

# Polynomially Adjusted Normal Approximation to the Null Distribution of Ansari-Bradley Statistic

Hyung-Tae Ha<sup>1</sup> · Wan-Youn Yang<sup>2</sup>

<sup>1</sup>Department of Applied Statistics, Kyungwon University

<sup>2</sup>Department of Applied Statistics, Kyungwon University

(Received April 2011; accepted November 2011)

---

## Abstract

The approximation for the distribution functions of nonparametric test statistics is a significant step in statistical inference. A rank sum test for dispersions proposed by Ansari and Bradley (1960), which is widely used to distinguish the variation between two populations, has been considered as one of the most popular nonparametric statistics. In this paper, the statistical tables for the null distribution of the nonparametric Ansari-Bradley statistic is produced by use of polynomially adjusted normal approximation as a semi parametric density approximation technique. Polynomial adjustment can significantly improve approximation precision from normal approximation. The normal-polynomial density approximation for Ansari-Bradley statistic under finite sample sizes is utilized to provide the statistical table for various combination of its sample sizes. In order to find the optimal degree of polynomial adjustment of the proposed technique, the sum of squared probability mass function(PMF) difference between the exact distribution and its approximant is measured. It was observed that the approximation utilizing only two more moments of Ansari-Bradley statistic (in addition to the first two moments for normal approximation provide) more accurate approximations for various combinations of parameters. For instance, four degree polynomially adjusted normal approximant is about 117 times more accurate than normal approximation with respect to the sum of the squared PMF difference.

**Keywords:** Ansari-Bradley statistic, null distribution, normal-polynomial approximation, moments.

---

## 1. Introduction

Nonparametric approaches have been extensively discussed over the course of many decades. Many authors, such as Barton and David (1958), Terry (1952) and Mood (1954) studied in connections with nonparametric tests on dispersions. The most important procedure for the nonparametric test statistics is to determine null distributions, from which the exact critical values of the test

---

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

<sup>2</sup>Corresponding author: Associate Professor, Department of Applied Statistics, Kyungwon University, Sungnam-ci, Kyunggi-Do 461-701, Republic of Korea. E-mail: [ywy@kyungwon.ac.kr](mailto:ywy@kyungwon.ac.kr)

statistics can be calculated. In order to obtain the probabilistic quantities of those tests, combinatorial methods are commonly utilized since the nonparametric statistics are distribution-free; however, the procedures to calculate the exact critical values to obtain the combinations of the possible outcomes are complicated and time consuming as the number of sample sizes increases. Under these circumstances, approximation methods to provide enough accuracy is necessary to allow computationally easy implementation and distribution approximation techniques, that include asymptotic approximation, have played essential roles in nonparametric testing. Recently, many researchers applied various higher order asymptotic approximation methods for nonparametric statistics. For instance, Froda and Eeden (2000) used a uniform saddlepoint expansion to the null distribution of the Wilcoxon-Mann-Whitney test, and Bean *et al.* (2004) compared several approximation methods such as saddlepoint, Edgeworth, uniform and normal approximations for the Wilcoxon-Mann-Whitney test under finite sample sizes.

For the approximation presented in this study, an approximation technique in distribution theory (called the normal-polynomial approximation) is utilized to provide an accurate approximation for the null distribution of the Ansari-Bradley statistic. Its main concept and computing procedures were proposed in Ha and Provost (2007). While Provost *et al.* (2009) explained that beta-polynomial approximation provides high accuracy for Ansari-Bradley statistic, this paper aims to show that normal-polynomial approximation is also very flexible and fast to adapt the features of the target distributions. We observed that the normal polynomial approximant requires only four moments in most cases to provide accuracy.

In Section 2 and Section 3, the generating function and moments of the Ansari and Bradley statistic and the brief introduction to normal-polynomial approximation are introduced, respectively. In Section 4, normal-polynomial approximation to the null distribution of the Ansari and Bradley statistic is applied in various combinations of two independent sample sizes and the approximation accuracy with respect to the degree of polynomial adjustment is compared in terms of the sum of squared probability mass function difference. The concluding remarks are provided in Section 5.

