Asian Journal of Statistical Sciences
Vol. 2, No-2, (2022), pp. 129-140

ARF INDIA
Academic Open Access Publishing
Journal homepage link:
http://arfjournals.com/ajss

Testing of Location Parameter with Optimum Choice of Sub-sample Size

Arun Kumar[a], Manish Goyal[b] and Narinder Kumar[c]

[ac] *Department of Statistics, Panjab University, Chandigarh, INDIA*
[b] *Institute of Integrated and Honors Studies, Kurukshetra University, Kurukshetra, INDIA*

ARTICLE HISTORY
Compiled March 25, 2023

Received 13 December 2022; Accepted 06 February 2023

ABSTRACT
Testing of location parameter is very important and is useful in many fields like agriculture, medical, social, economic etc. When the data does not follow the Normal distribution, the nonparametric tests are more robust and powerful than parametric tests. To address this problem, a new class of test statistic is proposed in this paper which is independent of any distribution. The proposed test is compared with existing nonparametric two sample location tests in the literature, using Pitman and Bahadur asymptotic relative efficiency for some underlying distributions. Optimum choice of sub-sample size is found so that asymptotic relative efficiency is maximized. A real life data example is provided to see the working of the proposed test. A Monte-Carlo simulation study is also applied to find power and level of significance of the proposed test.

KEYWORDS
Nonparametric test; asymptotic relative efficiency; Monte-Carlo simulation

1. Introduction

Let $X_1, X_2, \ldots, X_n$ and $Y_1, Y_2, \ldots, Y_m$ be the independent random samples of size n and m from two populations with absolutely continuous cumulative distribution functions $F(x)$ and $F(x-\Delta)$, respectively, and Δ is called the shift parameter. If $\Delta > 0$ then it means the $Y's$ are stochastically greater than the $X's$ and if $\Delta < 0$ then it means $Y's$ are stochastically smaller than the $X's$. Now our motive is to test the null hypothesis:

$$H_0 : \Delta = 0,$$

against the alternative hypothesis

$$H_1 : \Delta \neq 0.$$

CONTACT Author[c]. Email: nkumar@pu.ac.in

In this paper, to test such a problem, no parametric model assumptions are assumed regarding the $F(.)$.

In many real life situations, it is very important to test the equality of location parameters. For example, in the field of agriculture, if a researcher wants to test that two fertilizers have the same effect on the yield of a crop or not. In the field of medical, if a researcher wants to test out of two different drugs that they have equal effects to control the blood pressure level of human beings. When the data is not Normal, the most familiar non-parametric tests are given by [1] and [2], which was further generalized by the authors of [3]. Paper [4] considered the test statistic based on sub-sample median. In papers [5] and [6], the proposed test is based on order statistics. Authors of [7] developed the test statistic based on sub-sample mid range. In paper [8], the proposed test is based on sub-sample extremes and in paper [9], proposed test is based on minimum and median of the sub-samples.

The proposed class of U-statistics is defined in section 2. Its distribution is established in section 3. Comparisons of the proposed with some of the existing tests and optimum choice of sub-sample size are given in section 4. To test if the type of behavior has significant effect on cholesterol level or not, an illustrative example is provided in section 5. In section 6, a simulation study is carried out to see the performance of the proposed test.

2. Proposed Class of Tests

Let c, d, i and j be the fixed positive integer such that $2 \leq (c,d) \leq \min(n,m)$, $c+1 \geq 2i$ and $d+1 \geq 2j$. Now we define the following kernel:

$$\Phi(X_1, X_2, \ldots, X_c; Y_1, Y_2, \ldots, Y_d) =$$

$$= \begin{cases} 1 & \text{if } X_{i:c} \leq Y_{j:d} \text{ and } X_{c-i+1:c} \leq Y_{d-j+1:d} \\ 0 & \text{otherwise,} \end{cases}$$

where $X_{i:c}$ and $Y_{j:d}$ are the i^{th} and j^{th} order statistic from the sub-samples $(X_1, X_2, \ldots, X_c)$ and $(Y_1, Y_2, \ldots, Y_d)$ respectively. Similarly, $X_{c-i+1:c}$ and $Y_{d-j+1:d}$ are $(c-i+1)^{th}$ and $(d-j+1)^{th}$ order statistic from the sub-samples $(X_1, X_2, \ldots, X_c)$ and $(Y_1, Y_2, \ldots, Y_d)$ respectively.
The U-statistics associated with the kernel $\Phi(.)$ is defined as:

$$U_{c,d,i,j} = \left[\binom{n}{c}\binom{m}{d}\right]^{-1} \sum [\Phi(X_{w_1}, X_{w_2}, \ldots, X_{w_c}; Y_{z_1}, Y_{z_2}, \ldots, Y_{z_d})]$$

where the summation is extended over all possible combinations $(w_1, w_2, \ldots, w_c)$ of c integers chosen from $(1, \ldots, n)$ and all possible combinations $(z_1, z_2, \ldots, z_d)$ of d integers chosen from $(1, \ldots, m)$.

