Proceedings of the 2015 Winter Simulation Conference
L. Yilmaz, W. K. V. Chan, I. Moon, T. M. K. Roeder, C. Macal, and M. D. Rossetti, eds.

MODELING CUSTOMER DEMAND IN PRINT SERVICE ENVIRONMENTS USING BOOTSTRAPPING

Sudhendu Rai
Bo Hu

PARC- A Xerox Company
800 Phillips Road
Webster, NY 14450, USA

Ranjit Kumar Ettam

Xerox Business Services
Level 5 Aviator Building ITPB
Bangalore, KA 560066, India

ABSTRACT

For simulation modeling, what-if analysis and optimization studies of many service and production operations, demand models that are reliable statistical representations of current and future operating conditions are required. Current simulation tools allow demand modeling using known closed-form statistical distributions or raw demand data collected from operations. In many instances, demand data cannot be described by known closed-form statistical distributions and the raw data collected from operations is not representative of future demand. This paper describes an approach to demand modeling where historical demand data collected over a finite time period is combined with user-input using two-tier bootstrapping to produce synthetic demand data that preserves the statistical distribution of the original data but has overall metrics such as volume, workflow mix and individual task and job sizes that represent projected future state scenarios. When the customer demand data follows highly non-normal distributions, a modified procedure is presented.

1 INTRODUCTION

Simulation modeling, what-if analysis and simulation optimization of service operations requires accurate characterization of customer demand. These operations often exhibit high-variety workflows, varying demand by workflow type, randomness in arrival of service requests and variations in lead times. The statistical distributions describing these phenomenon often cannot be described using standard closed form distributions. Moreover, future state demand can have different overall volumes and task sizes when compared to the collected data. To determine a design configuration that is robust to these input variations, one needs to perform simulation studies using demand that is an accurate representation of these projected future states. At the same time, design of other ancillary processes such as inventory management, operator training, customer management are all dependent on being able to model the various customer demand scenarios that the service operation may experience in the future.

In recent years, the problem of modeling and forecasting of customer demand in service industry such as telephone call centers has received attention. Ibrahim et al (2012) used statistical models to forecast the incoming call volumes to make staffing decisions and build work schedules in telephone call centers. Weinberg, Brown and Stroud (2007), and Soyer and Tarimcilar (2008) use Bayesian techniques in their forecasts with application to call center data. Steinmann and De Freitas Filho (2013) have used simulation to generate data that can be used to evaluate the forecasting algorithms for inbound call center.

The literature on demand modeling and forecasting in service centers environments has been targeted towards building analytical models using historical data and the demand projection based on future market outlook has not been addressed. In this paper, we describe an integrated system to model future customer demand for service operations. We will use the printing service industry as an example but the underlying approach should be generalizable to other service operations. The system takes into account both available historical data and outlook of future market conditions, and produces synthetic customer demand using a two tier bootstrapping approach that preserves characteristics of historical demand pattern while adjusting for expected changes in market conditions. The characteristics of historical demand pattern are systematically evaluated and adjusted based on market outlook, allowing a more accurate representation of the future demand, to perform simulation what-if studies and optimization to develop robust print shop operations design. The approach described in this paper has been implemented within Xerox's Lean Document Production (LDP) suite Rai et al (2009) to enable evaluation of a wider range of design options using statistically consistent demand inputs that are representative of possible future demand scenarios.

In large print service centers, we observe heavy tailed characteristics of job quantity (Rai 2008). In such cases, the conventional bootstrapping methods fails to generate correct projection of customer demand distributions. Heavy tails are observed in many practical applications such as transaction processing, server farms, file size distributions on the web, internet traffic, CPU process lifetimes and other econometric applications, where the task size associated with performing various tasks are highly non-normal and sometimes heavy-tailed. In such cases, the bootstrap approach is modified using an iterative approach that generates representative distributions while preserving user-specified aggregate properties of the distributions.

This paper is organized as follows. In Section 2, an integrated system to model the customer demand in print service environments is described. Section 3 provides numerical illustrations. Section 4 describes some real world applications. Section 5 describes the conclusions and future work.

2 AN INTEGRATED SYSTEM TO MODEL CUSTOMER DEMAND IN PRINT SERVICE ENVIRONMENT

Within a Lean Document Production (LDP) toolkit, demand data collected from print shop operations is used to simulate and optimize print shop configurations (labor, equipment, operating policies and the like). This data collection process can be expensive and time-consuming in many instances. The demand data also exhibits variability and variety. Within a shop, multiple workflows or job types may simultaneously co-exist where each can have a different demand distribution. These demand distributions are often not amenable to description by known closed-form distributions and can only be described empirically. To generate the future state demand distributions while maintaining control on variations in each job type property (e.g. total volume, job size), a two-tiered bootstrapping approach is proposed (Hu and Rai 2011).

