S.Ü. Fen-Edebiyat Fakültesi Fen Dergisi Sayı 21 (2003) 99-108 KONYA

İki Örneklem Problemi İçin Yeni Bir Test İstatistiği ve Gücünün Diğer Testlerle Karşılaştırılması

Mehmet Fedai KAYA[1], Coşkun KUŞ[1]

Özet: Bu çalışmada, sürekli dağılımlar ailesi için dağılımdan bağımsız yeni bir istatistik önerilmiş, bu istatistiğin kesin dağılımı, $n \le 20$ için bilgisayar programı kullanılarak uygun durumların sayılmasıyla, $n = 25$ ve $n = 50$ için amprik dağılımı Monte Carlo simulasyon yöntemiyle elde edilmiştir. Testin gücü Monte Carlo simulasyon yöntemi kullanılarak Kolmogorov-Smirnov, Mann- Whitney Wilcoxon ve İşaret testlerinin gücü ile karşılaştırılmıştır.

Anahtar Kelimeler: Dağılımdan bağımsız istatistik, İki Örneklem Problemi, Monte Carlo simülasyon

A New Test Statistics For Two Sample Problem and Comparison of Power of the Test with Other Tests Via Simulation

Abstract: In this study, a new test statistic for two sample problem is proposed, exact distribution of new test statistic is obtained for $n \le 20$ using computer programming via count of suiatble situaition and emprical distribution of new test statistic is obtained for $n = 25$ and $n = 50$ using Monte Carlo simulation procedure. Power of the new test are compared with power of the Kolmogorov-Smirnov, Mann-Whitney Wilcoxon and Sign Tests using Monte Carlo procedure.

Keywords: Distribution free test statistics ,two sample problem, Monte Carlo Simulation

Giriş

$X_1, X_2, ..., X_m$ ve $Y_1, Y_2, ..., Y_n$ sırasıyla sürekli F ve G dağılım fonksiyonlarına sahip iki örneklem olmak üzere

$H_0 : F(x) = G(x), \forall x \in R$, hipotezinin

$H_1 : F(x) \ne G(x)$ veya $\left(F(x) > G(x) \text{ veya } F(x) < G(x)\right), \exists x \in R$

alternatif hipotezine karşı test edilmesi problemine iki örneklem problemi denir.

[1] Selçuk Üniversitesi Fen-Edebiyat Fak. İstatistik Böl. Öğretim üyesi

Burada $F(x) > G(x)$ gösterimi $P\{X \le x\} > P\{Y \le x\}$ anlamındadır ve Y rasgele değişkeni stokastik olarak X rasgele değişkeninden daha büyüktür denir.

İki örneklem problemi ile ilgili [1], [2], [4], [5], [6], [7] ve [8] gibi bir çok makale yayınlanmıştır. Literatürdeki testlerin çoğu Wilks'in boş bloklar veya Kolmogorov-Smirnov testlerinin bir modifikasyonudur. Daha ayrıntılı bilgi için [3]'e bakınız.

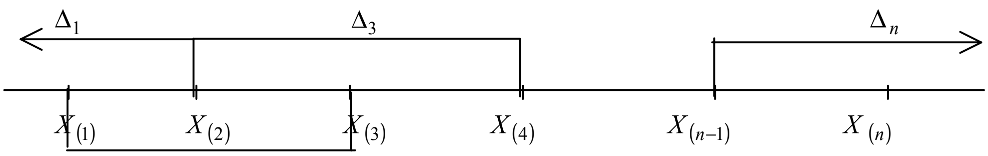
Yeni Test İstatistiği

$X_1, X_2, ..., X_n$ ve $Y_1, Y_2, ..., Y_n$ sırasıyla sürekli F ve G dağılım fonksiyonlarına sahip bağımsız iki örneklem ve $X_{(1)}, X_{(2)}, ..., X_{(n)}$ ve $Y_{(1)}, Y_{(2)}, ..., Y_{(n)}$ bu örneklemlerden oluşturulan sıra istatistikleri olsun. $X_{(0)} = -\infty$, $X_{(n+1)} = \infty$ ve $\Delta_r = \left(X_{(r-1)}, X_{(r+1)}\right)$, $r = 1,2,...,n$ olmak üzere

$$\xi_r = \begin{cases} 1 & , Y_{(r)} \in \Delta_r \\ 0 & , Y_{(r)} \notin \Delta_r \end{cases} \qquad \text{ve} \qquad S_n = \sum_{r=1}^{n} \xi_r$$

olsun.

