MANAGEMENT SCIENCE
Vol. 50, No. 7, July 2004, pp. 896–908
ISSN 0025-1909 | EISSN 1526-5501 | 04 | 5007 | 0896

informs

DOI 10.1287/mnsc.1040.0236

Modeling Daily Arrivals to a Telephone Call Center

Athanassios N. Avramidis, Alexandre Deslauriers, Pierre L'Ecuyer
GERAD and Département d'Informatique et de Recherche Opérationnelle, Université de Montréal,
C.P. 6128, succursale Centre-ville, Montréal, Québec, H3C 3J7 Canada,
avramidi@iro.umontreal.ca

We develop stochastic models of time-dependent arrivals, with focus on the application to call centers. Our models reproduce three essential features of call center arrivals observed in recent empirical studies: a variance larger than the mean for the number of arrivals in any given time interval, a time-varying arrival intensity over the course of a day, and nonzero correlation between the arrival counts in different periods within the same day. For each of the new models, we characterize the joint distribution of the vector of arrival counts, with particular focus on characterizing how the new models are more flexible than standard or previously proposed models. We report empirical results from a study on arrival data from a real-life call center, including the essential features of the arrival process, the goodness of fit of the estimated models, and the sensitivity of various simulated performance measures of the call center to the choice of arrival process model.

Key words: call center; arrival process; multivariate distribution; doubly stochastic Poisson process; input modeling; correlation
History: Accepted by Paul Glasserman, stochastic models and simulation; received February 27, 2003. This paper was with the authors 3 months for 2 revisions.

1. Introduction

Telephone call centers have become an integral part of the operations of many large organizations. With centers' growing presence and importance in the organization, managing their operations more efficiently has become an issue of significant economic interest (Gans et al. 2003). In modeling and analyzing call centers with quantitative methods, an important issue is the modeling of external customer call volume (we use equivalently the term *demand*). This demand involves uncertainty and should thus be studied with the appropriate statistical and stochastic techniques.

In this paper, we develop and study statistical models of the arrival process. Our development is influenced by two properties of call center arrivals observed in recent empirical studies, namely, a variance that is considerably higher than implied by Poisson arrivals (Jongbloed and Koole 2001) and strong positive association between the arrivals in different periods within the same day (Tanir and Booth 1999). The models extend those proposed recently by Jongbloed and Koole (2001) and Whitt (1999).

Brown and coauthors (2002) have also observed positive correlation between the demands of successive days in call centers and have developed models that account for this dependency via a time series component. This type of model can provide better short-term (day-to-day) forecasting of the demand than a model that assumes independence between days. Here we focus on intraday correlations only. The initial motivation for our models was to develop simulation tools for staffing and scheduling two weeks in advance in a call center. (For the centers we were involved with, union agreements specified that managers would provide working schedules two weeks in advance. Such constraints are frequent.) In this context, the demand of the 14 days that precede the target day is unknown, and a 14-day lag forecast of demand for the target day based on time-series models would have very little predictive power (time-series analysis of our data set has confirmed this assertion). In short, staffing decisions in call centers may have to be made far in advance so that the target day may be reasonably viewed as being independent of the current information set.

Our aim is to build models of call centers that permit one to estimate the expected value of certain performance measures (proportion of customers whose waiting time exceeds a given threshold, proportion of calls lost due to customer's abandonment, etc.) for a given daily staffing via simulation. We seek a valid model of the arrival process that is faithful enough to reproduce the behavior of interest in the system. A stronger positive correlation between demands of successive periods during the day favors queue buildup and therefore has an important impact on

the performance measures. On the other hand, queues do not build up across successive days. Modeling the dependencies at that level can also be important because it permits one to better forecast tomorrow's demand. But this is not our purpose here. On the other hand, our models can be combined with a time-series component for successive days, which we briefly discuss at the end of the paper.

We will provide both theoretical and empirical evidence of the improved modeling ability of the new models within a day. The theoretical evidence is a characterization of how the new models are more flexible in modeling the variances and correlations of the intraday arrival counts. Our empirical evidence is an improved ability of the new models to reasonably approximate the empirical means, variances, and correlations simultaneously on our data set and to match the behavior of a simulation model "bootstrapped" by our arrival data. We study empirically the sensitivity of various call center performance measures to the choice of arrival process model. The sensitivity analysis was executed via a simulation model of a Bell Canada call center that was developed, validated, and documented in Deslauriers (2003).

It is also important that the number of parameters that we have to estimate in the new models remains small, to avoid the danger of overfitting. As it turns out, our models have fewer parameters than the one proposed by Jongbloed and Koole (2001).

Although the model development was motivated by a specific call center application, stochastic models of arrival counts capturing simultaneously a time-varying arrival intensity and intraperiod correlation may be useful in many other application areas. As an example, we sketch the significance of such models for modeling and forecasting the arrival process of reservation requests for seats in a flight. Such arrival processes are typically observed by airlines as part of yield-management practice, which aims to manage (that is, sell at an attractive price) the inventory of seats. In typical airline yield-management practice, an important component is the model that forecasts future demand (reservation requests). The arrival process typically occurs over a fixed time horizon (typically the time period starting 360 days before the departure of a flight) and occurs repeatedly over many calendar days. In this setting, the feature of time-varying arrival intensity is well known and modeled by reservation profiles, which are estimates of the time-varying arrival rate. Such profiles are routinely computed by airline yield-management departments or airline yield-management system providers (Smith et al. 1992). The feature of correlation between the arrival counts in different time periods over the reservation horizon is less well understood and modeled, partly because of the increased modeling complexity required. However, it is clear that the presence of nonzero correlation can be leveraged for forecasting future demand based on the past (already realized) demand, in a manner similar to that discussed under *Further Model Properties* in §3.1.

Other potential applications are for modeling arrivals at ticket offices (for cinemas, museums, etc.), bus stops and subway stations, fast-food restaurants, and so on, where the arrival process is likely to behave in a similar way as for call centers.

This paper is organized as follows. Section 2 contains an introduction to some empirical aspects of call center arrival patterns and reviews previous work on modeling call center arrivals. In §3 we develop and study three new models. In §4 we describe a case study with arrival data from an actual call center, including various data-specific findings, the empirical quality of the fitted models, and the sensitivity of various performance measures of the call center to the choice of alternative models of the arrival process. Section 5 contains our conclusions and a perspective on future application of this work.

2. Background

Four properties of call center arrival processes have emerged in recent studies:

Property 1. *The total daily demand* (*number of calls*) *has overdispersion relative to the Poisson distribution* (*the variance is greater than the mean*) (*Jongbloed and Koole* 2001, *Deslauriers* 2003).

Property 2. *The arrival rate varies considerably with the time of day* (*see Tanir and Booth* 1999, *p*. 1643; *Deslauriers* 2003).

Property 3. *There is strong positive association* (*correlation*) *between arrival counts in a time partition of a day*.

Property 4. *There is significant dependency between arrival counts on successive days* (*Brown et al*. 2002).

The standard nonhomogeneous Poisson process (that is, a Poisson process with a deterministic arrival rate function), referred to in the future as model NHPP, is inconsistent with both Properties 1 and 3. In view of Property 1, Jongbloed and Koole (2001) proposed a doubly stochastic model under which arrivals follow a standard Poisson process with a random arrival rate. They model the rate as a gamma random variable, which results in the number of arrivals N being a negative binomial random variable. To model a time-varying arrival rate in their application, these authors estimated independent versions of the model for time periods having a priori different arrival rates. That is, in the model of Jongbloed and Koole (2001), referred to henceforth as Model 0, the different time periods are randomized separately by independent

gamma variables, and thus the correlations between the arrival counts in different time periods within the same day are constrained to be zero. On the issue of correlations, we quote the authors: "Details on the correlation between call volume in different intervals fall outside the scope of this paper" (Jongbloed and Koole 2001, p. 315).

To address the time-varying arrival rate while allowing nonzero correlations, Whitt (1999) proposed a doubly stochastic Poisson process model where the arrival rate function over a day is of the form $\Lambda(t) = Wf(t)$, where the only random quantity is W, a real-valued random variable. This W can be interpreted as the (unpredictable) "busyness" of a day, whereas $f(t)$ models the time-varying arrival intensity during the day. The presence and significance of Property 3 are less well known but became apparent in our case study. The new models introduced in the next section capture simultaneously Properties 1–3. Our models do not capture Property 4 because they are for a single day, as explained in the introduction.

