrival times

Lawrence Brown and Linda Zhao [1]

Department of Statistics

University of Pennsylvania

Haipeng Shen

Department of Statistics and Operations Research

University of North Carolina at Chapel Hill

Avishai Mandelaum

Department of Industrial Engineering and Management

Technion University

Draft

January 26, 2004

[1]Research supported in part by NSF grant and the Wharton Financial Institutions Center.

Abstract

It is usually assumed that the arrivals to a queue will follow a Poisson process. In its simplest form, such a process has a constant arrival rate. However this assumption is not always valid in practice. We develop statistical procedures to test a stochastic process is an inhomogeneous Poisson process, and show that call arrivals to a real-life call center follow such a Poisson process with an inhomogeneous arrival rate over time. We find that the inhomogeneous Poisson assumption is reasonably well satisfied. Then we derive statistical models that can be used to construct predictions of the inhomogeneous arrival rate, and provide estimates of the parameters in the models. The conclusion that the process is well modelled as an inhomogeneous Poisson process, together with a statistical model for the arrival rate in that process, could be used to enable realistic calculations or simulations of the performance of the queuing system. The models can also be used to predict future call volumes or workloads to the system.

1 Introduction

Call centers allow groups of agents to serve customers remotely via the telephone. They have become a primary contact point between customers and their service providers and, as such, play an increasingly significant role in more developed economies. While call centers are technology-intensive operations, often 70% or more of their operating costs are devoted to human resources, and to minimize costs their managers carefully track and seek to maximize agent utilization. Well-run call centers adhere to a sharply-defined balance between agent efficiency (measured by utilization level) and service quality (measured by the waiting time of calls). To achieve the balance, they use queueing-theoretic models. One of the key inputs to these mathematical models is the rate at which call arrives. As noted in Jongbloed and Koole (2001) and Gans, Koole and Mandelbaum (2003), uncertainty in call volumes is one of the main problems in managing a call center.

In this article we propose to model the arrival process to a call center as an inhomogeneous Poisson process. A testing procedure is also developed to verify the proposal. The process has a underlying smoothly varying rate function, λ, which depends on some covariates in the data like date, time of day, type and priority of the calls.

To make things a little complicated, the rate function λ is a hidden function that is not observed. That is, λ is not functionally determined by date, time of day and call-type information. This phenomenon is also observed before by Jongbloed and Koole (2001). Due to this feature, we develop models to predict statistically the number of arrivals as a function of only those repetitive features. The methods also allows us to construct confidence or prediction bounds in addition to predictions of future call volumes.

Call-arrival data gathered at an Israeli call center is used as motivation and illustration of the various problems and methodologies we discuss. We provide a very brief discussion in Section 2 of this data application.

The rest of the paper is organized as follows. In Section 2 we describe the call center data we will use as an example of an application of our methodology. Section 3 proposes a methodology to test whether a stochastic process can be modelled as an inhomogeneous Poisson process. The method is also illustrated on the Israeli call center data. After verifying the call arrival process is an inhomogeneous Poisson process, Section 4 introduces three different models to estimate the underlying Poisson arrival rate. The models can also be used to predict future call volumes. We conclude this article with a case study in Section 5, where our proposed tests and models are illustrated on arrival data from an Israeli call center.

2 Call Center Arrival Data

The data accompanying our study was gathered at a relatively small Israeli bank telephone call center in 1999. The portion of the data of interest to us here involves records of the arrival time of service-request calls to the center. These are calls in which the caller requests service from a call center representative. It is reasonable to conjecture that these arrival times are well modelled by an inhomogeneous Poisson process, as we will verify later in Section 3. The arrival rate for this process should depend on time of day, and perhaps other calendar related covariates such as month or day of the week. There are different categories of service that may be requested, and preliminary analysis clearly shows that this factor should also be considered since the arrival rate patterns differ considerably. For more information about various aspects of this data see Mandelbaum, Sakov and Zeltyn (2000). Different features of the data are investigated rather broadly by Brown, Gans, Mandelbaum, Sakov, Shen, Zeltyn and Zhao (2003).

3 Arrivals are inhomogeneous Poisson

Comments from Haipeng: We would like to describe a procedure that one can use to show a stochastic process is a Poisson process. To achieve that, we first show the arrivals do not depend on the exact time clock, then the arrivals are shown to be independent with each other. Finally we show that the inter-arrival times are exponentially distributed. The procedure is illustrated using the Israeli data.

Finally I added a section to show how the Poisson arrival rates are not easily "predictable", to motivate the notion of modelling the arrival rates randomly. Basically this section was in our JASA paper, but was took out due to the length requirements.

Classical theoretical models posit that arrivals form a Poisson process. It is well known that such a process results from the following behavior: there exist many potential, statistically identical callers to the call center; there is a very small yet non-negligible probability for each of them calling at any given minute, say, so that the average number of calls arriving within a minute is moderate; and callers decide whether or not to call independently of each other.

Common call-center practice assumes that the arrival process is Poisson with a rate that remains constant for blocks of time, often individual quarter-hours, half-hours or hours. A call center manager will then fit a separate queueing model for each block of time, and estimated performance measures may shift abruptly from one interval to the next.

A more natural model for capturing both stochastic and operational levels of detail is a time inhomogeneous Poisson process. Such a process is the result of time-varying probabilities that individual customers call, and it is completely characterized by its arrival-rate function λ. Knowledge that the arrivals follow an inhomogeneous Poisson process will be of use if it turns out that the arrival rate varies relatively slowly in time. Smooth variation of this sort is familiar in both theory and practice in a wide variety of contexts, and seems reasonable in call centers.

To be precise, we define an inhomogeneous Poisson process on $[0, \infty)$ as follows. Let $t \in [0,\ \infty)$. Let $\lambda(t)$ denote the arrival rate, assumed continuous for simplicity, and let $\Lambda(t) = \int_0^t \lambda(\tau)d\tau$. Let $\nu(t)$ denote the cumulative count process for a standard Poisson process with arrival rate $= 1$. Thus $P\{\nu(t) \leq k\} = e^{-k}/k! = Poiss(k;1)$. Then the cumulative number of counts for the inhomogeneous process is $N(t) = \nu(\Lambda(t))$. If $\lambda(t)$ is approximately a constant, λ, over some interval, say $(a,\ b)$ then $N(b) - N(a) \approx Poiss(\cdot, \lambda)$. The assumption that N is such a process can thus be tested by looking at short time intervals, estimating λ over the interval, and testing whether $N(t)$ follows a Poisson process law with constant rate λ over the interval. The constant, λ, in this description may depend on the interval.

We construct below two tests of call center arrival processes designed to determine whether the process is inhomogeneous Poisson with such a slowly varying arrival rate.

3.1 Testing no time dependence

Our first test is intended to discover whether there may be dependence on the exact clock time within a given short clock interval over a period of many days. (There seems no a-priori reason why such dependence should exist in our data; it might in principle occur if several customers had automatic phone systems designed to call the bank at the same time on every day.) Choose a given relatively short time interval over which $\lambda(t)$ can be presumed nearly constant on any given day. If $\lambda(t) = \lambda_{date}$ is a constant over this interval on each day, then the counts within this time interval over many days will be (approximately) uniformly distributed as a function of the clock time. This leads to a test of the null hypothesis.

We illustrate this test by analyzing arrivals on regular workdays from 8/1 through 12/31 during the time interval from 11:15 am to 11:30 am. (We have chosen for relatively arbitrary reasons to analyze only this 5-month period, rather than the whole year. Analysis of the entire year's data yielded qualitatively similar results.)

There were 3544 calls arriving to the queue during this period on such days. The histogram of these calls as a function of arrival time is shown in Figure 1.