Bootstrapping is a self-sustaining, non-parametric, computationally intensive approach to statistical inference used to produce voluminous data. The idea of bootstrapping is to resample (with replacement) from the sample at hand randomly assuming the sample at hand as surrogate population (Efron 1979; Efron and Tibshirani 1993). The adoption of a two-tier bootstrapping process enables independent control of job type and job properties. The ability of fine-tuning those parameters separately offers greater control over the modeling of customer demand patterns.

Our first step is to generate a stochastic process to model the timeline of job arrival events associated with customer demand (Tijms 2009; Taylor and Karlin 1998). We assume that the job arrival follows a revised Poisson process with the rate λ. With the revised Poisson process, the cumulative number of jobs at time t, $N(t)$, is Poisson distributed with mean $\lambda\xi$, i.e.

$$\{N(t) = k\} = e^{-\lambda\xi} \times \frac{(\lambda\xi)^k}{k!},$$

where ξ is the cumulative operation time units at time t. Different from the standard Poisson process, which is a function of continuous time, the revised Poisson process is a function of a set of discrete time periods spanning across the operating hours of a print shop. Using the standard Poisson process would greatly skew job arrival intervals and generate demand data that significantly deviate from actual print center operations. Second, we use a bootstrapping approach to model the job type for each arrival event. A job type is defined as a unique combination of required print shop functions and is sampled with replacement from a set of pre-determined job types with corresponding probabilities for each arrival event. Third, we use the same bootstrapping-type approach to model the job properties for each arrival event, given the job type determined in the second step. Job properties could be a set of specifications for a job, which include job duration (defined as the operating hours between job due time and its arrival time), job quantity, and quantities for each print shop function. A user-specified job properties database associates each job type with a number of possible job properties. In practice, a small sample of historical demand data is collected and analyzed to extract job types and establish the job properties database. Such information is then updated and augmented according to market condition forecast. The flow of information in the integrated data generating system is shown in Figure 1.

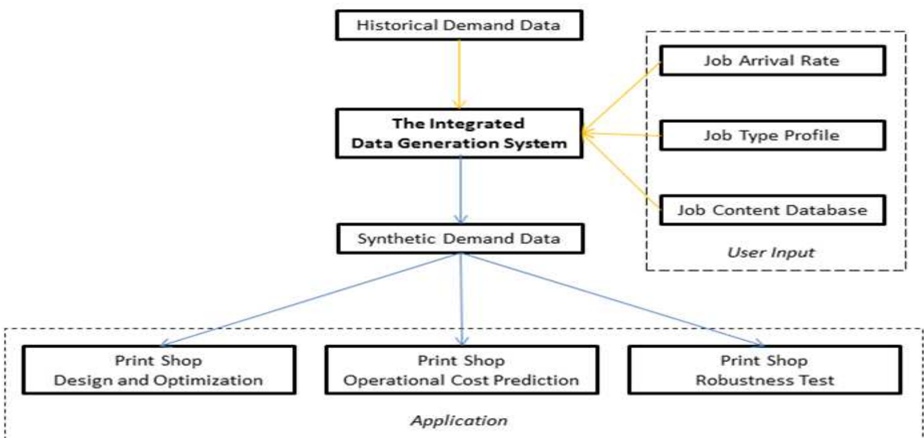

Figure 1: Illustrates the flow of information in the integrated customer data generation system.

The behavior of bootstrap method is consistent when the underlying distributions are normal or belongs to exponential family of distributions. Let F denote the cumulative distribution function of population and F_B be the cumulative distribution function of bootstrap sample. From the law of large number, the bootstrap behavior is consistent if $F_B \to F$ as $n \to \infty$. In large print shop operations, the customer demand distributions i.e., the job quantities are non-normal and the bootstrap method is inconsistent for such applications (Horowitz 2001). The study of alternative resampling techniques such as *m* out of *n* bootstrap method, subsampling, and parametric bootstrap etc., was described in the literature when the existing bootstrap behavior is inconsistent and distribution are non-normal. The m out of n bootstrap method for stable distributions was studied by Athreya (1987) which is based on drawing subsamples with or without replacement of size $m < n$ from the original data. But this method fails to provide reliable inference when the sample size is not very large. Romano and Wolf (1999) discussed the asymptotic inference for the mean when the data follows heavy tailed distributions. Cornea and Davidson (2015) have derived a parametric bootstrap for the purpose of inference on the expectation of a heavy tailed distribution.

