# Statistical Quality Control for the Service Sector

Jian Li[1] and Fugee Tsung[2,3]
[1]School of Management, Xi'an Jiaotong University, Shaanxi 710049, China
[2]Department of Industrial Engineering and Logistics Management,
Hong Kong University of Science and Technology,
Clear Water Bay, Kowloon, Hong Kong
[3]Corresponding author: Fugee Tsung, email: season@ust.hk

**Abstract**

The service industry has become increasingly important in our daily lives and global economies, and improving service quality will have a significant social and economical impact. The massive amount of data readily available provides us with opportunities to integrate advanced statistical methodologies with system knowledge to better model and control the quality of service systems. Monitoring and quick detection of abnormal activities have become critical in a variety of service industries that have complex operation systems and process huge amounts of data. However, most conventional statistical process control techniques cannot effectively handle the monitoring and surveillance of such a large amount of data on a real-time basis. This paper indicates the importance of and sheds some lights on developing statistical monitoring and control methodologies for the service sector. It focuses on three fundamental issues: (1) the monitoring of feedback data arising from customers' perception of service quality; (2) the monitoring and controlling metrics of operational performance in service systems; and (3) the monitoring of the functional relationship between operational determinants and the service quality perceived by end customers. In addition, we will elaborate on the first issue by reviewing some recently developed results.

Keywords: multivariate categorical processes, log-linear models, contingency table, nonparametric, profile monitoring

## 1 Introduction

In recent decades, the service sectors, such as the healthcare service, financial service, and after-sales service sectors, have drawn significant attention and become indispensable components of the global economy. In many developed countries, the service industry is a critical component, often constituting more than 70% of their respective nominal GDPs. Because of its importance, maintaining the healthy growth of the service sector deserves significant attention.

Similar to manufacturing enterprises, where good product quality can reduce production costs, improve reputation and eventually secure more profits, excellent service quality is noted as a key factor driving profit in the service sector. However, inherent challenges arise when we apply traditional statistical control methods in a service environment. Although there have been some research work on how to define and measure service quality (mostly from customers' perspectives), there still lacks a systematic way to monitor, control, and improve service quality, especially on the process/operational aspects. Targeting this gap, in this paper we intend to expand the methods and applications of statistical process control to a broader horizon by introducing novel strategies to monitor and control both the output: the customer service, and the input: service operations, of a service system.

We will discuss innovative methodologies for monitoring complex service systems. Both theoretical grounds and tools for practical applications will be investigated. In particular, our research will focus on three promising directions. (1) Multivariate categorical customer data monitoring: Categorical data such as customer survey and service ratings, contain important information from customers' perspective, and should be monitored to assure high service quality. The challenges are that categorical data streams are highly cross-correlated. (2) Multivariate nonparametric service operation monitoring: Targeting the unique characteristics of service operations, where non-normal data are prevalent, multivariate nonparametric monitoring methods for high-dimensional data should be developed to detect changes in service operations in a real-time manner. (3) Profile monitoring of service system relationships: Profiles mean the functional relationships between the system inputs and the system outputs, which are more complex than the input or output data themselves. We will discuss methods to detect and diagnose deviations of such relationships in the service sector.

The remaining of this paper is organized as follows. We first discuss the above-mentioned three fundamental research problems. Then for the first issue on monitoring customer-perceived quality, we review some of our newly proposed approaches in the literature. Finally, some concluding remarks ends this paper.

## 2 Research Problems

The successful marriage between statistical methods and engineering knowledge in quality monitoring and control can be found in manufacturing or production systems, where a variety of control chart tools (Zou and Tsung (2008)) and variation reduction techniques (Shi (2010)) and references therein) based on in-situ process data have been developed and have significantly improved the product and process quality. Unfortunately, as described in Koole and Mandelbaum (2002), the amount of research on statistical inference and quality monitoring in service systems like call centers is insufficient, while "so much is yet needed". Here we focus on three fundamental problems of quality improvement in the service sector.

### 2.1 Improving Customer-Perceived Quality

In the literature, there are research works on how to define and measure service quality in general. A number of frameworks, such as SERVQUAL (Parasuraman et al. (1988)) and GAP models (Parasuraman et al. (1985)), have been proposed and studied. These models mostly focus on the customers' perspective, defining service quality in the aspects of reliability, assurance, tangibles, empathy, and responsiveness. For example, in a typical call center, customers can rate their experience, say on a scale of 1 to 5, in areas including first call resolution (FCR), attitude of the customer service representative, the speed of the response, etc. These survey data from customers provide us with opportunities to benchmark and monitor the changes in the service quality from customers' perspective.