Note that a common practice in call center management is to divide the day into equal (e.g., 30-minute) periods; assume that the arrival rate over period i is $w\lambda_i$, where the λ_i are prespecified constants and the real number w is a "guess" (or "estimate") of the "busyness" of that day; and use Erlang formulas to approximate performance measures of the system over each period. These formulas effectively assume that the system is in steady state with arrival rate $w\lambda_i$ for each i. This is somewhat related to Whitt's model, but a major difference is that w is a constant factor in one case and a random variable in the other case.

3. New Models of Arrival Counts

3.1. Model 1: Doubly Stochastic Poisson Model

We consider here a special case of the model proposed by Whitt (1999) and introduced in §2, where we make the particular assumption that the factor W is gamma distributed. The motivation for considering this special case is that the joint distribution of arrival counts turns out to be analytically tractable; it is the negative multinomial distribution. This leads to straightforward model estimation, variate generation, and an analytical expression for the conditional mean function.

Arrivals follow a Poisson process with random arrival rate function $\Lambda(t)$, $t_S \le t \le t_E$, where t_S and t_E are the time points in a day when operations begin and end, respectively. Model 1 postulates that the arrival rate function is randomized by a gamma variable:

$$\text{Model 1:}\quad \Lambda(t) = Wf(t), \qquad W \sim \text{Gamma}(\gamma, 1), \quad (1)$$

where f is nonnegative and integrable on (t_S, t_E) and characterizes the time variation of the arrival rate over a day; Gamma$(\gamma, 1)$ is the gamma distribution with shape parameter $\gamma > 0$ and scale parameter 1. The function f captures the scale of the arrival rate, so there is no loss in generality by taking the gamma scale parameter as one.

Let $t_S = t_0 < t_1 < t_2 < \cdots < t_{k-1} < t_k = t_E$ denote a partition of (t_S, t_E). In applications, the partition will be chosen based on considerations such as data availability for model estimation, and the a priori known approximate behavior of the arrival intensity profile $f(\cdot)$. Define the random vector of arrival counts $\mathbf{X} = (X_1, X_2, \ldots, X_k)$, where X_i is the number of arrivals in the time interval $[t_{i-1}, t_i)$, $i = 1, \ldots, k$, and define the total daily demand $Y = \sum_{i=1}^k X_i$. Let

$$\lambda_i = \int_{t_{i-1}}^{t_i} f(t)\,dt, \quad i = 1, \ldots, k.$$

Proposition 1 characterizes the distribution of the vector $\mathbf{X}$ as a negative multinomial distribution with parameters $(\gamma, \lambda_1, \ldots, \lambda_k)$. The probability mass function is

$$P[\mathbf{X} = (x_1, x_2, \ldots, x_k)] = \frac{\Gamma(\gamma + \sum_{i=1}^k x_i)}{\Gamma(\gamma)\prod_{i=1}^k x_i!}\left(\frac{1}{1+\sum_{j=1}^k \lambda_j}\right)^{\gamma}\prod_{i=1}^k\left(\frac{\lambda_i}{1+\sum_{j=1}^k \lambda_j}\right)^{x_i} \quad (2)$$

for $(x_1, \ldots, x_k) \in \{0, 1, 2, \ldots\}^k$; see Johnson and Kotz (1969, p. 292) for an account of this distribution. The negative multinomial distribution is a multivariate generalization of the negative binomial distribution, where both distributions' most general definition allows the parameter γ to be positive real valued. For the negative binomial, we use the parameterization of Johnson and Kotz (1969) (which differs from that of Shao 1999, for example), so a negative binomial random variable with parameters γ and λ has mean $\gamma\lambda$ and variance $\gamma\lambda(1+\lambda)$. In the special case where γ is a positive integer, the negative multinomial has the following intuitive interpretation. Consider a sequence of independent trials, where each trial has $k+1$ possible outcomes and the probability of occurrence of outcome i is

$$p_i = \frac{\lambda_i}{1+\sum_{i=1}^k \lambda_i}, \qquad i = 1, \ldots, k, \quad (3)$$

and $p_{k+1} = 1 - \sum_{i=1}^k p_i$. We perform trials until exactly γ occurrences of outcome $k+1$ are observed, and we let X_i be the number of occurrences of outcome i, $i = 1, \ldots, k$. In this setting, we see immediately that the vector $(X_1, X_2, \ldots, X_k)$ has mass function (2). We denote the distribution defined in (2) as NegMult$(\gamma, p_1, p_2, \ldots, p_k, p_{k+1})$. Let CV$(Z)$ denote the coefficient of variation (CV) of a random variable Z.

PROPOSITION 1. (a) *Under Model* 1, *the arrival count vector* $\mathbf{X}$ *has the negative multinomial distribution with parameters* $(\gamma, \lambda_1, \ldots, \lambda_k)$. *The marginal distribution of* X_i *is negative binomial with parameters* (γ, λ_i) *and the marginal distribution of* Y *is negative binomial with parameters* $(\gamma, \sum_{i=1}^k \lambda_i)$.

(b) *The conditional distribution of* $\mathbf{X}$ *given* $Y = y$ *is multinomial with* y *trials and success probabilities* $\lambda_i / \sum_{j=1}^k \lambda_j$.

(c) *Regardless of the choice of model parameters* λ_i, *the coefficients of variation of the* X_i *and* Y *are constrained as follows*:

$$\mathrm{CV}^2(X_i) - \frac{1}{\mathrm{E}X_i} = \mathrm{CV}^2(Y) - \frac{1}{\mathrm{E}Y} = \frac{1}{\gamma} \quad \textit{for all } i. \tag{4}$$

(d) *The correlation between* X_i *and* X_j, *expressed in terms of the means* $\mathrm{E}X_i$, $\mathrm{E}X_j$, *and the parameter* γ *is*

$$\rho^{(1)}_{i,j}(\gamma) = \frac{1}{\sqrt{(1+\gamma(\mathrm{E}X_i)^{-1})(1+\gamma(\mathrm{E}X_j)^{-1})}}. \tag{5}$$

PROOF. Conditional on W, the components X_i of the vector $\mathbf{X}$ are independent Poisson random variables with rates $W\lambda_i$, respectively; this is a consequence of the property of independent increments of a Poisson process. The probability mass function of $\mathbf{X}$ can be written in closed form:

$$\begin{aligned} f^{(1)}(x_1, x_2, \ldots, x_k) &= \int_0^\infty \prod_{i=1}^k \left(\frac{(\lambda_i w)^{x_i} e^{-\lambda_i w}}{x_i!}\right)\left(\frac{w^{\gamma-1}e^{-w}}{\Gamma(\gamma)}\right) dw \\ &= \frac{\prod_{i=1}^k \lambda_i^{x_i}}{\Gamma(\gamma)\prod_{i=1}^k x_i!} \int_0^\infty w^{\sum_{i=1}^k x_i + \gamma - 1} e^{-w\left(\sum_{i=1}^k \lambda_i + 1\right)} dw \\ &= \frac{\left(\prod_{i=1}^k \lambda_i^{x_i}\right)\Gamma\left(\sum_{i=1}^k x_i + \gamma\right)}{\Gamma(\gamma)\left(\prod_{i=1}^k x_i!\right)\left(1+\sum_{i=1}^k \lambda_i\right)^{\sum_{i=1}^k x_i + \gamma}}, \end{aligned}$$

which can be rewritten as (2). We note that the integration argument above can be used to derive each of the marginals of X_i and Y, proving that they are all negative binomial. This proves (a). To prove (b), we record the negative binomial mass function of Y, where for notational simplicity we use the parameters p_i instead of λ_i:

$$f_Y(y) = \frac{\Gamma(\gamma+y)}{\Gamma(\gamma)y!}\left(\sum_{i=1}^k p_i\right)^y \left(1 - \sum_{i=1}^k p_i\right)^\gamma. \tag{6}$$

The required conditional distribution is the quotient of (2) over (6); that is:

$$\begin{aligned} f_{\mathbf{X}|Y}(x_1, \ldots, x_k \mid y) &= \frac{(\Gamma(\gamma+\sum_{i=1}^k x_i)/\Gamma(\gamma)\prod_{i=1}^k x_i!)\prod_{i=1}^k p_i^{x_i}(1-\sum_{i=1}^k p_i)^\gamma}{((\Gamma(\gamma+y))/(\Gamma(\gamma)y!))(\sum_{i=1}^k p_i)^y(1-\sum_{i=1}^k p_i)^\gamma} \\ &= \frac{y!}{\prod_{i=1}^k x_i!}\prod_{i=1}^k \left(\frac{p_i}{\sum_{i=1}^k p_i}\right)^{x_i}, \end{aligned} \tag{7}$$

proving (b). Result (c) follows by direct calculation invoking the mean and variance of the negative binomial distribution. Result (d) is a known property of the negative multinomial distribution (Johnson and Kotz 1969, p. 295). □

REMARK 1. Item (c) characterizes how Model 1 generalizes model NHPP in terms of the variances of X_i; under NHPP, the quantities in the left and middle of display (4) are constrained to equal zero. Item (d) characterizes how Model 1 generalizes Model 0 in terms of the correlation between X_i and X_j with $i \neq j$; under Model 0, this correlation is constrained to equal zero for all $i \neq j$.