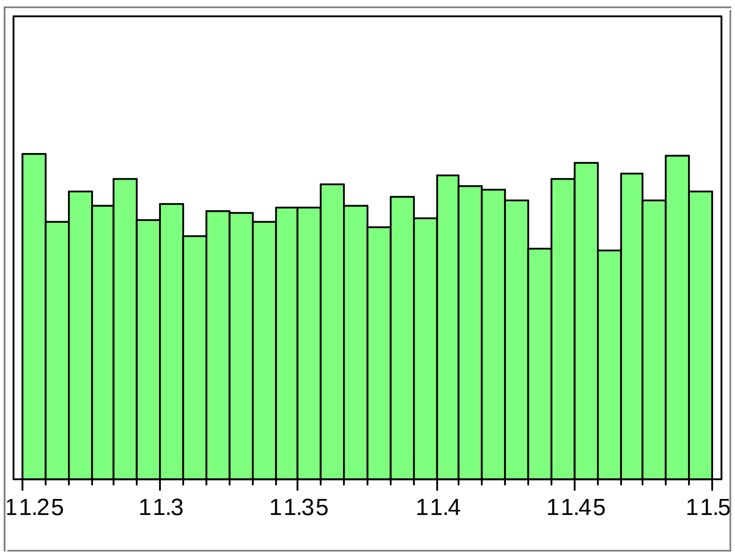

Figure 1: Histogram of call arrival time, for regular weekday calls from 8/1 through 12/31 between 11:15am and 11:30am. (Time is indicated in decimals of an hour, so $11.25 = 11:15$am.)

This histogram has a roughly uniform appearance. Standard tests of uniformity can be applied to this data. For example, the Chi-squared test of uniformity based on the 30 bins shown in the histogram is based on

$$\chi^2 = \sum \frac{(observed\,bin\,count - expected\,bin\,count)^2}{expected\,bin\,count},$$

where *expected bin count* =3544/30. This test statistic has 29 df under the null hypothesis. For the data shown in Figure 1 we computed $\chi^2 = 26.4$ and P-value = .60.

Alternatively, one could apply a one-sample Kolmogorov-Smirnov test of uniformity, based on the statistic

$$KS = \sup\left(\sqrt{n}\left|\hat{F}_n(t) - F(t)\right|\right),$$

where $\hat{F}_n$ denotes the sample cumulative distribution function. For the data shown in Figure 1 we computed $KS = 0.6727$ and P-value > 0.5.

We computed similar χ^2 and K-S tests for other 15-minute and 5-minute intervals. Most were not significant, as above. A few showed modestly significant results, with the most noticeable of these being for the interval from 2pm to 2:15pm, which had a χ^2 value of 57.3 on 29 df (P-value ≈ 0.001) and a similarly significant K-S value.

For the sake of completeness, Figure 2 shows the histogram for the data between 2pm and 2:15pm. As far as we can see there is no special pattern here to describe the deviations from the assumed uniformity of the null hypothesis:

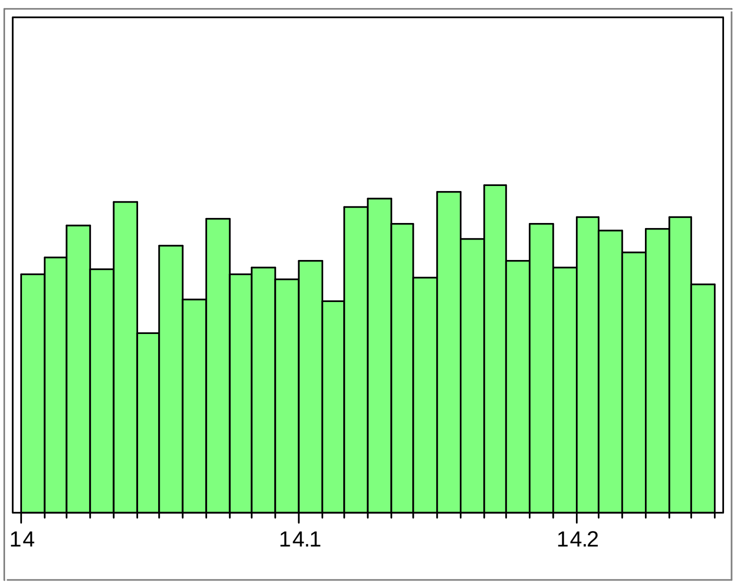

Figure 2: Histogram of call arrival time, for regular weekday calls from 8/1 through 12/31 between 2pm and 2:15pm. (Time is indicated in decimals of an hour, so 14.1 = 2:06pm.)

3.2 Testing no serial dependence

Comments from Haipeng: Are we trying to establish the independence of the arrival times here? If so, can we achieve that by showing the various transformations of the inter-arrival times are uncorrelated? I tried to put in some words below. Please feel free to modify.

Another possible violation of the Poisson process assumption would involve dependence among the call arrival times. We can establish the independence by looking at correlations of various transformations of the inter-arrival times.

3.3 Testing exponentiality of the inter-arrival times

Comments from Haipeng: How does the procedure in this section differ from the one we used in the JASA paper? If the same, should we stick with the current presentation or switch to the one in JASA?

Given that we can show that the arrivals do not depend on the exact clock and are independent among themselves, the Poissonaity of the arrival process can be investigated by examining the distribution of inter-arrival times. For a homogeneous Poisson process with given λ these inter-arrival times have an exponential distribution with scale $1/\lambda$. Thus we can expect the same to be approximately true if we examine inter-arrivals for our data over suitably short time intervals. However, the value of λ then depends on the interval, and care must be taken to properly normalize the data to take this into account. A description of our test procedure follows.

We chose to examine a particular (short) time interval over the period of regular workdays between 8/1 to 12/31. Let the interval be (a, b) of time length $b-a$. The analysis below shows results for the interval between 10am and 10:09am. (This was another interesting interval from the perspective of the Chi-square analysis described in Section 3.1. The computed value was $\chi^2 = 53.36$, where we again used 29 df. This corresponds to a P-value of 0.004.)

Let λ_{day} denote the latent Poisson rate over the given interval on the indicated day. (λ_{day} is assumed to be constant over the interval on each day.) Let $T_{day,j}$ denote the time of the j^{th} arrival on the indicated day. Set $T_{day,0} = a$, the beginning of the interval. Let $G_{day,j} = T_{day,j} - T_{day,j-1}$, $j = 1, \ldots J_{day}$ where J_{day} denotes the number of calls during the interval on that day. We define

$$H_{day,j} = \frac{J_{day} G_{day,j}}{b-a}.$$

Under the null hypothesis that the process is homogeneous Poisson over each interval (with a rate that depends on the interval) the values of $H_{day,j}$ will all be approximately exponentially distributed with rate = 1. This will be a good approximation so long as the values of J_{day} are not small. The values of $H_{day,j}$ will also be approximately independent. Hence the null hypothesis can be judged by applying a standard test to assess whether the observed values of H are exponential with rate 1. We chose to display the data on an exponential distribution (rate = 1) Q-Q plot and to use the corresponding K-S test for an exponential (1) distribution.

Figure 3 shows the Q-Q plot for the $\{H\}$ over the interval in question. The plot exhibits good visual fidelity to the 45^o line corresponding to the null hypothesis. The P-value of the K-S test here was 0.150. There is no significant evidence in this test that the process is not an inhomogeneous Poisson process. The result here is fairly typical of those obtained from the data for other time intervals. (The majority of those intervals had even larger P-values.)

