Managing uncertainty in call centers using Poisson mixtures

Geurt Jongbloed* and Ger Koole
Vrije Universiteit, Division of Mathematics and Computer Science
De Boelelaan 1081a, 1081 HV Amsterdam, The Netherlands
{geurt,koole}@cs.vu.nl

To appear in *Applied Stochastic Models in Business and Industry*

Abstract

We model a call center as a queueing model with Poisson arrivals having an unknown varying arrival rate. We show how to compute prediction intervals for the arrival rate, and use the Erlang formula for the waiting time to compute the consequences for the occupancy level of the call center. We compare it to the current practice of using a point estimate of the arrival rate (assumed constant) as forecast.

1 Introduction

A call center consists of equipment and people capable of delivering services by telephone. In the current paper we assume that these telephone calls are all inbound, i.e., they are generated by outside customers. The performance level of a call center is usually measured in terms of the waiting time of calls and the productivity of the call center employees, often called agents. One of the main problems in managing a call center is the uncertainty in call volume, and the fact that calls need to be answered quickly (on average between 10-20 seconds). The solution to this is to schedule some overcapacity with respect to the average call volume as to be able to handle peaks in incoming traffic. This overcapacity needs to be as small as possible, as overcapacity means unproductivity.

Uncertainty in call volume has various sources. Divide the period to be considered in fixed time intervals. It is often assumed that, during each time interval, arrivals follow a homogeneous Poisson process and that call handling times are exponentially distributed. This randomness creates short-term fluctuations in call volume. It is also often assumed that call handling times (service times) are constantly distributed in time. This assumption is justifiable to a considerable extent. However, assuming the arrival rate λ of the calls to be constant, is highly unrealistic. If the rate were constant, the expected waiting time could be estimated using the so-called Erlang formula. In section 2 we will describe this procedure and give interval estimates for the constant rate. Current practice is that call center planners differentiate between variables such as day of week, holidays or marketing activities, to obtain an estimate for the rate which is needed to apply the Erlang formula. Still, this differentiation

*Research supported by a grant from the Haak Bastiaanse Kuneman foundation of the Vrije Universiteit.

does not explain all variability in call volume. Adding more explanatory variables not only makes the analysis more complicated and time-consuming, it can also be useless: bad weather conditions can only be foreseen a few days in advance, while agent rosters often have to be published one or more weeks in advance. Hence, there is a need to model this extra variability in the call volume differently.

We propose to model the rate itself as a random variable. This could be done after or without differentiating between certain explanatory variables. In section 3 we will discuss the Poisson mixture model that corresponds to this situation and explain how the Erlang formula can be used if an estimate for the distribution of the rate (the mixing distribution) is available from the data. The problem of estimating the mixing distribution is addressed in section 4. A parametric as well as a nonparametric procedure is suggested.

Using our method, one gets a prediction interval for the arrival rate of calls, based on the data. Combined with the Erlang formula, this result can be used in two ways, that we will discuss both.

Assume that there is a certain service level, formulated as a function of the waiting time, that we have to adher to. There is a second performance measure, personnel costs, that has to be minimized. The first approach can be summarized as follows. We assume that we cannot adapt the workforce to changing conditions. This means that we have to take a worst-case scenario. By looking at a prediction interval for the arrival rate, we can give a stochastic guarantee for the service level: in $\alpha\%$ of the cases the service level is at least such that $p\%$ of the calls waits less than s seconds. This is based on applying the Erlang formula to the upper bound of the prediction interval. The lower bound can be used to show how far costs can be off, up to which level we can waste money for not needed personnel.

The second approach uses also the prediction interval, but in another way. Here we assume that the workforce can be adapted in a flexible way, for example by having flexible contracts that allow the call center to call for extra personnel if needed. The question is how many flexible and non-flexible agents should be scheduled. This is where the prediction interval comes in: the lower bound fed into the Erlang formula gives the number of fixed agents, the difference in needed agents between upper and lower bounds gives the number of agents with a flexible contract that might be needed.

Section 5 is devoted to a case study. The ideas and techniques of this paper are applied to data that were obtained at a call center of a Dutch insurance company.

We finish this introduction with a few references to the considerable call center literature. We would like to mention the modeling studies Brandt et al. [3] and Mandelbaum et al. [10] and the references in these papers. For a more managerial view on call centers we recommend Cleveland & Mayben [4].