Further Model Properties. In this paragraph we list further properties of the negative multinomial distribution. The reader interested in derivations or other properties not listed here may consult Johnson and Kotz (1969). We note that the variance of each marginal distribution (of the X_i as well as Y) is higher than the mean (Property 1); moreover, a variance less than or equal to the mean cannot be induced by this model. Note that the correlations are positive. The conditional distributions, given any subset of the variates, are also negative multinomials. Thus, Model 1 yields distributional forecasts of the remaining demand (rather than point forecasts), given the observed demand up to a given time point; such forecasts may have substantial value in short-term planning decisions (Gans et al. 2003). In particular, the mean of the conditional distribution of X_j given any subset of the X_i is a linear function of the sum of the X_i:

$$\mathrm{E}\left[X_j \mid X_{i_1}, X_{i_2}, \ldots, X_{i_m}\right] = \frac{\lambda_j}{1+\sum_{l=1}^m \lambda_{i_l}}\left(\gamma + \sum_{l=1}^m X_{i_l}\right) \quad \text{for } j \neq i_1, i_2, \ldots, i_m. \tag{8}$$

Parameter Estimation. Let $\{\mathbf{X}_j = (X_{1,j}, X_{2,j}, \ldots, X_{k,j})\}_{j=1}^n$ be a sample of independent and identically distributed observations of the vector $\mathbf{X}$. The maximum likelihood estimators (MLEs) of the parameters of the negative multinomial distribution satisfy the following equations (Johnson and Kotz 1969):

$$\sum_{l=1}^M (\hat{\gamma}+l-1)^{-1} F_l = \log\left(1 + \frac{1}{n\hat{\gamma}}\sum_{j=1}^n Y_j\right) \tag{9}$$

and

$$\hat{\lambda}_i = \frac{\sum_{j=1}^n X_{i,j}}{n\hat{\gamma}} \qquad \text{for } i = 1, \ldots, k, \tag{10}$$

where

$$Y_j = \sum_{i=1}^k X_{i,j} \qquad \text{for } j = 1, \ldots, n,$$

$$F_l = \frac{1}{n}\sum_{j=1}^n \mathbf{1}\{Y_j \geq l\} \qquad \text{for } l = 1, \ldots, M.$$

$\mathbf{1}\{\cdot\}$ denotes the indicator function, and $M=\max_j\{Y_j\}$. Solving the nonlinear equation (9) will typically require a numerical solver. Then the value of $\hat{\gamma}$ is simply plugged into (10).

3.2. Model 2: Seeking a More Flexible Covariance Matrix

In our case study, the correlations corresponding to the estimated Model 1 were too high relative to the sample correlations (see §4.2). This motivated the need to explore models that allow a richer covariance structure for $\mathbf{X}$.

We first considered modeling $\mathbf{X}$ by a multinomial distribution with a fixed number of trials and success probabilities that may be either constant or random. Under this type of model, the sum of the components is equal to the number of trials and is a parameter instead of being random, as required in our setting; moreover, the correlations are always negative (Mosimann 1963), which is inconsistent with the empirical evidence. A more general class of models is obtained by allowing the number of trials in the multinomial model to be random; Model 1 is a special case with analytically tractable properties, as Proposition 1 showed.

A second class of multivariate discrete distributions considered was the compound negative multinomial distribution (Mosimann 1963), which generalizes the negative multinomial distribution (i.e., it generalizes Model 1) by allowing the parameters p_i, $i=1,\ldots,k$, in (3) to be random. Under the compound negative multinomial distribution, the correlations supported are always positive (Mosimann 1963). The particular case where the vector $(p_1,p_2,\ldots,p_k,p_{k+1})$ has a Dirichlet distribution has been studied by Mosimann (1963), who derived the mass function and moments in closed form. We recall that the Dirichlet distribution $\mathcal{D}$ with parameters $\alpha_i>0$, $i=1,\ldots,k$, is a multivariate generalization of the beta distribution. Its density function is

$$f_{\mathbf{Q}}(q_1,\ldots,q_k)=\frac{\Gamma(\alpha_0)}{\prod_{i=1}^{k}\Gamma(\alpha_i)}\prod_{i=1}^{k}q_i^{\alpha_i-1} \tag{11}$$

over the simplex $\{(q_1,\ldots,q_k)\colon q_i\ge 0$ for each i and $q_1+\cdots+q_k=1\}$ and zero elsewhere, where $\alpha_0=\sum_{i=1}^{k}\alpha_i$. Its genesis is as follows. Let $Z_1,\ldots,Z_k$ be independent random variables where Z_i has the Gamma$(\alpha_i,2)$ distribution for each i (when α_i is integer, this is the χ^2 distribution with $2\alpha_i$ degrees of freedom). Then the distribution of $(Z_1,\ldots,Z_k)/\sum_{j=1}^{k}Z_j$ is Dirichlet with parameters $(\alpha_1,\ldots,\alpha_k)$. In particular, we have $\mathrm{E}Q_i=\alpha_i/\alpha_0$, $\mathrm{Var}(Q_i)=(\alpha_i/\alpha_0)(1-\alpha_i/\alpha_0)/(\alpha_0+1)$, and $\mathrm{Cov}(Q_i,Q_j)=-(\alpha_i\alpha_j)/(\alpha_0^2(\alpha_0+1))$. For a complete account of this distribution, see Johnson and Kotz (1969).

We define

$$\text{Model 2:}\quad \left.\begin{aligned}&(p_1,p_2,\ldots,p_k,p_{k+1})\sim\mathcal{D}(\beta_1,\ldots,\beta_{k+1}),\\ &\mathbf{X}\sim\mathrm{NegMult}(\nu,p_1,p_2,\ldots,p_k,p_{k+1}).\end{aligned}\right\} \tag{12}$$

The conditional distributions, given any subset of components of $\mathbf{X}$, are of the same form; that is, they are Dirichlet compound negative multinomial (see Johnson and Kotz 1969, p. 312). Thus the results of Mosimann (1963) yield closed-form expressions for the conditional means. To generate $\mathbf{X}$ for this model, once ν and the Dirichlet parameters have been estimated, first generate the vector $(p_1,p_2,\ldots,p_k,p_{k+1})$, compute the corresponding $\lambda_1,\ldots,\lambda_k$ by inverting (3) (this gives $\lambda_i=p_i/p_{k+1}$), then generate W and $\mathbf{X}$ just as in Model 1. (This actually gives a method for generating variates from a compound negative multinomial distribution.) Generating a vector from the Dirichlet distribution can be accomplished directly from the discussed genesis of the distribution following (11), if we assume the availability of a generator of gamma random variables.

3.3. Model 3: A Different Type of Flexibility

Our next model, denoted Model 3, will be shown to be more flexible than all other models discussed so far in certain aspects, including the range of induced correlations. A summary of this is contained in Table 1. To introduce Model 3, we start by defining the vector of ratios

$$\mathbf{Q}\equiv(Q_1,Q_2,\ldots,Q_k)\equiv(X_1/Y,X_2/Y,\ldots,X_k/Y). \tag{13}$$

We assume $\mathbf{Q}$ is independent of Y, effectively postulating that the assignment of total daily demand to time intervals follows a mechanism that is statistically independent from the daily volume. As a model for $\mathbf{Q}$, we use the Dirichlet distribution. This gives

$$\text{Model 3:}\quad \left.\begin{aligned}&Y\sim G,\\ &\mathbf{Q}\sim\mathcal{D}(\alpha_1,\ldots,\alpha_k)\quad\text{and}\\ &\qquad\text{independent of } Y,\\ &\widetilde{\mathbf{X}}\equiv(\widetilde{X}_i)_{i=1}^{k}=Y\mathbf{Q},\quad\text{and}\\ &\mathbf{X}\equiv(X_i)_{i=1}^{k}=[\widetilde{\mathbf{X}}],\end{aligned}\right\} \tag{14}$$

where G is an unspecified univariate distribution with mean μ_Y and variance σ_Y^2 and $[\widetilde{\mathbf{X}}]$ denotes the componentwise rounding of $\widetilde{\mathbf{X}}$ to the closest integer. This model has a substantially different genesis than Model 1: It does not arise as a Poisson process (standard or doubly stochastic) and only specifies the distribution of the arrival-count vector $\mathbf{X}$, without specifying a model for the interarrival times. As a model

for interarrival times, we adopt the following natural approach. Conditional on $\mathbf{X}$, the X_i arrivals occurring in interval i are distributed uniformly on $[t_{i-1}, t_i)$, as if the arrival count vector $\mathbf{X}$ were generated by a Poisson process.