In this paper we propose modification to the conventional bootstrap method when the customer demand data follows non-normal distributions. Let $X_1, X_2, X_3 \ldots X_n$ represent the job quantities of a job type in a time period *t* (where time period can be weekly, monthly or quarterly) and $S = X_1 + X_2 + X_3 +$

$\cdots + X_n$ denotes the sum of all job quantities. If the market research indicates that the future total job quantity of the job type in next time period t+1 is expected to increase by α %, then the desired total job quantity $S_{target} = \frac{S(100+\alpha)}{100}$.To generate the future synthetic jobs of the job type, the Efron's bootstrap method can be employed by randomly drawing jobs corresponding to the job type from job properties database (sample) until the total job quantity reaches S_{target} . Due to the heavy tailed nature of job demand process in large print service operations, the probability of random drawing of large observations is not negligible which may cause a significant increase in the future total job demand. For example, let $Y_1, Y_2, Y_3, \ldots Y_m$ be the job quantities drawn from the job properties database using Efron's bootstrap approach with replacement, the sum of first $m-1$ observations is given by $S_{m-1} = Y_1 + Y_2 + Y_3 + \cdots + Y_{m-1}$ and S_m denotes the sum of all the observations i.e. *m*. If Y_m is the m^{th} randomly drawn observation from the tail of the distribution, the % increase in total job quantity $\frac{S_m - S}{S}$ can be significantly larger than the desired % increase (α). To overcome this problem, we modify the conventional bootstrap approach by terminating the bootstrap resampling process when we are approximately closer to the desired % increase in total job quantity (α) i.e., we continue to sample until $\alpha(1-\psi) \leq \frac{S_m - S}{S} \times 100 \leq \alpha(1+\psi)$, where ψ is the acceptance limit specified by the user or a default value of 0.05. Otherwise, we restart the bootstrap process *r* more times, where *r* is user specified value. There is no guarantee that a bootstrap sample can be found in *r* iterations, but the probability of obtaining a bootstrap sample may increase by increasing *r* or ψ. Below, we describe the modified bootstrap algorithm mathematically.

Modified Bootstrapping Algorithm

1. Let $X_1, X_2, X_3 \ldots X_n$ represent the job quantities of a job type in time period *t* and $S = \sum_{i=1}^{n} X_n$
2. Get % increase in future job quantity (α) for next time period *t+1*, acceptance limit (ψ), maximum number of bootstrap samples to be drawn (*r*),
3. Set *i*=1, *bootstrapFlag=true*, and *NoSampleFound=true*
4. Do While $i \leq r$ AND *bootstrapFlag = true*
    - 4.1. Set *m*=1, S_0=0, and *sampleFlag = false*
    - 4.2. Do While *sampleFlag = false*
        - 4.2.1. Randomly select a job with replacement from the initial sample and set the job quantity as Y_m
        - 4.2.2. Compute $S_m = S_{m-1} + Y_m$
        - 4.2.3. If $S_m > S$ then
            - a. If $\alpha(1-\psi) \leq \frac{S_m - S}{S} \times 100 \leq \alpha(1+\psi)$ then
            - b. Set *sampleFlag =true, bootstrapFlag =false, NoSampleFound=false* and return the bootstrap sample
            - c. Else If $\frac{S_m - S}{S} \times 100 > \alpha(1+\psi)$
            - d. *sampleFlag =true ,bootstrapFlag =true*
            - e. Else
            - f. *m*=*m*+1
            - g. End If
        - 4.2.4. Else
        - 4.2.5. *m*=*m*+1
        - 4.2.6. End if
    - 4.3. End Loop
    - 4.4. *i*=*i*+1
5. End Loop

6. If *NoSampleFound* = *True* then
7. Return as bootstrap sample cannot be found in *r* iterations
8. End If

3 NUMERICAL ILLUSTRATION

In this section, we first illustrate the capability of integrated system for generating synthetic customer demand data using historical data and adjustments for market outlook. Next we illustrate the modified bootstrap procedure as a substitute to the conventional bootstrap method in the integrated system when the customer demand data follows non-normal distribution.