S_n rasgele değişkeninin sıfır değerini alması olasılığı sıfırdır. $P(S_n = 0) = P\left(Y_{(1)} \notin \Delta_1, Y_{(2)} \notin \Delta_2, ..., Y_{(n)} \notin \Delta_n\right) = 0$ olduğu aşağıdaki gibi gösterilebilir.



Şekil 1. $Y_{(1)}, Y_{(2)}, ..., Y_{(n)}$ rasgele değişkenlerinin düşebileceği bölgeler

$Y_{(n)}$, $Y_{(n)} < X_{(n-1)}$ olacak biçimde $1,2,...,n-1$'inci yere düşebilir,

$Y_{(n-1)}$, $Y_{(n-1)} < X_{(n)}$ olacak biçimde $1,2,...,n-2$'inci yere düşebilir,

$\vdots$

$Y_{(2)}$, $Y_{(2)} < X_{(3)}$ olacak biçimde 1'inci yere düşebilir ve

$Y_{(1)}$'e düşecek yer kalmadığından, $\left\{Y_{(1)} \notin \Delta_1, Y_{(2)} \notin \Delta_2, ..., Y_{(n)} \notin \Delta_n\right\}$ olayı hiçbir n için gerçeklenmez. Dolayısıyla $P(S_n = 0) = 0$'dır.

S_n istatistiğinin belli bir değeri alması $X_{(1)}, X_{(2)}, ..., X_{(n)}$ rasgele değişkenlerinin (sıra istatistiklerinin) Δ_r aralıklarına belli biçimde düşmesi sonucu olur. Bu düşmeler sonucunda X_i ve Y_i, $i = 1,...,n$ rasgele değişkenlerinin kendi aralarında bir sıralanmaları söz konusu olur ve bu sıralanmalar $(2n)!/(n!)^2$ kadardır. Her bir sıralanma diğerinden ayrık ve bileşkeleri tüm mümkün durumları verdiğinden bu sıralanmaların olasılıkları toplamı 1'dir. Diğer taraftan X'ler ve Y'ler aynı dağılımdan geldiğinden her bir sıralanma eşit olasılıklı ve aşağıdaki gibidir.

$$\frac{1}{\dfrac{(2n)!}{(n!)^2}} = \frac{(n!)^2}{(2n)!}$$

$n = 3$ için S_n istatistiğinin kesin dağılımını bulalım.

$$P(S_3 = 0) = 0$$

$$
\begin{aligned}
P(S_3=1) = {} & P\left(Y_{(1)} \in \Delta_1, Y_{(2)} \notin \Delta_2, Y_{(3)} \notin \Delta_3\right) + P\left(Y_{(1)} \notin \Delta_1, Y_{(2)} \in \Delta_2, Y_{(3)} \notin \Delta_3\right) \\
& + P\left(Y_{(1)} \notin \Delta_1, Y_{(2)} \notin \Delta_2, Y_{(3)} \in \Delta_3\right) \\
& = P\left\{Y_{(1)} \le Y_{(2)} \le Y_{(3)} \le X_{(1)}\right\} + P\left\{Y_{(1)} \le Y_{(2)} \le X_{(1)} \le Y_{(3)} \le X_{(2)}\right\} \\
& + P\{\phi\} + P\left\{X_{(2)} \le Y_{(1)} \le X_{(3)} \le Y_{(2)} \le Y_{(3)}\right\} + P\left\{X_{(3)} \le Y_{(1)} \le Y_{(2)} \le Y_{(3)}\right\}
\end{aligned}
$$

$$
\begin{aligned}
= {} & 3!\,3 \int_{-\infty}^{\infty}\int_{-\infty}^{y_3}\int_{-\infty}^{y_2}\int_{y_3}^{\infty} [1-F(x_1)]^2 f(x_1) f(y_1) f(y_2) f(y_3)\, dx_1 dy_1 dy_2 dy_3 \\
& + 3!\,3! \int_{-\infty}^{\infty}\int_{-\infty}^{y_3}\int_{-\infty}^{y_2}\int_{y_2}^{y_3}\int_{y_3}^{\infty} [1-F(x_2)] f(x_1) f(x_2) f(y_1) f(y_2) f(y_3)\, dx_2 dx_1 dy_1 dy_2 dy_3 \\
& + 3!\,3! \int_{-\infty}^{\infty}\int_{-\infty}^{y_3}\int_{-\infty}^{y_2}\int_{-\infty}^{y_1}\int_{y_1}^{y_2} F(x_2) f(x_2) f(x_3) f(y_1) f(y_2) f(y_3)\, dx_3 dx_2 dy_1 dy_2 dy_3 \\
& + 3!\,3 \int_{-\infty}^{\infty}\int_{-\infty}^{y_3}\int_{-\infty}^{y_2}\int_{-\infty}^{y_1} [F(x_3)]^2 f(x_3) f(y_1) f(y_2) f(y_3)\, dx_3 dy_1 dy_2 dy_3 \\
= {} & \frac{1}{20} + \frac{1}{20} + \frac{1}{20} + \frac{1}{20} = \frac{4}{20}
\end{aligned}
$$

