Bayesian Analysis of Call Center Arrival Data

Fabrizio Ruggeri\textsuperscript{1}, Refik Soyer\textsuperscript{2} and Murat Tarimcilar\textsuperscript{2}

\textsuperscript{1}Istituto di Matematica Applicata e Tecnologie Informatiche
Consiglio Nazionale delle Ricerche
Via Bassini 15, I-20133, Milano, Italy
fbrizio@mi.imati.cnr.it
www.mi.imati.cnr.it/\texttilde{}fbrizio/

\textsuperscript{2}George Washington University
Washington D.C., USA
OUTLINE OF THE TALK

• General issues on call centers data

• Some models for call centers data

• Efficacy of advertising campaigns

• Bayesian models

• Example

• Future research
CALL CENTER

- Centralised hub aimed to make or get calls to/from (prospective) customers

- (Often) primary point of contact between customers and businesses

- Major investment for many organisations

- 2.86 million operator positions in over 50,000 call centers in the US, with some locations employing over 1000 agents

- Not only businesses, but also governments, etc.

Some citations from Weinberg, Brown and Stroud, 2006
CALL CENTER OPERATIONS

• Different functions
  – Only inbound calls (e.g. requests for assistance)
  – Only outbound calls (e.g. promotions)
  – Both of them (possibly, outbound calls when idle from inbound ones, including answers to previous inbound calls)

• Different structures
  – All issues handled by equally trained agents
  – Various levels, e.g.
    * automated answers to simple issues
    * contact with lowly trained agents for ordinary issues
    * contact with highly trained agents (e.g. supervisors) for the most complex issues
CALL CENTER OPERATIONS

• Different technologies
  – Direct phone calls to agents
  – Phone calls to a computerised system, routing calls, e.g., to automated answers and different levels of agents
  – Computer-phone integration, allowing for identification of customers and immediate availability of his/her data (personal and past communications)
  – Contact center: Computer supporting agents via a full range of media (e-mail, fax, web pages, chat)

• Different policies
  – No customer must be lost
  – No premium customer must be lost
  – Trade between losses due to customer losses and center staffing
CALL CENTER QUALITY

• Qualitative measures, e.g.
  – Customer satisfaction about
    * user-friendly system
    * length of hold-in-line
    * effectiveness of answers
  – Company image affecting future businesses

• Quantitative measures, e.g.
  – Abandonment, as fraction of customers leaving the queue before service
  – Retrial, as average number of calls needed to solve a problem
  – Waiting, as its average or some percentiles of the waiting-time distribution
  – Profit
CALL CENTER DATA CLASSIFICATION

• Operational data
  – Typically, aggregated data over some periods (minutes/daily/weekly/yearly) from history of each call, e.g.
    * total number of calls served or abandoned
    * average waiting time
    * agents’ utilisation level

• Marketing data
  – Combination of phone data with customer’s profile and past history
  – Euro-figures for past sales and future marketing targets

• Psychological data
  – Surveys of customers, agents and managers about subjective perception of service level and working environment

Classification taken from Koole and Mandelbaum, 2001
CALL CENTER DATA USE

- Customer level
  - Future marketing target based on past transactions of existing customers
  - Identification of possible (and profitable) new customers

- Company level
  - Quality of service (lost calls, waiting times, etc.)
  - Staffing of call center
  - Monitoring of agents’ performance (*might be unlawful*)
  - Training of agents
CALL CENTER

(Statistical) interest in

- Forecasting demand in different periods (e.g. from hourly to yearly)
- Customer loss
- Optimal number of agents (possibly time-dependent)
- Customers’ (both current and perspective) profiles
- etc.
MODELS FOR CALL CENTER ARRIVAL DATA

Customer level

• Customer profile, using CRM (Customer Relations Management) and Data Mining

• Degree of satisfaction, mostly via surveys
MODELS FOR CALL CENTER ARRIVAL DATA

Company level

- Time series (especially ARIMA processes)
  - arrivals over periods of time
  - detection of different patterns in different periods (e.g. before and after Xmas)
  - normalised (w.r.t. total daily calls) arrivals during each day

- Queueing models
  - arrival time
  - system availability
  - (optimal) number of channels
  - service policy (e.g. premium customers)
MODELS FOR CALL CENTER ARRIVAL DATA