3.1 Generating Synthetic Customer Demand Data Using Integrated System

We collected historical data from a sample print service center that has a total of 306 jobs spanning 7 days of print shop operations. There are a total of 29 job types and, for the sake of conciseness, Table 1 shows only the 15 most populated types, which in aggregation represent more than 90% of total job volume in the historical data. Assume market research indicates that, compared to the historical data, the frequency of job arrival will increase by 10% and the percentage of Type 1 jobs will rise from 25% to 35%. To incorporate such information, we set the Poisson arrival rate $\lambda = \lambda_0(1 + 10\%)$ in the integrated system and adjust the job type profile to reflect the increase of Type 1 jobs, where $\lambda_0 = 0.266$ is the historical job arrival rate obtained by averaging the inter-arrival timings of historical jobs. We then use λ in the revised Poisson process to generate the job arrivals and populate the job type profile and the job properties database to reflect the characteristics of the historical data using the two tier bootstrapping approach. Because of the stochastic nature of our modeling process, we generate 5 data sets to examine the level of inherent variations in the synthetic data. As shown in Table 2, the job arrival frequency varies between 9.07% and 11.49% in the 5 synthetic data sets. To reduce the impact of variability in synthetic data on subsequent print shop performance, we average over simulation results using multiple data sets.

Table 1: Illustrates the percentage of jobs for each job type in the historical data set.

Job Type	Percentage(%)
Type 1 jobs	24.51
Type 2 jobs	16.67
Type 3 jobs	9.80
Type 4 jobs	9.48
Type 5 jobs	8.50
Type 6 jobs	8.17
Type 7 jobs	2.61
Type 8 jobs	2.29
Type 9 jobs	1.96
Type 10 jobs	1.96
Type 11 jobs	1.63
Type 12 jobs	1.63
Type 13 jobs	1.63
Type 14 jobs	1.31
Type 15 jobs	1.31

Table 2: Illustrates the job arrival frequency in five synthetic data sets.

	Historical	Target	Set 1	Set 2	Set 3	Set 4	Set 5
Mean operating hours between job arrivals	0.266	0.242	0.240	0.242	0.238	0.244	0.243
Increase in frequency	--	10.00%	10.59%	9.69%	11.49%	9.07%	9.37%

Similarly, Table 3 shows that the percentage of Type 1 jobs varying from 33.01% to 37.91% in the five synthetic data sets. Furthermore, we examine whether the synthetic data preserves the job type pattern in the historical data for all other job types. Ideally, the total percentage of other job types should scale down to 65% while maintaining the relative ratios among themselves. The ideal values are listed in Table 3 under “Target”. We applied Kolmogorov-Smirnov statistical tests to examine whether the job type distribution of each synthetic data set is not significantly different from that of the target (H0, null hypothesis). The resulting p-values are listed in the last row of Table 3. None of the null hypotheses can be rejected at significance level $\alpha=0.05$, suggesting that the synthetic job type mix successfully preserves the characteristics of historical job type mix while adjusting for changes in market conditions.

Table 3: Illustrates the synthetic job type mix with adjustments to market outlook.

	Historical	Target	Set 1	Set 2	Set 3	Set 4	Set 5
% of Type 1 jobs	24.51%	35.11%	36.93%	33.33%	37.91%	33.01%	37.91%
% of Type 2 jobs	16.67%	14.33%	16.34%	15.36%	19.28%	13.40%	13.73%
% of Type 3 jobs	9.80%	8.43%	9.80%	10.46%	7.84%	7.52%	6.54%
% of Type 4 jobs	9.48%	8.15%	8.17%	7.19%	5.56%	10.46%	6.86%
% of Type 5 jobs	8.50%	7.30%	7.52%	7.52%	5.23%	7.84%	6.86%
% of Type 6 jobs	8.17%	7.02%	5.88%	5.56%	6.54%	7.52%	8.50%
% of Type 7 jobs	2.61%	2.25%	2.94%	1.63%	1.63%	0.98%	1.63%
% of Type 8 jobs	2.29%	1.97%	1.63%	2.29%	1.63%	1.63%	2.29%
% of Type 9 jobs	1.96%	1.69%	1.63%	1.63%	1.31%	2.94%	1.63%
% of Type 10 jobs	1.96%	1.69%	0.98%	1.31%	1.63%	2.94%	1.28%
% of Type 11 jobs	1.63%	1.40%	1.33%	2.29%	1.96%	2.61%	1.18%
% of Type 12 jobs	1.63%	1.40%	1.63%	1.31%	1.63%	1.63%	1.26%
% of Type 13 jobs	1.63%	1.40%	1.31%	1.96%	1.31%	1.96%	1.68%
% of Type 14 jobs	1.31%	1.12%	0.33%	0.98%	0.65%	0.65%	0.98%
% of Type 15 jobs	1.31%	1.12%	0.33%	1.31%	0.00%	0.33%	0.98%
P-value	--	--	0.922	0.998	0.902	0.891	0.914

