Payphones, Parking-meters, Vending Machines and Optimal Bayesian Decisions on Collection-Times

Ranjan Maitra and Siddhartha R. Dalal *

Abstract

Payphones, parking meters and vending machines illustrate the modern business practice of substituting manpower with machinery. They do not eliminate manual labor completely, for it is still needed to replace full coin-boxes and to stock vending machines. Deciding when to replace a coin-box is important, with unequal losses for under- and over-estimation. This paper derives optimal methodology for this problem, by incorporating collection history and specifying common prior distributions over average daily fill-rate and standard deviation at each box. The approach is implemented and analyzed on collection records from 11,308 payphones over a large geographical region. When the loss from over-estimation is nineteen times that from under-estimation, our methods out-perform the one in current use at least 69.9% of the time, translating into average potential collection-cost reductions of over 21%.

1 Introduction

Technology is gradually replacing man with machine in the modern age, more so for tasks that are tedious and repetitive. Parking-meters in busy urban areas, vending or automated teller machines

*Ranjan Maitra is Assistant Professor, Department of Mathematics and Statistics, University of Maryland, Baltimore County, Baltimore, MD 21250 and Siddhartha R. Dalal is Chief Scientist and Executive Director of the Information Analysis & Services Research Department, Applied Research Area, Telcordia Technologies, formerly Bellcore, Morristown, NJ 07960. The authors thank an associate editor and four referees whose extensive comments greatly improved the presentation of this paper. They also acknowledge the help provided by Marvin Freedman, Mark Grossman, George McBride, Bart Mendez and Paul Tukey of Telcordia Technologies, at different stages in the retrieval and analysis of the dataset used in this paper.

and payphones provide common examples of this trend. These devices substantially obviate the need for personnel to provide services, making these available at times of the consumer's convenience, while also reducing associated variable costs. However, mechanization has not completely stamped out human intervention which is needed, for example, to collect and replace toll-filled coin-boxes and to stock items and cash into vending and automated teller machines. The goal of reducing variable costs associated with sending the business representative to replace full coin-boxes calls for an optimal number of trips — ideally whenever the coin-box is full. The trip route should also be chosen so as to minimize costs. Unfortunately however, the time taken for coin-boxes to fill is not deterministic so that a decision on when to replace them may not be error-free. There are two kinds of errors associated with faulty decisions. Collecting from a coin-box before it is full translates into a larger number of trips and avoidable variable costs expenditure while a full coin-box makes the facility unusable for the time for which the coin-box is full, resulting in lost revenue, frayed consumer confidence and possible punitive action by regulatory agencies. Thus, it is of practical importance to provide optimal dates for when a coin-box should be collected. Of similar spirit and import is the problem of restocking vending and automated teller machines. Once accurate collection times are available, the business can set up an optimal schedule for visiting the machines identified as needing attention on a particular day.

1.1 Payphone Coin-boxes: An Illustration

The case for optimal collection-date decisions is well made by the payphone which is today an integral part of the urban and rural landscape. This facility is inexpensive, always available, convenient to use, reasonably private, and so very popular among consumers. A very large volume of calls is handled each day by payphones. For instance, in the area served by a very large regional company in the United States, an average of $500,000 worth of calls in coins are made every day! Each payphone has a small cubic box, of the same size, for collecting toll. Once this is full, the payphone is unusable until it is emptied. This is done by a business representative who visits the payphone and replaces the old coin-box with an empty one. As the coin-box is removed, it seals automatically (to prevent pilferage) and is shipped under security to one of very few specialized collection centers. Here the box is unsealed and the collected amount ascertained. Collection cost is a variable cost: it would be optimal to collect when the box is full. However, the fill-time for a coin-box is not deterministic because of the random frequency of the random number of calls at different toll-rates made from the payphone. Further, there are two components to the collection cost: the compensation for the representative, and the more substantial cost of shipping the coin-box under security and processing it at the specialized collection center. Moreover, the fact that the coin-box seals automatically means that even ignoring trip cost, sending in the business representative repeatedly is impractical because the coin-box can not be put back even if it is not found to be full. So a replacement decision is irreversible and should be made optimally.

There are different costs associated with errors in collection-time decisions. Under-estimation translates into a larger number of trips and increased shipping and processing costs for the coin-boxes, leading to less-than-optimal allocation of resources and variable costs. Over-estimation, on the other hand, results in full coin-boxes before collection. During this time, the payphone is unusable and turns away consumers, depleting consumer confidence and causing losses in revenue. More damaging, aggrieved customers may lodge complaints with local regulatory mechanisms, resulting in very stiff fines. So the risk of over-estimating collection times for a payphone is higher than that of under-estimation.

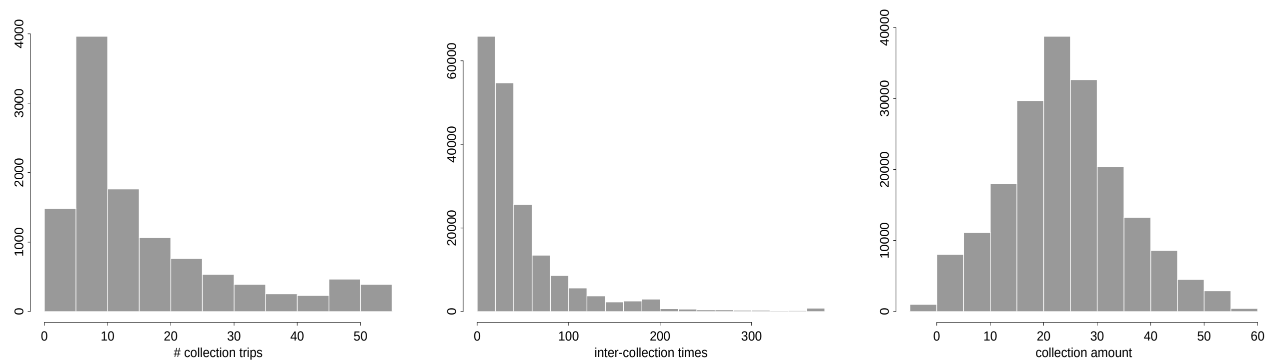
Payphones in the United States accept only three denominations of coins: the nickel, the dime and the quarter. These denominations are not just of different value, but also differ in size. This affects the amount in coins filling a coin-box. It is known that a coin-box filled with only nickels is worth around \$50 (Freedman, 1997). This figure is in itself random, because (circular) coins when deposited do not arrange systematically. The nickel is used as the unit of measure in calculating revenue in a payphone coin-box. All dimes and quarters are converted into *nickel equivalent* amounts using a formula (see Section 2) that accounts for both coin volume and value. The standardized measure of total amount in coins for a collection period is called the *nickel equivalent*. Because of the circular geometry of the coins, it is almost impossible to ascertain exactly how much a full box is worth in nickel equivalent: any collection of over \$50 in nickel equivalent is considered to be from a full box.

With its commercial implications, the problem of accurately deciding on the time for a coin-box collection trip has generated some interest. An in-house algorithm, COIN™ (see Appendix A for details), used by some large telecommunications firms, employs empirical reasoning to specify a formula for predicting fill-time of a coin-box. Experience shows that most of these predictions are unnecessarily conservative and result in collection trips when coin-boxes are far from full. Increasing inter-collection times without exceeding coin-box capacity is desirable in order to minimize collection costs. In this paper, we present an alternative approach to coin-box collection decisions, using data on payphone collection history over a large region in North America — see Section 2 — as an illustration. The goal is to optimally decide on when a coin-box should be collected. These decisions are needed as part of a larger decision-support system in which they will be used as input in order to devise an optimal trip schedule for visiting coin-boxes filling on a particular day. Our approach accounts for the stochastic nature of historical data in Section 3 and formulates collection-date decision in a likelihood and a Bayesian frame-work, under the assumption that the distribution of daily nickel equivalent is the same, everyday, at any given phone. Prior distributions on the mean daily nickel equivalent and standard deviation are specified and methodology to account for cases with censored fill-time data developed. Stochastic methods are used in our calculations, the exact implementation of which is described in Section 4. The results and performance evaluations are studied in Section 5. We conclude with a brief discussion on the results and useful pointers for further work.

2 The Data

As mentioned before, the dataset is a collection of records from all 76,017 pay-phones owned by a business entity in a very large region in North America. The records are of all collections from 1974 through 1996 and contain pay-phone identification number, collection date, nickel equivalent, and elapsed time (in days) between the previous and current collection dates. The *nickel equivalent* is the standardized measure accounting for both coin volume and value and is computed by first subtracting the amount collected in nickels from total revenue. This yields the *total silver* which is therefore the revenue accruing from dimes and quarters. The total silver upon division by 4 yields the *silver nickel equivalent*. The sum of the silver nickel equivalent and the revenue in nickels is defined to be the *nickel equivalent*.