In particular, when $c = d = i = j = 1$, the test statistic $U_{c,d,i,j}$ is same as that given by authors of [1] and [2].

3. The Distribution of Test Statistic

The expectation of $U_{c,d,i,j}$ is:

$$E(U_{c,d,i,j}) = \left[\binom{n}{c}\binom{m}{d}\right]^{-1} \sum E[\Phi(X_{w_1}, X_{w_2}, \ldots, X_{w_c}; Y_{z_1}, Y_{z_2}, \ldots, Y_{z_d})]$$

$$= P[X_{i:c} \leq Y_{j:d} \,\text{and}\, X_{c-i+1:c} \leq Y_{d-j+1:d}]$$

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{v} P[X_{i:c} \leq t \,\text{and}\, X_{c-i+1:c} \leq v] \,\mathrm{d}P[Y_{j:d} \leq t \,\text{and}\, Y_{d-j+1:d} \leq v].$$

Under the null hypothesis H_0, the expectation of $U_{c,d,i,j}$ is:

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{v} \sum_{s=c-i+1}^{c} \sum_{r=1}^{s} \frac{(c)!(d)!}{(r!)(s-r)!(c-s)!(d-2j)!((j-1)!)^2} (F(t))^{r+j-1}$$

$$\times (F(v) - F(t))^{s+d-r-2j} (1 - F(t))^{c-s+j-1} dF(t) dF(v).$$

After mathematical calculations, expectation of $U_{c,d,i,j}$ under the null hypothesis reduces to

$$E_{H_0}(U_{c,d,i,j}) = \sum_{s=c-i+1}^{c} \sum_{r=i}^{s} \frac{\binom{d+s-2j-r}{d-2j}\binom{c+j-s-1}{j-1}\binom{r+j-1}{r}}{\binom{c+d}{c}}.$$

The result of asymptotic distribution of U-statistic is given in [10]. Using this result, we find the asymptotic distribution of proposed test statistic in the following theorem.

Theorem 1: Let $N = n+m$. The asymptotic distribution of $\sqrt{N}[U_{c,d,i,j} - E(U_{c,d,i,j})]$, as $N \to \infty$ in such a way that $\frac{n}{N} \to \lambda$ and $0 \leq \lambda \leq 1$ is Normal with mean zero and variance $\sigma^2(U_{c,d,i,j})$, as

$$\sigma^2(U_{c,d,i,j}) = (c)^2 \frac{\xi_{10}}{\lambda} + (d)^2 \frac{\xi_{01}}{1-\lambda},$$

where

$$\xi_{10} = E[(\Phi(x_0, X_2, \ldots, X_c; Y_1, Y_2, \ldots, Y_d))^2] - [E(U_{c,d,i,j})]^2,$$

and

$$\xi_{01} = E[(\Phi(X_1, X_2, \ldots, X_c; y_0, Y_2, \ldots, Y_d))^2] - [E(U_{c,d,i,j})]^2,$$

with

$$\Phi(x_0, X_2, \ldots, X_c; Y_1, Y_2, \ldots, Y_d) =$$

$$= E[\Phi(X_1, X_2, \ldots, X_c; Y_1, Y_2, \ldots, Y_d) | X_1 = x_0],$$

and

$$\Phi(X_1, X_2, \ldots, X_c; y_0, Y_2, \ldots, Y_d) =$$

$$= E[\Phi(X_1, X_2, \ldots, X_c; Y_1, Y_2, \ldots, Y_d)|Y_1 = y_0].$$

Under H_0, asymptotic null variance of the test statistic, $\sigma_0^2(U_{c,d,i,j})$, after some computation is found as:

$$\sigma_0^2(U_{c,d,i,j}) = \frac{(c)^2\Psi_{c,d,i,j}}{\lambda(1-\lambda)},$$

where

$$\Psi_{c,d,i,j} = \begin{cases} \frac{1}{12}, & \text{for } c = d = 1 \\ A + B + C + D + E + F - (E_{H_0}(U_{c,d,i,j}))^2, & \text{for } c, d \geq 2, \end{cases}$$