2 Erlang formula for simple model

We consider a call center at a fixed period of time during a week day. The average call duration is $1/\mu$. It is our objective to determine the optimal number of agents to be scheduled. For the moment, we assume that there are c agents. As is customary in workforce management tools, we model the call center as a standard $M/M/c/\infty$ system. Then, for a fixed arrival rate λ, the waiting time can be approximated using the well known Erlang formula.

The Erlang formula requires that service times are exponential and that a stationary situation has been reached; both requirements are unrealistic in practice. However, it can be

seen that the Erlang formula gives an excellent approximation in most realistic cases. Let W denote the waiting time of an arbitrary customer. Define the load to the system $a = \lambda/\mu$. If $a \geq c$, then the load exceeds the service capacity, and the number of calls in the system grows to infinity. In this case $W = \infty$. We assume that $a < c$, the stability condition. Standard queueing theory (see, e.g., Gross and Harris [5], Section 2.3, Eq. (2.49)) tells us that

$$\mathbb{P}(W > t) = \begin{cases} C(c,a)\mathrm{e}^{-(c\mu-\lambda)t} & \text{if } c > a \\ 1 & \text{if } c \leq a \end{cases}$$

Here

$$C(c,a) = \frac{a^c}{(c-1)!\,(c-a)} \left[\sum_{j=0}^{c-1} \frac{a^j}{j!} + \frac{a^c}{(c-1)!\,(c-a)} \right]^{-1}.$$

is known as the probability of delay (indeed, $\mathbb{P}(W > 0) = C(c,a)$) and $a = \lambda/\mu$ as the offered load. The productivity is a number in $[0,1]$, defined as the mean number of busy agents divided by the number of agents, which is equal to $\rho = \lambda/(c\mu)$.

It is current practice to use point estimates for μ and λ to obtain an estimate for the waiting time distribution. In this simple constant-rate model, more can be said based on standard statistical theory. Suppose that we have K (comparable) periods in which respectively $x_1, x_2, \ldots, x_K$ calls came in. Moreover, suppose that these calls had durations $y_1, y_2, \ldots, y_n$. Here $n = \sum_{i=1}^{K} x_i$. Then the usual estimates for μ and λ are

$$\hat{\mu} = n\Big(\sum_{i=1}^{n} y_i\Big)^{-1} \text{ and } \hat{\lambda} = K^{-1}\sum_{i=1}^{K} x_i \tag{1}$$

However, since one can usually also use information on durations of calls from other periods to estimate μ, the estimate of μ is usually more accurate than that of λ. Moreover, an approximate $100(1-\beta)\%$ confidence interval for λ is given by

$$\left[\hat{\lambda} - u_{1-\beta/2}\sqrt{\frac{\hat{\lambda}}{K}}, \hat{\lambda} + u_{1-\beta/2}\sqrt{\frac{\hat{\lambda}}{K}} \right] \tag{2}$$

where u_β denotes the β-quantile of the standard normal distribution, i.e., $\Phi(u_\beta) = \beta$ where Φ is the cumulative probability distribution of a standard normal random variable.

The quantities of interest ρ and $\mathbb{P}(W > t)$ are both increasing in λ. (For ρ this is trivial, for $\mathbb{P}(W > t)$ this follows from a simple coupling argument, see, e.g., Ross [11], Section 8.2.) Therefore, the Erlang formula applied to the confidence limits for λ furnishes a confidence interval for $\mathbb{P}(W > t)$. Similarly a confidence interval for ρ can be derived. In these intervals the point estimate for μ can be used (because it is much more accurate than that of λ), but it is also possible to use the confidence limits for that. Then monotonicity of ρ and $\mathbb{P}(W > t)$ in μ is used.