Model 3, in modeling the discrete vector $\mathbf{X}$ by rounding the continuous vector $\widetilde{\mathbf{X}}$ to the closest integer, is not consistent with a counting process, and this is a theoretically unattractive feature. However, in many applications, the number of arrivals in time intervals of practical modeling interest is large, and in this case the distributions of $\mathbf{X}$ and $\widetilde{\mathbf{X}}$ will be indistinguishable for practical purposes. In the remainder of the analysis of Model 3, we do not distinguish between these two objects and denote them both as $\mathbf{X}$.

Model Properties. Regardless of the specification of the distribution G in (14), the marginal and conditional distributions of the X_i do not appear to correspond to any distribution with analytically known properties. However, the moments of X_i follow easily from the moments of the distribution G and those of the Dirichlet distribution. We have means

$$\mathrm{E}X_i = \mu_Y \frac{\alpha_i}{\alpha_0}, \tag{15}$$

variances

$$\mathrm{Var}(X_i) = \mathrm{E}Y^2 \frac{\alpha_i(\alpha_0-\alpha_i)}{\alpha_0^2(\alpha_0+1)} + \sigma_Y^2 \frac{\alpha_i^2}{\alpha_0^2}, \tag{16}$$

and covariances

$$\begin{aligned}\mathrm{Cov}(X_i, X_j) &= \mathrm{E}[\mathrm{Cov}(X_i, X_j \mid Y)] \\ &\quad + \mathrm{Cov}(\mathrm{E}[X_i \mid Y], \mathrm{E}[X_j \mid Y]) \\ &= \mathrm{E}\left[Y^2 \frac{-\alpha_i\alpha_j}{\alpha_0^2(\alpha_0+1)}\right] + \mathrm{Cov}\left(Y\frac{\alpha_i}{\alpha_0}, Y\frac{\alpha_j}{\alpha_0}\right) \\ &= \mathrm{E}Y^2 \frac{-\alpha_i\alpha_j}{\alpha_0^2(\alpha_0+1)} + \sigma_Y^2 \frac{\alpha_i\alpha_j}{\alpha_0^2} \\ &= \frac{\alpha_i\alpha_j}{\alpha_0^2}\left(\sigma_Y^2 - \frac{\mathrm{E}Y^2}{\alpha_0+1}\right), \quad i \neq j, \end{aligned} \tag{17}$$

where we used the known moments of the Dirichlet distribution. Proposition 2 below characterizes the marginal variances and the correlation structure of Model 3.

Proposition 2. *Under Model* 3:

(a) *Regardless of the choice of model parameters α_i, the coefficients of variation of the X_i are constrained as follows:*

$$\frac{\mathrm{CV}^2(X_i) - \mathrm{CV}^2(Y)}{1/\mathrm{E}X_i - 1/\mu_Y} = \frac{\mu_Y(1+\mathrm{CV}^2(Y))}{\alpha_0+1} \quad \textit{for all } i. \tag{18}$$

(b) *The correlation between X_i and X_j, expressed in terms of the means $\mathrm{E}X_i$ and $\mathrm{E}X_j$, is*

$$\begin{aligned}\rho_{i,j}^{(2)}(\alpha_0) = &\left(1 + \frac{\mu_Y \mathrm{E}Y^2/\mathrm{E}X_i}{(\alpha_0+1)\sigma_Y^2 - \mathrm{E}Y^2}\right)^{-1/2} \\ &\cdot \left(1 + \frac{\mu_Y \mathrm{E}Y^2/\mathrm{E}X_j}{(\alpha_0+1)\sigma_Y^2 - \mathrm{E}Y^2}\right)^{-1/2}\end{aligned} \tag{19}$$

for all $i \neq j$. The function $\rho_{i,j}^{(2)}(\alpha_0)$ is continuous in α_0, negative for $\alpha_0 \in (0, \mu_Y^2/\sigma_Y^2)$, and positive and strictly increasing for $\alpha_0 \in (\mu_Y^2/\sigma_Y^2, \infty)$. We have $\rho_{i,j}^{(2)}(\mu_Y^2/\sigma_Y^2) = 0$,

$$\begin{aligned}\lim_{\alpha_0 \to 0} \rho_{i,j}^{(2)}(\alpha_0) = &-\left[\frac{\sigma_Y^2}{\mu_Y^2} + \left(1 + \frac{\sigma_Y^2}{\mu_Y^2}\right)\left(\frac{\mu_Y}{\mathrm{E}X_i} - 1\right)\right]^{-1/2} \\ &\cdot \left[\frac{\sigma_Y^2}{\mu_Y^2} + \left(1 + \frac{\sigma_Y^2}{\mu_Y^2}\right)\left(\frac{\mu_Y}{\mathrm{E}X_j} - 1\right)\right]^{-1/2}\end{aligned}$$

and $\lim_{\alpha_0 \to \infty} \rho_{i,j}^{(2)}(\alpha_0) = 1$ for all $i \neq j$.

Proof. We prove Proposition 2 by direct manipulation of the moments in (15), (16), and (17) and standard calculus. □

Parameter Estimation. In view of the assumed independence of Y and the vector of ratios $(X_1/Y, \ldots, X_k/Y)$, the estimation problem for Model 3 decomposes into two separate estimation problems: estimation of the distribution G and estimation of the parameters $\alpha_1, \ldots, \alpha_k$. The Dirichlet density is given in (11), and maximum likelihood estimation based on a sample of independent, identically distributed observations $\{\mathbf{Q}_j\}_{j=1}^n$ is straightforward.

3.4. Comparison Between Models

Proposition 2(b) shows the increased flexibility of Model 3 compared to Models 1 and 2 in terms of correlations; in the latter models, the induced correlations are constrained to be nonnegative. Proposition 2(a) characterizes the increased flexibility of Model 3 compared to Models 1 and 2 in terms of variances, as we now explain. Under all three models regardless of the choice of model parameters, we have

$$\frac{\mathrm{CV}^2(X_i) - \mathrm{CV}^2(Y)}{1/\mathrm{E}X_i - 1/\mathrm{E}Y} = \delta \qquad \text{for all } i, \tag{20}$$

where δ is a constant that depends on the model parameters and whose explicit expression is given below. In other words, the ratios of excess coefficient of variation (CV) of each X_i relative to Y, normalized by the difference in the respective inverse means, are constrained to be equal across all i. Under Model 1, by rearranging terms in (4), we obtain $\delta = 1$. For Model 2, with $(p_1, p_2, \ldots, p_k, p_{k+1})$ distributed as $\mathcal{D}(l_1, \ldots, l_k, l_{k+1})$, a direct calculation based on the moments derived by Mosimann (1963) shows $\delta = (l_{k+1} - 1 + \nu)/(l_{k+1} - 2)$,

Table 1 **Comparison of Excess Dispersion of Y, Correlations, and δ, the Excess Dispersion of X_i Relative to Y, Under Various Models**

Model	$\mathrm{CV}^2(Y)-1/\mu_Y$	$\rho_{i,j}$	δ	Conditional distributions
NHPP	0	0	0	Indep. Poisson
Model 0	>0	0	1	Indep. negative binomial
Model 1	>0	>0	1	Negative multinomial
Model 2	>0	>0	>1	Dirichlet compound negative multinomial
Model 3	$>$, $=$, or <0	$>$, $=$, or <0	$>$, $=$, or <1, if $\mathrm{CV}^2(Y)-1/\mu_Y > -1$; <1, otherwise	No closed form

with the constraint $\delta > 1$ (the model's variances are finite only if $l_{k+1} > 2$). For Model 3, Proposition 2(a) established $\delta = \mu_Y(1+\mathrm{CV}^2(Y))/(\alpha_0+1)$. Thus, Model 3 with $\mu_Y(1+\mathrm{CV}^2(Y)) > 1$ allows, via α_0, values of δ on either side of one, and ranging anywhere in $(0, \mu_Y(1+\mathrm{CV}^2(Y)))$. Note that $\mu_Y(1+\mathrm{CV}^2(Y)) > 1$ is equivalent to $\mathrm{CV}^2(Y)-1/\mu_Y > -1$, which is a weaker condition than $\mathrm{CV}^2(Y)-1/\mu_Y > 0$, the condition on the distribution of Y that constrains both Models 1 and 2.