with,

$$A = \left(\sum_{s=c-i+1}^{c-1} \sum_{r=i}^{s} \sum_{p=0}^{d+s-r-2j} \sum_{q=0}^{c+j-2-s} \frac{d!(c-1)!(-1)^{p+q}}{r!(s-r)!(c-1-s)!((j-1)!)^2(p+r+j)} \right.$$

$$\left. \times \frac{1}{(d-2j)!(d+s-j+q+1)} \binom{d+s-r-2j}{p} \binom{c+j-2-s}{q} \right),$$

$$B = \sum_{r=i}^{c-i} \sum_{l=0}^{c+d-r-i-2j} \sum_{m=0}^{i+j-2} \sum_{p=i}^{c-i} \sum_{w=0}^{c+d-p-i-2j} \sum_{z=0}^{i+j-2} \frac{(d!(c-1!))^2(-1)^{m+l+w+z}}{p!r!(c-i-r)!(c-i-p)!((i-1)!)^2}$$

$$\times \frac{1}{((d-2j)!)^2(r+j+l)} \binom{i+j-2}{z} \binom{i+j-2}{m} \binom{c+d-r-i-2j}{l}$$

$$\times \frac{\binom{c+d-p-i-2j}{w}}{(c+d+z-i-j+1)((j-1)!)^4} \left(1 - \frac{1}{c+d+z-i-j+2} - \right.$$

$$\left. - \frac{1}{c+d+m-i-j+2} + \frac{1}{2c+2d+m+z-2i-2j+3} \right)$$

$$\times \frac{1}{(c+d+m-i-j+1)(p+j+w)},$$

$$C = \sum_{s=c-i}^{c-1} \sum_{p=0}^{d+s-i-2j+1} \sum_{q=0}^{c+j-s-2} \sum_{l=1}^{c-i} \sum_{m=0}^{d+l-i-2j+1} \sum_{w=0}^{c+j-l-2} \frac{((c-1)!)^2(-1)^{p+q+m+w}}{(s-i+1)!(c-1-s)!}$$

$$\times \binom{d+s-i-2j+1}{p} \binom{d+l-i-2j+1}{m} \binom{c+j-l-2}{w}$$

$$\times \frac{(d!)^2 \binom{c+j-s-2}{q}}{((i-1)!)^2(l-i+1)!(c-1-l)!((j-1)!)^4((d-2j)!)^2(j-1)!}$$

$$\times \frac{1}{(p+i+j-1)(m+i+j-1)} \left\{ \left(\left[\frac{1}{(d+s+q-j+1)(d+l+w-j+1)} \right] \right. \right.$$

$$\times \left[1 - \frac{1}{d+s+q-j+2} - \frac{1}{d+l+w-j+2} \right.$$

$$\left. \left. + \frac{1}{2d+s+l+q+w-2j+3} \right] \right) - \left(\left[\frac{1}{(d+s+q-i-2j-p+2)} \right. \right.$$

$$\left. \times \frac{1}{(d+l+w-j+1)} \right] \left[\frac{1}{i+j+p} - \frac{1}{d+l+w+i+p+1} - \right.$$

$$\left. \left. - \frac{1}{d+s+q-j+2} + \frac{1}{2d+s+l+q+w-2j+3} \right] \right)$$

$$- \left(\left[\frac{1}{(d+s+q-j+1)(d+l+w-i-2j-m+2)} \right] \left[\frac{1}{i+j+m} - \right. \right.$$

$$\left. \left. \frac{1}{d+s+q+i+m+1} - \frac{1}{d+l+w-j+2} + \frac{1}{2d+s+l+q+w-2j+3} \right] \right)$$

$$+ \left(\left[\frac{1}{(d+l+w-i-2j-m+2)(d+s+q-i-2j-p+2)} \right] \times \right.$$

$$\left[\frac{1}{2i+2j+p+m-1} - \frac{1}{d+l+w+i+p+1} - \frac{1}{d+s+q+i+m+1} \right.$$

$$+ \left.\left.\left.\frac{1}{2d+s+l+q+w-2j+3}\right]\right)\right\},$$

$$D = \left(\sum_{s=c-i+1}^{c-1} \sum_{r=i}^{s} \sum_{p=0}^{d+s-r-2j} \sum_{q=0}^{c+j-2-s} \binom{d+s-r-2j}{p} \binom{c+j-2-s}{q} 2(d!)^2(c-1)! \right.$$

$$\left. \times \frac{(-1)^{p+q}}{r!(s-r)!(c-1-s)!((j-1)!)^2(d-2j)!(p+r+j)(d+s-j+q+1)} \right)$$