Poisson processes with time dependent arrival rates

- Doubly stochastic Poisson models

- Nonhomogeneous Poisson processes (NHPPs)

Focus on forecasting models for optimal scheduling and staffing of telephone operators in a call center, using aggregate arrival data
BACKGROUND

- Consumer Electronics Producer
  - Limited variety of products
  - Long life cycle
  - Aging products, sporadic upgrades
  - Targeted advertising

- Sales
  - Average sale is around $500.00
  - Almost all the sales are through the sales call center
  - Call return is very low if first time the customer is denied the call

- Advertisement
  - Many media venues used, although print media is majority
  - Each ad is targeted and urging customers to place calls
OBJECTIVES

• Evaluation of different advertisement policies from a marketing viewpoint through call center arrival data

• Analysis of effects of relevant covariates, e.g.
  – Media expense (in $’s)
  – Venue type (monthly magazine, daily newspaper, etc.)
  – Ad format (full page, half page, colour, etc.)
  – Offer type (free shipment, payment schedule, etc. )
  – Seasonal indicators

• Prediction of calls volume generated by specific ad over any interval of interest

• Arrival data are not aggregated
## DATA

*Typical call arrival data for an ad*

<table>
<thead>
<tr>
<th>Time interval (in days)</th>
<th>Number of calls</th>
</tr>
</thead>
<tbody>
<tr>
<td>(0, 1]</td>
<td>6</td>
</tr>
<tr>
<td>(1, 2]</td>
<td>5</td>
</tr>
<tr>
<td>(2, 3]</td>
<td>1</td>
</tr>
<tr>
<td>(3, 4]</td>
<td>3</td>
</tr>
<tr>
<td>(4, 5]</td>
<td>2</td>
</tr>
<tr>
<td>(5, 6]</td>
<td>2</td>
</tr>
<tr>
<td>(6, 7]</td>
<td>3</td>
</tr>
<tr>
<td>(7, 8]</td>
<td>0</td>
</tr>
<tr>
<td>(8, 9]</td>
<td>2</td>
</tr>
<tr>
<td>(9, 10]</td>
<td>2</td>
</tr>
<tr>
<td>(10, 18]</td>
<td>0</td>
</tr>
</tbody>
</table>

- Number of calls for ad in unit intervals
- Features of the ad (media, cost, etc.)
- Corresponding ad known for each call *in earlier work*
GOALS OF THE ONGOING RESEARCH

- Handling calls unassigned to any ad
- Different models for covariates
- Bayesian nonparametrics
NOTATIONS

• $C_1, \ldots, C_m$: $m$ campaigns (ad)

• $T_1, \ldots, T_m$: campaigns starting times, $T_1 \leq T_2 \leq \ldots \leq T_m$

• Calls recorded as number of arrivals in $I_j = (t_{j-1}, t_j], \ j = 1, n$
  (starting times coincide with one of the endpoints of the intervals $I_j$)

• $n_{ij}$: number of calls in $I_j$ related to campaign $C_i$, $i = 1, m, j = 1, n$

• $u_j$: number of unallocated calls in $I_j$ (0 in earlier work)
  (in each $I_j$ consider only campaigns $C_i$ with $T_i \leq t_{j-1}$)
NONHOMOGENEOUS POISSON PROCESS

- $N_t, t \geq 0$ # events by time $t$
- $N(y, s)$ # events in $(y, s]$
- $\Lambda(t) = E N_t$ mean value function
- $\Lambda(y, s) = \Lambda(s) - \Lambda(y)$ expected # events in $(y, s]$

$N_t, t \geq 0$, NHPP with intensity function $\lambda(t)$ iff
1. $N_0 = 0$
2. independent increments
3. $P\{\text{# events in } (t, t+h) \geq 2\} = o(h)$
4. $P\{\text{# events in } (t, t+h) = 1\} = \lambda(t)h + o(h)$

$\Rightarrow P\{N(y, s) = k\} = \frac{\Lambda(y, s)^k}{k!} e^{-\Lambda(y, s)}, \forall k \in \mathbb{N}$
NONHOMOGENEOUS POISSON PROCESS

\[ \lambda(t) \equiv \lambda \forall t \Rightarrow \text{HPP} \]