3.4 The Poisson arrival rates are not easily "predictable"

The inhomogeneous Poisson process described above provides a stochastic regularity that can sometimes be exploited. However, this regularity is most valuable if the arrival rates are known, or can be predicted on the basis of observable covariates. The current section

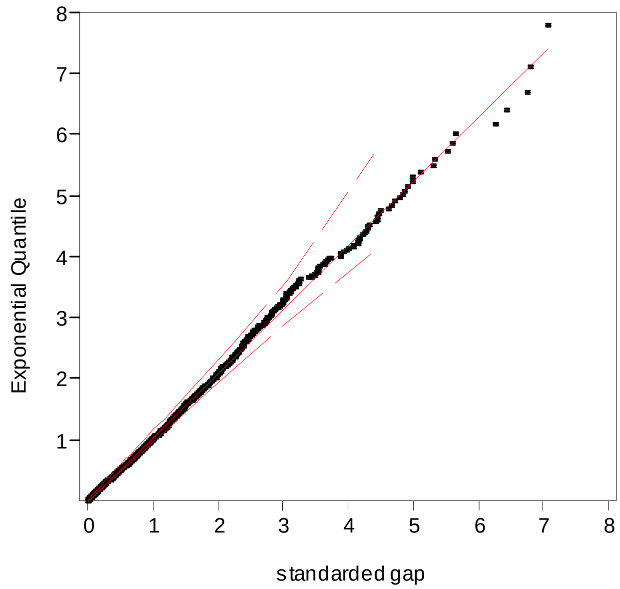

Figure 3: Exponential distribution Q-Q plot for the values of H over the interval from 10am to 10:09am on weekdays 8/1 – 12/31.

examines the hypothesis that the Poisson rates can be written as a function of the available covariates: call type, time-of-day and day-of-the-week. If this were the case, then these covariates could be used to provide valid stochastic predictions for the numbers of arrivals. But, as we now show, this is not the case.

The null hypothesis to be tested is, therefore, that the Poisson arrival rate is a (possibly unknown) function $\lambda_{type}(d,t)$, where $type \in \{PS, NE, NW, IN\}$ may be any one of the types of customers, $d \in \{\text{Sunday}, \ldots, \text{Thursday}\}$ is the day of the week, and $t \in [7, 24]$ is the time of day. For this discussion, let Δ be a specified calendar date (e.g. November 7th), and let $N_{type,\Delta}$ denote the observed number of calls requesting service of the given type on the specified date. Then, under the null hypothesis, the $N_{type,\Delta}$ are independent Poisson variables with respective parameters

$$\mathrm{E}[N_{type,\Delta}] = \int_{7}^{24} \lambda_{type}(d(\Delta), t)dt \, , \tag{1}$$

where $d(\Delta)$ denotes the day-of-week corresponding to the given date.

Under the null hypothesis, each set of samples for a given type and day-of-week, $\{N_{type,\Delta} \,|\, d(\Delta) = d\}$, should consist of independent draws from a common Poisson random variable. If so, then one would expect the sample variance to be approximately the same as the sample mean. For example, see Agresti (1990) and Jongbloed and Koole (2001) for possible tests.

Table 1 gives some summary statistics for the observed values of $\{N_{NE,\Delta} \,|\, d(\Delta) = d\}$ for weekdays in November and December. Note that there were 8 Sundays and 9 each of Monday through Thursday during this period. A glance at the data suggests that the $\{N_{NE,\Delta} \,|\, d(\Delta) = d\}$ are not samples from Poisson distributions. For example, the sample mean for Sunday is 163.38, and the sample variance is 475.41. This observation can be validated by a formal test procedure, as described in the following paragraphs.

Day-of-Week (d)	# of Dates (n_d)	Mean	Variance	Test Statistic (V_d)	P-value
Sunday	8	163.38	475.41	20.41	0.0047
Monday	9	188.78	1052.44	42.26	0.0000
Tuesday	9	199.67	904.00	38.64	0.0000
Wednesday	9	185.00	484.00	21.23	0.0066
Thursday	9	183.89	318.61	13.53	0.0947
ALL	44			$V_{all} = 136.07$	0.00005

Table 1: Summary statistics for observed values of $N_{NE,\Delta}$, weekdays in Nov. and Dec.

Brown and Zhao (2002) present a convenient test for fit to an assumption of independent Poisson variables. This is the test employed below. The background for this test is Anscombe's variance stabilizing transformation for the Poisson distribution (Anscombe, 1948).

To apply this test to the variables $\{N_{NE,\Delta} \,|\, d(\Delta) = d\}$, calculate the test statistic

$$V_d \;=\; 4 \;\;\times \sum_{\{\Delta \,|\, d(\Delta)=d\}} \left(\sqrt{N_{NE,\Delta} + 3/8} \;\;-\;\; \frac{1}{n_d} \sum_{\{\Delta \,|\, d(\Delta)=d\}} \sqrt{N_{NE,\Delta} + 3/8} \right)^2 ,$$

where n_d denotes the number of dates satisfying $d(\Delta) = d$. Under the null hypothesis that the $\{N_{NE,\Delta} \,|\, d(\Delta) = d\}$ are independent identically distributed Poisson variables, the statistic V_d has very nearly a Chi-squared distribution with $(n_d - 1)$ degrees of freedom. The null hypothesis should be rejected for large values of V_d. Table 1 gives the values of V_d for each value of d, along with the P-values for the respective tests. Note that for these five separate tests the null hypothesis is decisively rejected for all but the value $d =$ Thursday.

It is also possible to use the $\{V_d\}$ to construct a test of the pooled hypothesis that the $N_{NE,\Delta}$ are independent Poisson variables with parameters that depend only on $d(\Delta)$. This test uses $V_{all} = \sum V_d$. Under the null hypothesis this will have (very nearly) a Chi squared distribution with $\sum(n_d - 1)$ degrees of freedom, and the null hypothesis should be rejected for large values of V_{all}. The last row of Table 1 includes the value of V_{all}, and the P-value is less than or equal to 0.00005.

The qualitative pattern observed in Table 1 is fairly typical of those observed for various types of calls, over various periods of time. For example, a similar set of tests for type NW for November and December yields one non-significant P-value (P $= 0.2$ for $d(\Delta) =$ Sunday), and the remaining P-values are vanishingly small. A similar test for type PS (Regular) yields all vanishingly small P-values.

The tests above can also be used on time spans other than full days. For example, we have constructed similar tests for PS calls between 7am and 8am on weekdays in November and December. (A rationale for such an investigation would be a theory that early morning calls – before 8am – arrive in a more predictable fashion than those later in the day.) All of the test statistics are extremely highly significant: for example the value of V_{all} is 143 on 39 degrees of freedom. Again, the P-value is less than or equal to 0.00005.

In summary, we saw in Sections 3.1 to 3.3 that, for a given customer type, arrivals are inhomogeneous Poisson, with rates that depend on time of day as well as on other possible covariates. In Section 3.4 an attempt was made to characterize the exact form of this dependence, but ultimately the hypothesis was rejected that the Poisson rate was a function only of these covariates. For the operation of the call center it is desirable to have predictions, along with confidence bands, for this rate. We return to this issue in Section 4.

4 Modelling the Poisson rate

It is important to build a stochastic model for the arrival rate. A model of this type can be used to stochastically predict arrival patterns. Secondarily, it can also be used to more accurately identify whether current arrivals are consistent with previous experience, or whether they represent changes in the operational environment of the telephone service system. Our goal in this section is to present a suite of models for inhomogeneous Poisson arrival rates, and to describe the calculations needed to estimate the parameters in those models.