$$
\begin{aligned}
P(S_3=2) = {} & P\left(Y_{(1)} \in \Delta_1, Y_{(2)} \in \Delta_2, Y_{(3)} \notin \Delta_3\right) + P\left(Y_{(1)} \in \Delta_1, Y_{(2)} \notin \Delta_2, Y_{(3)} \in \Delta_3\right) \\
& + P\left(Y_{(1)} \notin \Delta_1, Y_{(2)} \in \Delta_2, Y_{(3)} \in \Delta_3\right) \\
& = P\left\{Y_{(1)} \le X_{(1)} \le Y_{(2)} \le Y_{(3)} \le X_{(2)}\right\} + P\left\{X_{(1)} \le Y_{(1)} \le Y_{(2)} \le Y_{(3)} \le X_{(2)}\right\} \\
& + P\left\{Y_{(2)} \le X_{(1)} \le X_{(2)} \le Y_{(3)} \le X_{(3)}\right\} + P\left\{Y_{(2)} \le X_{(1)} \le X_{(2)} \le X_{(3)} \le Y_{(3)}\right\} \\
& + P\left\{Y_{(1)} \le X_{(1)} \le X_{(2)} \le X_{(3)} \le Y_{(2)}\right\} + P\left\{X_{(1)} \le Y_{(1)} \le X_{(2)} \le X_{(3)} \le Y_{(2)}\right\} \\
& + P\left\{X_{(2)} \le Y_{(1)} \le Y_{(2)} \le Y_{(3)} \le X_{(3)}\right\} + P\left\{X_{(2)} \le Y_{(1)} \le Y_{(2)} \le X_{(3)} \le Y_{(3)}\right\}
\end{aligned}
$$

$$
\begin{aligned}
= {} & 3!\,3! \int_{-\infty}^{\infty}\int_{-\infty}^{y_3}\int_{-\infty}^{y_2}\int_{y_1}^{y_2}\int_{y_3}^{\infty} [1-F(x_2)] f(x_1) f(x_2) f(y_1) f(y_2) f(y_3)\, dx_2 dx_1 dy_1 dy_2 dy_3 \\
& + 3!\,3! \int_{-\infty}^{\infty}\int_{-\infty}^{y_3}\int_{y_2}^{y_3}\int_{x_1}^{y_3}\int_{y_3}^{\infty} [1-F(y_2)] f(x_1) f(x_2) f(x_3) f(y_2) f(y_3)\, dx_3 dx_2 dx_1 dy_2 dy_3 + \\
& + 3!\,3! \int_{-\infty}^{\infty}\int_{-\infty}^{y_3}\int_{y_2}^{y_3}\int_{x_1}^{y_3}\int_{x_2}^{y_3} [1-F(y_2)] f(x_1) f(x_2) f(x_3) f(y_2) f(y_3)\, dx_3 dx_2 dx_1 dy_2 dy_3 + \\
& + 3!\,3! \int_{-\infty}^{\infty}\int_{-\infty}^{y_2}\int_{y_1}^{y_2}\int_{x_1}^{y_2}\int_{x_2}^{y_2} [1-F(y_2)] f(x_1) f(x_2) f(x_3) f(y_1) f(y_2)\, dx_3 dx_2 dx_1 dy_1 dy_2 \\
& + 3! \int_{-\infty}^{\infty}\int_{-\infty}^{y_2}\int_{-\infty}^{y_1}\int_{y_1}^{y_2}\int_{x_2}^{y_2} [1-F(y_2)] f(x_1) f(x_2) f(x_3) f(y_1) f(y_2)\, dx_3 dx_2 dx_1 dy_1 dy_2 \\
& + 3! \int_{-\infty}^{\infty}\int_{-\infty}^{y_3}\int_{-\infty}^{y_2}\int_{-\infty}^{y_1}\int_{y_3}^{\infty} F(x_2) f(x_1) f(x_2) f(y_1) f(y_2) f(y_3)\, dx_2 dx_1 dy_1 dy_2 dy_3
\end{aligned}
$$