- \( \lambda(t) \): intensity function of \( N_t \)

- \( \lambda(t) := \lim_{\Delta \to 0} \frac{\mathbb{P}\{N(t, t + \Delta] \geq 1\}}{\Delta}, \forall t \geq 0 \)

- \( \mu(t) := \frac{d\Lambda(t)}{dt} \): Rocof (rate of occurrence of failures)

Property 3. \( \Rightarrow \mu(t) = \lambda(t) \) a.e. \( \Rightarrow \Lambda(y, s) = \int_y^s \lambda(t)dt \)
MODULATED NHPP MODEL

• NHPP’s with intensity dependent on covariates

\[ \lambda_i(t, Z_i) = \lambda_0(t) \exp\{\gamma' Z_i\} \]

• \( \lambda_0(t) \) baseline intensity and \( \gamma \) parameter

\[ \frac{\lambda_i(t,Z_i)}{\lambda_j(t,Z_j)} = \exp\{\gamma' (Z_i - Z_j)\} \]

⇒ Proportional Intensities Model

Cox, 1972
MODEL

Call center arrival data

• modeled as NHPP

• dependent on the campaign (i.e. covariates $Z_i$)

• dependent on the starting point $T_i$

$\Rightarrow$ for campaign $C_i$

• $\lambda_i(t) = \lambda_0(t - T_i) \exp\{\gamma'Z_i\}I_{[T_i,\infty)}(t)$

• $\Lambda_i(t) = \Lambda_0(t - T_i) \exp\{\gamma'Z_i\}I_{[T_i,\infty)}(t)$
MODEL

Superposition Theorem

*Sum of independent NHPP with intensity functions $\lambda_i(t)$ is still a NHPP with intensity function $\lambda(t) = \sum \lambda_i(t)$*
MODEL

Number of calls decreasing to zero $\Rightarrow$ Power Law Process (PLP)

- $\lambda(t) = M\beta t^{\beta-1}$

- $\Lambda(t) = Mt^\beta$

$M, \beta, t > 0$ but here $0 < \beta < 1 \Rightarrow \lambda(t) \downarrow 0$ for $t \rightarrow \infty$

Alternative: $\lambda(t) = \beta_0 \frac{\log(1+\beta_1 t)}{(1+\beta_1 t)}$

(increasing from 0 and then decreasing to 0)
MODEL

- Cumulative number of arrivals approximates $\Lambda$
- PLP $\Rightarrow \log \Lambda(t) = \log M + \beta \log t$
- $\Rightarrow$ plot of $\log \Lambda(t)$ v.s. $\log t$ to check appropriateness of PLP
MODEL WITH PERFECT LINKAGE OF CALLS

- No further details on the previous model to avoid repetitions with the model with unallocated calls

- Random effects model, accounting for differences unexplained by covariates
  - Model similar to the previous one, but
  - NHPPs with $\Lambda_i(t) = M_i t^\beta$
  - $\log M_i = \theta + \phi_i$, $\phi_i$ random effect terms
  - $\phi_i$'s conditional independent $\mathcal{N}(0, 1/\tau)$ and $\tau \sim \mathcal{G}(a_\tau, b_\tau)$
DIGRESSION: MODEL FOR UNASSIGNED CALLS

• Unassigned calls define a new process with $\Lambda_u(t) = \delta_u Mt^\beta$

• $\delta_u$ r.v. rescaling the baseline mean value function

• Model describes arrivals of unallocated calls but not their assignment to campaigns and, consequently, these data are less useful to measure effectiveness of campaigns
A POSSIBLE (BUT INCONVENIENT) MODEL

• Arrivals in the interval \( I_j, j = 1, n \):
  - \( n_{ij} \) related to campaign \( C_i, i = 1, m \)
  - \( u_j \) unallocated

• Interest in \( P(n_{1j} \in C_1, \ldots, n_{mj} \in C_m, u_j \in I_j) \)

• All possible allocations: \( u_j = \{u_{1j} \in C_1, \ldots, u_{mj} \in C_m\} \), with \( \sum_{l=1}^{m} u_{lj} = u_j \)

