Düzgün Dağılım Fonksiyonlarına Sahip Dağılımlar Ailesi İçin Majorant Vektörler Kullanılarak En Dar İnvaryant Güven Aralığının Elde Edilmesi[1]

Buğra SARAÇOĞLU[2], Mehmet Fedai KAYA[2]

Özet: Kaya ve ark. tarafından yapılan çalışmada düzgün dağılım fonksiyonları ailesi için sıra istatistikleri ve majorant vektörler yardımıyla oluşturulan $\left(\sum_{i=1}^{n} a_i X_{[i]}, \sum_{i=1}^{n} b_i X_{[i]}\right)$ rasgele aralığının $\alpha = \frac{1}{n+1}\sum_{i=1}^{n}(n-i+1)(b_i - a_i)$ seviyeli ana kitleyi kapsayan invaryant güven aralığı olduğu gösterilmişti [1]. Bu çalışmada, invaryant güven aralığının seviyesi $\alpha = 0.90$ olacak biçimde $\sum_{i=1}^{n} b_i X_{[i]} - \sum_{i=1}^{n} a_i X_{[i]}$ değerini en küçük yapan $\underline{a}$ ve $\underline{b}$ majorant vektörler arasındaki ilişkinin elde edilebilmesi için bir optimizasyon modeli kurulmuş ve bu modelin çözümü bilgisayar programı yardımıyla bulunmuştur.

Anahtar Sözcükler: Sıra istatistikleri, İnvaryant güven aralıkları, Majorant vektörler

Obtaining The Narrowest Invariant Confidence Interval For The Family Of Uniform Distribution Functions Using By Majorant Vectors

Abstract: A random interval $\left(\sum_{i=1}^{n} a_i X_{[i]}, \sum_{i=1}^{n} b_i X_{[i]}\right)$ that is constructed by order statistics and majorant vectors, was shown as invariant confidence intervals containing main mass for the family of uniform distribution with level $\alpha = \frac{1}{n+1}\sum_{i=1}^{n}(n-i+1)(b_i - a_i)$ given in the study by [1]. In this study, an optimization model is consructed for the relationship between $\underline{a}$ and $\underline{b}$ majorant vectors which minimizes $\sum_{i=1}^{n} b_i X_{[i]} - \sum_{i=1}^{n} a_i X_{[i]}$ such that $\alpha = 0.90$ and this model is solved via computer programme.

Key words: Order Statistics, Invariant Confidence Intervals, Majorant Vectors

---

[1] Bu makale Yüksek Lisans tezinin bir parçasıdır.

[2] Selçuk Üniversitesi Fen-Edebiyat Fakültesi İstatistik Bölümü 42075 Kampüs/KONYA

1. Giriş.

$\mathrm{P} = \{P_\theta, \theta \in \Theta\}$ parametrik dağılım fonksiyonlarının bir ailesi, $X_1, X_2, ..., X_n$, $P_\theta \in \mathrm{P}$ dağılımına sahip bir örneklem, f_1 ve f_2 ;

$$\forall (x_1, x_2, ..., x_n) \in R^n \text{ için } f_1(x_1, x_2, ..., x_n) \le f_2(x_1, x_2, ..., x_n) \tag{1.1}$$

koşuluna sahip iki borel ölçülebilir fonksiyon, X_{n+1} , $X_1, X_2, ..., X_n$ örnekleminden bağımsız ve aynı $P_\theta \in \mathrm{P}$ dağılımına sahip yeni bir rasgele değişken olmak üzere; $\forall \theta \in \Theta$ için,

$$P_\theta\{X_{n+1} \in (f_1(X_1, X_2, ..., X_n), f_2(X_1, X_2, ..., X_n))\} = \alpha \qquad , \quad \alpha \in R \tag{1.2}$$

oluyorsa $(f_1(X_1, X_2, ..., X_n), f_2(X_1, X_2, ..., X_n))$ rasgele aralığına P sınıfı için α güven seviyesine sahip ana kitleyi kapsayan invaryant güven aralığı denir.