$$+\,3!\int_{-\infty}^{\infty}\int_{-\infty}^{y_3}\int_{-\infty}^{y_2}\int_{-\infty}^{y_1}\int_{y_2}^{y_3} F(x_2)f(x_1)f(x_2)f(y_1)f(y_2)f(y_3)\,dx_2\,dx_1\,dy_1\,dy_2\,dy_3$$

$$=\frac{1}{20}+\frac{1}{20}+\frac{1}{20}+\frac{1}{20}+\frac{1}{20}+\frac{1}{20}+\frac{1}{20}+\frac{1}{20}=\frac{8}{20}$$

$$P(S_3=3)=P\left(Y_{(1)}\in\Delta_1, Y_{(2)}\in\Delta_2, Y_{(3)}\in\Delta_3\right)$$

$$=P\left\{Y_{(1)}\le X_{(1)}\le Y_{(2)}\le X_{(2)}\le Y_{(3)}\le X_{(3)}\right\}+P\left\{Y_{(1)}\le X_{(1)}\le Y_{(2)}\le X_{(2)}\le X_{(3)}\le Y_{(3)}\right\}$$
$$+P\left\{Y_{(1)}\le X_{(1)}\le X_{(2)}\le Y_{(2)}\le Y_{(3)}\le X_{(3)}\right\}+P\left\{Y_{(1)}\le X_{(1)}\le X_{(2)}\le Y_{(2)}\le X_{(3)}\le Y_{(3)}\right\}$$
$$+P\left\{X_{(1)}\le Y_{(1)}\le Y_{(2)}\le X_{(2)}\le Y_{(3)}\le X_{(3)}\right\}+P\left\{X_{(1)}\le Y_{(1)}\le Y_{(2)}\le X_{(2)}\le X_{(3)}\le Y_{(3)}\right\}$$
$$+P\left\{X_{(1)}\le Y_{(1)}\le X_{(2)}\le Y_{(2)}\le Y_{(3)}\le X_{(3)}\right\}+P\left\{X_{(1)}\le Y_{(1)}\le X_{(2)}\le Y_{(2)}\le X_{(3)}\le Y_{(3)}\right\}$$

$$=3!\,3!\int_{-\infty}^{\infty}\int_{-\infty}^{y_3}\int_{-\infty}^{y_2}\int_{y_1}^{y_2}\int_{y_2}^{y_3}\int_{y_3}^{\infty} f(x_1)f(x_2)f(x_3)f(y_1)f(y_2)f(y_3)\,dx_3\,dx_2\,dx_1\,dy_1\,dy_2\,dy_3$$

$$+\,3!\,3!\int_{-\infty}^{\infty}\int_{-\infty}^{y_3}\int_{-\infty}^{y_2}\int_{y_1}^{y_2}\int_{y_2}^{y_3}\int_{x_2}^{y_3} f(x_1)f(x_2)f(x_3)f(y_1)f(y_2)f(y_3)\,dx_3\,dx_2\,dx_1\,dy_1\,dy_2\,dy_3$$

$$+\,3!\,3!\int_{-\infty}^{\infty}\int_{-\infty}^{y_3}\int_{-\infty}^{y_2}\int_{y_1}^{y_2}\int_{x_1}^{y_2}\int_{y_3}^{\infty} f(x_1)f(x_2)f(x_3)f(y_1)f(y_2)f(y_3)\,dx_3\,dx_2\,dx_1\,dy_1\,dy_2\,dy_3$$

$$+\,3!\,3!\int_{-\infty}^{\infty}\int_{-\infty}^{y_3}\int_{-\infty}^{y_2}\int_{y_1}^{y_2}\int_{x_1}^{y_2}\int_{y_2}^{y_3} f(x_1)f(x_2)f(x_3)f(y_1)f(y_2)f(y_3)\,dx_3\,dx_2\,dx_1\,dy_1\,dy_2\,dy_3$$

$$+\,3!\,3!\int_{-\infty}^{\infty}\int_{-\infty}^{y_3}\int_{-\infty}^{y_2}\int_{-\infty}^{y_1}\int_{y_2}^{y_3}\int_{y_3}^{\infty} f(x_1)f(x_2)f(x_3)f(y_1)f(y_2)f(y_3)\,dx_3\,dx_2\,dx_1\,dy_1\,dy_2\,dy_3$$

$$+\,3!\,3!\int_{-\infty}^{\infty}\int_{-\infty}^{y_3}\int_{-\infty}^{y_2}\int_{-\infty}^{y_1}\int_{y_2}^{y_3}\int_{x_2}^{y_3} f(x_1)f(x_2)f(x_3)f(y_1)f(y_2)f(y_3)\,dx_3\,dx_2\,dx_1\,dy_1\,dy_2\,dy_3$$