3 The Poisson mixture model

The assumption of constant rate is unrealistic in many practical examples. For data to be generated independently by a single Poisson distribution, the variance and the mean should be approximately the same. Call volume data sometimes show a variance that substantially

dominates the mean. A method of dealing with this *overdispersion* of count data, is to use a Poisson mixture model for the data generating mechanism. Basically, each data point is being interpreted as generated in two steps. First, the rate λ of the Poisson variable is drawn from a distribution with distribution function H on $(0, \infty)$, and then a Poisson variable X with that rate is generated. This means that the distribution of X is a so-called Poisson mixture with mixing distribution H:

$$P_H(X = x) = \int_0^\infty \frac{\lambda^x}{x!} e^{-\lambda} \, dH(\lambda) \tag{3}$$

In the biomedical literature, Poisson mixtures are often used to account for overdispersion in Poisson count data. If the mixing distribution has finite support, it has the intuitive appeal that the population one draws from is heterogeneous, but consists of a finite number of homogeneous Poisson subpopulations. See e.g. Lindsay [8] and Böhning [2] for medical applications of Poisson mixtures.

We are given realized values $x_1, x_2, \ldots, x_K$ of the independent and identically distributed random variables $X_1, X_2, \ldots, X_K$, that are distributed as X. It is our objective to schedule agents (both for the models with and without flexible agents) in a statistically correct way, using the Erlang formula.

We want to apply the Erlang formula to a range of plausible λ's to assess the variability in the number of agents taking into account that λ is not fixed and known. To do this we propose to give $\underline{\lambda}$ and $\overline{\lambda}$ such that for the next draw Λ from H the following holds

$$P_H(\Lambda \in [\underline{\lambda}, \overline{\lambda}]) \geq 1 - \alpha \tag{4}$$

Now, if H were known, we could take $\underline{\lambda} = H^{-1}(\alpha/2)$ and $\overline{\lambda} = H^{-1}(1 - \alpha/2)$. This means that all uncertainty is contained in the randomness of Λ. As a special case, if we would know H to be a degenerate distribution on a known point, we could take this $\underline{\lambda} = \overline{\lambda} = \lambda$.

Of course, H is not known in practice, and has to be estimated from the data. If we would know H to be a degenerate distribution function assigning all its mass to one (unknown) point λ, we would estimate λ by $\hat{\lambda}$ as in (1). As argued in section 2, one can use the confidence set (2) of λ to get a confidence result in the spirit of (4): with confidence at least $1 - \alpha$, the true λ will be in the set $[\underline{\lambda}, \overline{\lambda}]$. If nothing is known about H, the first step is again to estimate the quantiles we are after. A next step would be to construct confidence intervals for these quantiles. Statistically, we therefore want to give point- or interval estimates of quantiles of the mixture distribution in a Poisson mixture. We restrict ourselves to point estimation in section 4.

4 Estimating the mixing distribution

How can we estimate (quantiles of) H? There are many ways, depending on the assumptions imposed. One approach is to estimate the distribution function H parametrically, and estimate the distribution by estimating its parameters. A well known drawback of this parametric procedure is that the choice of parametric family is rather arbitrary and usually only motivated by mathematical convenience. An advantage is that computations and statistical properties of the estimators can be dealt with relatively easily. Another way is to estimate the distribution function H nonparametrically (e.g., via maximum likelihood). We will discuss both approaches here.

Consider the demixing problem. If we take a known parametric family of distributions where H should belong to, we usually get a (less known) parametric family for the (discrete) sampling distribution. For specific classes of distributions for H, however, known discrete distributions emerge for X. For example, assume H to belong to the class of Gamma distribution functions, with densities

$$h_{s,r}(\lambda) = \frac{s^r}{\Gamma(r)} \lambda^{r-1} e^{-s\lambda} 1_{[0,\infty)}(\lambda). \tag{5}$$

Then (3) yields

$$\begin{aligned} P_{s,r}(X = x) &= \int_0^\infty \frac{\lambda^x}{x!} e^{-\lambda} h_{s,r}(\lambda)\, d\lambda = \frac{s^r}{x!\Gamma(r)} \int_0^\infty \lambda^{x+r-1} e^{-\lambda(1+s)}\, d\lambda \\ &= \frac{\Gamma(x+r)}{x!\Gamma(r)} \left(\frac{s}{1+s}\right)^r \left(1 - \frac{s}{1+s}\right)^x = \binom{r+x-1}{x} \left(\frac{s}{1+s}\right)^r \left(1 - \frac{s}{1+s}\right)^x \end{aligned}$$

This shows that the sampling distribution is the negative binomial distribution with success probability $s/(1+s)$ and (for $r \in \{1, 2, \ldots\}$) number of successes r. Estimating the mixing density thus boils down to estimating the parameters from a negative binomial distribution based on direct observations from this distribution, and plugging these estimates into expression (5). In section 6 we give details on this estimation problem. Having estimates $\hat{r}$ and $\hat{s}$ based on the sample, we can use the quantiles $H^{-1}_{\hat{r},\hat{s}}(\alpha/2)$ and $H^{-1}_{\hat{r},\hat{s}}(1-\alpha/2)$.

