Communications of the Korean Statistical Society
2012, Vol. 19, No. 3, 451–457

DOI: http://dx.doi.org/10.5351/CKSS.2012.19.3.451

Fourier Series Approximation for the Generalized Baumgartner Statistic

Hyung-Tae Ha[1,a]

[a]Department of Applied Statistics, Gachon University

Abstract

Baumgartner *et al.* (1998) proposed a novel statistical test for the null hypothesis that two independently drawn samples of data originate from the same population, and Murakami (2006) generalized the test statistic for more than two samples. Whereas the expressions of the exact density and distribution functions of the generalized Baumgartner statistic are not yet found, the characteristic function of its limiting distribution has been obtained. Due to the development of computational power, the Fourier series approximation can be readily utilized to accurately and efficiently approximate its density function based on its Laplace transform. Numerical examples show that the Fourier series method provides an accurate approximation for statistical quantities of the generalized Baumgartner statistic.

Keywords: Fourier series approximation, generalized Baumgartner statistic, Laplace transform, limiting distributions, saddlepoint approximation.

1. Introduction

Baumgartner *et al.* (1998) proposed a novel statistical test for a null hypothesis that two independently drawn samples of data originate from the same population, and Murakami (2006) generalized the test statistic for cases with more than two samples. One of the most important procedures for test statistics is to determine their density and distribution functions as well as their percentage points under a null hypothesis. Combinatorial methods have been widely utilized to calculate the statistical quantities of nonparametric test statistics since the distribution-free nonparametric statistics are very complicated and computationally heavy in most cases; in addition, obtaining the density and distribution functions of the generalized Baumgartner statistic is also difficult. But when the sample size is large, the characteristic function of its limiting distribution has been obtained in a simple form, whereas the limiting density and distribution functions are not simply expressed, see Murakami (2006). It is natural to consider appropriate approximation methods in such a case. Since the normal, gamma, or shifted F distributions cannot provide enough accuracy to approximate the limiting distribution, more advanced approximation techniques should be considered. Saddlepoint methods based on the cumulant generating function proposed by Daniels (1954, 1987) and Lugannani and Rice (1980) have been successfully utilized to approximate the limiting distribution of the generalized Baumgartner statistic in Murakami *et al.* (2009).

A numerical inversion through the use of a Fourier series can be a viable alternative when the generating functions cannot be exactly inverted or analytical approximations are crude; however, numerical methods based on the Fourier series have been relatively less highlighted to provide statistical

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (2011-0013791).

[1] Assistant Professor, Department of Applied Statistics, Gachon University, Sungnam-ci, Kyunggi-Do 461-701, Republic of Korea. E-mail: htha@gachon.ac.kr

values of nonparametric test statistics. Numerical and analytical approximation methods have their own merits and shortcomings. This paper utilizes a modification of the Fourier series method to calculate the probability density function via numerically inverting its Laplace transformation. The beauty of the Fourier series method is the Poisson summation to bound the discretization error in association with the trapezoidal rule. Many variants of the Fourier series method have been discussed in last several decades, see for instance Piessens (1971) and Biermé and Scheffler (2008). The utilized modification of the Fourier series method in this paper is remarkably easy to use and simple to program. It should be mentioned that the Fourier series method is a well known technique to accurately approximate tail probabilities and has been used with great success in many scientific fields such as queueing theory and finance; for excellent discussions of the theory and its applications see Abate and Whitt (1992a, 1992b, 1995).

The rest of this paper is organized as follows: In Section 2, the Baumgartner and generalized Baumgartner statistics are briefly introduced. In Section 3, we summarize the Fourier series method and discuss some practical aspects. In Section 4, we applied the Fourier series method to provide approximations for statistical quantities of the generalized Baumgartner statistic and addressed some computational issues. Concluding remarks are provided in Section 5.