Because of computer memory and disk-space limitations, we analyzed only collection records from between 1993 and 1996 on a random sample of 11,308 phones. There were a total of 189,433 such records on these phones.



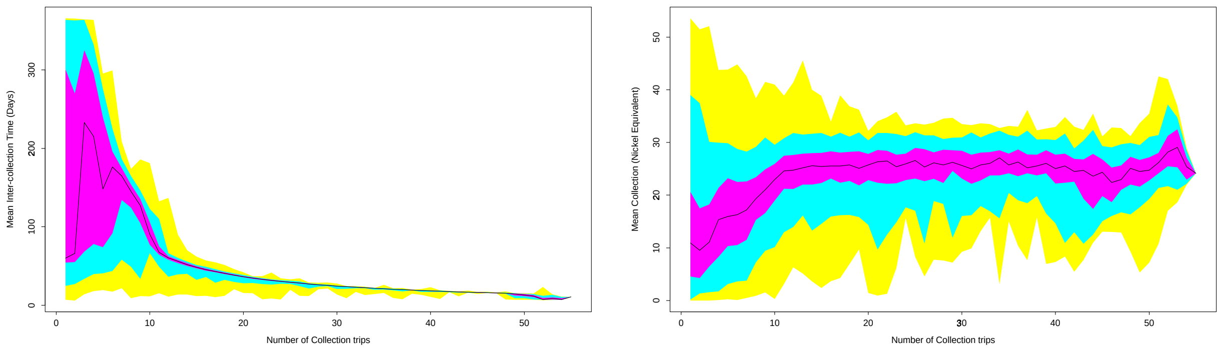
(a) (b) (c)

Figure 1: Histogram of (a) the total number of collections per phone and (b) the inter-collection times (in days) of all 189,433 collection records, and (c) the total amounts collected on each trip (in nickel equivalent).

Figure 1a displays a histogram of the total number of collection trips made to the pay-phones. The distribution is right-skewed, with a range from 1 to 55, and median and interquartile range of 11 and 14 trips, respectively. A large proportion of phones were clearly not visited for collection very often. However, there were a substantial number of phones (almost 15%) that were collected on the average of at least once a month.

Figure 1b is a histogram of the inter-collection times (in days) for the pay-phone coin-boxes. This distribution is also right-skewed, ranging from 1 day to 366 days with a median of 28 days, and an inter-quartile range of 42 days. In over 55% of the cases, collection records were from pay-phones that had been visited at least once a month. The histogram of the nickel equivalent in each trip (Figure 1c) indicates that almost 60% of the collection trips were made when the box was not even half-full. Indeed, the median collection was on the order of \$23.48 (in nickel equivalent) with an inter-quartile range of \$13.75. On the other hand, 3,334 (just under 2%) collection trips were made after the \$50 predicted threshold for a full coin-box had been met.

The averages of the inter-collection times and nickel equivalents for each pay-phone were computed and studied along with the number of collection trips for each phone. A few quantiles of these distributions for phones with same number of collection trips are presented in Figure 2. As expected, phones with more collection trips have lower inter-collection times on the average. It is also expected that these have higher mean nickel equivalent collections, because a good prediction



(a) (b)

Figure 2: Distribution of average – over all collection trips per phone – of (a) inter-collection times and (b) nickel equivalent, against the number of trips for each payphone. For both figures, the black bold line represents the median of the distribution, the extrema of the innermost bands (with dark shading) represent the two quartiles, the middle bands (with medium shading) the fifth and ninety-fifth percentiles, and the outermost bands (with the light shading) the minima and maxima of the respective averages.

algorithm should relate the number of trips to the daily nickel equivalent. Interestingly, this trend is only very slight. Further, there is more variability in both mean inter-collection times and mean nickel equivalents for phones having smaller number of collection trips.

For a more in-depth study, we chose 100 phones at random and studied the distributions of the nickel equivalent, the inter-collection times and the averaged daily nickel equivalent against the total number of collection trips. The patterns of Figure 2 were repeated for the first two distributions and are not reproduced here. There is no pattern for the nickel equivalent vis-a-vis the number of collection trips to the pay-phone (Figure 3). Ideally, a collection-trip should be made only when the coin-box is full; so the above observation is encouraging, but it is disconcerting that the median collection hovers well below the full-box threshold in all cases, and indeed is well below \$ 30 for most phones.

Figures 2 and 3 highlight the inadequacies of the current methodology. Payphones with similar daily nickel equivalent rates should have a similar number of trips which should also be directly related to the average daily nickel equivalent rate. This relationship is very slight. Further, the

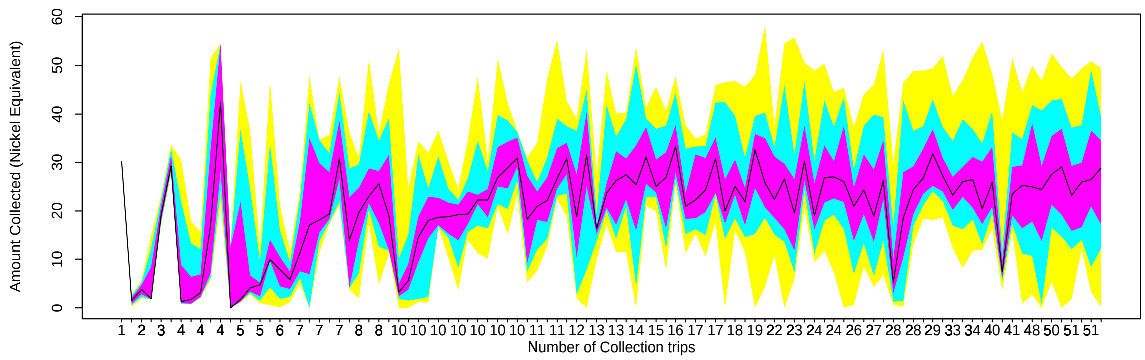


Figure 3: Distribution of nickel equivalent corresponding number of collection trips to each of 100 randomly selected payphones. The different quantiles are represented using the same descriptions as in Figure 2.

lack of pattern in the distribution of nickel equivalent vis-a-vis the number of collection trips and the fact that the nickel equivalent is in most cases well below $ 50 clearly indicates that it should be possible to increase inter-collection times without exceeding the threshold for a full coin-box.

Finally, the data were studied in order to see if the average daily nickel equivalent rate had a seasonal (monthly) component. Figure 4 displays a box-plot of the estimated daily nickel equivalent rate plotted against the month. The figure shows that there is not much difference between the middle part of any of the distributions. The only difference is in the tails — that is, while most of the tails are similar, the tails of the daily nickel equivalent measured in August is shorter than the others. Seasonal components were therefore excluded from our model formulations.

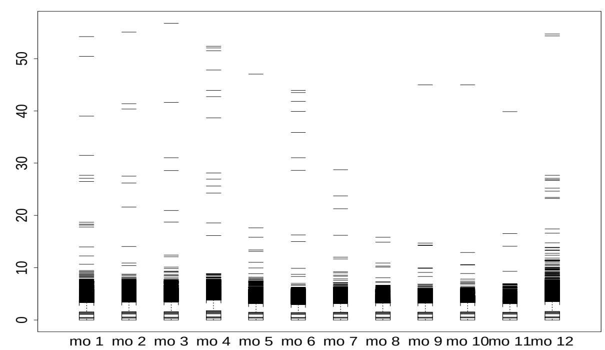


Figure 4: Boxplot of daily nickel equivalent over month, with "mo 1" denoting January, "mo 2" denoting February, and so on.

In our subsequent analyses, we have kept aside the last collection record for each payphone as our test set. The remaining 178,125 collection records, after some pre-processing (described in the next section) were used to make our predictions. The predictions were then tested using the test set, in order to evaluate predictive performance of our methodology in a realistic setting.

3 Statistical Methodology

3.1 Likelihood Modeling of Collection Amounts

The number of calls from a telephone has often been cited as a classical example of a Poisson Process (Feller, 1968; Karlin and Taylor, 1975). However, in our application, not all calls are of same duration or metered at the same rate. This is because toll for calls is based on number called and is sometimes charged at a fixed flat-rate for the first few minutes, followed by a varying per-minute rate. This means that a Poisson Process model for the coin-box collections may not be appropriate. Our proposed statistical model hinges on the premise that the daily collection at each phone has the same distribution every day for that phone. This assumption is reasonable in light of Figure 4. Formally, let θ denote the mean daily nickel equivalent for a given phone with σ the corresponding standard deviation. Further let X_{jk} denote the nickel equivalent collected on the kth day in the jth collection period at that phone and let $Y_j = \sum_{k=1}^{d_j} X_{jk}$ denote the total nickel equivalent from jth collection period of duration d_j days. Note that there is a hidden subscript i, denoting the ith phone, in all our notations which we ignore for the sake of notational simplicity. It follows from the above that Y_j is independent for each j and phone and for large d_j has an approximate normal distribution with mean $d_j\theta$ and variance $d_j\sigma^2$.