Jongbloed and Koole (2001) suggest a compound Gamma-Poisson model in a telephone context similar to ours in order to model the distribution of a one-way collection of counts like $\{N_{jk}: j = 1, \ldots, J\}$. Again N_{jk} denotes the number of counts on date j over a (short) time interval indexed by k. As they note, such models are a familiar tool in other areas of statistical applications. See for example Agresti (1990, problems 3.16-3.17). We first extend this idea in order to build a two-way fixed-effects model for the collection of counts $\{N_{jk}: j = 1, \ldots, J;\ k = 1, \ldots, K\}$. We then use a "square-root" transform to convert this model into an even easier to analyze and interpret Gaussian two-way model and fit the corresponding parameters. In order to improve the prediction ability of our model, we finally introduce an auto-regressive structure into the two-way Gaussian model in order to capture the intra-day dependence. Brown et al. (2003) suggest that there exists significant dependence between arrival counts on successive days.

4.1 Model 1: Gamma-Poisson models

Define the usual gamma distribution $\Gamma(r,s)$ to have density

$$f(x) = \frac{x^{r-1}e^{-x/s}}{\Gamma(r)s^r} \quad for \;\; x \in (0,\infty).$$

Note that in this parametrization $E(X) = rs$ and $Var(X) = rs^2 = sE(X)$. Also recall that if $X_1, \ldots, X_n$ are independent $\Gamma(r_1,s), \ldots, \Gamma(r_n,s)$ then $\sum X_j \sim \Gamma(\sum r_j, s)$. To define the two-way model, let Λ_{jk} be independent $\Gamma(\mu_{jk}/s, s)$ random variables with $\mu_{jk} = \alpha_j\beta_k$ and let

$$N_{jk} \sim Poiss(\Lambda_{jk}), \;\; \text{independent}, \;\; j = 1, \ldots, J, k = 1, \ldots, K.$$

In this model the variables Λ_{jk} are unobserved, latent variables with $E(\Lambda_{jk}) = \mu_{jk}$, $Var(\Lambda_{jk}) = \mu_{jk}s$. The parameters are $\alpha_1, \ldots, \alpha_J, \beta_1, \ldots, \beta_K$, and s. A feature of this model is that the N_{jk} are independent Gamma-Poisson variables. To correspond to this, we use the notation $N_{jk} \sim \Gamma\text{-P}(\mu_{jk}/s, s)$. (Alternatively – as is well-known – we can write the N_{jk} as Negative-binomial(p, ζ_{jk}) variables with $p = (1+s)^{-1}$ and index $\zeta_{jk} = \mu_{jk}/s$.) The model has the property that the marginal totals are also Gamma-Poisson. That is

$$N_{j+} \equiv \sum_k N_{jk} \sim \Gamma\text{-P}(\frac{\alpha_j\beta_+}{s}, s) \text{ where } \beta_+ = \sum \beta_k$$

and

$$N_{+k} \equiv \sum_j N_{jk} \sim \Gamma\text{-P}(\frac{\alpha_+\beta_k}{s}, s) \text{ where } \alpha_+ = \sum \alpha_j.$$

The parameters of this model can be numerically estimated by maximum-likelihood. See Jongbloed and Koole (2001) for some relevant remarks. We have done so, and the results are briefly reported in Section 5.

4.2 Model 2: Square-root Gaussian model

We wish to concentrate in this section on results from a closely related model that is even easier to fit, and for which standard regression diagnostics, tests, and prediction methods can be applied.

Our second, related, model is motivated by the fact that if $N \sim \Gamma\text{-P}(\theta/s, s)$ then $\sqrt{N+1/4}$ has approximately a normal distribution with mean $\sqrt{\theta}$ and variance $(1+s)/4$. Note that in this approximation the variance does not depend on θ. Concerning the mean a more precise statement is that the following approximation is very nearly an equality for small to moderate s, even for rather small values of θ,

$$\left[E(\sqrt{N+1/4})\right]^2 \approx \theta. \tag{2}$$

See Brown, Zhang and Zhao (2001) and Brown and Zhao (2002) for further comments about this approximation, including remarks about the choice here of the constant 1/4 under the square root sign. Hence we define

$$X_{jk} = \sqrt{N_{jk}+1/4},\ \rho_j = \sqrt{\alpha_j},\ \kappa_k = \sqrt{\beta_k},\ \sigma^2 = \frac{1+s}{4},$$

and we model the X_{jk} as independent normal variables with mean $\rho_j\kappa_k$ and variance σ^2. This is a multiplicative Gaussian model and the maximum likelihood estimates of the parameters can be obtained by a simple non-linear least squares routine. (Iteratively reweighted least squares provides an easy scheme. Fix initial $\{\rho_j\}$ and fit $\{\kappa_k\}$ by ordinary least squares, conditional on the given $\{\rho_j\}$. Then proceed similarly to fit $\{\rho_j\}$ given the $\{\kappa_k\}$ from that fit, and iterate the process until it converges. This converges within a few iterative steps as we fit the model to our data and the results are shown in Section 5.)

Either of the two above models are over-parameterized in the sense that the $J+K$ quantities $\{\alpha_j, \beta_k\}$ contain only $J+K-1$ independent parameters. One side condition needs to be imposed in order to get unique estimates, and so we assume $\sum \beta_k = 1 = \sum \kappa_k^2$. In this way the estimated values of β_k (or κ_k^2) become estimates of the conditional density of the number of observations at time interval k on day j given the overall volume N_{j+} on that day. This also makes the multiplicative structures in the models rather natural. In symbols, under Model 1,

$$\beta_k = E\left(\frac{N_{jk}}{N_{j+}}\middle| N_{j+}\right). \tag{3}$$

When Model 2 is used this expression is also very nearly correct in view of (2).

Both of the previous models are related to Quasi-likelihood solutions for a suitable Generalized Linear Model; but they are not the same. See, for example, McCullagh and Nelder (1989) or Agresti (1990, p457). As we will show in Section 5, these two models yield very similar results while some standard model diagnostic and prediction techniques can be easily implemented under the Gaussian model. Another advantage of using the Gaussian framework is that one can easily introduce a time series structure into the model to improve the forecasting. We will do so in Section 4.3

The constant 1/4 in (2) is the best choice if N is a Poisson(λ) variable. If $N \sim \Gamma\text{-P}(\lambda, s)$ as is the case in Model 2, then the best asymptotic choice of C is $C = (1+s)/4$ since a routine expansion yields

$$E_\lambda\left(\sqrt{N+C}\right) = \sqrt{\lambda} + \frac{1}{2\sqrt{\lambda}}\left(C - \frac{1+s}{4}\right) + O\left(\frac{1}{\lambda^{3/2}}\right).$$

For illustration purpose, Model 2 uses $C = 1/4$, which will work well if it turns out that s is small. For situations where $\hat{s}$ turns out to be large, we recommend redefining $X_{jk} = \sqrt{N_{jk} + C_{new}}$ where $C_{new} = (1+\hat{s})/4$. Then new estimates of the parameters $\{\alpha_j, \beta_k, s\}$ should be computed. If necessary, this process could be iterated, but we doubt that more than one iteration should be required for satisfactory numerical accuracy in ordinary applications. We will also show results from this modification in Section 5.

4.3 Model 3: Square-root Gaussian model with an AR structure

4.4 On connections with queueing theory applications

Comments from Avi: One should elaborate on the difference between the present arrival-process model (the arrival rate of the Poisson process is random) and standard queueing theory assumptions (Poisson arrival rate is either a constant, or a deterministic function in the case of time-inhomogeneous Poisson process) should be specified.

An approach to the calculation of performance measures (e.g. average waiting time) should be, at least, outlined. One possible approach is the following: Assume the arrival rate is constant during a specific time interval on a given day. Assume that its distribution on this interval over different days is known. Then the average performance characteristics for the interval can be obtained by the integration of the constant-rate steady-state formulae with respect to the distribution of the arrival rate.