3.2 Modified Bootstrap Approach for Non-Normal Customer Demand Distributions

In large print shop operations, the job sizes are highly non-normal and the modified bootstrap procedure is used as a substitute for generating jobs from the job properties database in the integrated data generating system. We illustrate numerically using the data collected from a sample large print service center for a time period of one month. During this period, the print service center receives 11940 jobs, of which 11923 jobs are print and insert job type and 17 jobs are insert job type. The total job quantity of print and insert job type is equal to 3558944. If market research indicates 20% (α_{target}) increase in future total job volume of print and insert job type, then $S_{\text{target}} = \frac{3558944\times(100+20)}{100}$ = 4270733. The job size distribution is shown in Figure 2 which has the heavy tail index of 0.4 (using theory of stable distributions).

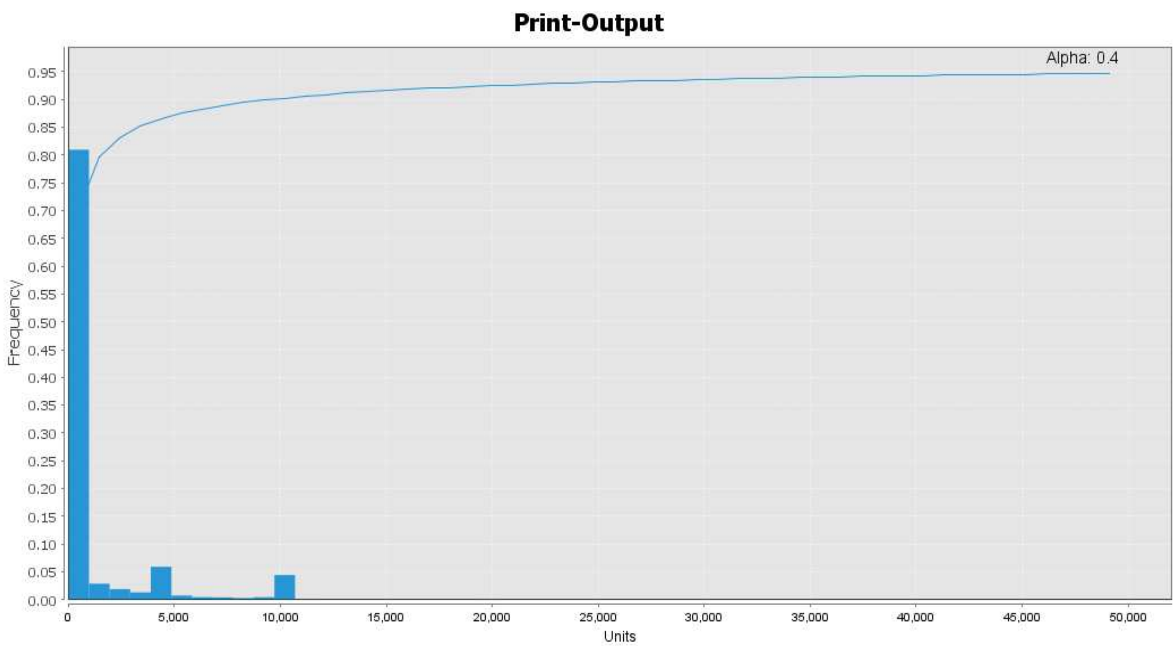

Figure 2: Illustrates the cumulative density function of job size for the sample print service center.

First we use the Efron's bootstrap approach for generating jobs from job properties database and compute the sum of total job quantity (S_m) and mean for each bootstrap sample. Figure 3 illustrates the achieved % increase of total job volume ($\alpha_{achieved}$) and sample mean for each bootstrap sample obtained using Efron's approach, where $\alpha_{Achieved} = \frac{|S_m - S|}{S} \times 100$. We observe that the $\alpha_{achieved}$ values frequently deviates from the target % increase in total job quantity (α_{target}=20%) and the oscillating behavior of sample means. Next, using the modified bootstrap procedure with the parameters α_{target}=20%, ψ=0.01, and r=5 we generate jobs from the job properties database for the print and insert job type. Figure 4 illustrates the achieved % increase in total job quantity($\alpha_{achieved}$) and the number of repetitions (r) of bootstrap sampling process using the modified bootstrap approach for 30 samples.