Our goal is to provide an optimal decision for the next collection time, d^* such that the probability of a full coin-box is limited (say, $1-p$). We choose d^* to be the smallest d such that $\mathbb{P}\left[Y_d^* \le 50\right] \le p$, where X_k^* is the nickel equivalent collected on the kth day after the last collection period, and $Y_d^* = \sum_{k=1}^{d} X_k^*$ is the amount collected within d days of the last collection. Under the approximate Gaussian assumption for Y_d^* and estimates $\hat{\theta}$ and $\hat{\sigma}$, we have:

$$\hat{d}^* = min\left\{d : \Phi\left[\frac{50 - d\hat{\theta}}{\sqrt{d}\hat{\sigma}}\right] \le p\right\} \tag{1}$$

as the replacement time for a coin-box. Here, Φ is the cumulative distribution function of the standard normal distribution.

3.2 A Bayesian Approach

The Bayesian approach specifies independent prior distributions $\pi(\theta, \sigma)$ on the box-specific mean daily nickel equivalent and standard deviation. Under this framework, our objective is to estimate the optimal replacement time for the payphone coin-box, *i.e.*,

$$\hat{d}^* = min\left[d : \mathrm{E}_\pi\left\{\Phi\left(\frac{50 - d\theta}{\sqrt{d}\sigma}\right) \mid Y_j; 1 \le j \le n\right\} \le p\right]. \tag{2}$$

Here n represents the number of uncensored records in the training dataset for the given coin-box.

3.3 A Decision-Theoretic Formulation

A decision-theoretic argument for (1) and (2) can be provided. In specifying a loss function, we note that given past historical data relating to collections over certain periods, the decision-problem is when to make the next collection trip. A longer collection period implies fewer overall trips and is attractive in terms of cost-cutting, but brings with it the severe risk of unhappy customers and punitive action by regulatory mechanisms in the event of a full coin-box. We specify the loss function for a given phone as:

$$L\left(d, \{\mathbf{Y}^*\}\right) = d^{-\frac{1}{2}}\left\{p(Y_d^* - M)1_{[Y_d^* > M]} + (1-p)(M - Y_d^*)1_{[Y_d^* < M]}\right\} \tag{3}$$

where $\{\mathbf{Y}^*\} = \{Y_d^* : d = 1, 2, \ldots\}$ is a listing of all the future cumulative collections made. M represents the full-box amount, and is equal to \$50 in nickel equivalent for the payphone application. Equation 3 very nicely captures the competing losses. To see this, first consider the expression within the braces. When the coin-box is not full by collection day d, $M - Y^*$ represents the loss in not waiting longer. On the other hand, if the coin-box is already full before d days, there is the possibility of fines and lost revenue. The first is assessed on the basis of the number of complainants, so it is reasonable to relate this to the amount that could not be collected because the payphone was unusable, or $Y^* - M$ as a surrogate for the (unavailable) number of affected consumers. Finally, the above is scaled by $\sqrt{d}$ such that errors caused when collecting in shorter time-intervals are penalized more severely than those caused when the collection time is large. The parameter p in (3) takes values between 0 and 1 and controls the trade-off between collecting early and collecting after the coin-box is full. For $p = 0.5$, the loss function reflects equal penalties for collecting early or late on a given day while for $p > 0.5$, there is greater penalty for collecting after fill-time than for collecting before.

In this scenario, (2) can be shown to be the Bayes rule minimizing the Bayes risk $r(\pi) = \int L(d, \{\mathbf{Y}^*\})\, \pi(\theta, \sigma \mid Y_j^*; 1 \leq j \leq n) d\theta d\sigma = I\!E_\pi(I\!E_{\theta,\sigma} L(d, \{\mathbf{Y}^*\}) \mid Y_j; j = 1, 2, \ldots, n)$.

Note that other loss functions such as linex loss (Zellner, 1986) may also be considered for this problem. However our choice of (3) provides us with the readily written solution in (2).

3.4 Incorporating Censoring

Our frame-work can be extended to include censored coin-box records (*i.e.*. those records that were full before collection). As before, let Y_js denote the uncensored data, d_j the corresponding elapsed times, and further let $W_j; j = 1, 2, \ldots, l$ denote the censored records at a given phone. In this setup, the likelihood function for the collections at each coin-box has the additional term

$$\prod_{k=1}^{l} I\!P_{\theta,\sigma}(Y_j > W_k) = \prod_{k=1}^{l} \left[1 - \Phi\left(\frac{W_k - d_k\theta}{\sigma\sqrt{d_k}}\right)\right]. \tag{4}$$

Implementation of both likelihood and Bayesian solutions to the problem is as before.

4 Implementation on Payphone Data

4.1 Preprocessing of the Training Data

The training dataset had a total of 7,219 phones with all collection intervals of at least 20 days. The asymptotic Gaussian likelihood assumption is reasonable here. On the other hand, there were 60,980 collection records (about 34.2% of the training dataset) with inter-collection times of less than 20 days. For the phones associated with these records, we merged adjacent records successively until we obtained a single merged record of at least 20 days. If any of the records being merged was censored, we considered the merged record to be also so. This strategy was possible for all but three phones: these three had been operational for less than 20 days in our training dataset. Consequently, we ignored the records for these three phones in our analysis. The processed training dataset now consisted of 140,854 collection records, of which 2,520 were censored. We used this dataset to implement and evaluate the methods described in this paper.

4.2 Maximum Likelihood Methods

Under the likelihood model in Section 3.1, the maximum likelihood estimators for θ and σ are:

$$\hat{\theta} = \sum_{j=1}^{n} Y_j / \sum_{j=1}^{n} d_j; \qquad \hat{\sigma}^2 = \frac{1}{n}\sum_{j=1}^{n}(Y_j - d_j\hat{\theta})^2/d_j. \tag{5}$$

Using these estimates in (1) with $p = 0.95$ provided $\hat{d}_i^*$, the optimal decision on date for collection after the previous trip.

The likelihood approach can only be implemented when $n > 1$: hence, the 370 payphones with $n \leq 1$ were discarded for this analysis. The residuals obtained from the uncensored collection records after centering and scaling the total nickel equivalent by estimated mean and standard deviation, were found to be symmetrically distributed. A Shapiro-Wilks' test (Royston, 1982) when performed on a sample of size 2,000 of these residuals reported a test statistic of 0.986, with a p-value of 0.999, indicating support for our likelihood model assumptions.

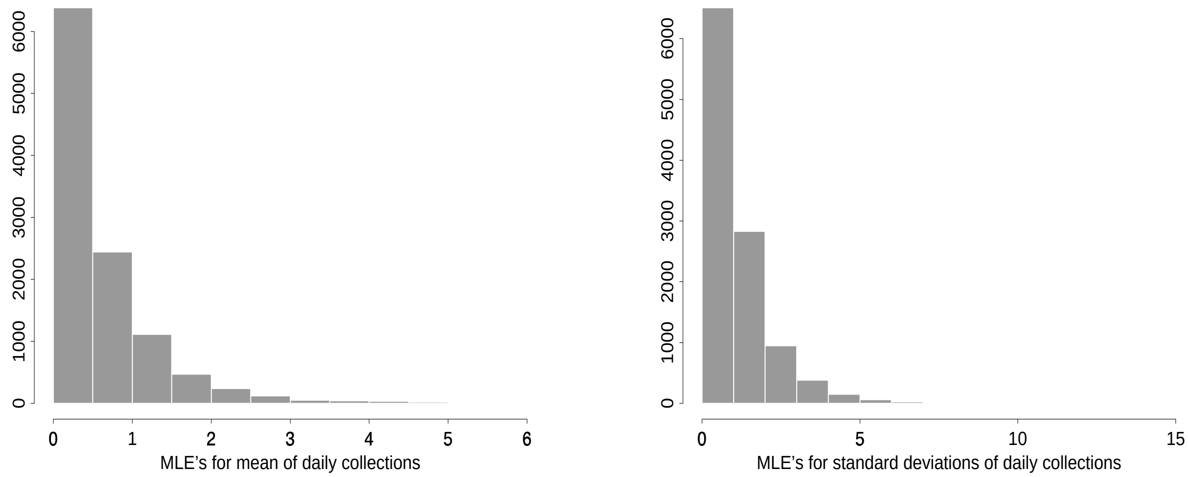
4.3 Bayesian Implementations

4.3.1 Prior Specification

To decide on a prior for θ and σ, we investigate the properties of the maximum likelihood estimates. Figure 5 displays the histograms of these estimates. While $\hat{\theta}$ and $\hat{\sigma}$ have the same basic distributional characteristics as gamma, both (and in particular $\hat{\sigma}$) are substantially heavier-tailed. Further the correlation coefficient between $\hat{\theta}$ and $\hat{\sigma}$ is 0.71. The high correlation can perhaps be explained by the fact that a good proportion of underlying phone calls are fixed-toll, short-duration local calls and therefore the corresponding nickel equivalent component can be modeled in terms of a scaled version of the Poisson distribution. We specify a simple prior distribution that captures the gamma-like shapes of the marginal distributions as in Figure 5 and also the interaction between θ and σ. Our specified prior is:

$$\pi(\theta, \sigma \mid \alpha, \beta, \gamma, \delta, \eta, \zeta) \propto \exp\left\{-\frac{\theta}{\beta} - \frac{\sigma}{\delta} - \eta \mid \theta - \zeta\sigma \mid\right\} \theta^{\alpha-1}\sigma^{\gamma-1}, \qquad 0 < \theta, \sigma < \infty \tag{6}$$

where the prior hyperparameters $\alpha > 0$, $\beta > 0$, $\gamma > 1$, $\delta > 0$, $\eta \geq 0$ and $\zeta > 0$ determine the constant of proportionality. The parameter η measures the strength of interaction between θ and σ. To see this, note that the prior prefers realizations where σ after scaling by the parameter ζ is



(a) (b)

Figure 5: Maximum-likelihood estimates of (a) mean daily nickel equivalents and (b) standard deviations

close to θ, and this preference is tempered by the interaction parameter η. Thus $\eta = 0$ implies that θ and σ are independently distributed as $\Gamma(\alpha, \beta)$ and $\Gamma(\gamma, \delta)$ respectively while for very large η, the prior is degenerate and puts mass where θ is a scaled version of σ. The choice of the absolute value of differences between θ and σ after scaling means that we do not penalize large deviations as severely as we would if we were using, say, squared differences.

We next checked for empirical evidence to support the use of (6). Regarding the $\hat{\theta}$s and $\hat{\sigma}$s as a sample from this distribution, we obtained maximum likelihood estimates of the hyperparameters to be $\hat{\alpha} = 0.791$, $\hat{\beta} = 0.729$, $\hat{\gamma} = 1.159$, $\hat{\delta} = 5.926$, $\hat{\eta} = 2.275$, and $\hat{\zeta} = 0.762$. (Note that the $\hat{\theta}$s and $\hat{\sigma}$s contain additional sampling variation on top of the between-coin-box variation, which the mixing distribution trying to estimate here. This means that the estimated mixing distribution will be overdispersed, and this overdispersion will be reflected in the hyper-parameter estimates.) With these estimates, we calculated the test statistic:

$$T(\pi, \hat{F}_n) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \mid \phi_\pi(t_1, t_2) - \phi_n(t_1, t_2) \mid dt_1 dt_2,$$

where $\phi_\pi(t_1, t_2)$ is the characteristic function for $\pi(\theta, \sigma \mid \hat{\alpha}, \hat{\beta}, \hat{\gamma}, \hat{\delta}, \hat{\eta}, \hat{\zeta})$ and $\phi_n(t_1, t_2) = \mathbb{E}_{\hat{F}_n}(e^{it_1\theta + it_2\sigma})$ is the characteristic function using the empirical joint cumulative distribution function $\hat{F}_n$ of the θs and σs. This test statistic had a value of 173,157.3. The reference distribution of $T(\pi, \hat{F}_n)$ was estimated by resampling. We obtained samples of size n from π and calculated the statistic

$$T(\pi, F^*) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \mid \phi_\pi(t_1,t_2) - \phi_n^*(t_1,t_2) \mid dt_1 dt_2,$$

for each sample. Here $\phi_n^*(t_1,t_2)$ is the characteristic function calculated using the new sample. Replicating this procedure 1,000 times gave us the reference distribution for our test statistic from which the p-value of our test statistic $T(\pi, \hat{F}_n)$ was computed to be 0.119, indicating support for (6) as our choice of prior.

The joint posterior distribution of θ and σ is independent of those for other phones and is given by:

$$\begin{aligned}\pi(\theta,&\sigma \mid Y_j; 1 \le j \le n) \\ &\propto \exp\left\{-\frac{\theta}{\beta} - \frac{\sigma}{\delta} - \eta \mid \theta - \zeta\sigma \mid - \sum_{j=1}^{n} \frac{(Y_j - d_j\theta)^2}{2 d_j \sigma^2}\right\} \theta^{\alpha-1}\sigma^{\gamma-n-1}, \qquad 0 < \theta, \sigma < \infty \end{aligned} \tag{7}$$

Note that the condition $\gamma > 1$ guarantees a proper posterior distribution when $n = 1$. The properties of (7) depend on the prior hyperparameters α, β, γ, δ, η and ζ. We treated these hyperparameters in two ways: one was to empirically estimate these from the observed training data while the other was to specify a hyperprior distribution on them.

4.3.2 Empirical Bayes Solution

The Empirical Bayes approach estimated the hyperparameters from the data by first marginalizing (7) over the θs and σs and then applying the principle of maximum likelihood. Using a multi-dimensional downhill simplex method (Nelder and Mead, 1982), these estimates were found to be $\hat{\alpha} = 0.675$, $\hat{\beta} = 0.517$, $\hat{\gamma} = 1.729$, $\hat{\delta} = 419.303$, $\hat{\eta} = 2.856$ and $\hat{\zeta} = 0.733$. Interestingly, δ obtained here is very different from that in Section 5.1, while the other hyperparameter estimates are similar. Analytic calculations of the expectation operator in (2) being intractable, we performed our computations using both numerical and stochastic methods.

For the numerical integration, a 15-point transformed Gauss-Kronrod rule (Piessens *et al*, 1983) was used for evaluating the infinite integrals. We also obtained upper bounds on the absolute error in numerical integration. The stochastic methods were performed using Markov Chain Monte Carlo simulation. The latter involves specifying full conditional distributions for the θs and σs and then generating realizations from an ergodic Markov Chain with the posterior distribution as the stationary distribution. Note that under the assumption of independent prior distributions for average collection rates and variances across payphones, the full conditionals for each θ and σ at one phone are independent of the rest. These are:

$$\log \pi(\theta \mid \sigma, Y_j; 1 \leq j \leq n) = const. + (\alpha - 1)\log\theta - \frac{\theta}{\beta} - \eta \mid \theta - \zeta\sigma \mid - \sum_{j=1}^{n} \frac{(Y_j - d_j\theta)^2}{2d_j\sigma^2} \tag{8}$$

$$\log \pi(\sigma \mid \theta, Y_j; 1 \leq j \leq n) = const. + (\gamma - n - 1)\log\sigma - \frac{\sigma}{\delta} - \eta \mid \theta - \zeta\sigma \mid - \sum_{j=1}^{n} \frac{(Y_j - d_j\theta)^2}{2d_j\sigma^2}. \tag{9}$$

Many versions of Markov Chain Monte Carlo methods can be used: for a detailed description, we refer to Besag *et al.* (1993), Tierney (1994), Chib and Greenberg (1995) and the references therein. A natural choice is the Gibbs sampler (Geman and Geman, 1984): however, given the computational demands placed by sampling from the full conditionals, we used a Hastings algorithm (Hastings, 1970). Given a current realization s, the proposal distribution for a new realization s^*, was a mixture of Gaussians specified as

$$q(s, s^*) = \sum_{k=1}^{3} \pi_k \phi(s^*; s, \tau_k(s)) \tag{10}$$

where $\phi(x; \mu, \lambda)$ denotes the mean-μ-variance-λ^2 Gaussian density evaluated at x, $\boldsymbol{\pi} = \{\pi_k; k = 1, 2, 3\}$ lies in the three-dimensional simplex and the elements of $\boldsymbol{\tau} = \{\tau_k; k = 1, 2, 3\}$ are all positive. We chose $\boldsymbol{\pi} = (0.35, 0.35, 0.30)$ for all θ- and σ-spaces. For the θ-space, the choice of $\tau_k = \xi_k 1_{[n=0]} + \upsilon_k 1_{[n>0]}(\sum_{j=1}^{n} Y_j + 0.01)/(\sum_{j=1}^{n} d_j)$, where $\boldsymbol{\xi} = \{\xi_1, \xi_2, \xi_3\} = \{0.5, 2.5, 5.0\}$ and $\boldsymbol{\upsilon} = \{\upsilon_1, \upsilon_2, \upsilon_3\} = \{0.01, 0.05, 8.0\}$. For the σ-space, τ_ks had the same form, with choices for $\boldsymbol{\xi}$ and $\boldsymbol{\upsilon}$ as $\{0.25, 3.0, 10.0\}$ and $\{0.1, 0.5, 10.0\}$ respectively. Note that the choice of (10) with the above $\boldsymbol{\tau}$ implies that the proposal distribution $q(s, s^*)$ is symmetric and the Hastings' algorithm reduces to a Metropolis algorithm (Metropolis *et al.*, 1953, Hammersley and Handscomb, 1964). Samples from the posterior distributions were generated by running the sampler from an initial random realization with a sequential visiting schedule and a burn-in period of 10,000 iterations. A sub-sample of size 500 – comprising every hundredth realization in the Markov Chain after burn-in – was stored and used to carry out the calculation of (2) in the predictions. The means of the posterior distributions of θs and σs were also estimated using this sample. Autocorrelation times and Monte Carlo standard errors of these estimates were calculated using the methods described in Sokal (1989), Besag and Green (1993) and Green and Han (1993).