One of the drawbacks of using a Gamma mixing distribution is that Gamma distributions are unimodal. If, e.g., one uses data for different time periods in the model (heterogeneous population), a very natural (though not necessary) consequence is that the mixing distribution is multimodal. The Gamma family is too rigid to allow this. One way out of this problem is to estimate the mixing distribution nonparametrically. A natural estimator for the mixing distribution is the (nonparametric) maximum likelihood estimator. Given data $x_1, x_2, \ldots, x_K$, we can write the loglikelihood function as

$$\begin{aligned} \phi(H) &= \sum_{i=1}^K \log P_H(X_i = x_i) = \sum_{i=1}^K \log \int_0^\infty \frac{\lambda^{x_i}}{x_i!} e^{-\lambda}\, dH(\lambda) \\ &= \sum_{j=0}^k m_j \log \int_0^\infty \frac{\lambda^j}{j!} e^{-\lambda}\, dH(\lambda) = \sum_{j=0}^k m_j \log \int_0^\infty \lambda^j e^{-\lambda}\, dH(\lambda) + C \end{aligned}$$

Here $k = \max_i x_i$ and m_j is the number of times j occurs in the sample $x_1, \ldots, x_K$. The constant C in this expression does not depend on the function H.

Some properties of the maximum likelihood estimator are known. In Lindsay [7] it is shown that it is a discrete distribution function with no more than K jumps. Moreover, as the sample size K tends to infinity, the estimator converges to the underlying mixing distribution uniformly in probability. See Kiefer and Wolfowitz [6].

There are many algorithms that can be used to compute the MLE. One specific program, developed by Böhning and Schlattmann and available as shareware on the web, is C.A.MAN (Computer Assisted Mixture ANalysis).

5 Case study

Now consider the data that were obtained from a call centre of a Dutch insurance company. In the period January 3 till February 9, 2000, the number of incoming calls were registered

each half hour. We consider the calls that arrived between nine o'clock and half past nine during the week days in this period (excluding Saturdays). This means we have a dataset with $K = 28$ and $n = 3278$ (part of the data is given in the appendix). Common practice would be to assume a Poisson distribution and estimate λ by $\hat{\lambda} = 117.1$. We could include information on the accuracy of this estimate and construct confidence interval (2). In the present situation this is given by $[113.7, 120.4]$.

First we will test for a (unicomponent) Poisson distribution. One way of doing so is using Neyman-Scott test for this situation. Following Lindsay [8], chapter 4, this procedure corresponds to using the test statistic

$$T_K = \sqrt{K/2}\left(\frac{S_K^2}{\overline{X}_K} - 1\right).$$

For K large, this statistic is approximately standard normally distributed under the null hypothesis of a unicomponent Poisson distribution. This can be seen using the Central Limit Theorem, the delta method and Slutsky's lemma. See Van der Vaart [12], example 3.4 for the reasoning in a related example. For our dataset, we have $t_K = 15.4$, with corresponding p-value of zero. Hence, the unicomponent Poisson distribution is clearly rejected. Since $t_K >> 2$, we say the data is *overdispersed*. In our Poisson situation this corresponds to the fact that the sample variance is too big compared to the sample mean.

Having rejected the unicomponent Poisson distribution, one could consider to use a graphical method discussed in Lindsay and Roeder [9], to asses the plausibility of a Poisson mixture. However, with our sample sizes in comparison with the range of the counts, these methods break down because of the many zeroes in the observed frequencies.

Assuming the mixing distribution to be a Gamma distribution, we obtained as estimates for the parameters s and r respectively $\hat{s} = 3.6$ and $\hat{r} = 32.3$. The estimated 5% and 95% quantiles of the mixing distribution are therefore given by 85.5 and 152.9. Using a bootstrap test based on the Kolmogorov Smirnov statistic (see appendix) to assess the goodness of fit of the Gamma mixing distribution, we obtained a p-value of approximately 0.15. Hence, based on the data of 9.00-9.30, the Gamma Poisson mixture model cannot be rejected.