$$+\,3!\,3!\int_{-\infty}^{\infty}\int_{-\infty}^{y_3}\int_{-\infty}^{y_2}\int_{-\infty}^{y_1}\int_{y_1}^{y_2}\int_{y_3}^{\infty} f(x_1)f(x_2)f(x_3)f(y_1)f(y_2)f(y_3)\,dx_3\,dx_2\,dx_1\,dy_1\,dy_2\,dy_3$$

$$+\,3!\,3!\int_{-\infty}^{\infty}\int_{-\infty}^{y_3}\int_{-\infty}^{y_2}\int_{-\infty}^{y_1}\int_{y_1}^{y_2}\int_{y_2}^{y_3} f(x_1)f(x_2)f(x_3)f(y_1)f(y_2)f(y_3)\,dx_3\,dx_2\,dx_1\,dy_1\,dy_2\,dy_3$$

$$=\frac{1}{20}+\frac{1}{20}+\frac{1}{20}+\frac{1}{20}+\frac{1}{20}+\frac{1}{20}+\frac{1}{20}+\frac{1}{20}=\frac{8}{20}$$

Diğer n değerleri için S_n istatistiğinin dağılımı yukarıdaki izlenen yöntemle elde edilebilir. $n=3$ durumunda S_n istatistiğinin dağılımını hesaplamada çıkan zorluklar $n>3$ durumunda artacağından Borland 5 C++ dilinde yazılmış bilgisayar program yardımıyla mümkün durumların saydırılması ile S_n istatistiğinin dağılımı, $n=12,\ldots,15$ ve 20 için elde edilmiş, $n=25$ ve 50 için de Monte Carlo simulasyon yöntemiyle 100000 deneme sonucunda tahmin edilmiştir. S_n istatistiğinin dağılımı Tablo 1,2,3 ve 4'te verilmiştir.

Aşağıda S_n istatistiğinin asimptotik dağılımı için gerekli olabilecek iki teorem verilmiştir. Ancak bu çalışmada S_n istatistiğinin asimptotik dağılımı üzerinde durulmayacaktır.

Teorem 1.

$$P\{Y_{(r)} \in \Delta_r\} = P\{X_{(r-1)} < Y_{(r)} < X_{(r+1)}\}$$
$$= \frac{n!}{(r-1)!(n-r)!} \sum_{t=r-1}^{r} C_n^t B(t+r, 2n-t-r+1)$$

İspat.

$$P\{X_{(r-1)} < Y_{(r)} < X_{(r+1)}\} = P\{X_{(r-1)} < Y_{(r)}\} - P\{X_{(r+1)} < Y_{(r)}\}$$
$$= \int_{-\infty}^{\infty} F_{r-1}(y) f_r(y) dy - \int_{-\infty}^{\infty} F_{r+1}(y) f_r(y) dy$$
$$= \int_{-\infty}^{\infty} \sum_{t=r-1}^{n} \binom{n}{t} F(y)^t (1-F(y))^{n-t} f_r(y) dy - \int_{-\infty}^{\infty} \sum_{t=r+1}^{n} \binom{n}{t} F(y)^t (1-F(y))^{n-t} f_r(y) dy$$
$$= \frac{n!}{(r-1)!(n-r)!} \sum_{t=r-1}^{n} \int_{-\infty}^{\infty} F^t(y)(1-F(y))^{n-t} F^{r-1}(y)(1-F(y))^{n-r} f(y) dy -$$
$$- \frac{n!}{(r-1)!(n-r)!} \sum_{t=r+1}^{n} \int_{-\infty}^{\infty} F^t(y)(1-F(y))^{n-t} F^{r-1}(y)(1-F(y))^{n-r} f(y) dy$$
$$= \frac{n!}{(r-1)!(n-r)!} \sum_{t=r-1}^{n} u^{r+t-1}(1-u)^{2n-r-t} du - \frac{n!}{(r-1)!(n-r)!} \sum_{t=r+1}^{n} u^{r+t-1}(1-u)^{2n-r-t} du$$
$$= \frac{n!}{(r-1)!(n-r)!} \sum_{t=r-1}^{n} B(r+t, 2n-t-r+1) - \frac{n!}{(r-1)!(n-r)!} \sum_{t=r+1}^{n} B(r+t, 2n-t-r+1)$$
$$= \frac{n!}{(r-1)!(n-r)!} \sum_{t=r-1}^{r} C_n^t B(r+t, 2n-t-r+1)$$

