

Extreme Skewness in Service Time Distributions: Its Threat to Evaluations of Material Equivalence of Service System Performance Using Large Sample Control of Type II Error Probabilities

Donald L. Pardew
Cybernetica Consulting, Inc.
425 Riverside Drive Suite 9D
New York, NY 10025
d.pardew@cybernetica-inc.com

Abstract: In evaluating the performance of service systems, it is necessary in some practical situations to conclude whether or not two service systems are equivalent, or at least not materially different. Statistical tests for evaluating equivalence however do not exist. One approach to evaluating material equivalence is to specify a materiality constant, δ , which is taken to represent material equivalence, then performing a standard test of hypothesis of the form $H_0 : \mu_1 - \mu_2 \leq \delta$ while controlling the type II error probability to a known, small value. Service times however, are in some cases severely skewed and without a specifiable distribution, making type II error probability uncertain. It is shown that the effect of skewness of any degree of severity, even in un-specifiable distributions, is a function of and controllable by, modest sample sizes.

Keywords: Skewness, Type II error, Service times, Statistical methods for evaluation.

1. Introduction and Statement of Problem

As is well understood, there do not exist any statistical procedures capable of testing for the equivalence of population parameters, only procedures for testing differences. Still, in some applications the interest is in knowing if two parameters are materially the same. One such application is in the comparison of the performance of service systems that are, by regulatory mandate, required to be equivalent for two customer populations on measures of service quality. A specific, “large data set” case of this problem arises in the case of deregulation of the telecommunications industry in the US. There, relatively few large telecommunication companies own the network infrastructure, access to which they are required, under deregulation rules, to provide to competing companies. But since the large companies that own the network must also service the customers of competing companies, regulation understandably requires that the owners of the network service the customers of competing companies at quality levels that are at least as good as they provide for their own customers. This creates the need for tests of equivalence, of which none exist. The best, though imperfect, solution seems to be the standard hypothesis testing procedure in which a materiality constant, δ , which is not to be exceeded is specified with the type II error probability well controlled by the choice

of sample size. The reasoning in these applications is that differences less than δ can, for all practical purposes, be considered equivalent. This back door approach to evaluating equivalence then boils down to controlling, in a standard hypothesis testing paradigm, the probability of a type II error to a level commensurate with the risk it presents in practical as opposed to statistical terms, and in these cases, the interest in the type II error probability generally exceeds that of the interest in the type I error probability. A particular instance of this problem arises in the context of regulation of the telecommunications industry regarding service times for customers of the network owners compared to those of its network-leasing competitors – e.g. installation times, repair times, etc. The problem however exists, throughout the service industry, where service times are a standard service quality measure. Service times are used both to compare a service system against a constant performance standard, δ , as in $H_0 : \mu \leq \delta$, as well against a reference system's performance, as in $H_0 : \mu_1 - \mu_2 \leq \delta$. The problem of service time equivalency testing has not been widely addressed, the general problem of skewed attribute metrics in quality control charting has been (Shore 1991, 2000). Given the importance of type II error probability to the information objective in these applications, it is critical to either be confident that the sampling distribution of the test statistic will conform to the central limit theorem and that the sample data are sufficient to warrant such confidence, or to use an alternative procedure. The choice of alternative procedures is generally limited to normalizing the data with an appropriate transformation or resorting to non parametric or exact probability procedures, e.g., permutation test, but with large sample data sets these non parametric procedures can become computationally clumsy. Service times however, are known to be skewed, and sometimes severely skewed. Consequently, in applications of these procedures, it may be the case that even large sample data sets or data transformations will not normalize the sampling distribution sufficiently to achieve control over the Type II error probability with the desired precision. A question therefore arises with service time distributions, as to how large a sample size is required to overcome the threat that computed and actual type II error probabilities may be materially different. Classical service times distributions are most typically exponential, and if we could be assured they were exponentially distributed, then there would be no need for concern about the possibility of non normality of the distribution of service time means. This is because it is known that means calculated on samples drawn from exponential service times are distributed as an Erlang- k distribution, where k is the sample size. The Erlang distribution converges rather quickly to a normal distribution, and would certainly do so for values of k in the neighborhood of 150, a typical sample size in applications of the type under consideration here. An inspection of the Erlang ($k=150$) distribution shows it to be, for all practical purposes, normal. Any minor departures from normality would be miniscule compared to the departure of actual distributions from normality rooted in measurement error and other unavoidable violations of data assumption, e.g., dependent clusters of data values. But, practical experience tells us that service times of the type in question here may be even more skewed than the classical case. Therefore, the “non normality” issue in the present case and its effect on type II error probabilities can be boiled down essentially to the effect of the third moment in the underlying service time distribution on the skew in its associated sampling distribution of means or mean differences.