2. The Generalized Baumgartner Statistic

2.1. The Baumgartner statistic

First, we consider the Baumgartner statistic that tests a hypothesis that two independently drawn samples of data originate from the same population. The Baumgartner statistic can be compared with nonparametric test statistics for the scale parameter such as the Wilcoxon, the Cramer-von Mises and the Kolmogorov-Smirnov tests. Denote $X = (X_1, \dots, X_n)$ and $Y = (Y_1, \dots, Y_m)$ be two random samples of n and m independent observations from two continuous distributions $F(x)$ and $G(y)$, respectively. In addition, the observations are combined and ranked in increasing order of magnitude. We are interested in testing the null hypothesis $H_0 : F = G$ against $H_A : F \neq G$. Further denoting $R_1 < \cdots < R_n$ and $P_1 < \cdots < P_m$ the combined-samples ranks of the X-value and Y-value, respectively, Baumgartner *et al.* (1998) defined a novel nonparametric two-sample test statistic as

$$B = \sum_{i=1}^{n} \frac{\left(R_i - \frac{n+m}{n} i\right)^2}{\frac{2i}{n+1}\left(1 - \frac{i}{n+1}\right) m(n+m)} + \sum_{j=1}^{m} \frac{\left(P_j - \frac{m+n}{m} j\right)^2}{\frac{2j}{m+1}\left(1 - \frac{j}{m+1}\right) n(m+n)}. \tag{2.1}$$

In addition, a modified Baumgartner statistic in a one-sided test was suggested by Neuhauser (2001) and Murakami (2006) proposed another modification of the Baumgartner statistic based on the exact mean and variance of R_i and P_j. Murakami's modification of the Baumgartner statistic was defined as

$$B^* = \sum_{i=1}^{n} \frac{\left(R_i - \frac{n+m+1}{n+1} i\right)^2}{\frac{2ni}{n+1}\left(1 - \frac{i}{n+1}\right) \frac{m(n+m+1)}{n+2}} + \sum_{j=1}^{m} \frac{\left(P_j - \frac{m+n+1}{m+1} j\right)^2}{\frac{2mj}{m+1}\left(1 - \frac{j}{m+1}\right) \frac{n(m+n+1)}{m+2}}. \tag{2.2}$$

Murakami (2006) claimed that the B^* statistic is more efficient than the B statistic for the location parameter when the sample sizes are unequal, whereas the powers of the B and B^* statistics are equivalent for the location, scale and location-scale parameters when the sample sizes are equal.

2.2. The generalized Baumgartner statistic

Murakami (2006) generalized two sample Baumgartner statistic into k-sample Baumgartner statistic based on the modified Baumgartner statistic. Denoting $\{X_{ij}|i = 1, \dots, k, j = 1, \dots, n_i\}$ k independent samples of size $n_1, \dots, n_k$ from continuous distributions function $F_i(x)$ and R_{ij} the increasing-order rank of X_{ij} in the combined $N = n_1 + \cdots + n_k$ samples, we are interested in testing the hypothesis $H_0 : F_1 = F_2 = \cdots = F_k$ against H_A : not H_0. A modification of the k-sample Baumgartner statistic, denoted by V_k, is of interest in this paper with expression as

$$V_k = \frac{k-1}{k} \sum_{i=1}^{k} \left(\frac{1}{n_i} \sum_{j=1}^{n_i} \frac{\left(R_{ij} - \frac{N+1}{n_i+1} j\right)^2}{\frac{j}{n_i+1} \left(1 - \frac{j}{n_i+1}\right) \frac{(N-n_i)(N+1)}{n_i+2}} \right). \tag{2.3}$$

The following characteristic function of its limiting distribution was provided in Murakami (2006):

$$C\mathcal{F}_k(u) = \left(\frac{-2\pi u I}{\cos\left(\frac{\pi}{2} \sqrt{1 + 8uI}\right)} \right)^{\frac{(k-1)}{2}}, \tag{2.4}$$

where $I \equiv \sqrt{-1}$ is one of the imaginary roots of $x^2 + 1 = 0$ and $e^{Ix} = \cos[x] + I \sin[x]$.

3. The Fourier Series Approximation

We illustrate a modification of the Fourier series method to approximate a probability density function of the generalized Baumgartner statistic via a numerical inversion of the Laplace transformation.

The Laplace transform for the limiting distribution of the V_k statistic can be easily obtained from Equation (2.4), that is,

$$\mathcal{L}_k(t) = \mathbb{E}\left[e^{-tV_k}\right] = \int_0^\infty e^{-tx} f_{V_k}(x)dx = \left(\frac{2\pi t}{\cos\left(\frac{\pi}{2} \sqrt{1 - 8t}\right)} \right)^{\frac{(k-1)}{2}}, \tag{3.1}$$

where t is in general regarded as a complex variable and $f_{V_k}(x)$ is the limiting density function of the k sample Baumgartner statistic. We are interested in finding the probability density function $f_{V_k}(x)$. The Fourier series method is an efficient approximation algorithm by inverting the Laplace transform where k is typically odd numbers. As mentioned in Abate and Whitt (1992b), Bromwich contour integral for inverting the Laplace transform $\mathcal{L}_k(t)$ can be expressed as the integral of a real-valued function of a real variable and the contour shall be any vertical line $t = \sigma$ such that $\mathcal{L}_k(t)$ has no singularities on or to the right of $\mathcal{L}_k(t)$. Then we obtain

$$f_{V_k}(t) = \frac{1}{2\pi I} \int_{\sigma - I\infty}^{\sigma + I\infty} e^{st} \mathcal{L}_k(s)ds. \tag{3.2}$$