5 Case Study: An Israeli Call Center

5.1 Preliminary Data Analysis and Outliers Detection

Comments: We want to show the time-varying behavior of the arrival process here. Maybe also high correlation among days.

We have already noted that the data exhibit rare short-term bursts and lulls in arrival activity. These can best be understood as outliers from a core stochastic model that applies during the normal operational environment. We will thus focus below on a model that applies during the period of normal operational environment that encompasses nearly all of the system's operation.

A stochastic model that incorporated the outliers could be built by constructing a mixture model that involved the period of normal operations (as described below) taken with probability near one, but with a small probability for a different type of environment that describes the activity during periods when outliers are observed. We do not pursue here the construction of such a mixture model. One reason is that the part of such a model involving outliers seems to require a different structure than the one for the normal environment. Also even with a year's data we have only a sparse amount of observed outlier activity on which to base a model for the non-usual environment. Thus a model for this part of the arrival process would probably need to rely heavily on expert a-priori evaluations about the type and frequency of outliers to be expected. Such an evaluation would, of course, also require understanding of the implications of the stochastic model we build below for the normal environment.

We continue to use the data for regular weekdays from 8/1 through 12/31, from 7am through midnight. Our model is built from binned data counts, N_{jk}, for short time intervals within this period. (The count, N_{jk}, is the number of calls arriving during the k^{th} time-interval period on the j^{th} day among those we analyzed.) For illustrative purposes we work with 15-minute time intervals. Intervals of about 10-15 minutes in length seem about right for this amount of data. With more data it would be preferable to use somewhat shorter intervals.

A first step in our analysis is to identify those time intervals that can be considered as involving outliers that should not be modelled by our normal-environment model. Figure 4 is a scatterplot showing the values N_{jk} by time of day. We subjectively identified 21 outliers among the 7,269 values. These are shown with either asterisks or open circles on the scatterplot.

If we were to remove just these 21 outlying values we would then be left with the analysis of an unbalanced model. This is feasible by the methods described below, but as is familiar in other statistical contexts balanced models are somewhat easier to analyze and interpret. Since our goal is primarily to illustrate the analysis, rather than to get the very best available estimates, we decided to delete entire days on which multiple outliers occurred. This leaves us with a balanced model with very few remaining outliers in the data. We thus removed five entire days of data from our analysis. The asterisks on the plot show the 15 outliers that were on those 5 removed days. (For example the four asterisks between times 15 (=3pm) and 16 all occurred on 12/30, the last regular working day of the year. The two asterisks above 90 at times 20.5 and 20.75 occurred on 8/09. In general there seemed nothing else out of the ordinary on the five removed days other than the values of N_{jk} shown as asterisks on the plot, and those for a few neighboring quarter-hour periods that had slightly elevated values. Since these days appeared fairly routine apart from the outlying observations, it appears that removing the entire days has roughly the same impact on the final result of the analysis as would have occurred from removing just the outlying values.

The six outliers shown as open circles on the plot occurred as isolated instances on six separate days. We decided not to remove these days from our analysis. Note the outlying value of 89 at 15.25 (on 8/16). This will remain throughout our analysis as a very noticeable outlying observation.

In addition there was one day (10/06) on which the data was missing for the final 7 quarter-hour periods of the day. We decided to also remove this day from our analysis, again in order to preserve the computational and interpretational simplicity of a balanced design. We are thus left with 6868 observations, to be indexed as $N_{jk}, j = 1, \ldots, 101, k = 1, \ldots, 68$ corresponding to 68 quarter-hour counts on 101 days.

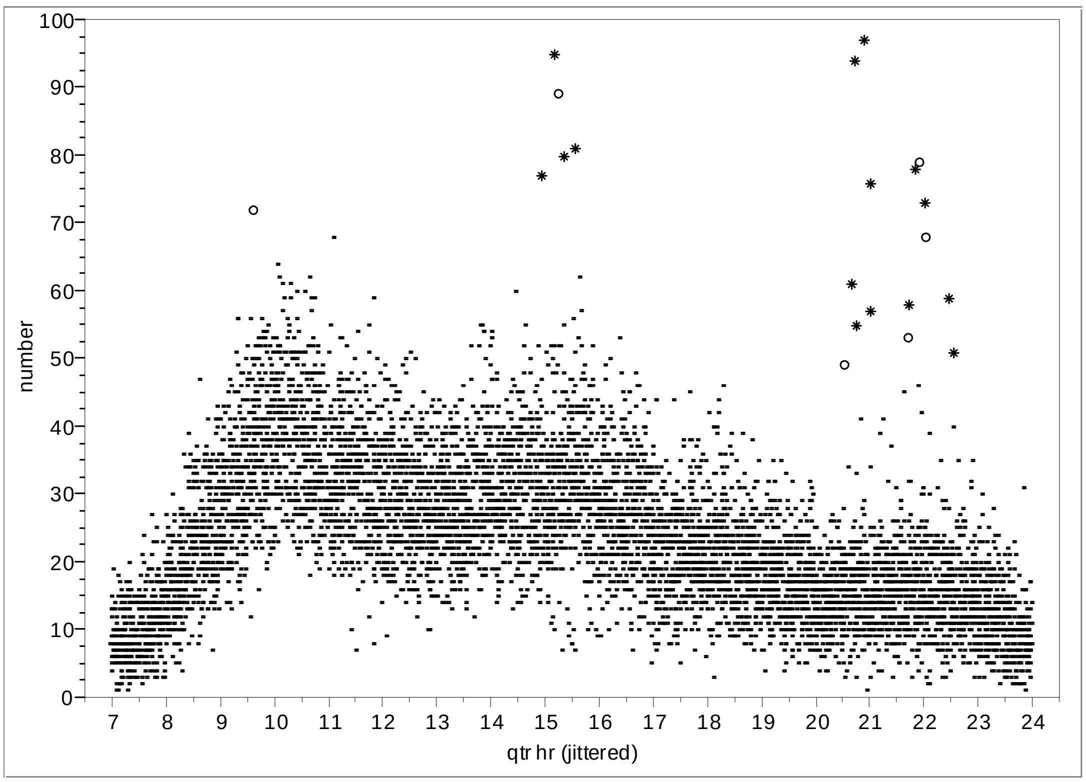

Figure 4: Scatterplot of values of N_{jk}. Asterisks show outliers on the 5 days later removed from the analysis. Open circles show other outliers.

5.2 Model Results

Figure 5 shows the values of the estimates $\hat{\kappa}_k^2 = \hat{\beta}_k$ derived from our data using Model 2. Since the second model has a Gaussian structure, it is easy to also derive (asymptotic) confidence limits for the estimates $\hat{\kappa}_k$, and then to convert these into confidence limits for

$$\hat{\beta}_k = \hat{\kappa}_k^2.$$

These confidence limits are also shown on Figure 5.