• Specify a distribution \( P(u_j) \) (e.g. multinomial) on the allocation

• Compute

\[
P(n_{1j} \in C_1, \ldots, n_{mj} \in C_m, u_j \in I_j) = \sum_{\text{all } u_j} P(n_{1j} + u_{1j} \in C_1, \ldots, n_{mj} + u_{mj} \in C_m | u_j) P(u_j)
\]
ASSUMPTIONS

- Latent variables $Y_j = (Y_{1j}, \ldots, Y_{mj}) \sim \text{Mult}(u_j, p_{1j}, \ldots, p_{mj})$, for each $I_j$ ⇒ unknown number of unallocated calls assigned to each campaign
- Drop the notations $\in C_i$ from the probabilities
- $\Delta_{ij} = \Lambda_i(t_j) - \Lambda_i(t_{j-1})$, for each $i, j$
- For each interval $I_j$, $\mathcal{P}(n_{1j} + Y_{ij}, \ldots, n_{mj} + Y_{mj}, Y_j, u_j) =
  \mathcal{P}(n_{1j} + Y_{ij}, \ldots, n_{mj} + Y_{mj} | Y_j, u_j) P(Y_j | u_j) P(u_j)
\approx \left\{ \prod_{i=1}^m \frac{\Delta_{ij}^{n_{ij} + Y_{ij}}}{n_{ij} + Y_{ij}!} e^{-\Delta_{ij}} \right\} \left\{ \left( \frac{u_j}{Y_{1j}, \ldots, Y_{mj}} \right) \prod_{i=1}^m p_{ij}^{Y_{ij}} \right\} P(u_j)
- No interest in $\mathcal{P}(u_j) \Rightarrow$ partial likelihood
PARTIAL LIKELIHOOD

For any NHPP

\[
\Pi_{j=1}^{n} \left( Y_{1j}, \ldots, Y_{mj} \right) \left\{ \prod_{i=1}^{m} \frac{\Delta_{ij}^{n_{ij}+Y_{ij}}}{n_{ij}+Y_{ij}!} e^{-\Delta_{ij}Y_{ij}} p_{ij} \right\} = e^{- \sum_{i=1}^{m} \Lambda(t_i)} \Pi_{j=1}^{n} \left( Y_{1j}, \ldots, Y_{mj} \right) \left\{ \prod_{i=1}^{m} \frac{\Delta_{ij}^{n_{ij}+Y_{ij}}}{n_{ij}+Y_{ij}!} p_{ij} \right\}
\]

For a PLP

- \( \sum_{i=1}^{m} n_{ij} = n_j \ (j = 1, n) \), \( \sum_{j=1}^{n} n_{ij} = N_i^* \) and \( \sum_{j=1}^{n} Y_{ij} = U_i^* \ (i = 1, m) \)
- \( \sum_{j=1}^{n} n_j = N \) and \( \sum_{j=1}^{n} Y_j = U \)

\[
M^{N+U} e^{-M \sum_{i=1}^{m} (t_n - T_i)^\beta e^{\gamma' z_i}} \prod_{i=1}^{m} e^{\gamma' Z_i (N_i^* + U_i^*)} \prod_{j=1}^{n} \left( u_j \right)_{Y_{1j}, \ldots, Y_{mj}} . \\
\prod_{i=1}^{m} \left\{ \left[ (t_j - T_i)^\beta - (t_{j-1} - T_i)^\beta \right] \right\}^{n_{ij}+Y_{ij}} \frac{Y_{ij}}{n_{ij}+Y_{ij}!} p_{ij} \}
\]
PRIORS

• Intervals $I_j, j = 1, n \Rightarrow$ Independent $p_j \sim \text{Dir}(\alpha_{1j}, \ldots, \alpha_{mj})$

• $\alpha_{ij} = \alpha e^{-\delta(t_j - T_i)} I_{(T_i, \infty)}(t_j)$
  - $\alpha_{ij} = 0$ for unstarted campaigns ($\Rightarrow$ degenerate Dirichlet)
  - decreasing $\alpha_{ij}$ from an interval to next ones
    (calls unlikely from older ads: $E_{pij} = \frac{\alpha_{ij}}{\sum_{i=1}^{m} \alpha_{im}}$)
  - $\alpha, \delta$: either known or prior

• $M \sim \mathcal{G}(a, b)$

• any $\pi(\beta)$ (nothing convenient for simulations)