The above discussion immediately suggests a test of the null hypothesis that an arbitrary random vector $\mathbf{X}$ is distributed under Model 1 with the parameters left unspecified. We define the statistical functional of the distribution of $\mathbf{X}$,

$$\theta \equiv \frac{1}{k}\sum_{i=1}^{k}\left(\frac{\mathrm{CV}^2(X_i)-\mathrm{CV}^2(Y)}{1/\mathrm{E}X_i - 1/\mathrm{E}Y} - 1\right), \tag{21}$$

which measures the aggregate excess dispersion of the X_i relative to Model 1 (under Model 1, $\theta = 0$). We have $\theta = \delta - 1$ under Models 1 to 3, but not for the general case where $\mathbf{X}$ has an arbitrary distribution. Let $\hat{\theta}$ be a straightforward estimator of θ, obtained by replacing the means and CVs in the expression of θ by their sample counterparts. This $\hat{\theta}$ can be used to test the null hypothesis H_0: $\theta = 0$ against the alternatives H_1: $\theta > 0$ and H_2: $\theta < 0$, corresponding to the cases that the components of $\mathbf{X}$ have overdispersion or underdispersion relative to Model 1, respectively. The distribution of $\hat{\theta}$ is unknown, but one approach to executing this test is via bootstrapping methods. See §4.2 for the results of the test in our case study.

Table 1 summarizes certain properties of models NHPP, Models 0, 1, 2, and 3, specifying the increased flexibility achieved by the latter models. Estimation and variate generation (for simulation studies) are easy for all the models.

4. Case Study: A Bell Canada Call Center

For the typical call center, the available data on customer arrivals is the aggregate number of arrivals observed over short intraday time intervals (Gans et al. 2003). For the example considered here, the data correspond to 25 half-hour intervals between t_S = 8:00 a.m. and t_E = 8:30 p.m. on each of the five working days of the week. The data cover a period of a little less than one year.

In §4.1 we discuss the preliminary data analysis that confirmed Properties 1–3. Section 4.2 contains the model estimation and the empirical quality of model fit. In §4.3 we study empirically the sensitivity of various performance measures of the call center to the choice of arrival process model.

4.1. Preliminary Data Analysis

The hypothesis that the arrivals follow a standard Poisson process (allowing the general nonhomogeneous case) was immediately rejected. Specifically, the arrival counts in all time intervals show significant overdispersion relative to the Poisson distribution (Property 1). As an indication of the overdispersion, the daily total number of arrivals, Y, had sample mean 1,201 and sample variance 53,419; the marginal distributions of the X_i exhibited similar overdispersion. Deslauriers (2003) provides further empirical analysis.

Given the a priori knowledge that the traffic pattern varies substantially across the days of the week, we began the statistical analysis with a multivariate analysis of variance test for the multivariate (25-dimensional) vector $\mathbf{X}$. The statistical decision problem is to cluster the five populations corresponding to each day of the working week so that different clusters have a different mean vector $\mathbf{X}$ (in the statistically significant sense). The test's main results were as follows: (a) There are three statistically different populations; and (b) the best clustering of the five populations to three clusters is Monday, Friday, and the aggregate Tuesday/Wednesday/Thursday. In the remainder of the paper, all reported results correspond to the aggregate population Tuesday/Wednesday/Thursday, unless otherwise indicated. Moreover, the arrival rate is clearly time varying (Property 2), as evidenced by a multiple comparisons test via Tukey's studentized

Figure 1 **Scatter Plot $(X_-(m), X_+(m))$ for $m=4$ (10:00 a.m.)**

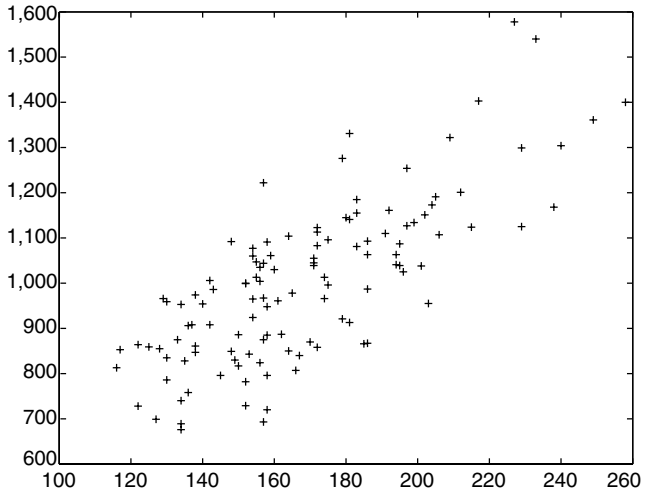
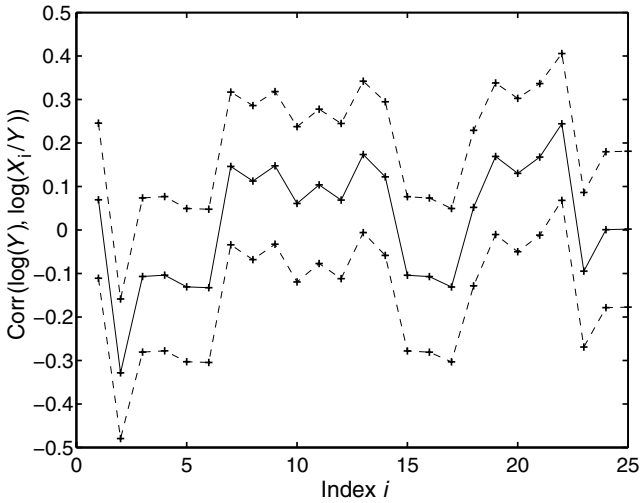


Figure 2 **Function $g(i) \equiv \text{Corr}(\log(Y), \log(X_i/Y))$, $i=1,\ldots,25$**



Note. (Solid line), point estimate; (dashed line), 95% confidence band. For each estimate $\hat{\rho}$, the confidence band is based on an asymptotic normal distribution of $0.5\log((1+\hat{\rho})/(1-\hat{\rho}))$, with an approximate variance equal to $1/(n-3)$, where n is the sample size.

range distribution. The details are omitted, in light of the well-known existence of this effect.

Figure 1 contains a scatter plot of the future demand $X_+(m) \equiv \sum_{i=m+1}^{25} X_i$ against the past demand $X_-(m) \equiv \sum_{i=1}^{m} X_i$ for $m=4$, corresponding to 10:00 a.m. The strong positive dependence between past and future demand within the same day is evident (Property 3). Scatter plots for other time points revealed a strong positive dependence of comparable magnitude.