## 2. Ansari-Bradley Statistic

On denoting  $X_1, \dots, X_m$  and  $Y_1, \dots, Y_n$  two independent samples of sizes  $m$  and  $n$  of independent observations, where  $m$  is less than  $n$ , from two populations with continuous cumulative distribution functions,  $A(t)$  and  $B(t)$ , respectively, we assume that the difference in location parameters of the two populations is known to be zero. The location parameters of the two populations need not to be known. Then  $A(t)$  and  $B(t)$  are assumed to be of the same form and to differ at most in the value of a scale parameter  $\alpha$ , so that  $B(t) = A(\alpha t)$ . Ansari and Bradley (1960) developed a rank-order test of the null hypothesis,

$$H_0 : \alpha = 1, \quad i.e., \quad B(t) = A(t), \quad (2.1)$$

against either one-sided or two sided alternatives to  $H_0$ . The two independent samples of sizes  $m$  and  $n$  of independent observations are ranked or ordered in a combined array represented by

$$Z_1, \dots, Z_{m+n}. \quad (2.2)$$

If the difference of the two population means exists, one may adjust by the mean difference without loss of generality. The ranks are assigned from both ends of (2.2), beginning with unity and working towards the center. The ranks can differ due to the sum of two sample sizes. If  $m + n$  is even, we

have the ranks

$$1, 2, 3, \dots, \frac{m+n}{2}, \frac{m+n}{2}, \dots, 3, 2, 1. \quad (2.3)$$

That is, the ranks denoted by  $R(Z)$  can be assigned as

$$\begin{aligned} R(Z_1) = 1, R(Z_2) = 2, \dots, R\left(Z_{\frac{m+n}{2}}\right) &= \frac{m+n}{2}, \\ R\left(Z_{\frac{m+n}{2}+1}\right) &= \frac{m+n}{2}, \dots, R(Z_{m+n-1}) = 2, R(Z_{m+n}) = 1. \end{aligned} \quad (2.4)$$

In addition, if  $m+n$  is odd, we have the symmetric array of ranks with the center of  $(m+n+1)/2$

$$1, 2, 3, \dots, \frac{m+n-1}{2}, \frac{m+n+1}{2}, \frac{m+n-1}{2}, \dots, 3, 2, 1. \quad (2.5)$$

The ranks in this case can be assigned in the similar way of the case that  $m+n$  is even. The Ansari and Bradley nonparametric test statistic is the sum of the ranks associated with the  $X$ -sample, that is,

$$W = \sum_X R(Z). \quad (2.6)$$

If the Ansari-Bradley statistics for  $X$  and  $Y$  samples are different, it means that the scale parameters of two populations can more likely be considered different to each other.

We are interested in the null distribution of the Ansari-Bradley statistic  $W$ . It is tedious to carry on the statistical tables for the exact null distributions of all possible combinations of sample sizes. But interestingly, its frequency generating function, which fully facilitates consideration of its null distributions, can be simply expressed. As derived in Ansari and Bradley (1960), its frequency generating function of  $W$  given the sample sizes  $m$  and  $n$ , denoted by  $g_W(u, v | m, n)$ , is

$$g_W(u, v | m, n) = \begin{cases} \prod_{i=1}^N (1 + u^i v)^2, & \text{if } m+n = 2N, \\ (1 + u^{N+1} v) \prod_{i=1}^N (1 + u^i v)^2, & \text{if } m+n = 2N+1, \end{cases} \quad (2.7)$$

the corresponding probability denoted by  $P(W = w | m, n)$  is the coefficient of  $u^w v^m$  in the expansions of the above Equation (2.7) over Binomial( $m+n, m$ ). And Ansari and Bradley (1960) kindly provided the explicit expressions for the important statistical quantities such as the mean, variance and skewness and kurtosis, which are the main information to be used in the proposed technique. In the case of  $m+n = 2N$ ,

$$E(W) = \mu_W = \frac{m(m+n+2)}{4} \quad (2.8)$$

and

$$\sigma_W^2 = \frac{mn(m+n-2)(m+n+2)}{48(m+n-1)}. \quad (2.9)$$

The third and fourth moments can be explicitly obtained as follows.