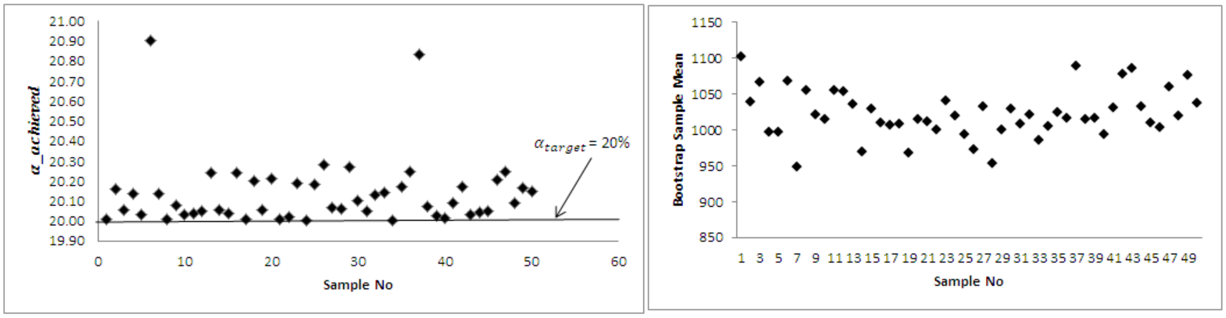

Figure 3: Illustrates $\alpha_{achieved}$ and sample mean values for each bootstrap sample obtained using Efron's bootstrapping method.

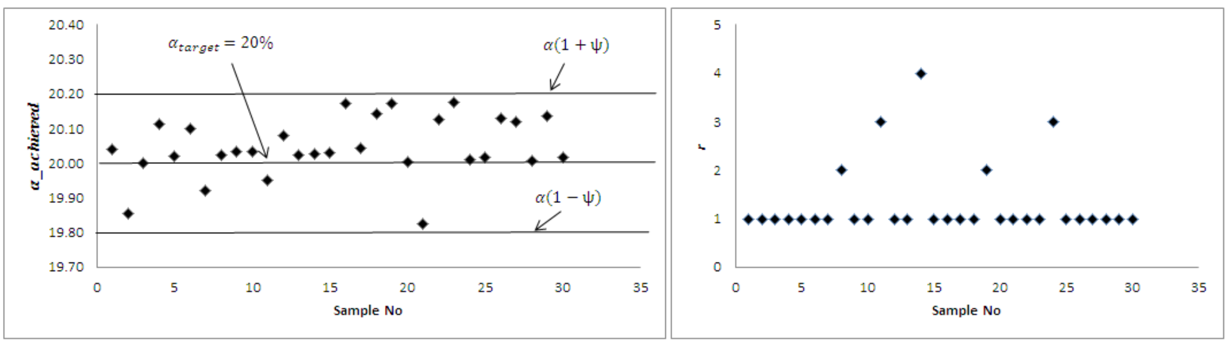

Figure 4: Illustrates the $\alpha_{achieved}$ and the number of repetitions (r) using modified bootstrap approach.

4 APPLICATIONS

We have identified three potential applications of the integrated system. First, the system may be applied in print shop design and optimization to reduce data collection cost and lead time, and to incorporate changes in market outlook. In current practice, if customer demand pattern varies over time, the data collection phase has to cover all patterns throughout the time horizon, which are often measured in months. The large volume of data and lengthy collection process necessitates a labor-intensive and error-prone data cleansing and verification process after data collection is completed. Our system can generate voluminous customer demand data that preserves the characteristics of historical data of limited size, while market condition changes can be readily incorporated during the demand modeling process. Thus, the application of the integrated system can realize significant cost savings by eliminating massive data collection effort and reducing lead time. Furthermore, the system's ability to model future demand based on market research enables a more robust and dynamic print shop design and optimization process, which cannot be accomplished by using historical data alone.

Second, the system may be used in combination with print shop design tools, such as the LDP suite as shown in Figure 5, to predict future operational costs, which are essential for resource planning and contract pricing. Because the efficiency of print shop operations is sensitive to not only demand volume but also demand types and job arrival patterns, it is unlikely that future operational costs can be successfully predicted based on a re-scaled version of historical data alone. As a result, printing service contract that is profitable at the initial operating conditions can become unprofitable when market condition changes. Using the integrated system, future demand can be properly modeled under multiple scenarios of future market conditions, which offers accurate predictions of future operational costs.