4.2. Model Estimation and Empirical Quality of Fit

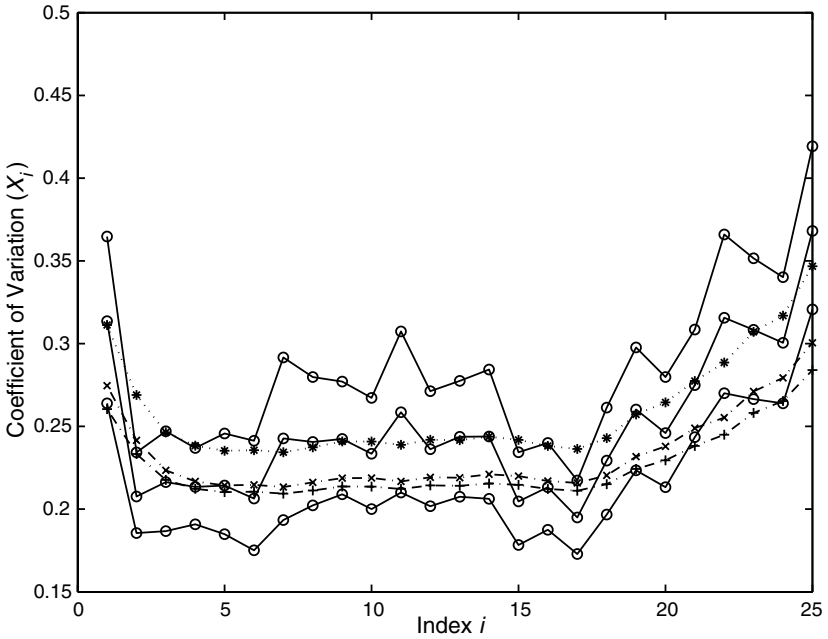
The hypothesis test discussed after Proposition 2 allows a modeler to test the appropriateness of Model 1 before proceeding into parameter estimation. This test was executed by inverting a nonparametric confidence interval for θ obtained by the hybrid bootstrap method (Shao 1999, p. 456). Based on a bootstrap sample of size 2,000, a 95% confidence interval for θ is (0.239, 0.568). (We verified by Monte Carlo simulation that the sample size 2,000 is sufficient for approximating the distribution of the bootstrapped statistic with negligible error.) The estimated positive θ suggests that the individual arrival counts X_i are more dispersed relative to the total arrival count Y than as predicted under Model 1. This gives some evidence against Model 1.

Model estimation was done for all three models via maximum likelihood (we completed the specification of Model 3 by taking G as the gamma distribution; fitting a negative binomial gave essentially identical results). We attempted to validate the hypothesis of independence between Y and $\mathbf{Q}$ as follows. Figure 2 depicts the sample correlation function $g(i) \equiv \text{Corr}(\log(Y), \log(X_i/Y))$, $i=1,\ldots,25$, with a 95% confidence interval at each value of i. (Note that the confidence band depicted does not correspond to a simultaneous confidence interval for all 25 values of i; such a simultaneous interval could be computed via the Bonferroni inequality and would be wider. This also applies to the other figures where confidence bands are depicted. The lines joining the observations are only visual artefacts to improve readability.) Two of those 25 confidence intervals (at $i=2$ and $i=22$) fail to cover zero. Thus, there is mild evidence against the independence assumption, but it is not strongly violated.

We compared the quality of model fit as follows. First, as expected, the fitted means under all models were essentially indistinguishable from the corresponding sample means. We thus concentrated on assessing the quality of fit by the CVs and the correlations of the X_i. Figure 3 compares the sample coefficients of variation of X_i, $i=1,\ldots,25$ (including a nonparametric 95% confidence interval for each i) to the exact values under the estimated models based on the same data set. The data provide evidence that Models 1 and 2 underestimate the CVs toward the end of the day, whereas Model 3 appears to overestimate the CVs at the beginning of the day, but to a lesser extent. Overall, no model is a clear winner with respect to "fitting" the coefficients of variation of the X_i.

The primary motivation for introducing the new models was in fact to better fit the correlation structure of the X_i. To assess the effectiveness of models in fitting these correlations we focus on the function $\rho(m) \equiv \text{Corr}(X_-(m), X_+(m))$, $m=1,\ldots,25$, that is, the correlation between past demand $X_-(m)$ and future demand $X_+(m)$ as a function of the "observation"

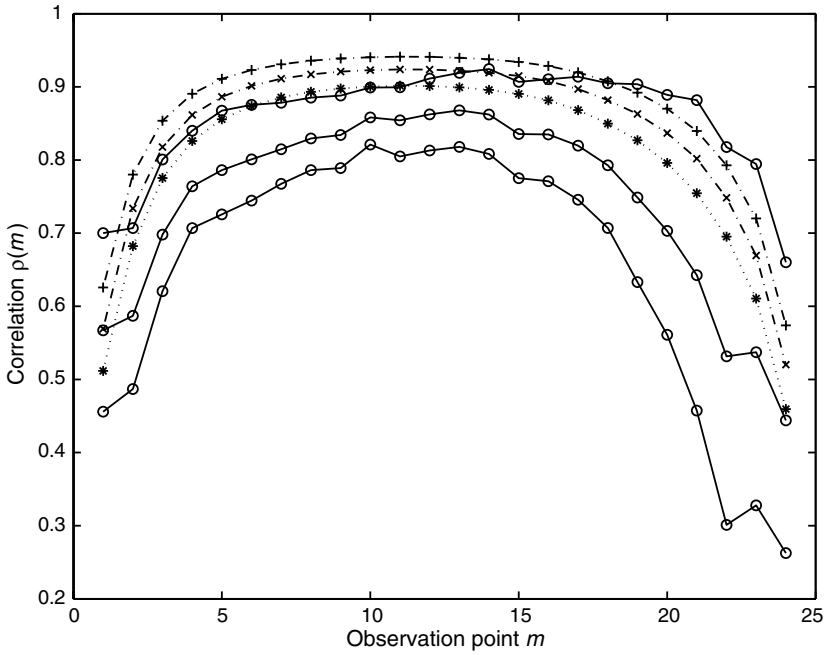
Figure 3 Comparison of Sample Coefficients of Variation of X_i, $i=1,\ldots,25$, to the Exact Values Under Estimated Models



Notes. Point estimate and 95% confidence band (solid lines with ∘) and exact values under estimated models Model 1 (dash-dotted line with +), Model 2 (dash-dotted line with ×), and Model 3 (dotted line with ∗). Confidence intervals are obtained by the hybrid bootstrap method.

time point m $(1\le m\le 25-1)$. This approach simply reduces the number of correlations examined (instead of examining the $k\times(k-1)/2$ correlations between the X_i, we examine only $k-1$ correlations). Figure 4 compares the sample function $\hat{\rho}(m)$ (estimated by the entire data set, with a 95% confidence interval for each m) to the exact function $\rho(m)$ under the estimated Models 1 to 3 (based on the same data set). It is seen that all three models tend to overestimate the correlations but are enormous improvements over Model 0, which assumes zero correlation. Model 3 is the empirical winner, followed by Model 2 and then Model 1.

Figure 4 Comparison of Sample Correlation Function $\rho(m)\equiv\mathrm{Corr}(X_{-}(m),X_{+}(m))$, $m=1,\ldots,25$ to the Exact Function Under Estimated Models



Notes. Point estimate and 95% confidence band (solid lines with ∘) and exact function under estimated models Model 1 (dash-dotted line with +), Model 2 (dash-dotted line with ×), and Model 3 (dotted line with ∗). Confidence intervals are obtained by the hybrid bootstrap method.

Table 2 Performance of Estimated Models 2 and 3 with Respect to the Measure θ of Overdispersion Relative to Model 1

Cluster	Nonparametric estimate $\hat{\theta}$	95% Conf. interval for θ	θ of estimated Model 2	θ of estimated Model 3
Monday	0.285	(0.150, 0.481)	0.110	0.536
Tuesday–Thursday	0.406	(0.239, 0.568)	0.234	0.709
Friday	0.465	(0.266, 0.738)	0.257	0.602

Finally, we report the empirical fit of Models 2 and 3 with respect to θ, the measure of overdispersion relative to Model 1. For each of the three different clusters of days, Table 2 shows the estimate $\hat{\theta}$ discussed after the definition (21), a 95% confidence interval for θ, and the value of θ under the estimated Models 2 and 3. The estimated Models 2 and 3, while removing the modeling constraint $\theta=0$ of Model 1, appear to underestimate and overestimate θ, respectively.

4.3. Sensitivity of Call Center Performance to Arrival Process Model

In this section we study empirically the sensitivity of various call center performance measures to the choice of alternative models of the arrival process. The sensitivity analysis was executed via a simulation model of the Bell Canada call center (Deslauriers 2003). The call center handles two types of calls, inbound and outbound, and is staffed by two types of agents, inbound only and blend.

The center handles only inbound calls during the period 8:00 a.m.–2:00 p.m. (inbound mode) and handles both inbound and outbound calls during the 2:00 p.m.–8:30 p.m. (blend mode). Tables 3 and 4 contain the arrival process and all other model parameters, respectively.