4.3.3 Fully Bayesian Solution

In the hierarchical fully Bayesian framework, we specified a density on the hyperparameters which was flexible enough to model the uncertainty in the hyperparameter values. Since α, γ and β, δ play a similar role to the shape and scale parameters of the gamma distribution, we chose non-informative hyperpriors corresponding to those obtained by using Jeffreys' principle (Jeffreys, 1939) for the gamma distribution. These were, therefore specified as:

$$\pi(\alpha) \propto \mid \psi(\alpha) \mid^{\frac{1}{2}}, \qquad \pi(\gamma) \propto \mid \psi(\gamma) \mid^{\frac{1}{2}}, \qquad \pi(\beta) \propto \frac{1}{\beta}, \qquad \pi(\delta) \propto \frac{1}{\delta}, \qquad \alpha, \beta, \delta > 0, \gamma > 1,$$

where $\psi(\cdot)$ is the first derivative of the digamma function. Hyperpriors for η, ζ were also non-informative and chosen to be uniformly distributed over the positive quadrant.

Analytic or numerical methods for calculating (2) being again impractical, required computations were performed by sampling, via Markov Chain Monte Carlo, from the full conditionals of α, β, γ, δ, η, ζ and the θs and σs. The full conditionals of the θs and σs are the same as in (8) and (9) while those for the prior hyperparameters are presented in Appendix B. These were used to obtain Markov Chain Monte Carlo samples from the joint posterior distribution of α, β, γ, δ, η, ζ, θs and σs. Restricting attention only to the θs and σs among these realizations provided a sample from the marginal posterior distribution of θs and σs.

Metropolis algorithms using the same proposal distributions as before were employed while sampling from the full conditionals of θs and σs. We also used Metropolis while sampling from the full conditionals of α, β, γ, η and ζ. The proposal distribution in each case was Gaussian and centered at the current state, with standard deviations of 0.01, 0.01, 0.01, 0.05 and 0.005 respectively. While sampling from the full conditional of δ, we used a Gaussian proposal distribution with both mean and standard deviation as the current state – note that this is no longer a symmetric proposal but a general case of the Hastings' algorithm. The simulations were run with a random initial realization and a random visiting schedule. Other characteristics such as burn-in period, chain length, sub-sampling rate were same as before.

4.4 Incorporating Censored Records

There were 2,520 censored records in the training dataset, corresponding to observations made on 1,838 phones. Eight phones had all records censored. These 2,520 censored records together with the 138,334 complete collection records were used in the computations. We only reported fully Bayesian solutions for this problem, which were computed using the same proposal distributions and burn-in, chain length and sub-sampling rate as in Section 4.3.3.

5 Results and Performance Evaluations

In this section, we analyze the results of the implementation of our suggested approaches on the payphone test dataset. For any coin-box, we denote the nickel equivalent in the test dataset by Y_{n+1} and the corresponding collection time as d_{n+1}. The average nickel equivalent collected here was \$ 29.66 while the average inter-collection time was 116.82 days. About 5.84% of the coin-boxes exceeded full-box threshold of \$ 50 in nickel equivalent before collection.

All decisions on collection times were made using $p = 0.95$ in (3). This means that the ratio of loss due to over-estimation was set to be 19 times that due to under-estimation. Further, the optimal $\hat{d}^*$ satisfying (1) and (2) were constrained to be below 366 days, implying a decision to collect the coin-box at least once a year, which is reasonable.

While reporting results on the performance evaluations, we use the term coin-box *current nickel equivalent* and *current collection date* to denote Y_{n+1} and d_{n+1} respectively. We use $\hat{d}^*$ to denote the optimal collection date obtained using the method being evaluated. The term Y^* is the corresponding predicted nickel equivalent if the coin-box collection was made on the $\hat{d}^*$th day. For each approach, we evaluated performance using a series of measures. We calculated the proportion of coin-boxes for which $\hat{d}^*$ was longer than d_{n+1}. Next, given Y_{n+1} and d_{n+1}, we calculated the conditional probability of a full coin-box by day $\hat{d}^*$. Thus, we calculated $\mathbb{P}_{\hat{\theta},\hat{\sigma}}(Y^* > 50 \mid Y_{n+1})$ for the maximum likelihood approach, whereas for the Bayesian formulations, we used

$$\mathbb{E}_{\theta,\sigma}\left[\mathbb{P}(Y^* > 50 \mid Y_j, j = 1, 2, \ldots, n+1)\right], \tag{11}$$

which we call the *conditional posterior full-box probability measure.* A similar measure is calculated for the case that includes the censored records. We also calculated the *mean increase in collection intervals* by averaging $\hat{d}^* - d_{n+1}$ over all payphones and the corresponding *mean per cent relative increase*, obtained by averaging $100(\hat{d}^* - d_{n+1})/d_{n+1}$. The latter measure is also the potential per cent reduction in collection costs.

Remark: We use the qualifier potential when describing the possible per cent reduction in collection costs because the optimal route decided on by the coin-collection authority may not always mirror the locational sequence of our optimal collection-time recommendations. However, given that trip costs are a fraction of collection costs which also include shipping and processing at the specialized collection center, the actual per cent savings should be close to the potential value.

5.1 Maximum Likelihood Methods

The estimated $\hat{\theta}$s and $\hat{\sigma}$s had means of 0.626 and 1.136 respectively. Their histograms were presented in Figure 5. The first column in Table 1 summarizes the performance of this approach. Optimal collection intervals $\hat{d}^*$ were larger than d_{n+1} in 73.8% of the cases. Within these cases, Y_{n+1} was greater than 50 (so our decisions are guaranteed to be worse) in 533 (or 4.9 %) cases. For cases where our suggested intervals were wider, the distribution of $\mathbb{P}_{\hat{\theta},\hat{\sigma}}(Y^* > 50 \mid Y_{n+1})$ was considerably right-skewed, with a zero median and an upper quartile of 0.20. Out of the 2,594 cases for which $\hat{d}^* < d_{n+1}$, Y_{n+1} was less than \$ 50 in 2,399 cases, which means that in these cases our decisions were unnecessarily conservative. For the remaining 195 cases however, the distribution of the computed $\mathbb{P}_{\hat{\theta},\hat{\sigma}}(Y^* > 50 \mid Y_{n+1})$ had a median of 0.33, and an inter-quartile range of 0.55. The overall increase in inter-collection times was on the order of 21 days, translating into a potential per cent reduction in collection costs of around 29.8%.