$$\times \left(\sum_{r=i}^{c-i} \left[\frac{(c-1)!}{r!(c-i-r)!(c-1-c+i)!(j-1)!^2(d-2j)!} \right] \right.$$

$$\left.\times \sum_{l=0}^{c+d-r-i-2j} \sum_{m=0}^{i+j-2} \frac{\binom{c+d-r-i-2j}{l}\binom{i+j-2}{m}(-1)^{p+q}}{(r+j+l)(c+d+m-i-j+1)} \left(1 - \frac{1}{c+d+m-i-j+2} \right) \right),$$

$$E = \left(\sum_{s=c-i+1}^{c-1} \sum_{r=i}^{s} \sum_{p=0}^{d+s-r-2j} \sum_{q=0}^{c+j-2-s} \binom{d+s-r-2j}{p} \binom{c+j-2-s}{q} \frac{((c-1)!)^2}{r!(s-r)!} \right.$$

$$\left. \times \frac{2((d!))^2(-1)^{p+q}}{(c-1-s)!((j-1)!)^2((d-2j)!)^2(p+r+j)(i-1)!(d+s-j+q+1)} \right)$$

$$\times \left(\sum_{s=c-i}^{c-1} \sum_{p=0}^{d+s-i-2j+1} \sum_{q=0}^{c+j-s-2} \frac{\binom{d+s-i-2j+1}{p}\binom{c+j-s-2}{q}(-1)^{p+q}}{(s-i+1)!(c-1-s)!(j-1)!^2(p+i+j-1)} \right.$$

$$\times \left[\frac{1}{d+s+q-j+1} \left\{ 1 - \frac{1}{d+s+q-j+2} \right\} - \frac{1}{d+s+q-i-2j-p+2} \right.$$

$$\left. \times \left\{ \frac{1}{i+j+p} - \frac{1}{d+s+q-j+2} \right\} \right],$$

and

$$F = \sum_{r=i}^{c-i} \sum_{l=0}^{c+d-r-i-2j} \sum_{m=0}^{i+j-2} \sum_{s=c-i}^{c-1} \sum_{p=0}^{d+s-i-2j+1} \sum_{q=0}^{c+j-s-2} \frac{2(d!)((c-1)!)^2(-1)^{m+l+p+q}}{(c-i-r)!(c-1-c+i)!}$$

$$\times \frac{\binom{i+j-2}{m}\binom{c+d-r-i-2j}{l}\binom{c+j-s-2}{q}\binom{d+s-i-2j+1}{p}}{(r+j+l)(c+d+m-i-j+1)(i-1)!(s-i+1)!((j-1)!)^4 r!}$$

$$\times \frac{1}{(c-1-s)!(p+i+j-1)(r+j+l)((d-2j)!)^2}$$

$$\times \Bigg(\Bigg[\frac{1}{d+s+q-j+1} \Bigg\{ 1 - \frac{1}{c+d+m-i-j+2} - \frac{1}{d+s+q-j+2}$$

$$+ \frac{1}{c+2d+m+s+q-i-2j+3} \Bigg\} \Bigg] - \frac{1}{d+s+q-i-2j-p+2} \Bigg[\frac{1}{i+j-p}$$

$$- \frac{1}{d+s+q-j+2} - \frac{1}{c+d+m+p+1} + \frac{1}{c+2d+m+s+q-i-2j+3} \Bigg] \Bigg).$$

Now, we found the value of expectation and variance of $U_{c,d,i,j}$ under null hypothesis for some values of c, d, i and j given in Table 1.

Table 1. Expectation and Variance of $U_{c,d,i,j}$ under null hypothesis without λ

(c,d,i,j)	(2,2,1,1)	(2,3,1,1)	(3,2,1,1)	(3,3,1,1)	(3,4,1,1)	(4,3,1,1)	(4,4,1,1)	(4,4,2,2)
$E_o(U_{c,d,i,j})$	0.3333	0.3000	0.3000	0.3000	0.2857	0.2857	0.2857	0.3714
$\sigma_o^2(U_{c,d,i,j})$	0.1206	0.1273	0.1273	0.1467	0.1566	0.1566	0.1740	0.2496

4. Asymptotic Relative Efficiency

Asymptotic relative efficiency (ARE) describes that how well one test performs relative to another test with increase or decrease in sample and sub-sample sizes. In this section, we compare the proposed test with other tests using asymptotic relative efficiencies as Pitman asymptotic relative efficiency and Bahadur asymptotic relative efficiency. The comparisons are given in subsections 4.1 and 4.2, using Pitman and Bahadur asymptotic relative efficiency, respectively.