For NHPP, the number of arrivals in period i is Poisson with rate λ_i^0. For Model 0, the number of arrivals is negative binomial with parameters r_i, s_i, where r_i is the number of successes and $s_i=1/p_i-1$, where p_i is the success probability. Arrival process parameters, in addition to those presented in Table 3 are for Model 1, $\hat{\gamma}=36.49$; for Model 2, $\hat{\nu}=48.47$ and $\hat{\beta}_{26}=213.55$; for Model 3, the estimated moments of the Gamma distribution of Y are $\hat{\mu}_Y=1{,}169.95$ and $\hat{\sigma}_Y^2=38{,}655$. Each customer abandons with probability 0.005 upon being asked to wait; otherwise, he or she joins the queue and abandons if his or her patience time is exceeded. The patience times are i.i.d. exponentially distributed with mean $1/\eta_i$ in period i. The inbound service times have the gamma distribution with parameters (δ_i, ζ_i) in period i. For the outbound service times (only), we used kernel density estimation based on a sample of more than 50,000 individual observations. In blend mode, when the total number of idle agents is at least 4 and $z \geq 1$ of them are blend

Table 3 Arrival Process Parameters

Period i	NHPP	Model 0		Model 1	Model 2	Model 3
	λ_i^0	r_i	s_i	λ_i	β_i	α_i
1	24.7	15.6	1.57	0.67	108.2	14.6
2	37.0	60.2	0.61	1.01	164.4	22.5
3	50.0	36.8	1.35	1.37	221.0	30.2
4	56.6	36.7	1.54	1.55	251.3	34.1
5	59.4	36.4	1.63	1.62	264.0	35.9
6	59.1	40.9	1.44	1.61	262.0	35.7
7	60.7	25.0	2.42	1.66	269.8	36.4
8	57.9	24.6	2.35	1.58	254.0	34.7
9	54.7	24.8	2.20	1.50	241.5	32.8
10	54.8	25.2	2.17	1.50	241.1	32.8
11	56.7	21.7	2.60	1.55	251.4	33.9
12	53.9	25.1	2.15	1.47	239.1	32.3
13	54.2	22.8	2.37	1.48	240.2	32.4
14	52.5	23.3	2.25	1.43	231.1	31.4
15	53.4	42.8	1.24	1.46	235.7	32.3
16	56.7	36.1	1.56	1.55	250.0	34.2
17	58.2	47.7	1.22	1.59	255.9	35.3
18	53.0	27.7	1.91	1.45	234.1	31.8
19	43.6	21.0	2.07	1.19	191.4	26.0
20	39.5	26.4	1.49	1.08	173.6	23.7
21	34.0	21.3	1.59	0.93	148.3	20.4
22	30.6	14.8	2.06	0.84	136.2	18.1
23	25.5	16.7	1.52	0.69	112.6	15.1
24	23.2	19.1	1.21	0.63	102.9	13.9
25	18.7	10.9	1.71	0.51	83.4	10.9

Table 4 Parameters of the Simulation Model

Period i	Outbound success prob. κ_i	Mean patience time $1/\eta_i$ (sec)	Inbound serv. time (sec)		# Inbound agents	# Blend agents
			δ_i	ζ_i		
1	0	400	0.729	817.0	10	0
2	0	400	0.729	817.0	16	0
3	0	400	0.729	817.0	21	0
4	0	700	0.729	817.0	23	0
5	0	700	0.729	817.0	24	0
6	0	600	0.729	817.0	24	0
7	0	600	0.729	817.0	24	0
8	0	600	0.729	817.0	24	0
9	0	600	0.620	927.6	22	0
10	0	600	0.620	927.6	22	0
11	0	500	0.620	927.6	28	0
12	0	500	0.620	927.6	26	0
13	0.27	500	0.620	927.6	23	5
14	0.27	500	0.620	927.6	22	11
15	0.28	500	0.755	753.8	22	15
16	0.29	500	0.755	753.8	22	17
17	0.29	500	0.755	753.8	20	16
18	0.30	500	0.553	996.9	17	14
19	0.33	500	0.553	996.9	15	11
20	0.37	500	0.553	996.9	8	16
21	0.40	500	0.553	996.9	4	17
22	0.38	500	0.518	981.6	3	16
23	0.41	500	0.518	981.6	3	15
24	0.41	100	0.518	981.6	3	17
25	0.41	50	0.518	981.6	3	15

Table 5 Sensitivity of Various Performance Measures to Choice of Alternative Arrival Process Models (Entire Day's Operation)

	NHPP	Model 0	Model 1	Model 2	Model 3	Data
Quality of service	$92\pm\epsilon$	$89.1\pm\epsilon$	$88.7\pm\epsilon$	$88.1\pm\epsilon$	$87.7\pm\epsilon$	88.1 ± 0.2
Abandonments per day	15.6 ± 0.1	23.4 ± 0.1	24.4 ± 0.2	26.4 ± 0.3	27.8 ± 0.3	26.9 ± 6.8
Calls served per day	1154.1 ± 0.3	1146.3 ± 0.4	1144.6 ± 1.4	1152 ± 1.3	1141 ± 1.4	1143.1 ± 30.6
Agent utilization (%)	$71.2\pm\epsilon$	$70.7\pm\epsilon$	$70.6\pm\epsilon$	$70.8\pm\epsilon$	$70.4\pm\epsilon$	70.5 ± 1
Mean waiting, all calls (sec)	$5.1\pm\epsilon$	8.8 ± 0.1	9.6 ± 0.1	10.4 ± 0.1	11.2 ± 0.1	10.6 ± 2.3
Mean waiting, queued calls (sec)	44.9 ± 0.2	60.1 ± 0.2	63.2 ± 0.3	65.6 ± 0.3	68.7 ± 0.3	68 ± 5.2

Note. The performance measure quality of service is defined as the percentage of calls having waited less than 20 seconds before being answered.

agents, the system dials $2z$ outbound calls in parallel; the number of successful outbound calls is a binomial random variable with $2z$ trials and success probability κ_i for period i; outbound calls that cannot be immediately answered are lost. The time required for the dialer to start the call is exponentially distributed, with a mean of 2 seconds.

We compare the estimated performance measures for the input models discussed in the paper and for a data-driven model to be explained later. Table 5 summarizes the results for the entire day of operation, whereas Table 6 provides results for the inbound-only mode of operation (8:00 a.m.–2:00 p.m.). The results are more sensitive to the arrival model in the inbound-only mode than in the blend mode, because in the latter mode the outbound calls "smooth out" the workload variations. For each performance measure, we report the estimated daily mean and the half-width of a 95% confidence interval (half-width entries ϵ correspond to values less than 0.05), based on 60,000 independent replications for models NHPP to Model 3. First, we observe that certain performance measures are not very sensitive to the input model (quality of service, agent utilization, and calls served per day), while the number of abandonments and customer waiting times appear to be more sensitive. The most striking evidence of the influence of arrival process model on estimated performance is offered by the mean waiting time of all calls, which is more than doubled as we go from NHPP to Model 3. Second, system performance (across all performance measures) is decreasing in the order NHPP, Models 0, 1, 2, and 3. This last result is not surprising, in view of the following: $\mathbf{X}$ has substantially higher marginal variances under Models 0, 1, 2, and 3 relative to NHPP; $\mathbf{X}$ has positive covariances under Models 1, 2, and 3 against zero covariances under Model 0; $\mathbf{X}$ has increasing variances in the order Models 1, 2, and 3, as seen in Figure 3. We conclude that insights obtained from a simulation model of the call center are sensitive to the choice of arrival process model.

The data-driven simulation experiment (column "data") in the tables has been performed as follows. Our data set has observations for 120 days corresponding to the cluster Tuesday/Wednesday/Thursday. For each of these 120 days, we made 500 simulation runs of the model with exactly the same arrival count as in the data for each half hour. These arrivals were randomized uniformly over the half hour. These randomizations and all other random variables in the simulation model were independent across the 500 runs. This gave a total of 60,000 runs, which were dependent because many of them used the same arrival counts. We then took the average of each performance measure of interest over the 500 runs associated with each of the 120 days in

Table 6 Sensitivity of Various Performance Measures to Choice of Alternative Arrival Process Models (Inbound Operation Only 8:00 a.m.–2:00 p.m.)