Figure 1 shows the empirical distribution function of the counts with the distribution functions based on the unicomponent Poisson model and the Gamma Poisson-mixture.

Table 1 gives for the different methods the bounds for λ. In the unicomponent case, this is the 90% confidence set, whereas in the other situations these are the estimated quantiles of the mixing distribution. The period 9.00-9.30 corresponds to the situation described above. The other periods were analyzed in exactly the same way.

Table 1 shows that the Gamma mixture of Poisson distributions furnishes an acceptable model for most time periods. The p-values are usually above 0.05. Exception is the period $09.30 - 10.00$. There the empirical distribution of the counts differs more from the maximum likelihood fit based on the Gamma mixture of Poisson distributions than would be plausible if the true distribution would be a Gamma mixture of Poisson distributions.

Using the available data, we estimate the mean length of a call by $\hat{\mu} = 419$ seconds. Table 2 shows the consequences for the lower and upper bound of people to be scheduled when based on the different methods discussed in this paper by applying the Erlang formula to the bounds on the Poisson rates.

The results can be interpreted as follows. Let us consider the 9.00-9.30 interval. It is common practice in call centers to schedule based on a simple point estimate. Then

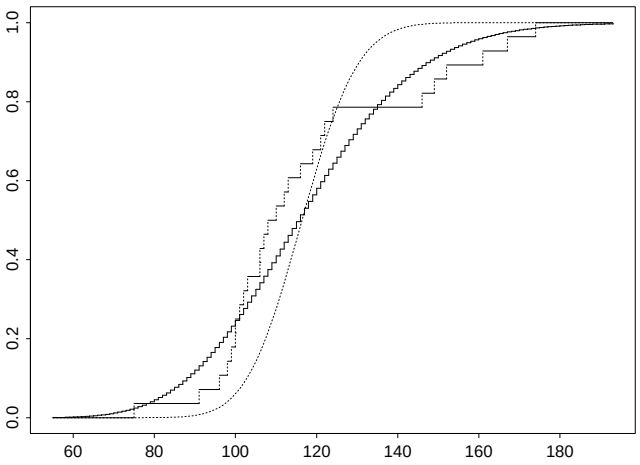

Figure 1: *Empirical distribution function of the data with unicomponent fit (dotted line) and Gamma mixture fit*

there should be 34 agents available for call handling from 9.00 to 9.30. The (rejected!) unicomponent model gives confidence bounds 33 and 34. Call centers using this method experience highly fluctuating service levels. The reason lies in the fact that the unicomponent model is rejected; instead a model such as the Poisson mixture could be used. This model shows that within reasonable bounds, the arrival rate can vary between 85.4 and 152.9, with corresponding occupation levels 25 and 43. Thus, using the monotonicity of the Erlang formula discussed earlier, if we were to schedule 43 agents, then with probability 0.95 we satisfy the service level constraint that 80% of the customers has a waiting time shorter than 20 seconds. On the other hand, if we schedule only 25 agents, then the service level will only be attained with probability 0.05, corresponding to the fact that the Λ drawn from H will be sufficiently small.

A call center manager should manage his or her call center in such a way that it is possible to vary, on a relatively short term basis, between 25 and 43 agents. A number of ways to do this are:
- Let scheduled agents do other work such as outgoing calls or handling incoming emails;
- Have supervisors or other employees on standby to join the agents;
- Have agents with flexible contracts on standby.

For the 9.00 to 9.30 interval for example, we might have 30 scheduled agents in the call center and 5 additional agents that are supposed to handle emails, faxes, and letters. There are 2 supervisors, and at the beginning of the day there are 6 agents that can be called to join the call center. (Based on expectations and call volume between 8.00 and 8.30 these agents are called; details on the correlation between call volume in different intervals fall outside the scope of this paper.) We assume that there are enough emails and other tasks to work on to employ 5 additional agents. It is easily seen that the number of agents between 9.00 and 9.30, in this realistic situation, can indeed vary between 25 and 43.