4.1. Pitman Asymptotic Relative Efficiency

In this section, Pitman efficacy of the $U_{c,d,i,j}$ test is calculated and compared with relative tests in the terms of Pitman asymptotic relative efficiency (ARE). The limiting efficacy of the test $U_{c,d,i,j}$ under local alternatives $\Delta_N = \frac{\Delta}{\sqrt{N}}$ is given as:

$$e^2(U_{c,d,i,j}) = \lim_{N\to\infty} \frac{\frac{d}{d\Delta_N}\left[E(U_{c,d,i,j})|\Delta_N = 0\right]^2}{N\sigma_0^2(U_{c,d,i,j})}.$$

For $c = d = 1$,

$$\frac{d}{d\Delta_N}\left[E(U_{c,d,i,j})|\Delta_N = 0\right] = \sqrt{N}\int_{-\infty}^{\infty}(f(x))^2 dx,$$

and for $c, d \geq 2$,

$\frac{d}{d\Delta_N}\left[E(U_{c,d,i,j})|\Delta_N = 0\right] =$

$$= \sqrt{N}\int_{-\infty}^{\infty}\int_{-\infty}^{y}\sum_{s=c-i+1}^{c}\sum_{r=i}^{s}\sum_{p=0}^{s-r}\sum_{q=0}^{c-s}\frac{c!d!(-1)^{p+q}(F(y)-F(x))^{d-2j}(F(x))^{j-1}}{r!(s-r)!(c-s)!(j-1)!(d-2j)!}$$

$$\times((F(y))^{s-r-p+q}(r+p)(F(x))^{r+p-1}f(x)+(F(y))^{s-r-p+q-1}(s-r-p+q)(F(x))^{r+p}f(y))$$

$$\times\binom{s-r}{p}\binom{c-s}{q}(1-F(y))^{j-1}f(y)f(x)dxdy.$$

Now, we found the Pitman asymptotic relative efficacy for all choices of c, d, i and j with maximum value of $c = d = 10$ accordingly with condition to the value of i and j. The value of c, d, i and j which gives the maximum value of Pitman efficacy for all the considered distributions is the optimum choice of sub-sample size. Column two in Table 2 gives such optimal choice of sub-sample size. Using the optimum sub-sample size, we compare the proposed test with test given by [1] and [2] named as (WMW), test given by [4] and named as (K_m), test given by [7] and named as $(OZ_{r,s})$. Table 2 comprises the Pitman ARE of the proposed test with optimum sub-sample size.

Table 2. Pitman ARE of $U_{c,d,i,j}$ test with respect to different tests

Distribution	Optimum sub-sample size c,d,i,j	Tests						
			K_m			$OZ_{r,s}$		
		WMW	$m=1$	$m=2$	$m=3$	$r,s=1,3$	$r,s=2,3$	$r,s=3,3$
Uniform	10,10,1,1	6.500	11.058	12.798	13.775	5.293	4.271	3.531
U-quadratic	10,10,1,1	15.646	51.074	81.793	109.450	10.705	7.361	5.383
Beta(1,2)	6,10,1,1	5.393	8.500	9.677	10.342	4.520	3.756	3.187
Beta(2,2)	10,10,1,1	1.713	2.167	2.361	2.476	1.573	1.443	1.345
Normal	3,3,1,1	1.032	1.129	1.184	1.221	1.010	0.997	1.000
Cauchy	10,10,5,5	1.403	1.117	1.042	1.011	1.589	1.903	2.343
Exponential	4,10,1,1	7.775	13.228	15.309	16.477	6.331	5.110	4.224
Gumbel	2,5,1,1	1.266	1.416	1.495	1.546	1.227	1.198	1.187
Logistic	1,1,1,1	1.000	1.029	1.059	1.081	1.005	1.027	1.065
Rayleigh	4,10,1,1	1.421	1.676	1.792	1.863	1.345	1.277	1.230

From Table 2, it can be observed that proposed test with optimum choice of sub-sample sizes performs better than (WMW) test for all the considered distributions except for Logistic distribution. The proposed test always performs better than (K_m) test when we used optimum sub-sample sizes of the proposed test. The proposed test with optimum choice of sub-sample sizes performs better than $(OZ_{r,s})$ test except for Normal distribution.