Let $s = \sigma + I\,u$ where u is a real number, then the probability density function is equivalent to

$$f_{V_k}(t) = \frac{e^{\sigma t}}{2\pi} \int_{-\infty}^{\infty} (\cos[ut] + I\sin[ut])\, \mathcal{L}_k(\sigma + Iu)du. \tag{3.3}$$

Since the probability density function $f_{V_k}(t)$ is a real valued function, Equation (3.3) can be reduced to

$$f_{V_k}(t) = \frac{e^{\sigma t}}{2\pi} \int_{-\infty}^{\infty} (\mathrm{Re}[\mathcal{L}_k(\sigma + Iu)]\cos[tu] - \mathrm{Im}[\mathcal{L}_k(\sigma + Iu)]\sin[ut])\, du, \tag{3.4}$$

where Re[·] and Im[·] are real and imaginary numbers, respectively.

The Fourier series approximation can be explained as a method to numerically calculate a standard inversion integral. The mathematical beauty of the Fourier series method is the Poisson summation to bound the discretization error in association with the trapezoidal rule. The integral can be replaced by an infinite summation with a discretized interval π/T. That is,

$$f_{V_k}(t) = \frac{e^{\sigma t}}{2\pi} \sum_{j=-\infty}^{\infty} \left(\mathrm{Re}\left[\mathcal{L}_k\left(\sigma + I\,\frac{j\pi}{T}\right)\right]\cos\left[t\,\frac{j\pi}{T}\right] - \mathrm{Im}\left[\mathcal{L}_k\left(\sigma + I\,\frac{j\pi}{T}\right)\right]\sin\left[\frac{j\pi}{T}\,t\right]\right). \tag{3.5}$$

The difficulty to calculate this infinite series can be solved from appropriate truncation of the infinite series from a Laplace transform based on estimates or bounds. The numerical integration based on Laplace transforms can produce a nearly alternating series and the convergence can be accelerated by techniques such as Euler summation. Then, finally, the selected Fourier series approximation with three parameters σ and T and M, denoted by $f_F(t\,;\sigma,T,M)$, uses the following formula to perform the inversion of a Laplace transformation on the positive real value:

$$f_F(t;\sigma,T,M) = \frac{e^{\sigma t}}{T}\left(\frac{1}{2}\mathcal{L}_k(\sigma)+\sum_{j=1}^{M}\left(\mathrm{Re}\left[\mathcal{L}_k\left[\sigma+\frac{j\pi I}{T}\right]\right]\cos\left[\frac{j\pi t}{T}\right] - \mathrm{Im}\left[\mathcal{L}_k\left[\sigma+\frac{j\pi I}{T}\right]\right]\sin\left[\frac{j\pi t}{T}\right]\right)\right). \tag{3.6}$$

The appropriate values of two parameters σ and T should be selected for an accurate inverse. The choice for the parameters σ and T must be such that $2\,T > \mathrm{Max}[t]$ and $\sigma = \sigma_0 - \log[\mathbf{E}]/(2\,T)$, where $\mathbf{E}$ is a measure of the maximum relative error and σ_0 is the exponential order of the target density function $f_{V_k}(t)$. It should be mentioned that the Fourier series density approximant $f_F(t\,;\sigma,T,M)$ is an integrable function such that the corresponding Fourier series distribution approximant can easily be obtainable via integration.

4. Numerical Examples

In this section, we compute the Fourier series approximation for the generalized Baumgartner statistic and obtained the approximated significant levels and limiting density and distribution functions. Saddlepoint approximation is also utilized as a benchmark to illustrate the performance of the Fourier series approximation. We revisit the cases when $k = 3$ and 5 discussed in Murakami *et al.* (2009) to make numerical comparisons in terms of the significance levels and percentage points. The infinite sum in Equation (3.6) was truncated at $M = 500$, which is large enough to guarantee consistent accuracy for all the numerical examples in this Section.