$$\mu_3 = 0, \quad (2.10)$$

and

$$\begin{aligned} \mu_4 = & \frac{mn(m+n+2)}{3840(m+n-3)(m+n-2)(m+n-1)} \{5mn(m+n)^4 - 2(m^2 + 19m^4n + 52m^3n^2 \\ & + 52m^2n^3 + 19mn^4 + n^5) + 4(3m^4 + 16m^3n + 26m^2n^2 + 16mn^3 + 3n^4) \\ & - 4(6m^3 - 34m^2n - 34mn^2 + 6n^3) - 16(2m^2 + 25mn + 2n^2) + 96(m+n)\}. \end{aligned} \quad (2.11)$$

Although the higher moments to determine the tail probabilities are expressed in complicated form, they can be computed based on their generating functions.

When  $m+n = 2N+1$ , the slightly different mean and variance were also be derived as

$$E(W)(= \mu_W) = \frac{m(m+n+1)^2}{4(m+n)} \quad (2.12)$$

and

$$\sigma_W^2 = \frac{mn(m+n+1)(3+(m+n)^2)}{48(m+n)^2}. \quad (2.13)$$

The third and fourth moments are

$$\mu_3 = \frac{mn(n-m)(m+n-1)(m+n+1)^2}{32(m+n-2)(m+n)^3} \quad (2.14)$$

and

$$\begin{aligned} \mu_4 = & \frac{mn(m+n+2)}{3840(m+n-3)(m+n-2)(m+n-1)} \{5mn(m+n)^6 - (2m^7 + 17m^6n + 57m^5n^2 \\ & + 100m^4n^3 + 100m^3n^4 + 57m^2n^5 + 17mn^6 + 2n^7) + 2(m^6 + 14m^5n + 47m^4n^2 + 68m^3n^3 \\ & + 47m^2n^4 + 14mn^5 + n^6) + 2(2m^5 - 35m^4n - 115m^3n^2 - 115m^2n^3 - 35mn^4 + 2n^5) \\ & + 15(4m^4 - m^3n - 10m^2n^2 - mn^3 + 4n^4) + 15(2m^3 + 9m^2n + 9mn^2 + 2n^3) - 30(m^2 \\ & - mn + n^2)\}. \end{aligned} \quad (2.15)$$

### 3. Polynomially Adjusted Normal Approximations

Provost *et al.* (2009) used a beta-polynomial approximant to the target distribution. In this paper, we utilize a special type of general semi parametric approach (the normal-polynomial approximation) proposed in Ha and Provost (2007). Let  $W$  be a discrete random variable whose support is the closed interval and let its raw moments  $E(W^h)$  be denoted by  $\mu_W(h)$ ,  $h = 0, 1, \dots$ . We are interested in approximating the probability mass functions of the discrete random variable  $W$ . A polynomially adjusted normal approximation of degree  $d$ , denoted by  $r_d(w)$ , is

$$r_d(w) = s(w) \sum_{i=0}^d \theta_i w^i. \quad (3.1)$$

Note that this density approximant is expressed as the product of normal approximation,  $s(w)$ , which can be considered as an initial approximation, and a polynomial adjustment,  $\sum_{i=0}^d \theta_i w^i$ . That is, the normal approximation is

$$s(w) = \frac{1}{\sqrt{2\pi\sigma_s^2}} \text{Exp} \left( -\frac{(w - \mu_s)^2}{2\sigma_s^2} \right). \quad (3.2)$$