4.2. Bahadur Asymptotic Relative Efficiency

The approximate Bahadur slope of the test statistics $U_{c,d,i,j}$ is given by

$$C(U_{c,d,i,j}) = \frac{1}{\sigma_0^2(U_{c,d,i,j})}[E(U_{c,d,i,j}) - E_{H_0}(U_{c,d,i,j})]^2.$$

Bahadur asymptotic efficiency of $U_{c,d,i,j}$ with respect to WMW is given by

$$B(U_{c,d,i,j}) = \frac{C(U_{c,d,i,j})}{C(WMW)},$$

Similarly, one can find Bahadur asymptotic efficiency for other tests also. We found the asymptotic relative efficacy for fixed value of $\Delta(\text{shift}) = 0.01$, 0.05 and 0.1 and for all the choice of c, d, i and j with maximum value of $c = d = 10$ accordingly with condition to the value of i and j. The value of c, d, i and j which gives the maximum value of Bahadur efficacy for all the considered distributions is the optimum choice of sub-sample size. Column two in Tables 3, 4 and 5 give the optimal choice of sub-sample size. Now, using the optimum sub-sample size, we compare the performance of the proposed test with respect to relative competitors of two sample location problem in terms of the Bahadur asymptotic relative efficiency. Namely, we compare the proposed test with test given by [1] and [2] named as (WMW), test given by [4] and named as (K_m), test given by [7] and named as $(OZ_{r,s})$. Tables 3, 4 and 5 comprise the Bahadur ARE of the proposed test with optimum sub-sample size.

Table 3. Bahadur ARE of $U_{c,d,i,j}$ test with respect to different tests when $\Delta(\text{shift}) = 0.01$

Distribution	Optimum sub-sample size c,d,i,j	Tests						
			K_m			$OZ_{r,s}$		
		WMW	$m=1$	$m=2$	$m=3$	$r,s=1,3$	$r,s=2,3$	$r,s=3,3$
Uniform	10,10,1,1	7.151	12.170	14.084	15.162	5.823	4.699	3.885
U-quadratic	10,10,1,1	20.228	66.296	106.062	141.916	13.841	9.507	6.939
Beta(1,2)	5,10,1,1	5.161	8.020	9.131	9.761	4.327	3.633	3.124
Beta(2,2)	10,10,1,1	150.269	190.123	207.215	217.761	5.568	5.144	4.814
Normal	3,3,1,1	1.037	1.135	1.191	1.228	1.015	1.003	1.006
Cauchy	10,5,5,1	10.287	8.188	7.643	7.413	11.648	13.954	17.178
Exponential	4,10,1,1	7.573	12.759	14.765	15.892	6.167	5.010	4.176
Gumbel	10,10,1,1	1.274	1.425	1.505	1.556	1.234	1.205	1.195
Logistic	1,1,1,1	1.000	2.550	1.059	1.081	1.006	1.027	1.065
Rayleigh	4,10,1,1	1.447	1.707	1.825	1.897	1.368	1.299	1.252

Table 4. Bahadur ARE of $U_{c,d,i,j}$ test with respect to different tests when $\Delta(\text{shift})= 0.05$

Distribution	Optimum sub-sample size c,d,i,j	Tests						
			K_m			$OZ_{r,s}$		
		WMW	$m=1$	$m=2$	$m=3$	$r,s=1,3$	$r,s=2,3$	$r,s=3,3$
Uniform	10,10,1,1	10.556	14.190	21.809	22.525	8.596	6.937	5.719
U-quadratic	10,10,1,1	73.617	266.680	412.904	552.978	13.354	54.522	158.352
Beta(1,2)	10,2,1,1	5.135	7.642	8.731	9.385	4.085	3.638	3.179
Beta(2,2)	10,10,1,1	6.059	7.672	8.455	8.931	3.834	3.915	3.772
Normal	3,3,1,1	1.060	3.586	2.282	1.935	1.037	1.025	1.028
Cauchy	10,5,5,1	2.217	18.346	7.362	4.846	2.510	3.008	3.703
Exponential	2,10,1,1	7.103	11.551	13.360	14.387	6.061	4.978	4.151
Gumbel	10,10,1,1	1.309	1.465	1.547	1.601	1.274	1.243	1.228
Logistic	2,2,1,1	1.006	1.035	1.065	1.088	1.033	1.033	1.071
Rayleigh	4,10,1,1	1.547	1.826	1.955	2.035	1.457	1.386	1.341

Table 5. Bahadur ARE of $U_{c,d,i,j}$ test with respect to different tests when Δ(shift)= 0.1