• any $\pi(\gamma)$ (nothing convenient for simulations)
POSTERIORS

- $n^{(j)} = (n_{1j}, \ldots, n_{mj})$, for interval $I_j$, $j = 1, n$

- $n = (n^{(1)}, \ldots, n^{(n)})$

- $p_j | Y_j, n \sim Dir(\alpha_{1j} + Y_{1j}, \ldots, \alpha_{mj} + Y_{mj})$, $j = 1, n$

- $\gamma | M, n, \beta, Y_j: \pi(\gamma) e^{-\sum_{i=1}^{m} \{M(t_n-T_i)^{\beta}-(N_i^{*}+U_i^{*})\}} e^{\gamma'}Z_i$
POSTERIORS

- $M|\beta, n, \gamma \sim G(a + N + U, b + \sum_{i=1}^{m}(t_n - T_i)^{\beta}e^{\gamma Z_i})$

- $\beta|M, n, \gamma, Y_{-j} \propto \pi(\beta)e^{-M\sum_{i=1}^{m}(t_n-T_i)^{\beta}e^{\gamma Z_i}}\prod_{j=1}^{n}\prod_{i=1}^{m}[(t_j-T_i)^{\beta}-(t_{j-1}-T_i)^{\beta}]^{n_{ij}+Y_{ij}}$

- $P(Y_j|\beta, M, \gamma, n) \propto \left(Y_{1j},...,Y_{mj}\right)\prod_{i=1}^{n}\frac{\Delta_{ij}^{n_{ij}+Y_{ij}}}{(n_{ij}+Y_{ij})!}\Delta_{ij}Y_{ij}, j = 1, n$

$\Rightarrow$ MCMC simulation (Gibbs with Metropolis steps within)
EXAMPLE

- Weekly time intervals
- 10 campaigns
- Cost of the campaign (in $1000) as unique covariate
- Actual data considered and missing links randomly assigned
- Interest in interval 4 with 3 active campaigns and 4 missing links
EXAMPLE

Posterior probabilities

<table>
<thead>
<tr>
<th>$Y_{ij}$</th>
<th>Ad1</th>
<th>Ad2</th>
<th>Ad3</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.1362</td>
<td>0.5070</td>
<td>0.2766</td>
</tr>
<tr>
<td>1</td>
<td>0.2732</td>
<td>0.3414</td>
<td>0.2654</td>
</tr>
<tr>
<td>2</td>
<td>0.2756</td>
<td>0.1246</td>
<td>0.2406</td>
</tr>
<tr>
<td>3</td>
<td>0.2208</td>
<td>0.0228</td>
<td>0.1556</td>
</tr>
<tr>
<td>4</td>
<td>0.0942</td>
<td>0.0042</td>
<td>0.0618</td>
</tr>
</tbody>
</table>

Posterior means and actual values

<table>
<thead>
<tr>
<th></th>
<th>Ad1</th>
<th>Ad2</th>
<th>Ad3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>1.86</td>
<td>0.67</td>
<td>1.50</td>
</tr>
<tr>
<td>Actual</td>
<td>1</td>
<td>0</td>
<td>3</td>
</tr>
</tbody>
</table>

Predictive distribution of calls in future intervals
EXAMPLE

Posterior density of $\gamma$ (covariate coefficient)

$\text{density.default}(x = \text{gnsfr}[200:2272])$

- $Z = 0$ for campaign cost above a threshold and $Z = 1$ under it
- High probability of negative $\gamma$
- $\Rightarrow$ higher $\lambda(t)$ for expensive campaign $\Rightarrow$ more calls for it
EXAMPLE

Posterior density of $M$ (PLP parameter: $\lambda(t) = Mt^\beta$)

density.default(x = Ms[200:2272])

$N = 2073$  Bandwidth = 0.2556
EXAMPLE

Posterior density of $\beta$ (PLP parameter: $\lambda(t) = Mt^\beta$)
FUTURE RESEARCH

- More detailed data analysis
- Other models including covariates
- Nonparametric approach
HIERARCHICAL NHPP MODELS

• $m$ campaigns $C_i \Rightarrow m$ PLPs with parameters $(M_i, \beta_i)$, $i = 1, m$