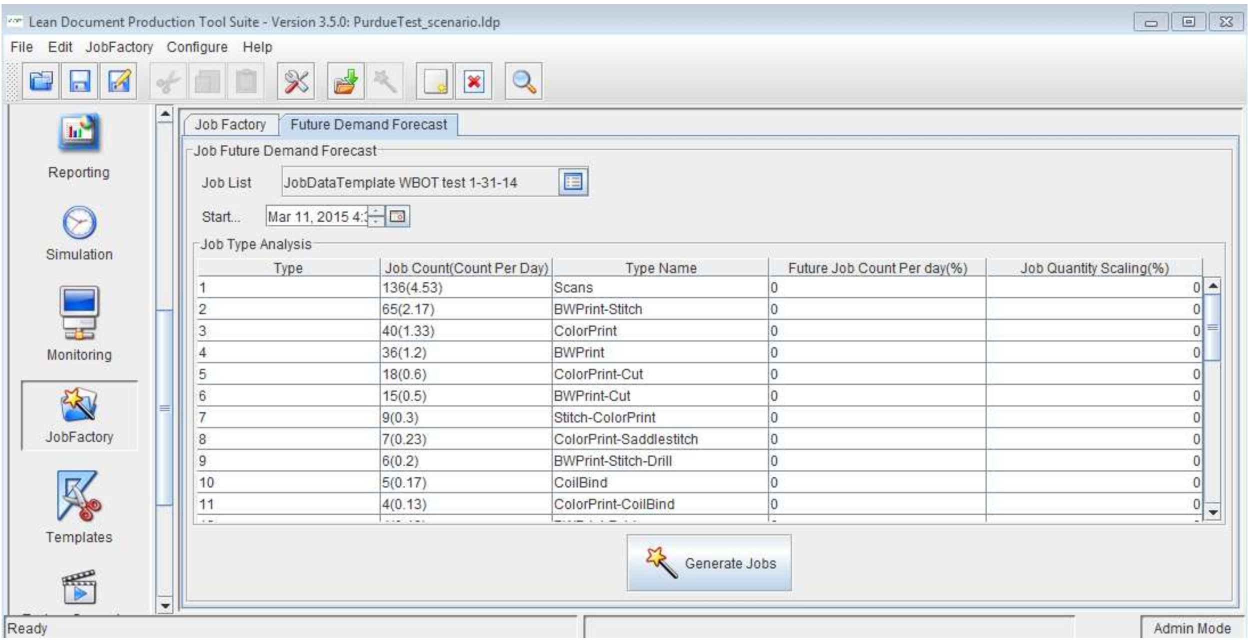

Figure 5: Illustrates the user interface for the integrated system to generate the synthetic customer data in LDP suite.

Third, the integrated system may be used in the robustness test of a print shop design and in new customer account management. The goal of a robustness test is to identify scenarios where the print shop operational efficiency could be severely compromised. For example, the following scenarios may be examined during the test: changes in job arrival frequency, changes in job type mix, changes in job properties etc. In daily operations, the robustness test may be used to determine whether or not to take a new customer account. The account manager may use the system to generate a set of demand data taking into account jobs from the prospective customer and simulate print shop operations in the LDP suite. The resulting performance measurements indicate whether the print shop can bear the additional demand without significant decrease in overall efficiency.

5 CONCLUSIONS AND FUTURE WORK

In this paper, we described an integrated approach to model future customer demand for print service operations. The system takes into account both available historical data and outlook of future market conditions, and produces synthetic customer demand using a two tier bootstrapping approach. The techniques involved in this integrated system are rather complex and delicate, e.g. the installation of the revised Poisson arrival process and the use of two-tier bootstrapping. For the system to produce accurate and useful demand data sets, it requires careful design of computational steps supported by logical reasoning of available information. This integrated system can be used in a broad range of contexts, including commercial print shops, managed printing services, etc. Once the system is established and theoretical techniques automated, the user inputs needed for generating appropriate data sets are easy to understand (e.g. job arrival rate, job type mix, and job properties, etc.). Therefore, the automated system can be adopted by print shop operators, performance analysts, and sale personnel with minimum training cost.

When the demand distribution is non-normal, the Efron's bootstrap method is inconsistent and unreliable, the modified bootstrap procedure is used as an alternative for generating customer demand in the integrated system. The efficacy of the modified bootstrap procedure is empirically proved using data from large scale print service operations. Future work includes comparing the proposed modified bootstrap approach with other resampling techniques such as Subsampling, Parametric bootstrap and Bayesian bootstrap etc. Finally, although the system is developed in the assumed environment of printing service industry, it can also be used to generate demand scenarios of other service processes.