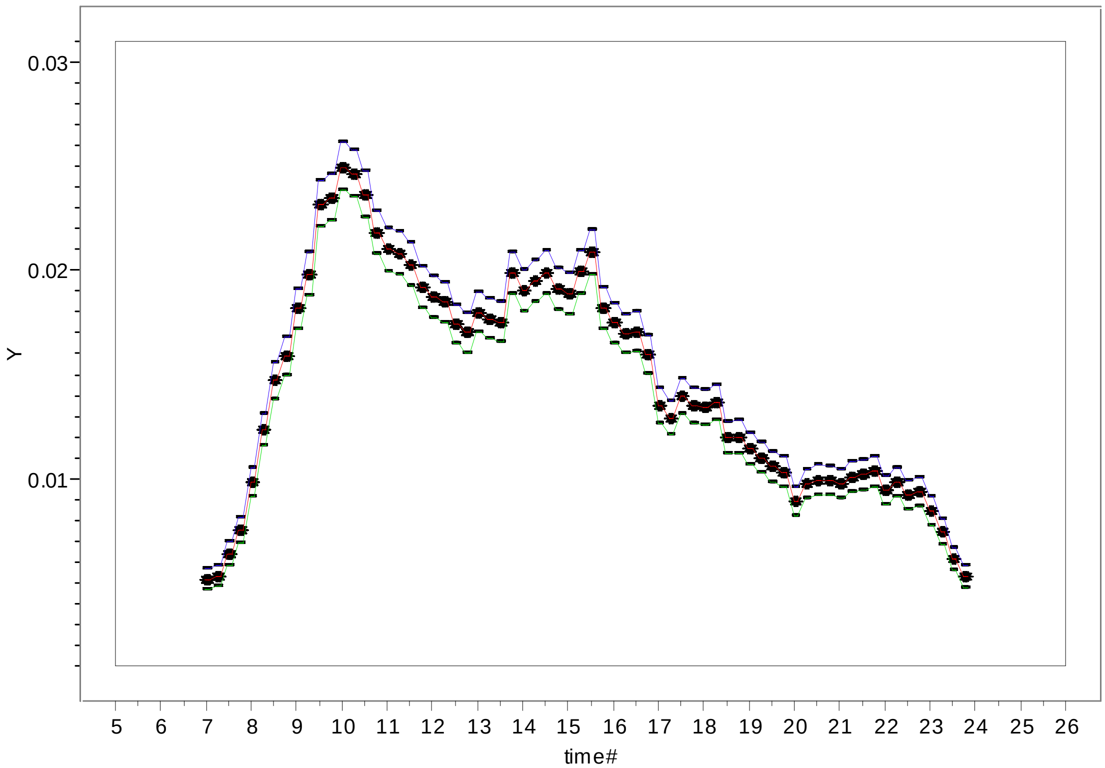

Figure 5: Estimated conditional density by time, from model 2; and lower and upper 95% confidence limits. (Data for weekdays 8/1 – 12/30 with 6 days removed.)

In addition we note that the average of the squared residuals is 0.358. Hence $\hat{s}^{(2)} = 4 \times (0.358 - 0.25) = 0.432$. (As one should expect this value is very nearly the same estimate as is derived for s in an analysis of our Model 1 – the Gamma-Poisson model. See Table 2 We also note that the value of s in either model is a function of the basic time intervals used.

For example, if we were to have used 10 minute intervals (2/3 as long) then we would expect to find $\hat{s} \approx (2/3)^2 \times 0.432 = 0.192$. Thus, as the time intervals become shorter it becomes harder to distinguish between the Gamma-Poisson model and a pure Poisson model.)

Because Model 2 is Gaussian, it is possible to use well-known diagnostic techniques in order to judge how well the data corresponds to the model. Figure 6 is a plot of residuals versus time. There is no particular evidence of strong heteroscedasticity as a function of time. The observation from time 15 $^1/_4$ on 8/16 does stand out as an outlier, as one should expect. The residual plot versus day (not shown here) is similarly benign.

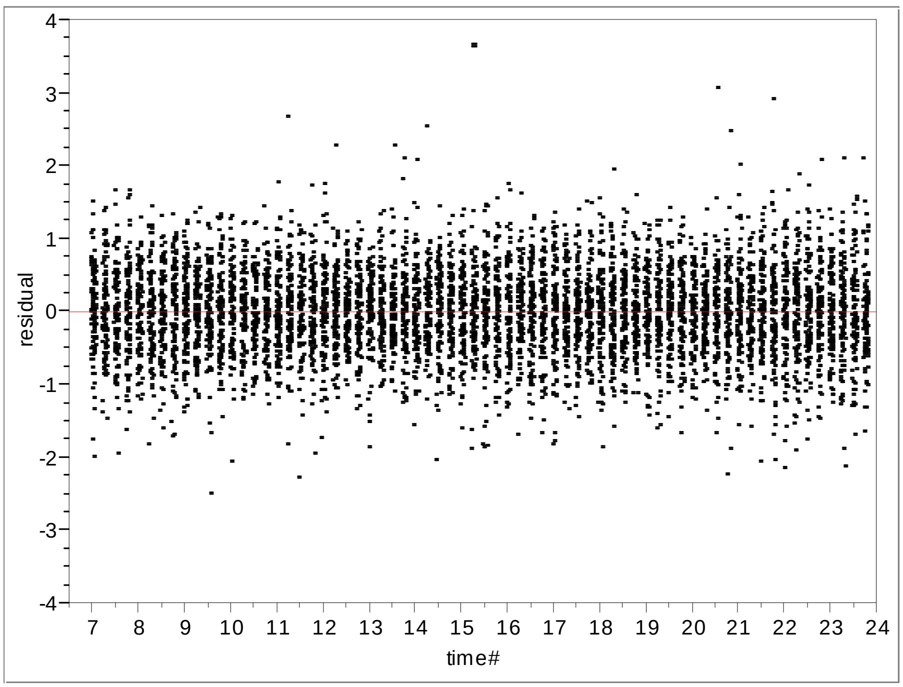

Figure 6: Residual plot for the data in Figure 5.

It is further desirable to check whether the residuals satisfy the model's assumption of normality. As usual this can be done visually by constructing a normal quantile plot of the residuals. This is shown in Figure 7.

Apart from about a dozen too-large residuals (especially that on 8/16 mentioned above) this plot visually agrees quite well with the assumed normal distribution. A Kolmogorov-

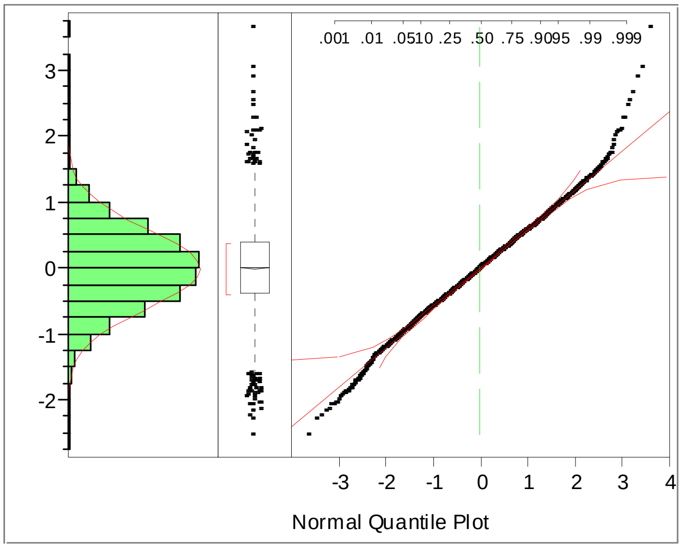

Figure 7: Histogram and normal quantile plot of the residuals for the data in Figure 5.

Smirnov test of normality yields the K-S statistic $\sup\left\{\sqrt{n}\left|\hat{F}_n(t) - \Phi(t)\right|\right\} = 0.8681$(where $\hat{F}_n$ denotes the sample cdf of the normalized residuals). This has a P-value of 0.56, a remarkably insignificant value considering the large sample size of $n = 101 \times 68 = 6868$. We conclude that our Model 2 provides a very suitable model for this data. (The Gamma-Poisson model described in Section 4.1 is closely related to Model 2, and so is presumably also a very suitable model.)