• Possible priors
  
  $- \beta_i | \rho, \delta \sim \mathcal{G}(\rho \exp\{X_i^T \delta\}, \rho)$
  
  $- M_i | \nu, \sigma \sim \mathcal{G}(\nu \exp\{X_i^T \sigma\}, \nu)$

• $\rho, \nu, \delta, \sigma$
  
  $- \text{Priors}$
  
  $- \text{Empirical Bayes}$
NONPARAMETRIC APPROACH

# events in \([T_0, T_1]\) \sim \mathcal{P}(\Lambda[T_0, T_1]), \Lambda[T_0, T_1] = \Lambda(T_1) - \Lambda(T_0)

Parametric case: \(\Lambda[T_0, T_1] = \int_{T_0}^{T_1} \lambda(t)dt\)

Nonparametric case: \(\Lambda[T_0, T_1] \sim \mathcal{G}(\cdot, \cdot)\)

\(\Rightarrow\) \(\Lambda\) d.f. of the random measure \(M\)

**Notation:** \(\mu B := \mu(B)\)

**Definition 1** Let \(\alpha\) be a finite, \(\sigma\)-additive measure on \((\mathcal{S}, \mathcal{S})\). The random measure \(\mu\) follows a **Standard Gamma** distribution with shape \(\alpha\) (denoted by \(\mu \sim GG(\alpha, 1)\)) if, for any family \(\{S_j, j = 1, \ldots, k\}\) of disjoint, measurable subsets of \(\mathcal{S}\), the random variables \(\mu S_j\) are independent and such that \(\mu S_j \sim G(\alpha S_j, 1)\), for \(j = 1, \ldots, k\).

**Definition 2** Let \(\beta\) be an \(\alpha\)-integrable function and \(\mu \sim GG(\alpha, 1)\). The random measure \(M = \beta \mu\), s.t. \(\beta \mu(A) = \int_A \beta(x)\mu(dx), \forall A \in \mathcal{S}\), follows a **Generalised Gamma** distribution, with shape \(\alpha\) and scale \(\beta\) (denoted by \(M \sim GG(\alpha, \beta)\)).
NONPARAMETRIC APPROACH

Consequences:

• $\mu \sim \mathcal{P}_{\alpha,1}$, $\mathcal{P}_{\alpha,1}$ unique p.m. on $(\Omega, M)$, space of finite measures on $(\mathbb{S}, S)$, with these finite dimensional distributions

• $M \sim \mathcal{P}_{\alpha,\beta}$, **weighted random measure**, with $\mathcal{P}_{\alpha,\beta}$ p.m. induced by $\mathcal{P}_{\alpha,1}$

• $EM = \beta \alpha$, i.e. $\int_{\Omega} M(A) \mathcal{P}_{\alpha,\beta}(dM) = \int_A \beta(x) \alpha(dx), \forall A \in S$
NONPARAMETRIC APPROACH

**Theorem 1** Let $\xi = (\xi_1, \ldots, \xi_n)$ be $n$ Poisson processes with intensity measure $M$. If $M \sim \mathcal{GG}(\alpha, \beta)$ a priori, then $M \sim \mathcal{GG}(\alpha + \sum_{i=1}^{n} \xi_i, \beta/(1 + n\beta))$ a posteriori.

**Data:** $\{y_{ij}, i = 1 \ldots k_j\}_{j=1}^{n}$ from $\xi = (\xi_1, \ldots, \xi_n)$

**Bayesian estimator of** $M$: measure $\tilde{M}$ s.t., $\forall S \in S$,

$$\tilde{M}S = \int_{S} \frac{\beta(x)}{1 + n\beta(x)} \alpha(dx) + \sum_{j=1}^{n} \sum_{i=1}^{k_j} \frac{\beta(y_{ij})}{1 + n\beta(y_{ij})} I_S(y_{ij})$$

Constant $\beta \Rightarrow \tilde{M}S = \frac{\beta}{1 + n\beta} [\alpha S + \sum_{j=1}^{n} \sum_{i=1}^{k_j} I_S(y_{ij})]$. 

NONPARAMETRIC APPROACH

Data (calls) recorded as number of events in disjoint intervals

- Comparison between parametric and nonparametric models
- Update of the Gamma process