Period	$\hat{\lambda}$	Confidence bounds	$\hat{r}$	$\hat{s}$	Gamma quantiles	p-val
08.00-08.30	12.1	$[11.0, 13.2]$	16.5	0.7	$[7.6, 17.3]$	0.07
08.30-09.00	41.8	$[39.7, 43.8]$	24.3	1.7	$[28.9, 56.6]$	0.71
09.00-09.30	117.1	$[113.7, 120.4]$	32.3	3.6	$[85.4, 152.9]$	0.15
09.30-10.00	155.5	$[151.7, 159.4]$	21.1	7.4	$[104.3, 215.1]$	0.00
10.00-10.30	158.4	$[154.5, 162.3]$	23.6	6.7	$[108.9, 215.5]$	0.51
10.30-11.00	160.2	$[156.2, 164.1]$	26.7	6.0	$[112.8, 214.3]$	0.23
11.00-11.30	157.4	$[153.5, 161.3]$	25.2	6.2	$[109.6, 212.2]$	0.12
11.30-12.00	156.3	$[152.4, 160.1]$	34.7	4.5	$[115.3, 202.3]$	0.24
12.00-12.30	131.1	$[127.6, 134.7]$	30.0	4.4	$[94.3, 172.8]$	0.23

Table 1: *Estimate of λ with associated confidence set in unicomponent model. The p-values for the unicomponent model were zero invariably. Also estimates of parameters of Gamma mixing distribution and resulting quantiles are given. The final column gives the p-values of the goodness of fit test for the Gamma mixing distribution based on* 500 *simulations (see section 6.2).*

6 Appendix

6.1 The negative binomial distribution

In this section we show how the maximum likelihood estimators for the parameters in the negative binomial model can be estimated. Consider the probability density given by

$$P_{p,r}(X = x) = \frac{\Gamma(x + r)}{x!\Gamma(r)} p^r (1 - p)^x$$

for $r > 0$ and $0 < p < 1$. Our aim is to estimate the parameter vector (p, r) based on a sample $x_1, \ldots, x_K$ of this negative binomial distribution. The loglikelihood function is given by

$$\phi(p, r) = \frac{1}{K} \sum_{i=1}^{K} \log(P_{p,r}(X = x_i)) = \frac{1}{K} \sum_{i=1}^{K} \left(\log \frac{\Gamma(x_i + r)}{x_i!\Gamma(r)} + r \log p + x_i \log(1 - p) \right)$$

For fixed values of r, this loglikelihood can easily be maximized over p: its maximum is attained in $\hat{p}_r = r/(r + \bar{x}_K)$. Hence, the *profile loglikelihood* (apart from constants not depending on r in the transition $\doteq$) becomes

$$\begin{aligned} \tilde{\phi}(r) &= \phi(\hat{p}_r, r) \doteq \frac{1}{K} \sum_{i=1}^{K} \log \frac{\Gamma(x_i + r)}{\Gamma(r)} + r \log r - (r + \bar{x}_K) \log(r + \bar{x}_K) \\ &= \frac{1}{K} \sum_{i=1}^{K} \sum_{j=1}^{x_i} \log(r + j - 1) + r \log r - (r + \bar{x}_K) \log(r + \bar{x}_K) \end{aligned}$$

This function can be maximized using for example Newton's method. After having found $\hat{r}$, we get $\hat{p} = \hat{r}/(\hat{r} + \bar{x}_K)$ and in terms of the Gamma parameters, $\hat{s} = \hat{p}/(1 - \hat{p})$.

6.2 Bootstrap test on Gamma mixture

To test for the goodness of fit of the Gamma-Poisson mixture model, the following procedure can be followed. The null hypothesis is $H_0 : H \equiv H_{s,r}$ for *some* $s, r > 0$, where $H_{s,r}$

Period	$N(\hat{\lambda})$	Unicomponent $[\underline{N}, \overline{N}]$	Gamma $[\underline{N}, \overline{N}]$
08.00-08.30	5	$[5, 6]$	$[4, 7]$
08.30-09.00	14	$[13, 14]$	$[10, 18]$
09.00-09.30	34	$[33, 34]$	$[25, 43]$
09.30-10.00	43	$[42, 44]$	$[30, 58]$
10.00-10.30	44	$[43, 45]$	$[30, 59]$
10.30-11.00	45	$[44, 46]$	$[33, 58]$
11.00-11.30	44	$[43, 45]$	$[32, 58]$
11.30-12.00	44	$[43, 45]$	$[33, 55]$
12.00-12.30	37	$[36, 38]$	$[28, 48]$