The two parameters  $\mu_s$  and  $\sigma_s^2$  of the normal approximation are estimated from the first two moments of target distribution as follows:

$$\mu_s = \mu_W \quad \text{and} \quad \sigma_s^2 = \sigma_W^2. \tag{3.3}$$

The  $h^{th}$  raw moment of the normal approximation can be expressed as

$$\mu_s(h) = \frac{2^{\frac{h}{2}} (1 + (-1)^h) \Gamma(\frac{1+h}{2})}{4\sqrt{\pi}} + \sum_{j=0}^h \binom{h}{j} \frac{3^j 4^{h-j-1} 2^{\frac{j}{2}} (1 + (-1)^j) \Gamma(\frac{1+j}{2})}{\sqrt{\pi}}. \tag{3.4}$$

From the moment matching technique between the moments of the target distribution and the normal approximation, we can obtain the coefficients  $\theta_i$  of the polynomial adjustment. That is, the coefficients  $\theta_i$  satisfy the following matrix form equation,

$$\begin{pmatrix} \theta_0 \\ \theta_1 \\ \vdots \\ \theta_d \end{pmatrix} = \begin{pmatrix} \mu_s(0) & \mu_s(1) & \cdots & \mu_s(d-1) & \mu_s(d) \\ \mu_s(1) & \mu_s(2) & \cdots & \mu_s(d) & \mu_s(d+1) \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \mu_s(d) & \mu_s(d+1) & \cdots & \mu_s(2d-1) & \mu_s(2d) \end{pmatrix}^{-1} \begin{pmatrix} 1 \\ \mu_W(1) \\ \vdots \\ \mu_W(d) \end{pmatrix}. \tag{3.5}$$

In order to obtain distribution approximant  $R_d(w)$ , one can simply use numerical integration, that is  $\int_0^t r_d(w)dw$ . The continuity correction should be used to approximate discrete distributions via continuous function. That is, the approximation to the distribution of a nonparametric statistic will be corrected as  $R_d(t) = \int_0^{t+1/2} r_d(w)dw$ .

An important step for the proposed method is the determination of optimal degree of polynomial adjustment. A suitable choice of optimal degree can be decided via minimizing the sum of squared differences between the exact null and approximated probability mass function. The sum of squared difference over the entire range between the exact and approximated PMF's can be measured as

$$DPMF = \sum_{i=MinW}^{MaxW} (P(W = i | m, n) - r_d(i))^2, \tag{3.6}$$

where  $MinW$  and  $MaxW$  are, respectively, the possible minimum and maximum values of Ansari-Bradley statistic. For instance, if  $m$  and  $n$  are even numbers,  $MinW = (m^2 + 2m)/4$  and  $MaxW = (m^2 + 2mn + 2m)/4$ . The difference with respect to probability distribution function can be measured as

$$DDF = \sum_{i=MinW}^{MaxW} \left( \sum_{j=MinW}^i P(W = j | m, n) - R_d(i) \right)^2. \tag{3.7}$$

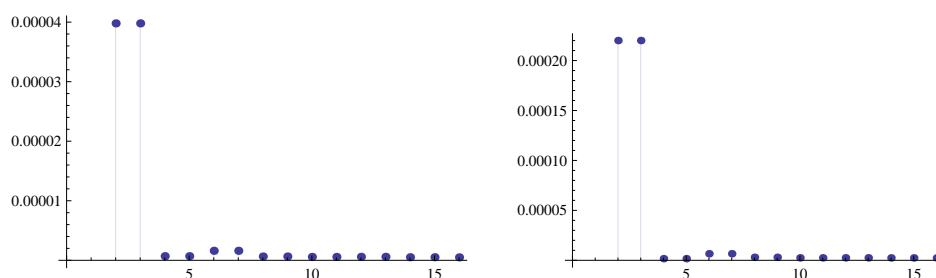
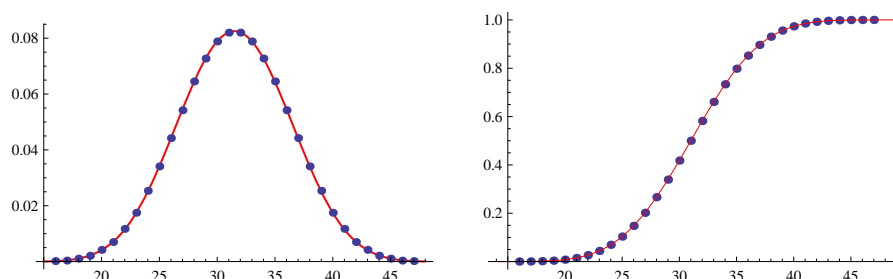
### 4. Numerical Results