REFERENCES

Athreya, K.B. 1987. "Bootstrap of the Mean in the Infinite Variance Case." *The Annals of Statistics* 15:724-731.

Cornea, A., and R. Davidson. 2015. "A Parametric Bootstrap for Heavy Tailed Distributions." *Econometric Theory* 31: 449-470.

Efron, B. 1979. "Bootstrap Methods: Another Look At The Jackknife." *The Annals of Statistics* 7:1-26.

Efron, B., and R.J. Tibshirani. 1993. *An Introduction to the Bootstrap*.1st ed. New York: Chapman & Hall.

Horowitz, J. L . 2001. *Handbook of Econometrics*. Elsevier Science.

Hu, B., S. Rai. 2011. "System and Method of Modeling Customer Demand in Print Production Environment." US Patent No. 20120188591, issued July 2012.

Ibrahim, R., P. L'Ecuyer, N. Regnard, and H. Shen. 2012. "On the Modeling and Forecasting of Call Center Arrivals." In *Proceedings of the 2012 Winter Simulation Conference*, edited by C. Laroque, J. Himmelspach, R. Pasupathy, O. Rose, and A. M. Uhrmacher, 256-267. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Rai, S. 2008. "Fat Tail Inputs in Manufacturing Systems." In *Proceedings of the 2008 Industrial Engineering Research and Conference*, edited by J. Flower and S. Mason, Vancouver, Institute of Industrial Engineers, Norcross, GA.

Rai, S., C. B. Duke, V. Lowe, C. Q. Trotter, and T. Scheermesser. 2009. "LDP Lean Document Production – O. R.- Enhanced Productivity Improvements for the Printing Industry." *Interfaces* 39: 69-90.

Romano, J.P., and M. Wolf. 1999. "Subsampling Inference for the Mean in the Heavy Tailed Case." *Metrika* 50:55-69.

Soyer, R., and M. M. Tarimcilar. 2008. "Modeling and Analysis of Call Center Arrival Data: A Bayesian Approach." *Management Science* 54 (2): 266–278.

Steinmann, G., and P. J. De Freitas Filho. 2013. "Using Simulation to Evaluate Call Forecasting Algorithms for Inbound Call Center." In *Proceedings of the 2013 Winter Simulation Conference*, edited by R. Pasupathy, S.-H. Kim, A. Tolk, R. Hill, and M. E. Kuhl, 1132-1139. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Taylor, H.M., and S. Karlin.1998. *An Introduction to Stochastic Modeling* .3rd ed. Academic Press.

Tijms, H.C. 2003. *A First Course in Stochastic Models*. John Wiley.

Weinberg, J., L. D. Brown, and J. R. Stroud. 2007. "Bayesian Forecasting of an Inhomogeneous Poisson Process with Applications to Call Center Data." *Journal of the American Statistical Association* 102 (480): 1185–1198.

AUTHOR BIOGRAPHIES

SUDHENDU RAI is a Fellow at the Palo Alto Research Center – A Xerox Company. He received his PhD. from MIT (1993), MS from Caltech (1989), and BTech from IIT, Kanpur (India) in 1988 – all in Mechanical Engineering. He is the lead inventor of the LDP Lean Document Production® solution that was a finalist at the 2008 Franz Edelman competition and got first place at IIE Lean Best Practice Award (2011). He holds 68 patents with 30 additional pending. He is a member of ASME, INFORMS, IIE and a senior member of IEEE. He was a finalist for the Rochester Engineer of the Year award in 2007.His e-mail address is Sudhendu.Rai@parc.com.

RANJIT KUMAR ETTAM is a lead engineer at the Xerox Business Services, Bangalore, India. He received his MTech from IIT Kharagpur (2007) in Industrial Engineering and Management, and BTech (2005) from Jawaharlal Nehru Technological University (JNTU), Anantapur, India in Chemical Engineering. His e-mail address is Ranjit.Kumar2@xerox.com.

BO HU is a research scientist at the Palo Alto Research Center – A Xerox Company. She received her PhD. in Operations Management in 2010 and M.S. in Business Administration in 2006 – both from University of Rochester, M.A. in Management Science from Tsukuba University, Japan in 2002, and B.E. in Applied Economics from Xi'an Jiaotong University, China in 1997. Her research interests are in data-driven business process modeling and optimization, data analytics, and behavioral economics. Her e-mail address is Bo.Hu@parc.com.