Due to the limitations of human perception these data are often expressed as a vector of categorical responses (e.g., the Likert-type scale), which makes the large body of existing multivariate statistical monitoring techniques for continuous data (see Bersimis et al. (2007) and references therein) not applicable here. By contrast, only limited literature can be found on the monitoring of multivariate categorical data. Patel (1973) proposed a Hotelling's $T^2$ type $\chi^2$-chart for multivariate binomial populations. In Marcucci (1985), a generalized $p$-chart was developed for monitoring multinomial processes, which extends the $p$-chart to three or more attribute levels. However, it applies to only one factor. There is no appropriate approach for the case of multiple factors, in which at least one has more than two attribute levels. Moreover, most of the existing methods suffer from a drawback in that they are designed for monitoring the marginal sums of cross-classification probabilities for each factor, and they may not be sensitive to other types of shifts, such as the changes in two-factor or higher-order interaction effects. There seems to be a severe lack of general methodologies to deal with the monitoring of multivariate categorical processes.

## 2.2 Monitoring Operational Performance in Service Systems

Different from quality data from customers' perspective, which can be considered as the output performance of the service systems, operational or process data may enable us to monitor the systems and detect changes at an early stage before lower service quality can be perceived by end customers. As a result, in addition to monitoring categorical customer data, we should develop methods for early warning and detection from the service operation perspective. For example, in call centers there are many process performance measures, such as Average Speed of Anaswer (ASA), queue time, abandonment rate, and service levels. These are in fact operational determinants of caller satisfaction in call centers (Feinberg et al. (2000)). By evaluating these quality metrics and comparing them with benchmarks, we are able to detect systematic changes early and more efficiently than monitoring customer feedback at the end of the process. It is worth noting that many of these quality metrics may not be directly observable, and need to be estimated from individual transactional data. Moreover, they usually do not follow normal distributions, in contrast with quality data in a manufacturing environment (McNeill et al. (2005) and Shore (2006)). Therefore, existing multivariate control charts (see Bersimis et al. (2007)) cannot be properly applied here. Such unique data characteristics motivate us to develop novel multivariate nonparametric control charts using high dimensional and skewed distributed transactional level data.

## 2.3 Profile Monitoring of Service Operations

With the aforementioned techniques for monitoring customer feedback data and service operation data in place, we may further explore the functional relationship between these two types of quality metrics. By relating quality perceived by customers to those metrics collected in service operations, we are able to target the cause of changes in the process resulting in customer dissatisfaction, and make correction and improvement accordingly. Take call centers for example. Intuitively, as verified in the literature, queue time, abandonment rate, etc. do have negative relationships with caller satisfaction, while FCR and ASA have positive relationships (Feinberg et al. (2000)). Adding to the complexity, different customers also have different quality standards. Paid customers or customers in premium classes often have lower tolerance on waiting time. Customers calling for a refund, or replacement, etc. often demand higher resolution rate on the first call than those calling with general inquiries. Other factors like age, gender, occupation, may also influence customer feedback. Therefore, a thorough understanding of the relationship between customer-perceived service quality and customer profiles and service operational performance would lead to a more accurate diagnosis of system changes and more efficient improvement of service quality.

In the literature, it is commonly assumed that linear profiles with normally distributed random errors can be used to describe the functional relationships. However, it is well recognized that the underlying process distribution in many service applications is not normal, and the functional relationship is not necessarily linear (e.g., the congestion curves in queueing). To this end, a non-parametric methodology for monitoring the nonlinear profiles, including the regression coefficients and profile variations, is highly desired. Therefore, we should also devote ourselves to developing novel methods for robust profile monitoring.

# 3 Multivariate Categorical Monitoring of Service Quality

The unique features of service systems and the limitations of traditional approaches in service contexts motivate us to develop new methodologies, and promote a different mindset in quality monitoring and control in service enterprises. Here in this paper, we elaborate on the first issue of monitoring customer-perceived quality by reviewing some recently proposed approaches.