We also carried out an analysis of the data of Figure 5 using the Gamma-Poisson model of Section 4.1. The results are strikingly similar to those from Model 2. Table 2 displays this similarity. It gives various quantiles from the distribution of $\{\hat{\alpha}_j^{(2)}/\hat{\alpha}_j\}$ and of $\{\hat{\beta}_j^{(2)}/\hat{\beta}_j\}$ where $\hat{\alpha}_j$, $\hat{\beta}_j$ are the estimates from the Γ-P model and $\hat{\alpha}_j^{(2)}$, $\hat{\beta}_j^{(2)}$ are the estimates from Model 2.

In addition we recomputed results from the modified Model 2 type of analysis, beginning with

$$X_{jk} = \sqrt{N_{jk} + (1+\hat{s})/4}$$

where we took $\hat{s}$ from the Γ-P estimates. Table 2 also displays results from this recomputed analysis.

Note that the results from the recomputed analysis are even closer to those from the Γ-P model. The most noticeable difference is that the first Model 2 analysis tends to slightly underestimate the α_j coefficients that correspond to daily volume, and this systematic underestimation is eliminated in the recomputed model.

The estimated values of s from the three models were

$$\hat{s} = 0.4329,\ \hat{s}^{(2)} = 0.4336,\ \hat{s}^{(3)} = 0.4244.$$

(The recomputed model gives a slightly smaller value of $\hat{s}$; but its sample variance of residuals, $(N_{jk} - \hat{N}_{jk})$, is only very slightly less (32.864 vs 32.688).)

Percentile	Percentile Name	$\hat{\alpha}_j^{(2)}/\hat{\alpha}_j$	$\hat{\alpha}_j^{(3)}/\hat{\alpha}_j$	$\hat{\beta}_k^{(2)}/\hat{\beta}_k$	$\hat{\beta}_k^{(3)}/\hat{\beta}_k$
100.0%	maximum	0.9981	1.003	1.003	1.003
90.0%		0.9972	1.002	1.002	1.002
75.0%	quartile	0.9967	1.001	1.001	1.001
50.0%	median	0.9958	1.000	1.000	1.000
25.0%	quartile	0.9946	0.999	0.997	0.999
10.0%		0.9919	0.997	0.993	0.998
0.0%	minimum	0.9849	0.991	0.989	0.995

Table 2: Ratios of coefficients of Gamma-Poisson and Model 2 Analyses: $\hat{\alpha}_j^{(2)}$, $\hat{\beta}_k^{(2)}$ correspond to $X_{jk} = \sqrt{N_{jk} + 1/4}$; $\hat{\alpha}_j^{(3)}$, $\hat{\beta}_k^{(3)}$ correspond to $X_{jk} = \sqrt{N_{jk} + (1+\hat{s})/4}$.

5.3 Prediction with Model 2

For planning of facilities and staffing it is important to have prediction confidence statements of arrivals at particular times of day. The overall objective would then be to combine these in a queuing model along with stochastic information about service times and customer patience in order to derive realistically tuned service models. In order to form such statements we need to also study the distribution of $\alpha_j = \rho_j^2$. Figure 8 is a histogram and normal quantile plot of the values of ρ_j.

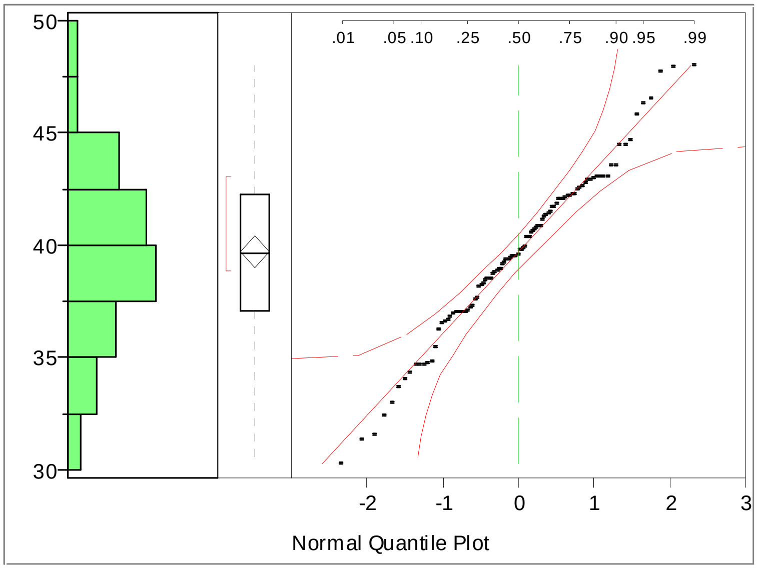

Figure 8: Histogram and normal quantile Plot of ρ for the data in Figure 5

It is evident that the daily effects, $\{\rho_j\}$, in this model are well modelled by a normal distribution with a mean $\hat{\mu}_\rho = 39.709$ and a standard deviation $\hat{\sigma}_\rho = 3.633$. (Correspondingly, we should expect that if one adopts the Gamma-Poisson model the $\{\alpha_j\}$ in that model are approximately distributed as $\Gamma\text{-P}(39.709^2, 4 \times (3.633^2 - 0.25)) = \Gamma\text{-P}(1580, 51.8)$. (The $\{\alpha_j\}$ in turn can be thought of as adjusted estimates of the $\{E(N_{j+})\}$. The mean and standard deviation of the $\Gamma\text{-P}(1580, 51.8)$ distribution are $\mu = 1580$ and $\sigma = 288.8$. The observed sample mean and standard deviation of $\{N_{j+}\}$ are 1597.2 and 285.4, respectively, which agrees fairly well with the Γ-P mean and variance just described.)

Let X_{*k}^* denote a future observation at time period k. We have the following distributions, or estimated distributions. (In the following $N(\mu, \sigma^2)$ denotes the normal distribution with the indicated mean and variance, and the symbol $\sim$ indicates an approximate (asymptotic) distributional assertion and $\approx$ an approximate (asymptotic) equality.)

$$N^*_{*k} = (X^*_{*k})^2 - 1/4, X^*_{*k} = \rho_j \kappa_k + \varepsilon^*_k \;\; \mathit{where} \; \varepsilon^*_k \sim N(0, \sigma*^2) \;\; ,$$
$$\hat{\kappa}_k \sim N\left(\kappa_k, \sigma^2_{\kappa,k}\right) \;\; \left(\text{where } \sigma^2_{\kappa,k} = \frac{\sigma *^2 \left(1 - \kappa_k^2\right)}{\sum \rho_j^2} \right) \;\; ,$$
$$\hat{\rho} \sim N(\mu_\rho, \sigma^2_\rho) \;\; .$$

(The value for $\sigma^2_{\kappa,k}$ is the result of a routine asymptotic calculation of Fisher information for this model.)