Table 2: *Numbers of agents associated to the numbers in table 1, obtained by the Erlang formula with $\mu = 419$, for 80% in 20s. service level. E.g. $\underline{N}$ in the unicomponent case is obtained by applying the Erlang formula to the lower confidence limit for λ given in table 1 and $\underline{N}$ in the Gamma case is obtained by applying the Erlang formula to the estimated lower quantile of the mixing distribution. The values $\overline{N}$ are obtained similarly.*

denotes the distribution function with density function (5). As test statistic, one can take the supremum distance between the empirical distribution function $\hat{F}$ of the counts and the maximum likelihood fit $F_{\hat{r},\hat{s}}$ of the negative binomial distribution:

$$T = \sup_{x>0} |\hat{F}(x) - F_{\hat{r},\hat{s}}|.$$

Big values of this test statistic contradict the validity of the null hypothesis. Since the test statistic is not distribution free over the null hypothesis, we suggest to approximate its distribution using the bootstrap and Monte Carlo simulation.

Given our data $x_1, \ldots, x_K$, we estimate r and s via maximum likelihood. The bootstrap approximation is that we approximate the distribution of T under the null hypothesis by the distribution of T based on a sample from the (negative binomial) distribution $F_{\hat{r},\hat{s}}$. The Monte Carlo approximation to this distribution is obtained by estimating this distribution by the empirical distribution function of a large sample (of size B) from this distribution.

To draw *one* observation from this distribution, the following recipe can be used. First draw a sample of size K from $F_{\hat{r},\hat{s}}$, and denote this sample by $x_1^*, x_2^*, \ldots, x_K^*$. Then compute $\hat{r}^*$ and $\hat{s}^*$ based on this generated sample and compute the supremum distance between the empirical distribution function of $x_1^*, x_2^*, \ldots, x_K^*$ and the distribution function $F_{\hat{r}^*,\hat{s}^*}$. By repeating this procedure B times, one gets realizations $t_1^*, t_2^*, \ldots, t_B^*$ from the bootstrap approximation to the null distribution of T. Denoting by t the realization of T based on the original sample, the bootstrap approximation to the p-value can then be computed as

$$p\text{-value} \approx \frac{\#\{1 \leq i \leq B \,:\, t_i^* \geq t\}}{B}$$

where $\#V$ denotes the number of elements in a set V. Figure 2 shows the empirical distribution function of the 500 T^*-values together with the observed value $t = 0.145$ of T for the period 9.00-9.30.

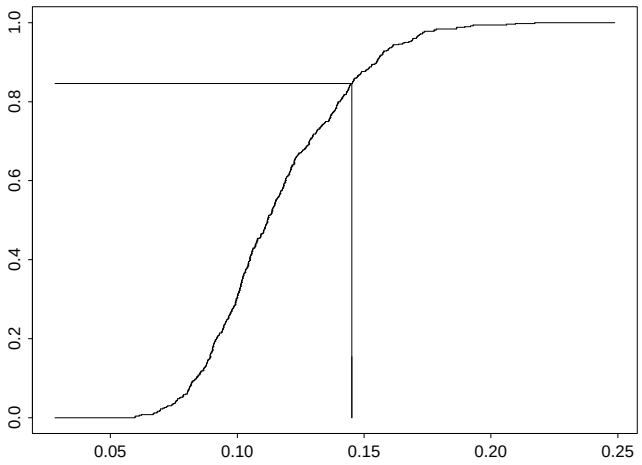

Figure 2: *Empirical distribution function of the bootstrap sample. The approximate p-value is* 0.154.

6.3 The data

Table 3 gives the dataset used in the case study. Figure 3 gives the boxplot of the columns of table 3.