Customer-perceived service quality feedback data are usually multivariate categorical, and they are usually summarized in a multi-way contingency table. There is a clear need to model the relationship between each cell count and factor levels associated with it. This resembles standard multi-way ANOVA, where responses to all factor level combinations are also placed in a multi-way table. Note that in our case the response (cell count) is not normally distributed as is in ANOVA. We may employ a log-linear model by taking the logarithms of cell count expectations in a multi-way contingency table. A detailed discussion of it is given by Bishop et al. (2007). For a three-way contingency table of size $h_1 \times h_2 \times h_3$ that involves three factors, given the sample size $N$, the log-linear model characterizing the relationship between the cell count expectation $m_{ijk}$ ($i = 1, \ldots, h_1$; $j = 1, \ldots, h_2$; $k = 1, \ldots, h_3$) or further the cell probability $p_{ijk} = m_{ijk}/N$ in cell$(i, j, k)$, and the factor levels $i, j, k$, is

$$\ln p_{ijk} = u^{(0)} + u_i^{(1)} + u_j^{(2)} + u_k^{(3)} + u_{i,j}^{(1,2)} + u_{i,k}^{(1,3)} + u_{j,k}^{(2,3)} + u_{i,j,k}^{(1,2,3)},$$

where the $u$-terms are the main or factor interaction effects.

Combined with some identifiability constraints, the log-linear model for a general $p$-way contingency table with $p$ categorical factors can be expressed as

$$\ln \mathbf{p} = \mathbf{1}\beta_0 + \sum_{i=1}^{2^p-1} \mathbf{X}_i \boldsymbol{\beta}_i = \mathbf{1}\beta_0 + \mathbf{X}\boldsymbol{\beta},$$

where $\mathbf{p}$ of size $h \times 1$ ($h = h_1 \times \ldots \times h_p$) is the cell probability vector, $\mathbf{X}_i$ is the design submatrix corresponding to the $i$th main or interaction effect, and $\boldsymbol{\beta}_i$ is the coefficient subvector. There is a one-to-one correspondence between the $i$th effect, the design submatrix $\mathbf{X}_i$, and the coefficient subvector $\boldsymbol{\beta}_i$ ($i = 1, \ldots, 2^p - 1$). We see that totally there are $2^p - 1$ effects, including $p$ main effects and all interaction effects. In fact, a main effect determines the marginal distribution of its corresponding factor, whereas an interaction effect determines the dependence among the several factors included in it. So the cell probability vector is essentially determined by the coefficient subvectors $\boldsymbol{\beta}_i$ ($i = 1, \ldots, 2^p - 1$).

Denote the log-linear model by $F(\mathbf{X}; \boldsymbol{\beta})$. It is assumed that the $j$th multivariate sampling observation vector $\boldsymbol{n}_j$ is collected over time following the change-point model:

$$\boldsymbol{n}_j \overset{\text{i.i.d.}}{\sim} \begin{cases} F(\mathbf{X}; \boldsymbol{\beta}^{(0)}), & \text{for} \quad j = 1, \ldots, \tau, \\ F(\mathbf{X}; \boldsymbol{\beta}^{(1)}), & \text{for} \quad j = \tau + 1, \ldots, \end{cases}$$

where $\tau$ is the unknown change-point (i.e., when the change of customer perception occurs), and $\boldsymbol{\beta}^{(0)} \neq \boldsymbol{\beta}^{(1)}$ are the known in-control (IC) and unknown out-of-control (OC) coefficient vectors (that categorize and characterize the customer service), respectively.

According to the one-to-one correspondence between factor effects and coefficient subvectors, shifts in the marginal distribution of a factor or the dependence among multiple factors, which appear in the form of deviations of its main effect or their interaction effect, respectively, are reflected by the changes of corresponding coefficient subvectors. As a result, the quality monitoring task is to test if $\boldsymbol{\beta} = \boldsymbol{\beta}^{(0)}$. By the likelihood ratio test, Li et al. (2013) proposed a log-linear multivariate binomial/multinomial (LMBM) control chart based on this hypothesis. The LMBM chart can be actualized as a general SPC tool for the monitoring of multivariate/univariate binomial/multinomial data. By numerical simulations, Li et al. (2013) demonstrated that at certain expenses of sensitivity to changes in main effects, this LMBM chart is much more robust than traditional ones, which take only marginal cell probability sums into account. Instead, the LMC chart provides much higher detection ability to possible shifts in interaction effects of multiple factors which represent their dependence.