	NHPP	Model 0	Model 1	Model 2	Model 3	data
Quality of service	$89.8\pm\epsilon$	$85.3\pm\epsilon$	$84.5\pm\epsilon$	$83.5\pm\epsilon$	$83\pm\epsilon$	83.8 ± 0.3
Abandonments per day	10.3 ± 0.1	16.9 ± 0.1	18.1 ± 0.2	19.8 ± 0.2	21 ± 0.2	20.1 ± 6.1
Calls served per day	615.7 ± 0.2	609.1 ± 0.3	607.6 ± 0.7	612 ± 0.7	605.3 ± 0.7	606 ± 15
Agent utilization (%)	$66.7\pm\epsilon$	$66\pm\epsilon$	65.8 ± 0.1	66.3 ± 0.1	65.6 ± 0.1	65.6 ± 1.6
Mean waiting, all calls (sec)	6.7 ± 0.1	12.8 ± 0.1	14.5 ± 0.1	15.9 ± 0.2	17.1 ± 0.2	16.2 ± 3.7
Mean waiting, queued calls (sec)	51.3 ± 0.3	71.6 ± 0.3	77.5 ± 0.3	80 ± 0.3	84.2 ± 0.4	83.9 ± 6.3

Note. The performance measure quality of service is defined as the percentage of calls having waited less than 20 seconds before being answered.

the data, in order to obtain 120 "independent" observations and then compute confidence intervals in a standard way. These confidence intervals turn out to be rather wide, because the variance of the arrival counts of the 120 days in the data set is large, and no matter how many simulation runs we perform, this part of the variance is never reduced. Models 2 and 3 provide performance measures that are closer to the averages provided by the data-driven simulation than the other models, especially for the number of abandonments and waiting times. However, the results of Model 1—and of Model 0 for many of the performance measures—are also inside the wide confidence intervals of the data-driven model. Thus, there is too much variance in our data to conclude that any of the Models 1 to 3 produces an output that behaves closer to the data-driven model, with statistical significance. It seems clear, on the other hand, that these three models behave better than NHPP and Model 0.

5. Conclusion and Application Perspective

We developed and studied models that simultaneously capture three features of an arrival process observed repeatedly over a fixed finite horizon (that is, a day), namely, overdispersion compared with a Poisson process, a time-varying arrival intensity over the course of the horizon, and nonzero covariance between the arrival counts in different time periods within the horizon. Our study of the arrival process to a Bell Canada call center has confirmed the simultaneous presence of these properties, and a strong positive association between arrival counts was observed. Simple-to-use models such as NHPP (nonhomogeneous Poisson process) and Model 0 of Jongbloed and Koole (2001), while capturing a time-varying arrival intensity, do not support correlation between arrival counts in different time periods within the arrival horizon. Moreover, we have shown in §4.3 that simulation-based call center performance measurement is sensitive to the arrival-process model and more particularly to the presence of correlation. This establishes the need for more advanced modeling of the arrival process for future applications.

Models 1 and 2 are particular cases of doubly stochastic Poisson processes that are especially appealing in light of their easy-to-use parameter estimation, variate generation, and forecasting. We have identified and characterized one aspect of the lack of fit of Model 1 via the statistical functional θ measuring the degree of overdispersion of the interval-level arrival counts relative to the total (daily) arrival count, where overdispersion is with respect to a reference Model 1. We developed the Dirichlet Model 3 and characterized how it increases the flexibility of induced moments relative to Models 1 and 2. We have also documented the superiority of Models 1 to 3 relative to NHPP and Model 0 in our case study.

We note that the new models can be easily adapted to handle distinct classes of arriving "jobs," where jobs may include distinct classes of calls and possibly other types of requests, for example, electronic mail or chat (Koole and Mandelbaum 2002). One simple approach to such adaptation is to model the aggregate arriving jobs with the standard models presented here and then assign each arriving job item to a particular class by sampling a discrete distribution corresponding to the different job items. This approach preserves the features of time-varying arrival rate and induced correlations for each distinct job class.

Our Models 1 to 3 are for a single day of a call center's operation. This does not mean that all days must be assumed to be independent and to have the same distribution parameters. For example, the distribution of W for Model 1 can be different for different days of the week or may be on any type of available information (that is, it is the first working day after a holiday, a new promotional campaign has just been launched for a given service, etc.). In particular, the values of W for successive days may be statistically dependent and may be modeled by a time series. In all cases, it suffices to replace the distribution of W in Model 1 by its conditional distribution given the available information. If the model is built so that this conditional distribution is gamma, then our development for Model 1 applies for each day. Similarly, Model 3 can be extended to allow a conditional distribution of Y given available information. Such models that incorporate day-to-day dependency are certainly worthy of further development and investigation, but this is beyond the scope of the present paper. Developments in that direction can be found in Brown et al. (2002).

In the context of call center management with respect to operational efficiency, we envision two uses of the models developed. First and foremost, in simulation or analytical studies of call centers, the models aim to be valid, faithful representations of the arrival process. Second, by capturing the strong dependence between the arrival counts in different time periods within the same day, the models yield as a natural byproduct a predictive distribution of future demand within the day, given the observed demand so far. Such short-term forecasts (possibly a vector of forecasts of future arrivals by time of day) may prove useful in short-term, within-the-day planning. For example, when the forecast of $X_+(m)$ is low relative to the current staffing level, actions can be taken to improve agent utilization, such as initiating outbound

calls or scheduling agent training or meetings (Gans et al. 2003).

We conclude with some suggestions for future work. Models 1 to 3 introduced here are parsimonious: They have only one or two parameters in addition to the mean arrival rate over each period. A good topic for further research would be to design and study models with a few more parameters that could better fit the correlations and/or individual interval dispersions. In our case study, the correlations for Models 1 to 3 were systematically higher than those in the data. Conceivably, this could be improved by a single additional parameter that would control the overall amount of correlation. With respect to dispersions, we note that Models 2 and 3 allowed finer control of the dispersion of the individual time-interval demands relative to the total daily demand, with the constraint that the dispersion can be adjusted (relative to Model 1) either upward or downward for all time intervals.

A possible approach for allowing a different direction of dispersion adjustment across time intervals is a hierarchical model. At the first level, one models the multivariate demand over aggregated intervals (level-1 demand); then, conditionally on the level-1 demand, one assigns (probabilistically) this demand to the target intervals (level-2 demand). In this modeling approach, Model 3 appears interesting because of its ability to adjust the dispersions of level-2 demand, relative to level-1 demand in either direction. Moreover, this hierarchical approach facilitates control of the number of model parameters. Another aspect that would require further work is the experimentation of the proposed models with different sets of real-life data from telephone call centers and from other types of systems where arrival processes are likely to behave in a similar way.

Acknowledgments

This research was supported by grants OGP-0110050 and CRDPJ-251320 from NSERC-Canada, a grant from Bell Canada (via the "Laboratoires Universitaires Bell"), and grant 00ER3218 from NATEQ-Québec to the third author. The work of the second author was supported by an NSERC-Canada scholarship. The authors thank Eric Buist for his help in running the simulations and two anonymous referees whose comments helped improve the paper.

References

Brown, L., N. Gans, A. Mandelbaum, A. Sakov, H. Shen, S. Zeltyn, L. Zhao. 2002. Statistical analysis of a telephone call center: A queueing-science perspective. Technical report, The Wharton School, University of Pennsylvania, Philadelphia, PA.

Deslauriers, A. 2003. Modélisation et simulation d'um centre d'appels téléphoniques dans un environnement mixte. Master's thesis, Department of Computer Science and Operations Research, University of Montréal, Montréal, Québec, Canada.

Gans, N., G. Koole, A. Mandelbaum. 2003. Telephone call centers: Tutorial, review, and research prospects. *Manufacturing Service Oper. Management* **5** 79–141.

Johnson, N. L., S. Kotz. 1969. *Distributions in Statistics: Discrete Distributions.* Houghton Mifflin, Boston, MA.

Jongbloed, G., G. Koole. 2001. Managing uncertainty in call centers using Poisson mixtures. *Appl. Stochastic Models Bus. Indust.* **17** 307–318.

Koole, G., A. Mandelbaum. 2002. Queueing models of call centers: An introduction. *Ann. Oper. Res.* **113** 41–59.

Mosimann, J. E. 1963. On the compound negative multinomial distribution and correlations among inversely sampled pollen counts. *Biometrika* **50** 47–54.

Shao, J. 1999. *Mathematical Statistics.* Springer, New York.

Smith, B. C., J. F. Leimkuhler, R. M. Darrow. 1992. Yeild management at American Airlines. *Interfaces* **22** 8–31.

Tanir, O., R. J. Booth. 1999. Call center simulation in Bell Canada. P. A. Farrington, H. B. Nemhard, D. T. Sturrock, G. W. Evans, eds. *Proc. 1999 Winter Simulation Conf.*, www.informs-cs.org, 1640–1647.

Whitt, W. 1999. Dynamic staffing in a telephone call center aiming to immediately answer all calls. *Oper. Res. Lett.* **24** 205–212.