Teorem 2. $P\{Y_{(r)} \in \Delta_r\} = P\{Y_{(n-r+1)} \in \Delta_{n-r+1}\}, \ \ r = 1, \dots, n$

İspat.

$$P\{Y_{(r)} \in \Delta_r\} = \frac{n!}{(r-1)!(n-r)!} \sum_{t=r-1}^{r} C_n^t B(t+r, 2n-t-r+1)$$
$$= \frac{n!}{(n-r)!(r-1)!} \left[C_n^{r-1} B(2r-1, 2n-2r+2) + C_n^r B(2r, 2n-2r+1)\right]$$
$$= \frac{n!}{(r-1)!(n-r)!} \left[C_n^{n-r} B(2n-2r+1, 2r) + C_n^{n-r+1} B(2n-2r+2, 2r-1)\right]$$
$$= \frac{n!}{(n-r)!(r-1)!} \sum_{t=n-r}^{n-r+1} C_n^t B(t+n-r+1, n-t+r)$$
$$= P\{Y_{(n-(r-1))} \in \Delta_{n-(r-1)}\}$$

Tablo 1. $P\{S_n = s\}$, $n = 1,2,\ldots,15$ kesin olasılıkları

n/S_n	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
2	$\frac{2}{6}$	$\frac{4}{6}$													
3	$\frac{4}{20}$	$\frac{8}{20}$	$\frac{8}{20}$												
4	$\frac{10}{70}$	$\frac{20}{70}$	$\frac{24}{70}$	$\frac{16}{70}$											
5	$\frac{28}{252}$	$\frac{56}{252}$	$\frac{72}{252}$	$\frac{64}{252}$	$\frac{32}{252}$										
6	$\frac{84}{924}$	$\frac{168}{924}$	$\frac{224}{924}$	$\frac{224}{924}$	$\frac{160}{924}$	$\frac{64}{924}$									
7	$\frac{264}{3432}$	$\frac{528}{3432}$	$\frac{720}{3432}$	$\frac{768}{3432}$	$\frac{640}{3432}$	$\frac{384}{3432}$	$\frac{128}{3432}$								
8	$\frac{858}{12870}$	$\frac{1716}{12870}$	$\frac{2376}{12870}$	$\frac{2640}{12870}$	$\frac{2400}{12870}$	$\frac{1728}{12870}$	$\frac{896}{12870}$	$\frac{256}{12870}$							
9	$\frac{2860}{48620}$	$\frac{5720}{48620}$	$\frac{8008}{48620}$	$\frac{9152}{48620}$	$\frac{8800}{48620}$	$\frac{7040}{48620}$	$\frac{4480}{48620}$	$\frac{2048}{48620}$	$\frac{512}{48620}$						
10	$\frac{9724}{184756}$	$\frac{19448}{184756}$	$\frac{27456}{184756}$	$\frac{32032}{184756}$	$\frac{32032}{184756}$	$\frac{27456}{184756}$	$\frac{19712}{184756}$	$\frac{11264}{184756}$	$\frac{4608}{184756}$	$\frac{1024}{184756}$					
11	$\frac{33592}{705432}$	$\frac{67184}{705432}$	$\frac{95472}{705432}$	$\frac{113152}{705432}$	$\frac{116480}{705432}$	$\frac{104832}{705432}$	$\frac{81536}{705432}$	$\frac{53248}{705432}$	$\frac{27648}{705432}$	$\frac{10240}{705432}$	$\frac{2048}{705432}$				
12	$\frac{117572}{2704156}$	$\frac{235144}{2704156}$	$\frac{335920}{2704156}$	$\frac{403104}{2704156}$	$\frac{424320}{2704156}$	$\frac{396032}{2704156}$	$\frac{326144}{2704156}$	$\frac{232960}{2704156}$	$\frac{139776}{2704156}$	$\frac{66560}{2704156}$	$\frac{22528}{2704156}$	$\frac{4096}{2704156}$			
13	$\frac{416024}{10400600}$	$\frac{832048}{10400600}$	$\frac{1193808}{10400600}$	$\frac{1447040}{10400600}$	$\frac{1550400}{10400600}$	$\frac{1488384}{10400600}$	$\frac{1279488}{10400600}$	$\frac{974848}{10400600}$	$\frac{645120}{10400600}$	$\frac{358400}{10400600}$	$\frac{157696}{10400600}$	$\frac{49152}{10400600}$	$\frac{8192}{10400600}$		
14	$\frac{1485800}{40116600}$	$\frac{2971600}{40116600}$	$\frac{4279104}{40116600}$	$\frac{5230016}{40116600}$	$\frac{5684800}{40116600}$	$\frac{5581440}{40116600}$	$\frac{4961280}{40116600}$	$\frac{3969024}{40116600}$	$\frac{2820096}{40116600}$	$\frac{1740800}{40116600}$	$\frac{901120}{40116600}$	$\frac{368640}{40116600}$	$\frac{106496}{40116600}$	$\frac{16384}{40116600}$	
15	$\frac{5348880}{155117520}$	$\frac{10697760}{155117520}$	$\frac{15452320}{155117520}$	$\frac{19018240}{155117520}$	$\frac{20920064}{155117520}$	$\frac{20920064}{155117520}$	$\frac{19100928}{155117520}$	$\frac{15876096}{155117520}$	$\frac{11907072}{155117520}$	$\frac{7938048}{155117520}$	$\frac{4595712}{155117520}$	$\frac{2228224}{155117520}$	$\frac{851968}{155117520}$	$\frac{229376}{155117520}$	$\frac{32768}{155117520}$