#### Ratio of Difference Measures for Various Combinations

We consider approximating null distributions of Ansari-Bradley statistic with various combinations of its independent two sample sizes. Table 4.1 shows the optimal degree(OD) of polynomial adjustment and the ratios(ROD and RODC) of difference measures for normal and polynomially adjusted approximations. The  $4^{th}$  degree is chosen as the optimal degree in most cases except for very small

**Table 4.1.** Ratios of difference measures for various combinations

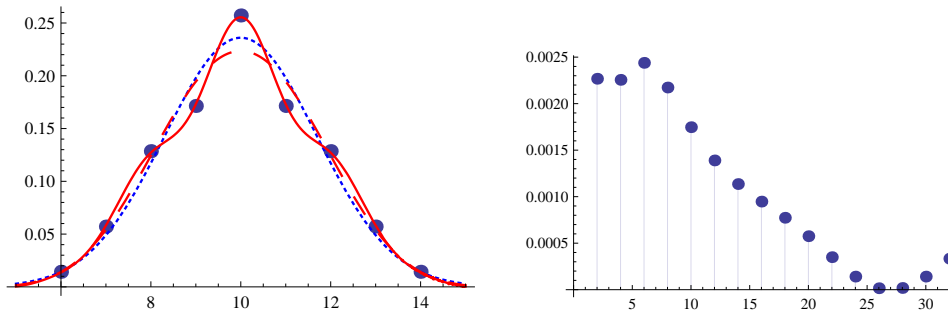
	4				6			
	OD	DPMF	ODC	DDF	OD	DPMF	DDF	RODC
4	16	2.39459	16	1.9838				
6	4	4.96275	4	8.38059	4	2.5191	4	9.339
8	4	2.58864	4	10.5679	4	31.2161	4	67.778
10	4	9.56948	4	48.41	4	17.0416	4	108.847
12					4	108.939	4	230.89

**Figure 4.1.** The PMF difference measures according to the degrees of polynomial adjustment when  $m = 7$  and  $n = 9$ (left panel); The distribution difference measures(right panel)**Figure 4.2.**  $4^{th}$  degree Normal-Polynomial Approximant(solid) and Exact PMF(dotted) when  $m = 7$  and  $n = 9$ (left panel);  $4^{th}$  degree Normal-Polynomial Distribution Approximant(solid) and Exact Distribution (dotted) when  $m = 7$  and  $n = 9$ (right panel)

sample sizes, for instance,  $m = 4$  and  $n = 4$ . The accuracy of polynomially adjusted normal approximations is observed to consistently and severely be improved when two sample sizes are not small. For instance, when  $m = 6$  and  $n = 12$ , the polynomially adjusted normal approximation is over 108 and 230 times more accurate than normal approximation in terms of the proposed difference measures in defined (3.6) and (3.7), respectively.