2. Analytic Results

It is not possible to know with certainty if any given sample size will be adequate to counteract pronounced skew in the service time distribution unless the exact distribution of the service times is known, which it almost always is not. However, we can make some general observations about the effect of sample size on overcoming the non normalizing effect of skew in the service time distribution.

Consider first, a simple case of a sample of 2 observations, x_1 and x_2 where the sample mean is given by $\bar{x} = (x_1 + x_2) / 2$

We assume that the sample is random and therefore x_1 and x_2 are independently and identically distributed with mean, μ , standard deviation σ , third central moment = S

And where skewness, Sk , is defined as: $Sk = (S^2 / \sigma^6)$

We can express $\bar{x} - \mu$ as $(x_1 - \mu) / 2 + (x_2 - \mu) / 2$

From this, the third moment around the origin can be computed as follows:

$$E(\bar{x} - \mu)^3 = E \left[\frac{(x_1 - \mu)^3 + (x_2 - \mu)^3}{8} + \frac{3(x_1 - \mu)^2(x_2 - \mu)}{4} + \frac{3(x_2 - \mu)^2(x_1 - \mu)}{4} \right]$$

The second and third terms are 0 because of independence, and the first term gives

$\frac{S}{4}$ where S , again, is the third central moment of the variable. We also know that the variance of the sample mean = $\sigma^2 / 2$. Therefore, the skewness of the sample mean is given by

$$\frac{Sk}{2} = \frac{S^2 / 16}{\sigma^6 / 8}$$

In other words, in the above case, the skewness of the means is a half of the skewness of the original variable. If the sample size is increased to n , the skewness of the sample mean will be reduced by a factor of n . This is because

$$E(\bar{x} - \mu)^3 = E \left[\frac{\sum_i (x_i - \mu)^3}{n^3} + \frac{3 \sum_{i,j \neq i} (x_i - \mu)^2 (x_j - \mu)}{n^2} \right] = \frac{S}{n^2}$$

Moreover, the variance of the sample mean is given by σ^2 / n . Therefore, the skewness of the sample mean is given by:

$$\frac{Sk}{n} = \frac{[S / n^2]^2}{[\sigma^6 / n^3]^2}$$

This provides a general argument that the skewness reduces with sample size n by a factor of n . Applying the result to a hypothetical case, if it is considered that a sample of 30 is known to be sufficient to adequately normalize (i.e., drive the third moment to nearly zero) the means of exponentially distributed service times, then this implies that a sample of 150, can normalize a distribution of service times with about 5.0 times the skewness, i.e. $150/30 = 5.0$. Similar ratios of skewness to sample size can be computed using the same logic. As such computations will show, it requires very extreme skewness in the parent distribution of service times to materially effect the skewness in the sampling distribution of mean service times, or said another way, it does not require extremely large sample sizes to prevent extreme skewness in the service time distribution from materially effecting the skewness of the sampling distribution of mean service times.

3. Summary and Conclusion

Service times are known to be positively skewed, so therefore the only material departure from normality likely to be exhibited by non normalized distributions of service time means is also a positive skew. This will happen only when the underlying distribution of service times are severely skewed, sample sizes are small and no effort is made to normalize the service times comprising the sample through an appropriate transformation. It has been demonstrated that sample size is effective in normalizing the sampling distribution of service time means drawn from a population of severely skewed service times, and that a sample of size 150 will be effective in normalizing all but the most extreme cases of positive skew, and that this effectiveness will exist even without exercising the other measures available to normalize the distributions, e.g., data transformations.

Service times are a common performance or service quality measure of service systems, and, as described earlier there are many situations in which it is desired to evaluate if two service systems are performing at materially equivalent levels. From a customer's point of view, it is usually the case that only differences in service times that are perceptually detectable are important. These perceptually detectable differences, or just noticeable differences (jnd's) can be determined empirically and then specified as the materiality constant. Then, the standard hypothesis testing procedure, in which care is taken to control the type II error probability can be used to achieve as adequate, if imperfect, test of material equivalence.

References

- Shore, H., (2000) General control charts for attributes, *IIE Transactions*, 32, 1149-1160.
- Shore, H., (1991) New control limits that preserve the skewness of the monitoring statistic, *Transactions of the ASQC 45th Annual Quality Congress*, ASQC, Inc., Milwaukee, WI. 77-83.