5.2 Empirical Bayes Method

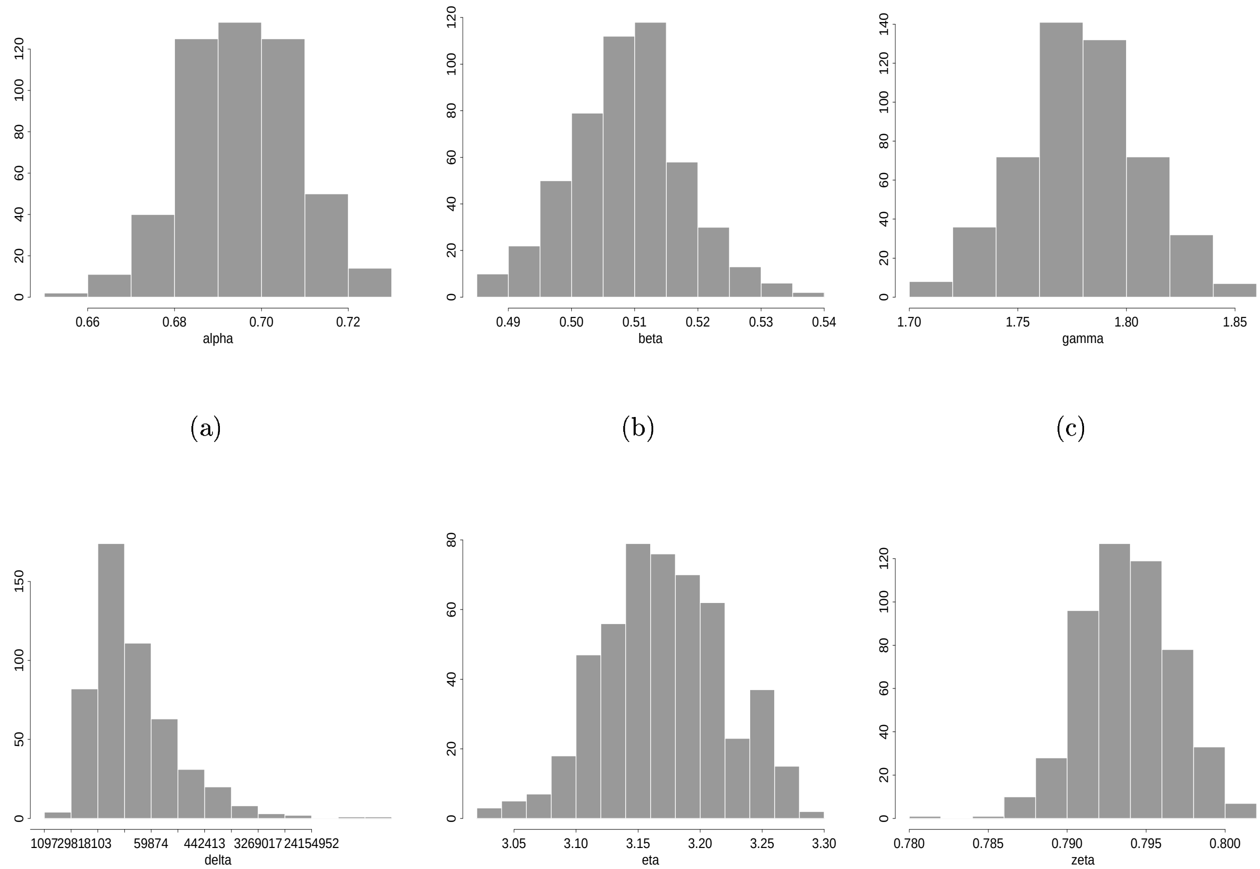
Our Metropolis algorithms had acceptance rates ranging from 25.7% to 99.3%. The chain did not suffer from problems that arise from strong interactions or multi-modality. Median autocorrelation times while estimating the posterior means of θs and σs were both 4. Monte Carlo standard errors, relative to the estimated posterior means of θs and σs had medians of 0.45% of 1.06% respectively, and corresponding inter-quartile ranges of 0.71% and 0.58%. The upper and lower 2.5% points of the distribution of these per cent relative standard errors were at (0.15, 4.23) for θs and (0.51, 4.63) for σs respectively. Within the limits of sampling variability, the stochastic and numerical methods of estimating the posterior means of θs and σs were indistinguishable for

Table 1: Summary of performance of computations done using (a) maximum likelihood (ML), (b) empirical Bayesian, (c) fully Bayesian and (d) fully Bayesian methods with censored records included. ML predictions are possible only for 10,938 cases for which $n > 1$. In row 3, we report $\mathbb{P}_{\hat{\theta},\hat{\sigma}}(Y^* > 50 \mid Y_{n+1})$ for the ML approach and (11) for the other approaches. The medians are reported with inter-quartile ranges in parenthesis.

Performance Measure	**ML**	**Emp. Bayes**	**Fully Bayes**	**With Censoring**
# cases	10,938	11,308	11,308	11,308
% $(\hat{d}^* > d_{n+1})$	73.8	71.2	72.2	69.8
$\mathbb{P}(Y^* > 50 \mid Y_{n+1})$	0.000 (0.228)	0.0001 (0.120)	0.0001 (0.127)	0.000 (0.094)
Mean $(\hat{d}^* - d_{n+1})$	20.44	11.34	12.68	10.56
Mean % rel. increase	29.78	23.98	25.34	20.97

about 20% of the cases. However, these were significantly different for a large number of cases. Interestingly, these cases were also those where the numerical integrals had high relative upper bounds on their approximation errors. Indeed, relative to the values of the numerical integral, these upper bounds had medians of 152% and 149% for the posterior means of the distributions of θs and σs, and corresponding inter-quartile ranges of 22% and 24%. The high upper bounds on the numerical approximation errors point to the possible difficulty of numerically evaluating these integrals and caution us while using these results. However, it satisfying to note that for cases where the numerical integrals are evaluated with guaranteed high precision (*i.e.* those with low upper bounds on the relative approximation error), the differences in MCMC-estimated and numerically-evaluated integrals are not statistically significant. We thus report here only results obtained using stochastic methods.

The posterior means of θs and σs ranged from 0.002 to 5.623 and 0.032 to 8.75 respectively, with medians and inter-quartile ranges of 0.376 and 0.605 for the θs and 0.933 and 0.901 for the σs. Table 1 (second column) summarizes the performance of the empirical Bayes approach. Optimal $\hat{d}^*$ were larger than d_{n+1} in 71.2% cases. Within these, Y_{n+1} was greater than 50 for only 455 (5.6%) payphones. Overall, for cases where $\hat{d}^* > d_{n+1}$, the calculations done using (11) had a right-skewed distribution with median 0.005 and an upper quartile of 0.199. For the 3,258 phones for which $\hat{d}^* < d_{n+1}$, Y_{n+1} was below 50 for 3,046 cases. For the other 212 cases, the distribution



(a) (b) (c)

(d) (e) (f)

Figure 6: Histogram of estimated marginal posterior distributions of the hyperparameters: (a) α, (b) β, (c) γ, (d) δ, (e) η and (f) ζ. The histogram for δ is in the log-scale.

of the conditional full-box posterior probability measures (11) was right-skewed with a median of 0.327 and an inter-quartile range of 0.498. Overall, as reported in Table 1, the increase in optimal collection intervals over current methods averaged 11.34 days, translating into a potential relative reduction in variable collection costs of 23.4%.

5.3 Fully Bayesian Approach

Acceptance rates of our Metropolis-Hastings algorithms ranged from 26.2% to 99.9%. The chain was also free from problems caused by multi-modality or from phenomena such as critical slowing down. Posterior distributions of the hyperparameters are summarized in Figure 6 and Table 2

Table 2: Descriptive measures of the posterior distributions of the hyperparameters using only complete observations (left-block) and all observations (right block). Monte Carlo standard errors of estimated means are in parenthesis.

Hyper-parameter	First Quartile	Median	Mean (Monte Carlo SE)	Third Quartile	First Quartile	Median	Mean (Monte Carlo SE)	Third Quartile
α	0.686	0.695	0.695 (0.0007)	0.704	0.697	0.707	0.707 (0.0006)	0.716
β	0.503	0.509	0.509 (0.0004)	0.514	0.528	0.534	0.534 (0.0004)	0.540
γ	1.762	1.780	1.779 (0.0015)	1.797	1.672	1.688	1.688 (0.0012)	1.705
δ	543.98	1000.97	45160.0 (35441.05)	3145.62	696.0	1548.4	36206.8 (16835.7)	4462.6
η	3.137	3.169	3.170 (0.0029)	3.203	2.834	2.867	2.865 (0.0019)	2.894
ζ	0.781	0.792	0.794 (0.0001)	0.796	0.780	0.782	0.782 (0.0001)	0.784

(left block). The posterior marginal distributions of α, γ and η are more or less symmetric while they are mildly skewed for β and ζ. The posterior marginal distribution of δ is, however, very right-skewed. The estimated posterior hyperparameter means of α, β, γ, η and ζ are significantly different from the values empirically estimated from the data in Section 4.3.2. But the difference between the estimated mean of the posterior distribution of δ and the empirical Bayes estimate is not statistically significant and can be explained by sampling variability.

Per cent relative Monte Carlo standard errors of posterior means of θs and σs estimated from our Markov Chain had medians of 0.49% and 1.09% and inter-quartile ranges of 0.74% and 0.80% respectively. Median autocorrelation times were on the order of 5 and 4 respectively. The posterior means for θs ranged from 0.0005 to 5.768 with medians and interquartile ranges of 0.377 and 0.606 while posterior means for σs ranged from 0.016 to 8.386 with a median of 0.908 and an inter-quartile range of 0.881. The posterior means of the θs obtained using the fully Bayesian approach were significantly different from those obtained using the empirical Bayesian approach in only 859 (or about 7.6%) phones. Disagreement was more pronounced in the case of the posterior means of σs, with 4,158 (or about 36.8%) cases reporting significant statistical differences.