The LMBM chart actually considers the most general case that in the OC state the OC process coefficient vector does not equal the IC vector $\boldsymbol{\beta}^{(0)}$. In practice, it is reasonable to assume that any changes involve only a few coefficient subvectors or only a few coefficients in the appropriate model. Suppose that we have some *a priori* knowledge that in the OC state only one coefficient $\beta_i$ ($1 \leq i \leq h - 1$) is incremented by an unknown constant $\delta_i$, but its index or location $i$ is unknown. The alternative hypothesis for describing the OC state reduces to

$$H_1: \ \boldsymbol{\beta} = \boldsymbol{\beta}^{(0)} + \mathbf{d}_1\delta_1 \text{ or } \ \boldsymbol{\beta} = \boldsymbol{\beta}^{(0)} + \mathbf{d}_2\delta_2 \ldots \text{ or } \ \boldsymbol{\beta} = \boldsymbol{\beta}^{(0)} + \mathbf{d}_{h-1}\delta_{h-1},$$

where $\delta_i$ ($i = 1, 2, \ldots, h - 1$) are the unknown shift magnitudes, and the possible shift directions $\mathbf{d}_1$, $\mathbf{d}_2$, $\ldots$, $\mathbf{d}_{h-1}$ are defined in a similar way, applying to $\beta_1$, $\beta_2$, $\ldots$, $\beta_{h-1}$, respectively. Based on this hypothesis, Li et al. (2012) proposed a log-linear directional (LLD) control chart by the generalized likelihood ratio test, which specializes in detecting one-coefficient shifts and considers more information. This should be powerful if the real shift indeed occurs in only one coefficient in the log-linear model.

The LMBM chart and the LLD chart are for monitoring multivariate categorical data, which appear often in customer-perceived quality. However, they have different objectives, in that the former is a general detection tool whereas the latter is devised for detecting directional shifts.

## 4 Conclusion

This paper addresses SPC for the service industry. While we elaborate on the first fundamental problem of monitoring customer-perceived quality in the form of feedback data that are usually multivariate categorical, the other two issues of service operational performance surveillance and service profile monitoring should be equally emphasized and deployed. Our ongoing research are now focusing on these two fundamental problems. In a word, service quality improvement has great social and economical impact, which can be applied into activity monitoring in call center companies, fraud detection in telecommunications marketing, customer relationship management in logistics, health and financial surveillance, and so forth.

## References

Bersimis, S., Psarakis, S. and Panaretos, J. (2007) "Multivariate statistical process control charts: an overview," *Quality and Reliability Engineering International*, 23, 517–543.

Bishop, Y. M. M., Fienberg, S. E. and Holland, P. W. (2007) *Discrete Multivariate Analysis*, Springer, New York.

Feinberg, R. A., Kim, I.-S., Hokama, L., Ruyter, K. d. and Keen, C. (2000) "Operational determinants of caller satisfaction in the call center," *International Journal of Service Industry Management*, 11, 131–141.

Koole, G. and Mandelbaum, A. (2002) "Queueing models of call centers: an introduction," *Annals of Operations Research*, 113, 41–59.

Li, J., Tsung, F., and Zou, C. (2012) "Directional control schemes for multivariate categorical processes," *Journal of Quality Technology*, 44, 136-154.

Li, J., Tsung, F., and Zou, C. (2013) "Multivariate binomial/multinomial control chart," *IIE Transactions*, accepted.

Marcucci, M. (1985) "Monitoring multinomial processes," *Journal of Quality Technology*, 17, 86–91.

McNeill, J. E., Fowler, J. W., Mackulak, G. T. and Nelson, B. L. (2005) "Cycle-time quantile estimation in manufacturing systems employing dispatching rules," *WSC'05: Proceedings of the 37th Conference on Winter Simulation*, Orlando, Florida, 749–755.

Parasuraman, A., Zeithaml, V. A. and Berry, L. L. (1985) "A conceptual model of service quality and its implications for future research," *The Journal of Marketing*, 49, 41–50.

Parasuraman, A., Zeithaml, V. A. and Berry, L. L. (1988) "SERVQUAL: a multiple item scale for measuring consumer perceptions of service quality," *Journal of Retailing*, 64, 12–40.

Patel, H. I. (1973) "Quality control methods for multivariate binomial and Poisson distributions," *Technometrics*, 15, 103–112.

Shore, H. (2006) "Control charts for the queue length in a G/G/S system," *IIE Transactions*, 38, 1117–1130.

Shi, J. (2010) *Stream of Variation Modeling and Analysis for Multistage Manufacturing*, Taylor & Francis, .

Zou, C. and Tsung, F. (2008) "Directional MEWMA schemes for multistage process monitoring and diagnosis," *Journal of Quality Technology*, 40, 407–427.