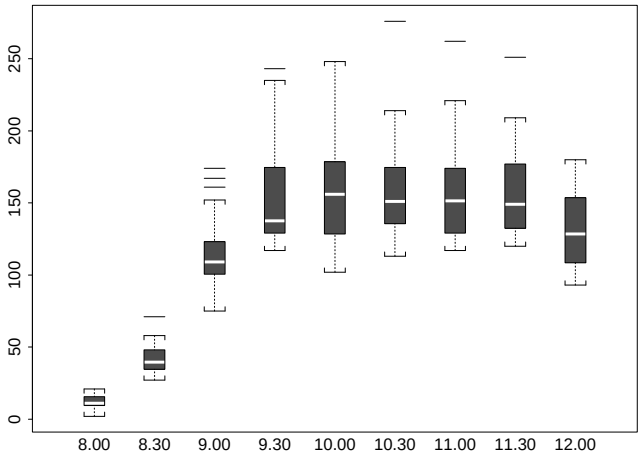

Figure 3: *Each boxplot corresponds to the half-hour period starting at the time below it.*

8.00	8.30	9.00	9.30	10.00	10.30	11.00	11.30	12.00
18	44	161	218	248	276	262	251	180
9	27	119	173	183	177	185	209	180
14	35	121	129	168	148	152	167	164
8	42	100	176	158	154	154	159	134
12	31	75	117	140	116	161	152	131
21	58	146	200	212	195	213	184	166
10	35	110	138	159	158	151	170	130
2	37	96	123	116	125	117	151	122
4	37	108	119	102	136	127	120	97
9	41	106	137	128	139	152	124	101
16	51	167	235	197	205	188	184	134
19	32	112	140	166	135	128	140	104
7	34	101	130	127	155	141	131	130
11	38	99	129	133	148	117	137	98
10	27	106	133	115	132	121	127	93
17	71	174	209	192	204	221	185	168
12	43	124	137	174	167	147	152	115
10	38	103	126	136	132	163	147	117
11	28	98	146	126	127	117	121	93
10	28	100	134	121	139	137	133	108
15	58	152	243	218	206	204	196	174
9	42	122	144	161	157	152	142	144
11	57	91	124	154	144	141	139	127
17	50	102	130	129	113	130	143	127
13	43	107	125	132	146	120	132	121
19	58	149	231	213	214	199	197	163
14	46	116	145	174	165	159	157	141
10	38	113	164	153	172	148	125	109

Table 3: *Dataset containing numbers of incoming calls on* 28 *weekdays during different time periods on the day.*

References

[1] D. Böhning. Numerical estimation of a probability measure. Journal of Statistical Planning and Inference 11, 57–69, 1985

[2] D. Böhning, P. Schlattmann and B. Lindsay. Computer-assisted analysis of mixtures (C.A.MAN): statistical algorithms. Biometrics 48, 283–303, 1992. (`http://www.medizin.fu-berlin.de/sozmed/caman.html`)

[3] A. Brandt, M. Brandt, G. Spahl, and D. Weber. Modelling and optimization of call distribution systems. In V. Ramaswami and P.E. Wirth, editors, *Proceedings of the 15th International Teletraffic Conference*, pages 133–144. Elsevier, 1997.

[4] B. Cleveland and J. Mayben. *Call Center Management on Fast Forward*. Call Center Press, 1997.

[5] D. Gross and C.M. Harris. *Fundamentals of Queueing Theory*. Wiley, 2nd edition, 1985.

[6] J. Kiefer and J. Wolfowitz. Consistency of the maximum likelihood estimator in the presence of infinitely many incidental parameters. Annals of Mathematical Statistics 27, 886–906, 1956.

[7] B.G. Lindsay. The geometry of mixture likelihoods: a general theory. The Annals of Statistics 11, 86–94, 1983.

[8] B.G. Lindsay. *Mixture models: theory, geometry and applications*. NSF-CBMS Regional Conference Series in Probability and Statistics. Volume 5, 1995.

[9] B.G. Lindsay and K. Roeder. Residual diagnostics in the mixture model. Journal of the American Statistical Association 87, 785–795, 1992.

[10] A. Mandelbaum, W.A. Massey, M.I. Reiman, and R. Rider. Time varying multiserver queues with abandonments and retrials. In P. Key and D. Smith, editors, *Proceedings of the 16th International Teletraffic Conference*, 1999.

[11] S.M. Ross. *Stochastic Processes*. Wiley, 1983.

[12] A.W. van der Vaart. *Asymptotic Statistics*. Cambridge University Press, 1998.