Table 1 (third column) summarizes the performance of the fully Bayesian approach. Our optimal collection times $\hat{d}^*$ were longer than d_{n+1} in 8,164 or 72.2% of the cases. Of these, 457 cases were guaranteed to be worse, since Y_{n+1} was greater than 50. For the remaining 7,707 cases, the distribution of the conditional full-box probability measures (11) had a median of 0.0049 and an upper quartile of 0.203. For the 3,144 phones for which $\hat{d}^*$ was less than d_{n+1}, Y_{n+1} was below 50 in 2,937 of the cases. For the other 207 phones, computed values of (11) had a right-skewed distribution with a median of 0.332 and an inter-quartile range of 0.435. On the average, our optimal collection times were 12.68 days longer than current ones, translating into a per cent potential reduction in collection costs of 25.3%.

5.4 Incorporating Censored Records

Acceptance rates for the Metropolis-Hastings' algorithms ranged from 25.8% to 99.9%. The histograms for the posterior distributions of the hyperparameters had shapes similar to those in Figure 6. However, the numerical descriptive measures for these posterior distributions, presented in Table 2 (right block) indicate differences. Posterior means of α, β, γ, η and ζ obtained here were significantly different from those in Section 5.3, while the difference in the posterior means of δ obtained using the two methods was not statistically significant.

Per cent relative Monte Carlo standard errors of the estimated means of θs and σs had medians of 0.52% and 1.09% and inter-quartile ranges of 0.76% and 0.82% respectively. Median autocorrelation times were on the order of 5 and 4. Posterior means of θs ranged from 0.0005 to 6.596 with a median of 0.382 and an interquartile range of 0.621 while those for σs ranged from 0.016 to 9.242 with a median of 0.95 and an interquartile range of 0.955. Differences in posterior means were statistically significant from those in Section 4.3.3 in 2,280 (20.1%) cases for the θs and 2,796 (24.7%) cases for the σs. The performance of our approach is summarized in Table 1 (fourth column). Our optimal $\hat{d}^*$s were longer than d_{n+1} in 7,888 (69.8%) cases, of which Y_{n+1} exceeded 50 for 436 (or 5.6%) phones. For the other 7,452 cases, the calculations of (11) were right-skewed with a median of 0 and an upper quartile of 0.09. Of the 3,420 cases for which $\hat{d}^* < d_{n+1}$, Y_{n+1} is below 50 in 3,120 cases, so our algorithm was unnecessarily conservative for these cases. For the other 230 phones for which $\hat{d}^* < d_{n+1}$, our calculations of (11) were right-skewed with a median of 0.321 and an interquartile range of 0.495. On the average, our $\hat{d}^*$s were 10.56 days longer than the d_{n+1}s, translating into a per cent potential reduction of collection costs by 21.0%.

Remarks:

1. Using a statistical model improves decisions on collection times and reduces collection costs.

2. Average potential collection cost reductions using a maximum likelihood approach (Section 5.1) was 29.8% while those using the Bayesian approaches ranged from 21.0% and 25.3%. This is not disappointing: the calculations for the Bayesian solutions can be applied even for phones with little historical information, *i.e.* $n \leq 1$. It is reasonable to expect that predicting fill-dates for phones with little historical information would make the problem harder. It is encouraging that the test dataset validates that the probability of a full-box using d^* is low for the Bayesian solutions and lowest when including censored records. Further, under all approaches, we did worse in around 5% of the cases (since $Y_{n+1} > 50$ at time d_{n+1} and we recommended $d^* > d_{n+1}$) which is our desired error in precision.

3. Interestingly, both empirical and fully Bayesian approaches provided close results. But accounting for uncertainty in the prior hyper-parameters improved predictive performance, as shown by improved performance of the fully Bayesian approach.

4. Including censored collection records did not substantially reduce collection cost. However, the conditional probability of a full-box using the test dataset was reduced further.

5. The exact choice in functional form to represent the interaction between the θs and the σs in (6) did not influence the results; replacing the absolute values of the scaled difference between θ and σ with the corresponding squared values yielded similar results.

6 Discussion

This paper exhibits an alternative approach to optimal decisions on collection times for coin-boxes of machines such as payphones, vending-machines and parking-meters. Bayesian solutions are developed and tested on data that are records of collections at different payphones in a geographically large region in North America. Our results indicate good performance and if adopted, have the potential of substantially reducing variable costs associated with collection trips. In particular, the possible variable cost reductions on the average are at least 21%. The methodology presented here is perhaps better appreciated when the cost reduction is placed in the context of the hundreds of millions of dollars annually at stake in the payphone business. Though the solution to the problem has been implemented and tested here in the frame-work of predicting fill-times for payphone coin-boxes, the methodology is general enough to incorporate covariate information and also be readily extended to applications such as predicting collection times for parking-meters or vending machines.

A few issues need to be discussed. With the technological advances, so-called “smart” payphones and vending-machines have been developed. These machines ring up a control center when a certain proportion of their coin-boxes are full. However, while it will probably be a very long time before every payphone gets replaced by its smarter cousin, it is also our view that the problem of deciding optimal collection times for coin-boxes remains and simply gets translated in scale. In order to reduce the risk of punitive action by regulatory agencies (which may occur, for instance, if a customer wants to use a payphone in between the time it has filled and the time the representative replaces it), the smart machine would possibly have to be set up to ring up the control center after, say, 80% of the coin-box has filled. Then the problem would be to decide on the time taken for the remaining proportion (20%) of the coin-box to fill. The solutions presented here would also be applicable in this revised scenario.

A second issue concerns the problem of predicting the time for replenishing stocks at vending or automated teller machines. This problem is complementary in scope to the one discussed in this paper and the methods detailed here are also applicable there. Instead of only predicting the fill-time for the coin-boxes, we will also be predicting the time taken for stocks to be sold.

Another issue concerns incorporating optimality of visiting schedules of the coin-collection authority. We have here provided the decision-makers with the wherewithal for optimal decisions of collecting from coin-boxes. On top of this, the decision-makers need to formulate an optimal visiting schedule incorporating distances between the payphones, accessibility, and proximity of the payphone locations. This would mean, as explained in the introduction, the setting up of some sort of an optimal visiting scheme, such as the traveling salesman algorithm, for the range of coin-boxes identified as needing attention on a particular date.

A final comment pertains to the choice of priors for the different coin-boxes. In our application, we chose independent distributions. In some applications, it may be more realistic to specify a spatial dependence structure for which our methodology could be extended with some modifications. Therefore, while the approaches suggested in this paper offer improved fill-time prediction for a coin-box, there are at a least a few interesting issues requiring further attention.

Appendix A: The COIN™ Prediction Algorithm

In this section, we describe briefly the COIN™ algorithm that is currently in use in predicting the next time for collection in a coin-box. The prediction formulae are described by a set of mathematical equations and comprise a set of four equations:

1. *The Predicted Daily Nickel Equivalent Equation* is the predicted daily nickel equivalent collected in a pay-phone coin-box and is denoted for the kth collection period as:

$$\hat{X}_k = 0.7X_{k-1} + 0.3H_{k-1},$$

where X_{k-1} is the daily nickel equivalent collected at the pay-phone in the $(k-1)$th (*i.e.*, previous) collection period. Here H_{k-1} is the auxiliary variable at the $(k-1)$th collection period and is obtained from the auxiliary variable equation.

2. *The Auxiliary Variable Equation* is defined as:

$$H_k = 0.8X_{k-1} + H_{k-1} - 0.8\hat{X}_{k-1}.$$

The role of the auxiliary variable equation is to act as a buffer and to make $\hat{X}_k$ conform more closely with the normal revenue trend of the pay-phone. Note that $H_0 = 0$.

3. *The Coefficient of Variation Equation* reacts to differences between the predicted daily nickel equivalent and the actual collected. The idea is that coin stations with relatively stable daily nickel equivalent rates for successive collections should cause the coefficient of variation V_k to decrease while those with high variations should cause it to increase. This is defined as:

$$V_k = 0.25(1 - \frac{\hat{X}_{k-1}}{X_{k-1}}) + 0.8V_{k-1}.$$

For pay-phones with no previous collection history, $V_k = 0.3$. Note that the above has nothing in common with the more common statistical term denoted by the same.