Tablo 2. $P\{S_{20}=s\}$ kesin olasılıkları

s	$P\{S_{20}=s\}$	s	$P\{S_{20}=s\}$	s	$P\{S_{20}=s\}$	s	$P\{S_{20}=s\}$
1	$\frac{3534526380}{137846528820}$	6	$\frac{15721136640}{137846528820}$	11	$\frac{7779932160}{137846528820}$	16	$\frac{464257024}{137846528820}$
2	$\frac{7069052760}{137846528820}$	7	$\frac{15562337280}{137846528820}$	12	$\frac{5456056320}{137846528820}$	17	$\frac{171573248}{137846528820}$
3	$\frac{10316995920}{137846528820}$	8	$\frac{14450741760}{137846528820}$	13	$\frac{3502653440}{137846528820}$	18	$\frac{49545216}{137846528820}$
4	$\frac{12991772640}{137846528820}$	9	$\frac{12586129920}{137846528820}$	14	$\frac{2031124480}{137846528820}$	19	$\frac{9961472}{137846528820}$
5	$\frac{14847740160}{137846528820}$	10	$\frac{10255365120}{137846528820}$	15	$\frac{1044578304}{137846528820}$	20	$\frac{1048576}{137846528820}$

Tablo 3. $n=25$ için S_n test istatistiğinin olasılık dağılımının tahmini

s	$P\{S_{25}=s\}$	s	$P\{S_{25}=s\}$	s	$P\{S_{25}=s\}$	s	$P\{S_{25}=s\}$
1	0.020263	8	0.098635	15	0.020240	22	0.000084
2	0.040590	9	0.091746	16	0.012782	23	0.000019
3	0.060085	10	0.081389	17	0.007507	24	0.000003
4	0.076361	11	0.068927	18	0.003945	25	0.000000
5	0.088975	12	0.055612	19	0.001848		
6	0.097240	13	0.042098	20	0.000803		
7	0.100399	14	0.030177	21	0.000272		

Tablo 4. $n=50$ için S_n test istatistiğinin olasılık dağılımının tahmini

s	$P\{S_{50}=s\}$	s	$P\{S_{50}=s\}$	s	$P\{S_{50}=s\}$	s	$P\{S_{50}=s\}$
1	0.010270	14	0.056840	27	0.003150	40	0.000000
2	0.019470	15	0.051210	28	0.002040	41	0.000000
3	0.029900	16	0.045190	29	0.001050	42	0.000000
4	0.039970	17	0.039940	30	0.000790	43	0.000000
5	0.047470	18	0.033160	31	0.000390	44	0.000000
6	0.055310	19	0.028420	32	0.000210	45	0.000000
7	0.059770	20	0.022690	33	0.000080	46	0.000000
8	0.062490	21	0.019270	34	0.000070	47	0.000000
9	0.067270	22	0.014780	35	0.000040	48	0.000000
10	0.067930	23	0.010290	36	0.000010	49	0.000000
11	0.067780	24	0.008040	37	0.000000	50	0.000000
12	0.064020	25	0.005780	38	0.000000		
13	0.061060	26	0.003850	39	0.000000		

Tablo 5. $H_0 : F = G$ hipotezini $H_1 : F \neq G$ hipotezine karşı $\alpha = 0.1$ anlam seviyesinde test etme probleminde S_n, Mann-Whitney, Kolmogorov-Smirnov ve İşaret Testlerinin farklı durumlardaki güçlerinin 10000 deneme yapılarak elde edilen tahminleri