Distribution	Optimum sub-sample size c, d, i, j	Tests						
			K_m			$OZ_{r,s}$		
		WMW	$m = 1$	$m = 2$	$m = 3$	$r, s = 1, 3$	$r, s = 2, 3$	$r, s = 3, 3$
Uniform	10,10,1,1	17.430	18.979	42.414	36.820	14.193	11.454	9.365
U-quadratic	10,10,1,1	514.977	2721.105	3326.484	4801.519	44.724	145.877	306.929
Beta(1,2)	5,10,1,1	11.373	16.423	19.067	20.846	8.372	8.039	7.255
Beta(2,2)	10,10,5,1	5.975	7.437	8.620	9.333	2.732	3.062	3.053
Normal	3,3,1,1	1.087	3.489	2.251	1.925	1.063	1.052	1.055
Cauchy	10,5,5,1	1.680	13.058	5.269	3.483	1.901	2.279	2.806
Exponential	7,10,2,1	7.290	11.433	13.213	14.252	6.483	5.427	4.566
Gumbel	10,10,1,1	1.353	1.515	1.602	1.659	1.320	1.288	1.271
Logistic	2,2,1,1	1.017	1.047	1.078	1.102	1.022	1.044	1.083
Rayleigh	10,10,5,1	1.786	2.114	2.270	2.371	1.673	1.598	1.554

From the Tables 3, 4 and 5, it can be observed that the proposed test using optimum sub-sample size always performs better than (WMW) (k_m) and $(OZ_{r,s})$ tests in terms of Bahadur asymptotic relative efficiency.

5. A Real Life Example

To see the execution of proposed test, we worked out the test on two different real life examples as given in subsections 5.1 and 5.2.

5.1. Example Based on Effect of Behavior Type on Cholesterol Level

In California, a study is carried out where middle-aged men's were investigated to study the relationship between behavior pattern and the risk of coronary heart disease [11]. The particular data were obtained for the 40 heaviest men in the study (all weighing at least 225 pounds) and record cholesterol measurements (mg per 100 ml), and behavior type on a twofold categorization. In general terms, type A behavior is characterized by urgency, aggression and ambition, where as type B behavior is relaxed, non-competitive and less hurried. The question of interest is to test that cholesterol level is same or not in two different types of behavior for heavy middle-aged men's.

Let H_0 : Cholesterol levels in both behaviors' are same for heavy middle-aged men's.

Vs

H_1 : Cholesterol levels in both behaviors' are not same for heavy middle-aged men's.

Using Kolmogorov-Smirnov test, we see that data set follows the Cauchy distribution so the optimal choice of the sub-sample size from Table 2 is $c = d = 10$ and $i = j = 5$.

We find test statistic of this data set as: $U_{10,10,5,5} = 0.9689$ and the corresponding P-value $= 0.0079$. As the P-value is less than 0.05 (at 5%level of significance) so the null hypothesis is rejected. This implies that cholesterol levels in both behaviors' are not same for heavy middle-aged men's.

5.2. Example Based on Effect of Neuroleptic Treatment on Dopamine β-Hydroxylase

The relationship of Dopamine β-hydroxylase activity (DBH) in the cerebrospinal fluid (CSF) and responsiveness to neuroleptic treatment is examined by [12]. To test such a problem they divided the patients in two groups i.e. 15 patients who became nonpsychotic after chronic treatment with antipsychotic medication compared and the 10 patients who remained psychotic after treatment. Now,

Let H_0 : Dopamine β-hydroxylase activity (DBH) in the cerebrospinal fluid (CSF) is same for both groups of patients. Vs

H_1 : Dopamine β-hydroxylase activity (DBH) in the cerebrospinal fluid (CSF) is not same for both groups of patients.

Using Kolmogorov-Smirnov test, we see that data set follows the Uniform distribution so the optimal choice of the sub-sample size from Table 2 is $c = d = 10$ and $i = j = 1$.

We find test statistic of this data set as: $U_{10,10,1,1} = 1.00$ and the corresponding P-value $= 0.0006$. As the P-value is less than 0.05 (at 5% level of significance) so the null hypothesis is rejected. This implies that Dopamine β-hydroxylase activity (DBH) in the cerebrospinal fluid (CSF) is not same for both the group of patients.

6. Simulation Study

In this section, Monte Carlo simulation study is carried out to estimate power and level of significance using different sample sizes. Ten thousands random samples of sizes 10(10)30 from two Normal distributed populations are generated. The power of the proposed test is computed for Normal distribution at 5% and 1% level of significance. The estimated power (shift > 0) and level of significance (shift $=0$) of the proposed test with optimum choice of sub-sample size is given in Tables 6 and 7.