4.1. Approximating the significance levels and percentage points

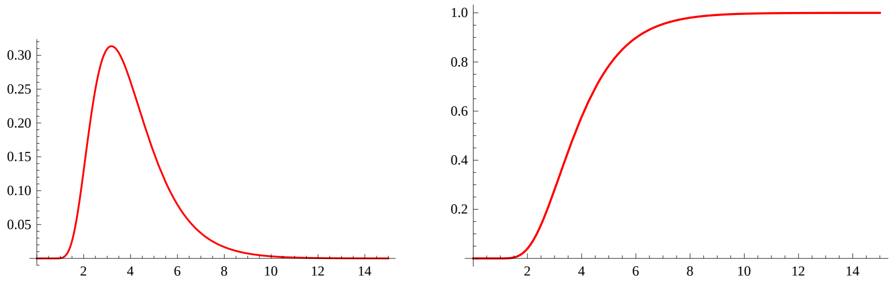
In Table 1, the significant levels of the limiting distributions of the generalized Baumgartner statistic when $k = 3$ are obtained by making use of the exact limiting distribution, the saddlepoint approximation and the Fourier series approximation. The significant levels via the Fourier series approximation are expressed in seventh decimals. It is shown in Table 1 that, for example, when the exact right tail probability at the point of 3.39934 is 0.10, the approximated probabilities of the Lugannani and Rice saddlepoint and the Fourier series approximants at the point of 3.39934 are respectively 0.0997 and 0.0999776. Although the saddlepoint method proposed in Murakami *et al.* (2009) provided accurate approximations, one can obtain even greater precision by making use of the Fourier series approximation for the limiting distributions of the generalized Baumgartner statistic when $k = 3$. We used

Table 1: The significant levels When $k = 3$

Critical value when $k = 3$	Exact Limiting Distribution	Saddlepoint Method	Fourier Series Approximation
5.70380	0.0100	0.0101	0.0099745
4.78738	0.0250	0.0251	0.0249785
4.09389	0.0500	0.0501	0.0499731
3.39934	0.1000	0.0997	0.0999776

Table 2: The percentage points when $k = 5$

Significance Level	Saddlepoint Approximation	Estimated by Simulation	Fourier Series Approximation
0.0100	8.80090	8.854	8.79750
0.0250	7.72522	7.763	7.72330
0.0500	6.88781	6.916	6.88726
0.1000	6.01961	6.039	6.02076



-1ex

Figure 1: *Fourier series density approximation (left panel); Fourier series distribution approximation (right panel)*

$\sigma = 0.01$ and $T = 50$ for computing significant levels in Table 1. In Table 2, the simulated and approximated percentage points for the generalized Baumgartner statistic when $k = 5$ were obtained. Table 2 shows that, for instance, 6.039 was obtained as the tenth percentage point by simulation in Murakami *et al.* (2009), and the approximated tenth percentage points via the saddlepoint and Fourier series approximants are respectively 6.01961 and 6.02076. Although we cannot determine which approximation method provides better accuracy since the simulated results contain errors to find the exact ones, we observed that the approximated percentage points obtained from both approximation methods are in excellent agreement with the simulated ones.

4.2. Approximating density and distribution functions

The Fourier series density approximant was obtained from Equation (3.6) for the generalized Baumgatner statistic when $k = 5$, and the corresponding Fourier series distribution approximant was easily obtained via the integration of the Fourier series density approximant. Figure 1 shows the Fourier series density (left panel) and distribution (right panel) approximants with two parameters $\sigma = 0.005$ and $T = 100$.

4.3. Computational issues

The symbolic computational package *Mathematica* was utilized to obtain the numerical examples. In addition to the accurate approximation in the entire range over the target distribution, the Fourier series approximation is computationally efficient and easy to program. The *Mathematica* program

Table 3: Computational times in seconds

	$T = 10$	$T = 20$	$T = 50$	$T = 100$	$T = 500$
$\sigma = 0.001$	5.141	5.438	5.468	5.531	5.453
$\sigma = 0.01$	5.468	5.547	5.531	5.500	5.453
$\sigma = 0.1$	5.390	5.437	5.469	5.515	5.500
$\sigma = 1$	6.171	6.142	6.344	6.437	6.219
$\sigma = 3$	6.141	18.125	18.484	18.500	6.328
$\sigma = 5$	45.625	80.970	56.984	13.343	6.313