4. *The Predicted Interval Equation* is the predicted time till the next collection. Assuming a full coin-box threshold of \$ 50 nickel equivalent, this is given by:

$$\hat{d}_k = 50\frac{(1 - 0.7V_k)^2}{\hat{X}_k}.$$

Appendix B: The Optimal Decision Rule – Proof

The risk function is given by $R(d, \{\mathbf{Y}^*\}) = d^{-\frac{1}{2}}\mathbb{E}_{\theta,\sigma}\left[p(Y_d^* - M)1_{\{Y_d^*>M\}} + (1-p)(M - Y_d^*)1_{\{Y_d^*<M\}}\right]$. Then, writing $Y_d^* = d\theta + \sigma\sqrt{d}Z$ with $Z \sim N(0,1)$, and denoting $(M - d\theta)/\sigma\sqrt{d}$ by m_d yields $R(d, \{\mathbf{Y}^*\}) = \mathbb{E}\left[p(Z - m_d)1_{\{Z>m_d\}} + (1-p)(m_d - Z)1_{\{Z<m_d\}}\right] = \mathbb{E}(Z - m_d)1_{\{Z>m_d\}} + (1-p)m_d$. For $d < d^*$, $m_d > m_d^*$ and $\mathbb{P}(Z > m_d) < (1-p)$, so that $R(d, \{\mathbf{Y}^*\}) - R(d^*, \{\mathbf{Y}^*\}) = \mathbb{E}(Z - m_d^*)1_{\{m_d^*<Z<m_d\}} + (m_d - m_{d^*})\{(1-p) - \mathbb{P}(Z > m_d)\} \geq 0$. For $d > d^*$, $m_d < m_d^*$, and since $\mathbb{P}(Z > m_d) \geq (1-p)$, we have $R(d, \{\mathbf{Y}\}) - R(d^*, \{\mathbf{Y}\}) = \mathbb{E}(Z - m_d)1_{\{m_d<Z<m_{d^*}\}} + (m_{d^*} - m_d)\{\mathbb{P}(Z > m_{d^*}) - (1-p)\} \geq 0$. Thus for $d \neq d^*$, $R(d, \{\mathbf{Y}^*\}) - R(d^*, \{\mathbf{Y}^*\}) \geq 0$. For the Bayesian risk, note that $r(\pi) = \mathbb{E}_\pi R(d, \{\mathbf{Y}^*\} \mid Y_j; j = 1, 2, \ldots, n)$ where the expectation operator $\mathbb{E}_\pi$ is taken with respect to the posterior distribution of θ and σ (marginalized over the hyper-prior distributions for the fully Bayesian case) and arguments similar to the above case hold, inside $\mathbb{E}_\pi$. The result follows.

Appendix C: Full Conditional Distributions of the Hyper-parameters

The full conditionals of the prior-hyperparameters α, β, γ, δ, η and ζ are given by the following set of equations:

$$\begin{aligned}
&\log\pi(\alpha \mid \beta, \gamma, \delta, \eta, \zeta, \theta_i, \sigma_i, Y_{ij}; 1 \le j \le n_i, 1 \le i \le N) \\
&\quad = (\alpha - 1)\sum_{i=1}^{N} \log\theta_i - N\log c(\alpha, \beta, \gamma, \delta, \eta, \zeta) + \frac{1}{2}\log\psi(\alpha), \qquad \alpha > 0
\end{aligned}$$

$$\begin{aligned}
&\log\pi(\beta \mid \alpha, \gamma, \delta, \eta, \zeta, \theta_i, \sigma_i, Y_{ij}; 1 \le j \le n_i, 1 \le i \le N) \\
&\quad = -\frac{1}{\beta}\sum_{i=1}^{N} \theta_i - N\log c(\alpha, \beta, \gamma, \delta, \eta, \zeta) - \log\beta, \qquad \beta > 0
\end{aligned}$$

$$\begin{aligned}
&\log\pi(\gamma \mid \alpha, \beta, \delta, \eta, \zeta, \theta_i, \sigma_i, Y_{ij}; 1 \le j \le n_i, 1 \le i \le N) \\
&\quad = (\gamma - 1)\sum_{i=1}^{N} \log\sigma_i - N\log c(\alpha, \beta, \gamma, \delta, \eta, \zeta) + \frac{1}{2}\log\psi(\gamma), \qquad \gamma > 1
\end{aligned}$$

$$\begin{aligned}
&\log\pi(\delta \mid \alpha, \beta, \gamma, \eta, \zeta, \theta_i, \sigma_i, Y_{ij}; 1 \le j \le n_i, 1 \le i \le N) \\
&\quad = -\frac{1}{\delta}\sum_{i=1}^{N} \sigma_i - N\log c(\alpha, \beta, \gamma, \delta, \eta, \zeta) - \log\delta, \qquad \delta > 0
\end{aligned}$$

$$\begin{aligned}
&\log\pi(\eta \mid \alpha, \beta, \gamma, \delta, \zeta, \theta_i, \sigma_i, Y_{ij}; 1 \le j \le n_i, 1 \le i \le N) \\
&\quad = -\eta\sum_{i=1}^{N} \mid \theta_i - \zeta\sigma_i \mid - N\log c(\alpha, \beta, \gamma, \delta, \eta, \zeta), \qquad \eta > 0
\end{aligned}$$

$$\begin{aligned}
&\log\pi(\zeta \mid \alpha, \beta, \gamma, \delta, \eta, \theta_i, \sigma_i, Y_{ij}; 1 \le j \le n_i, 1 \le i \le N) \\
&\quad = -\eta\sum_{i=1}^{N} \mid \theta_i - \zeta\sigma_i \mid - N\log c(\alpha, \beta, \gamma, \delta, \eta, \zeta), \qquad \zeta > 0
\end{aligned}$$

where

$$c(\alpha, \beta, \gamma, \delta, \eta, \zeta) = \int_0^\infty \int_0^\infty \pi(\theta, \sigma \mid \alpha, \beta, \gamma, \delta, \eta, \zeta) d\theta d\sigma.$$

References

[1] Besag, J. E., Green, P., Higdon, D. and Mengersen, K. (1995), "Bayesian computation and stochastic Systems" (with discussion), *Stat. Sci.*, 10, 3-41.

[2] Besag, J. E. and Green, P. J. (1993), "Spatial statistics and Bayesian computation" (with discussion), *J. Roy. Stat. Soc. Ser. B*, 55, 25-37.

[3] Chib, S. and Greenberg, E. (1995), "Understanding the Metropolis-Hastings algorithm", *American Statistician*, 49, 327-35.

[4] Feller, W. F. (1968), *An Introduction to Probability Theory and Its Applications"*, Vol. 1, Third Edition, Wiley.

[5] Freedman, M. (1997), private communication.

[6] Geman, S. and Geman, D. (1984), "Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images", *IEEE Trans. Pattern. Anal. Machine Intell.*, 6, 721-741.

[7] Green, P. J. and Han, X.-L. (1992), "Metropolis methods, Gaussian proposals and antithetic variables", In *Stochastic Models, Statistical Methods and Algorithms in Image Analysis*, 74:142-64, Springer-Verlag, Berlin.

[8] Hastings, W. K. (1970), "Monte Carlo sampling methods using Markov Chains and their applications", *Biometrika*, 57, 97-109.

[9] Jeffreys, H. (1939), *Theory of Probability*, Oxford University Press, Oxford.

[10] Karlin, S. and Taylor, H. M. (1975), *A first course in stochastic processes*, Second Edition, Acad. Press, New York.

[11] Metropolis, N., Rosenbluth, A., Rosenbluth, M., Teller, A. and Teller, E. (1953), "Equations of state calculations by fast computing machines", *J. Chem. Phys.*, 21, 1087-1091.

[12] Nelder, J. A., and Mead, R. (1965), "A simplex method for function minimization" *Computer Journal*, 7, 308-313.

[13] Piessens, R., deDoncker-Kapenga, E.,Überhuber, C. and Kahaner, D. (1983), *QUADPACK: A Subroutine Package for Automatic Integration*, Springer-Verlag, New York.

[14] Royston, J. P. (1982), "An extension of Shapiro and Wilk's W Test for normality to large samples", *Appl. Stat.*, 31, 2, 115-124.

[15] Sokal, A. D. (1989), "Monte Carlo methods in statistical mechanics: foundations and new algorithms", *Cours de Troisiéme Cycle de la Physique en Suisse Romande*, Lausanne.

[16] Tierney, L. (1994), "Markov chains for exploring posterior distributions," (with discussion), *Ann. Stat.*, 22, 1701-1762.

[17] Zellner, A. (1986), "Bayesian estimation and prediction using asymmetric loss functions", *J. Amer. Stat. Assoc.*, 81, 446-451.