Dağılımlar		$X \sim U(0,1)$ $Y \sim U(0,1)$	$X \sim U(0,1)$ $Y \sim U(.2,1.2)$	$X \sim U(0,1)$ $Y \sim Üstel(1)$	$X \sim U(0,1)$ $Y \sim Üstel(2)$	$X \sim Üstel(1)$ $Y \sim Üstel(2)$	$X \sim N(0,1)$ $Y \sim N(1,1)$	$X \sim N(0,1)$ $Y \sim N(0,4)$	$X \sim N(0,1)$ $Y \sim N(0,9)$	$X \sim N(0,1)$ $Y \sim N(.5,4)$
Örnek Hacmi	Testler									
n=25	S Testi	.1012	.7598	.3282	.7258	.5384	.8945	.3656	.6346	.402
	Mann-Whitney	.1006	.7342	.4978	.9774	.6741	.9603	.1116	.1228	.3014
	Kolmogorov-Smirnov	.0988	.5706	.7416	.9977	.5998	.9172	.346	.6802	.5238
	İşaret Testi	.1004	.568	.3783	.9248	.507	.8665	.1008	.0979	.2324
n=50	S Testi	.0988	.9413	.4062	.8547	.7652	.9886	.6446	.8652	.6426
	Mann-Whitney	.1074	.9450	.7396	1	.8997	1	.1180	.1243	.464
	Kolmogorov-Smirnov	.099	.866	.9846	1	.8456	.9966	.6293	.9753	.8396
	İşaret Testi	.1023	.8234	.5917	.9966	.772	.9875	.0948	.1041	.3451

Sonuç

$X_1, X_2, ..., X_n \sim F$ ve $Y_1, Y_2, ..., Y_n \sim G$ bağımsız iki örneklem olmak üzere, bu çalışmada $H_0 : F = G$ hipotezini $H_1 : F \neq G$ hipotezine karşı test etmek için Kolmogorov-Smirnov, Mann Whitney-Wilcoxon ve İşaret testine alternatif yeni test istatistiği önerildi. Yeni istatistiğin gücü, yukarıda adı geçen testlerin güçleriyle Monte Carlo simulasyon yöntemi kullanılarak çeşitli dağılımlar ve örneklem büyüklükleri için karşılaştırıldı. Elde edilen test istatistiklerin ampirik gücü Tablo 5'de verilmiştir.

$X_1, X_2, ..., X_n \sim U(0,1)$ ve $Y_1, Y_2, ..., Y_n \sim U(0.2,1.2)$ durumunda S_n testinin gücünün diğer testlerden daha iyi olduğu görülmekte ve $X_1, X_2, ..., X_n \sim N(0,1)$ ve $Y_1, Y_2, ..., Y_n \sim N(0,9)$ durumunda Mann Whitney-Wilcoxon ve İşaret testinin gücü son derece kötü olmasına karşın, S_n ve Kolmogorov-Smirnov tetlerinin gücünün çok daha iyi olduğu Tablo 5'den anlaşılmaktadır. Sonuç olarak örneklem varyanslarının tahmini arasındaki fark büyük olduğu durumlarda Mann Whitney-Wilcoxon ve İşaret testi yerine S_n ve Kolmogorov-Smirnov testlerinin kullanılması daha uygun olacaktır.

Kaynaklar

1- Bairamov, I.G., Özkaya N., **On The Nonparametric Test For Two Sample Problem Based On Spacings.** Journal Of Applied Statistical Science, 10, 1: 57-68 (2000).

2- Borovkov, A.A., **Asymptotically Optimal Tests for Compound Hypotheses.** Theor.Prob. Appl., 20,1:447-469 (1975).

3- Borovkov, A.A., **Matematical Statistics.** Nauka, Moskow (1984).

4- Dixen, W.J., **A Criterion for Testing the hypothesis thet Two Samples are From the same Population.** Ann. Math. Stat., 11:199-204 (1940).

5- Siddiqui M., Gürler Ü., **A Two Sample Matching Test.** Order Statistics and Nonparametrics:Theory and Appl., 237-243 (1992).

6- Smirnov, N., **On the Estimation of the Discrepancy Between Emprical Curves of Distribution for Two Independent Samples.** Bull. Math. Univ. Moscow, 2,2:3-16 (1939).

7- Wald, A., Wolfowitz J., **On a Test Whether Two Samples are From the same Population.** Ann. Math. Stat., 11:147-162 (1940).

8- Wilks, S.S., **A Combinatorial Tests for the Problem of Two Samples are From Continuous Distributions.** Proc. Fourth Berkeley Symp. on Math. Stat. and Prob., University of California Press (1962).