The relevant parameters can be estimated here from the data as

$$\sigma*^2 \approx \hat{\sigma}*^2 = 0.358, \text{ the mean square residual,}$$
$$\sigma^2_{\kappa,k} \approx \hat{\sigma}^2_{\kappa,k} = 3.3114 \times 10^{-6} \times (1 - \hat{\kappa}^2_k),$$
$$\mu_\rho \approx \hat{\mu}_\rho = 39.709, \text{ the sample mean of } \{\hat{\rho}_{\mathrm{j}}\},$$
$$\sigma^2_\rho \approx \hat{\sigma}^2_\rho = 13.201, \text{ the sample variance of } \{\hat{\rho}_{\mathrm{j}}\}.$$

We write $\hat{\kappa}_k \sim \kappa_k + \varepsilon_{\kappa,k}, \;\; \hat{\mu}_\rho \sim \mu_\rho + \varepsilon_\rho$ where the $\varepsilon_{\kappa,k}, \quad \varepsilon_\rho$, and $\varepsilon*$ are independent normal variables with mean 0 and the indicated variances. Then

$$\Pr(N^*_{*k} \le C^2) \approx \Pr(X^*_{*k} \le C) \approx \Pr\left((\hat{\mu}_\rho - \varepsilon_\rho)(\hat{\kappa}_k - \varepsilon_{\kappa,k}) + \varepsilon* \le C\right).$$

The probability on the right is easy to simulate, but numerically this is not needed since $\varepsilon_\rho \varepsilon_{\kappa k}$ is a nearly negligible quantity. We can thus write,

$$\Pr(X^*_{*k} \le C) \approx \Phi\left(\frac{C - \hat{\mu}\hat{\kappa}_k}{(\hat{\mu}^2_\rho \sigma^2_{\kappa,k} + \kappa^2_k \sigma^2_\rho + \sigma*^2)^{1/2}} \right).$$

This yields $100(1-\alpha)\%$ prediction limits of the form

$$N^*_{*k} \in \left(\hat{\mu}_\rho \hat{\kappa}_k \pm z_{1-\alpha} \sqrt{\hat{\mu}^2_\rho \sigma^2_{\kappa k} + \kappa^2_k \sigma^2_\rho + \sigma*^2} \right), \tag{4}$$

where $z_{1-\alpha}$ is the indicated normal quantile.

It is also possible to use this same line of reasoning to develop confidence limits for the parameter $E(N^*_{*k})$. This yields $100(1-\alpha)\%$ limits of the form

$$E(N^*_{*k}) \in \left(\hat{\mu}_\rho \hat{\kappa}_k \pm z_{1-\alpha}\sqrt{\hat{\mu}_\rho^2 \sigma_{\kappa k}^2 + \kappa_k^2 \sigma_\rho^2}\right). \tag{5}$$

Figure 9 displays the result of using these formulas to yield prediction limits and prediction-confidence limits for future values of N^*_{*k} and $E(N^*_{*k})$. As is familiar from other applications contexts these prediction bands are much wider than are confidence bands for corresponding expectations. More striking is that these bands are both much wider than those in Figure 5 for the conditional density given N_{+k}. This is because the prediction accuracy is strongly influenced by the variability of the $\{N_{+k}\}$, and this is qualitatively much larger than the variability of the conditional variables N_{jk} given N_{+k}.

5.4 Prediction with Model 3

6 Conclusion

References

[1] Agresti, A. (1990). *Categorical Data Analysis*, Wiley and Sons.

[2] Anscombe, F. J. (1948). The transformation of Poisson, binomial and negative-binomial data. *Biometrika* **35**, 246–254.

[3] Brown, L. D., Gans, N., Mandelbaum, A., Sakov, A., Shen, H., Zeltyn, S. and Zhao, L. (2003). Statistical Analysis of a Telephone Call Center: A Queueing-Science Perspective. Revised for *JASA*.

[4] Brown, L.D., Zhang, R. and Zhao, L. (2001). Root un-root methodology for nonparametric density estimation. *Technical Report*.

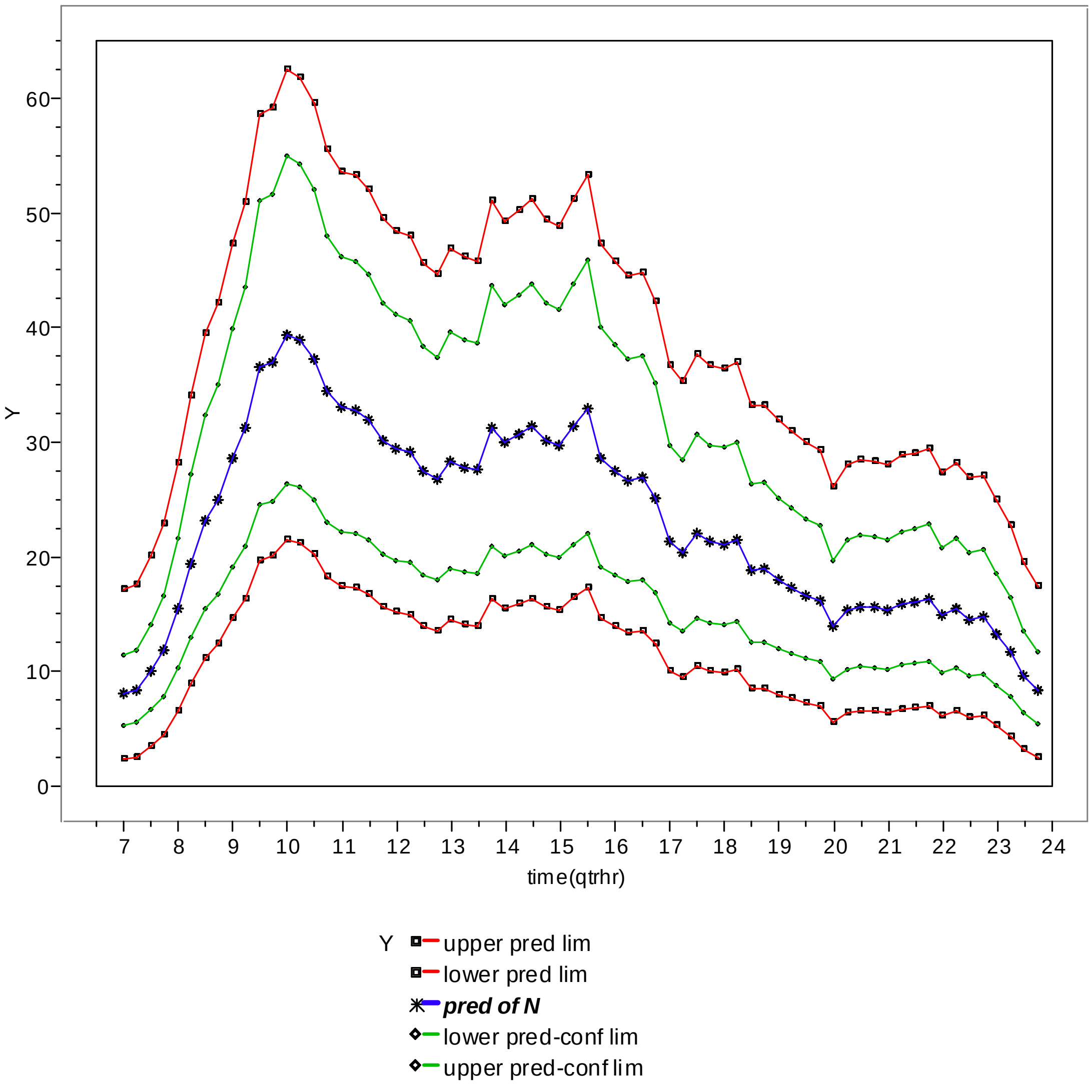

Figure 9: 95% Prediction bands derived from the data in Figure 5. Inner bands are from (4) and outer ones from (5).

[5] Brown, L. D. and Zhao, L. (2002). A test for the Poisson distribution. *Sankhya*, **64**, 611-625.

[6] Gans, N., Koole, G. and Mandelbaum, A. (2003). Telephone Call Centers: Tutorial, Review, and Research Prospects. *M&SOM* **5**, 79–141.

[7] Jongbloed, G. and Koole, G.M. (2001). Managing uncertainty in call centers using Poisson mixtures. *Applied Stochastic Models in Business and Industry*, **17**, 307-318.

[8] Mandelbaum, A., Sakov, A. and Zeltyn, S. (2000). Empirical Analysis of a Call Center. *Technical Report*.

[9] McCullagh, P. and Nelder, J.A. (1989). *Generalized Linear Models*, Second Edition. Chapman Hall, New York.