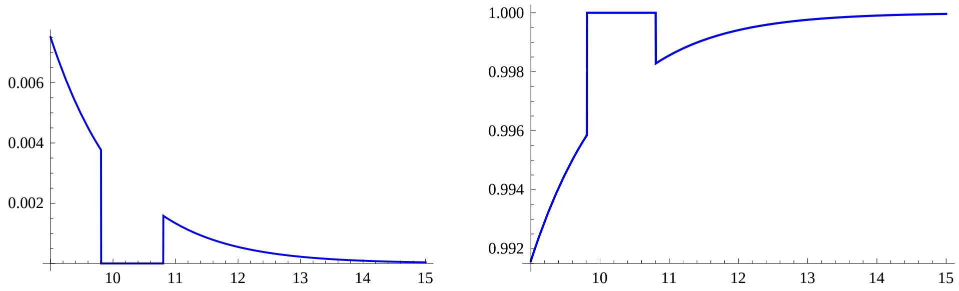


Figure 2: *Saddlepoint density approximation (left panel); Lugananni-Rice saddlepoint approximation (right panel)*

for the Fourier series approximation is less than 10 lines to obtain approximations to the density and distribution functions. Table 3 shows computational times in seconds that are used to obtain tail probability starting from 5.71234 of the Fourier series approximation for the generalized Baumgartner statistic when $k = 3$, from which one shall have brief ideas about the computational workload in various combinations of the two parameters σ and T. We used a Core2Duo E6300(1.86GHz) CPU in a personal computer for these computations.

It should be mentioned that saddlepoint approximation can be crude because of a difficulty of numerical computation to obtain the appropriate saddlepoint. The starting point of Newton-Raphson algorithm requires serious consideration to find the appropriate saddlepoint. Figure 2 shows the awful features of the saddlepoint density and distribution approximants for the limiting the distribution of the 5 sample Baumgartner statistic when the saddlepoint was calculated via the Newton-Raphson method with the starting point of 0.01.

5. Concluding Remarks

By using Fourier series approximation, the density and distribution functions, and critical values of the generalized Baumgartner statistic could be very accurately and easily obtained. Combinatorial and saddlepoint methods are arguably ones of the most widely used methods to exactly or approximately evaluate statistical quantities of nonparametric statistics. However, the Fourier series approximation should also be considered as a viable alternative.

Acknowledgement

The author wishes to express his gratitude to the journal editor and three referees. The detailed comments of three referees led to improvements on the quality of this paper.

References

Abate, J. and Whitt, W. (1992a). Numerical inversion of probability generating functions, *Operational Research Letters*, **12**, 245–251.

Abate, J. and Whitt, W. (1992b). The Fourier-series method for inverting transforms of probability distributions, *Queueing Systems*, **10**, 5–88.

Abate, J. and Whitt, W. (1995). Numerical inversion of Laplace transforms of probability distributions, *ORSA Journal on Computing*, **7**, 36–43.

Baumgartner, W., Weiß, P. and Schindler, H. (1998). A nonparametric test for the general two-sample problem, *Biometrics*, **54**, 1129–1135.

Biermé, H. and Scheffler, H. (2008). Fourier series approximation of linear fractional stable motion, *Journal of Fourier Analysis and Applications*, **14**, 180–202.

Daniels, H. E. (1954). Saddlepoint approximations in statistics, *Annals of Mathematical Statistics*, **25**, 631–650.

Daniels, H. E. (1987). Tail probability approximations, *International Statistical Review*, **55**, 37–48.

Lugannani, R. and Rice, S. O. (1980). Saddlepoint approximation for the distribution of the sum of independent random variables, *Advances in Applied Probability*, **12**, 475–490.

Murakami, H. (2006). A *k*-sample rank test based on modified Baumgartner statistic and its power comparison, *Journal of the Japanese Society of Computational Statistics*, **19**, 1–13.

Murakami, H., Kamakura, T. and Taniguchi, M. (2009). A saddlepoint approximation to the limiting distribution of a *k*-sample Baumgartner statistic, *Journal of the Japanese Statistical Society*, **39**, 133–141.

Neuhauser, M. (2001). One-side two-sample and trend tests based on a modified Baumgartner-Weiss-Schindler statistic, *Journal of Nonparametric Statistics*, **13**, 729–739

Piessens, R. (1971). Gaussian quadrature formulas for the numerical integration of Bromwich's integral and the inversion of the Laplace transform, *Journal of Engineering Mathematics*, **5**, 1–9.

Received February 14, 2012; Revised March 26, 2012; Accepted April 19, 2012