#### Case of $m = 7$ and $n = 9$

We consider an example of Ansari-Bradley statistic where its sample sizes are seven and nine, that is  $m = 7$  and  $n = 9$ . Figure 4.1 shows the PMF difference measures defined in (3.6) according to the degrees of Polynomial Adjustment in left panel and distribution difference measures defined in (3.7) in right panel. As can be seen in the Figure,  $4^{th}$  degree polynomially adjusted normal approximation, which uses only two more moments than normal approximation, dramatically reduces the approximation error and provides high approximation accuracy. This is the benefit to use



**Figure 4.3.** Normal Approximation (tiny dashing),  $4^{th}$  (large dashing) and  $26^{th}$  degree Normal-Polynomial Approximations (solid) and Exact PMF (dotted) When  $m = 4$  and  $n = 4$  (left panel); The Proposed Difference Measures According to the Degrees of Polynomial Adjustment When  $m = 4$  and  $n = 4$  (right panel).

polynomially adjusted normal approximation. Figure 4.2 shows the graphical representation that  $4^{th}$  degree normal-polynomial density and distribution approximants (solid) in the left and right panel, respectively, show close agreements with the exact null distribution (dotted).

#### Case of $m = 4$ and $n = 4$

The null distribution of Ansari-Bradley with  $m = 4$  and  $n = 4$  is an interesting and challenging case since normal approximation may fail to provide enough accuracy, and one can see that the proposed methodology is flexible to adapt the unique feature and provide high approximation accuracy. Due to the unusual feature, it is rare that the  $26^{th}$  degree as an optimal degree of the proposed approximation is much larger than the usual  $4^{th}$  degree. As can be seen in the left panel of Figure 4.3, while  $4^{th}$  degree normal approximation cannot adapt the behavior of the target PMF (dotted) in mode area,  $26^{th}$  degree normal polynomial approximant can interestingly accurately follow the target.

### 5. Concluding Remarks

The normal-polynomial approximation to the distribution of the Ansari-Bradley statistic is considered in this paper. The proposed approximation provides significant accuracy in approximations for statistical quantities of the Ansari-Bradley statistic. From the numerical results, it is determined that the approximation can provide accurate critical values of the target distribution. It should be noted that the normal-polynomial approximation also provides the explicit representation for the approximated density. Selection for the optimal degree of the polynomial adjustment was also proposed to obtain a suitable approximation. The continuity correction should be used to approximate a discrete distribution via a continuous function. The symbolic computational package *Mathematica* was utilized to obtain the numerical examples.

### Acknowledgements

The authors wish to express the gratitude to the journal editor and two referees. The suggestions and comments of two referees improved the quality of this manuscript in many aspects including the characteristics of the proposed method and numerical examples.

## References

- Ansari, A. R. and Bradley, R. A. (1960). Rank-sum tests for dispersions, *The Annals of Mathematical Statistics*, **31**, 1174–1189.
- Barton, D. E. and David, F. N. (1958). A test for birth order effects, *Annals of Human Genetics*, **22**, 250–257.
- Bean, R., Froda, S. and van Eeden, C. (2004). The normal, edgeworth, saddlepoint, uniform approximations to the Wilcoxon-Mann-Whitney null-distribution: A numerical comparison, *Journal of Nonparametric Statistics*, **16**, 279–288.
- Froda, S. and Eeden, C. V. (2000). A uniform saddlepoint expansion for the null-distribution of the Wilcoxon-Mann-Whitney statistic, *The Canadian Journal of Statistics*, **28**, 137–149.
- Ha, H-T. and Provost, S. B. (2007). A viable alternative to resorting to statistical tables, *Communication in Statistics: Simulation and Computation*, **36**, 1135–1151.
- Lehmann, E. L. (1951). Consistency and unbiasedness of certain nonparametric tests, *Annals of Mathematical Statistics*, **22**, 165–179.
- Mood, A. M. (1954). On the asymptotic efficiency of certain nonparametric two-sample tests, *Annals of Mathematical Statistics*, **25**, 514–522.
- Provost, S. B., Jing, M. and Ha, H-T. (2009). Moment-based approximations of probability mass functions with applications involving order statistics, *Communication in Statistics: Theory and Method*, **38**, 1969–1981.
- Terry, M. E. (1952). Some rank order tests which are most powerful against specific parametric alternatives, *Annals of Mathematical Statistics*, **23**, 346–366.