Table 6. Estimated power and level of significance of $U_{c,d,i,j}$ at 5% level of significance

Sample	Optimum Sub-sample size: (c, d, i, j)=(3,3,1,1)						
size n, m	Shift=0	Shift=0.25	Shift=0.5	Shift=0.75	Shift=1	Shift=1.25	Shift=1.5
10,10	0.054	0.136	0.284	0.473	0.665	0.820	0.916
10,20	0.052	0.148	0.327	0.556	0.768	0.915	0.967
10,30	0.046	0.150	0.351	0.605	0.816	0.935	0.970
20,20	0.050	0.168	0.417	0.701	0.900	0.980	0.980
20,30	0.050	0.187	0.466	0.765	0.942	0.983	0.989
30,30	0.050	0.203	0.543	0.845	0.978	0.998	0.999

Table 7. Estimated power and level of significance of $U_{c,d,i,j}$ at 1% level of significance

Sample	Optimum Sub-sample size: (c, d, i, j)=(3,3,1,1)						
size n, m	Shift=0	Shift=0.25	Shift=0.5	Shift=0.75	Shift=1	Shift=1.25	Shift=1.5
10,10	0.020	0.060	0.150	0.288	0.472	0.655	0.817
10,20	0.014	0.060	0.182	0.373	0.580	0.780	0.914
10,30	0.015	0.069	0.193	0.399	0.636	0.844	0.946
20,20	0.012	0.059	0.231	0.512	0.770	0.926	0.966
20,30	0.010	0.081	0.282	0.587	0.867	0.974	0.980
30,30	0.010	0.106	0.335	0.678	0.915	0.989	0.997

The following observations can be made from the Tables 6 and 7:

(i) From Table 6, it can be observed that at 5% level of significance, reasonable power of the proposed test is achieved at $n = 10$ and $m = 20$. While from Table 7, it can be seen that at 1% level of significance, reasonable power of the proposed test is achieved at $n = 30$ and $m = 30$. As the sample size increases power of the proposed test also increases.

(ii) When the sample size $n = m = 30$, we achieved the estimated level of significance of the proposed test that is 5% and 1% as given in Tables 6 and 7, respectively.

Acknowledgement(s)

The authors are grateful to the Editor-in-Chief and the anonymous reviewer for their insightful comments on an earlier version of this manuscript which have led to a substantial improvement in the presentation of the paper.

References

[1] Wilcoxon, F. (1945). Individual comparisons by rank methods. Biometrics Bulletin, 1, 80-83. DOI: 10.2307/3001968.

[2] Mann, H.B. and Whitney, D.R. (1947). On a test of whether one of the random variables is stochastically greater than the other. Annals of Mathematical Statistics, 18(1), 50–60.

[3] Deshpande, J.V. and Kochar, S.C. (1982). Some competitors of Wilcoxon-Mann Whitney test for location alternative. Journal of Indian Statistical Association, 19, 9-18.

[4] Kumar, N. (1997). A class of two-sample tests for location based on sub-sample medians. Communications in Statistics - Theory and Methods, 26(4), 943-951. DOI: 10.1080/03610929708831960.

[5] Öztürk, Ö. (2001). A generalization of Ahmad's class of Mann-Whitney-Wilcoxon Statistics. Australian and New Zealand Journal of Statistics, 43(1), 67–74. DOI: 10.1111/1467-842X.00155.

[6] Kumar, N. (2015). A class of nonparametric tests based on sample quantiles. Journal of Combinatorics, Information and System Sciences, 40(1-4), 211-223.

[7] Kumar, N. and Goyal, M. (2018). Testing of location parameter based on sub-sample midrange. Far East Journal of Theoretical statistics, 54(1), 1-20. DOI: 10.17654/TS054010001.

[8] Kumar, N. and Kumar, A. (2020). Two sample statistical test for location parameters. Accepted for publication in Journal of Modern Applied Statistical Methods.

[9] Kumar, A., Goyal, M. and Kumar, N. (2021). A class of tests based jointly on sub-sample minimum and median. Journal of the Kerala Statistical Association, 32(1), 1-17.

[10] Randles, R.H. and Wolfe, D.A. (1979). Introduction to the Theory of Nonparametric Statistics. John Wiley and Sons, New York.

[11] Selvin, S. (1991). Statistical analysis of epidemiological data. Oxford University Press, New York.

[12] Papas, T.S., Rushlow, T.E., Lautenberger, J.A., Baluda, M.A., B. Perbal, B., Chirikjian, J.G., and Reddy, E.P. (1982). Schizophrenia: Dopamine, β-Hydroxylase Activity and Treatment Response. Science, 216, 1423-1425. DOI: 10.1126/science.6124036.