$\forall \underline{a} = (a_1, a_2, ..., a_n) \in R^n, \forall \underline{b} = (b_1, b_2, ..., b_n) \in R^n$ için $D^+ = \{(u_1, u_2, ..., u_n) : u_1 \ge u_2 \ge ... \ge u_n\}$ biçiminde tanımlanan bir küme olsun. $a_{[1]} \ge a_{[2]} \ge ... \ge a_{[n]}$ ve $b_{[1]} \ge b_{[2]} \ge ... \ge b_{[n]}$; $\underline{a} \in R^n$ ve $\underline{b} \in R^n$ vektörlerinin elemanlarının büyüklük sırasına göre dizilmiş hallerini göstermek üzere;

$$\begin{aligned} &1.\ \sum_{i=1}^{n} a_{[i]} = \sum_{i=1}^{n} b_{[i]} \\ &2.\ \sum_{i=1}^{k} a_{[i]} \le \sum_{i=1}^{k} b_{[i]} \qquad k = 1, 2, \ldots, n-1 \end{aligned} \tag{1.3}$$

koşulları sağlanıyorsa $\underline{a}$ ve $\underline{b}$ vektörlerine majorant vektörlerdir denir ve $\underline{a} \prec \underline{b}$ biçiminde gösterilir [2].

$\underline{a}$ ve $\underline{b}$ vektörlerinin birbirleriyle majorant olmaları için gerek ve yeter koşul; $\forall \underline{u} = (u_1, ..., u_n) \in D^+$ için ;

$$\sum_{i=1}^{n} a_i u_i \le \sum_{i=1}^{n} b_i u_i$$

olmasıdır [2].

2. Düzgün Dağılım Fonksiyonlarına Sahip Dağılımlar Ailesi için İnvaryant Güven Aralıkları

Teorem 2.1. $X_1, X_2, ..., X_n ; F(x) \in \mathrm{P} = \left\{ F(x) : F(x) = \dfrac{x-c}{d-c} \quad , \ c < x < d \ , \ F(x) = x \right\}$ dağılımına sahip bir örneklem, X_{n+1}, bu örneklemden bağımsız ve aynı $F(x)$ dağılımına sahip yeni bir rasgele değişken ve $\underline{a} \prec \underline{b}$ olmak üzere;

$$(f_1(X_1, X_2, .., X_n), f_2(X_1, X_2, .., X_n)) = \left(\sum_{i=1}^{n} a_i X_{[i]}, \sum_{i=1}^{n} b_i X_{[i]} \right)$$

rasgele aralığını göz önüne alalım.

$$\begin{aligned} P\{X_{n+1} \in (f_1(X_1, X_2, .., X_n), f_2(X_1, X_2, .., X_n))\} &= P\left\{ X_{n+1} \in \left(\sum_{i=1}^{n} a_i X_{[i]}, \sum_{i=1}^{n} b_i X_{[i]} \right) \right\} \\ &= \frac{1}{n+1} \sum_{i=1}^{n} (n-i+1)(b_i - a_i) \qquad (2.1) \\ &= a_1, ..., a_n ; b_1, ..., b_n) \equiv \alpha \end{aligned}$$

dır. Yani, $\left(\sum_{i=1}^{n} a_i X_{[i]}, \sum_{i=1}^{n} b_i X_{[i]} \right)$ rasgele aralığı P sınıfı için α seviyeli invaryant güven aralığıdır [1].

İspat.

$G=\{(x_1,x_2...,x_n,x_{n+1}) : x_{n+1}\in (f_1(x_1,x_2...,x_n), f_2(x_1,x_2...,x_n)),\ c<x_n<x_{n-1}<...<x_1<d\}$
olmak üzere;

$$P\{X_{n+1}\in (f_1(X_1,X_2,..,X_n), f_2(X_1,X_2,..,X_n))\}=P\{X_1,X_2,..,X_n,X_{n+1}\in G\}$$

$$=\int_c^d\int_c^{x_1}...\int_c^{x_{n-1}}\int_{f_1(x_1,...x_n)}^{f_2(x_1,...x_n)} n!dF(x_{n+1})dF(x_1)...dF(x_n)$$

$$=n!\int_c^d\int_c^{x_1}...\int_c^{x_{n-1}}[F(f_2(x_1,x_2,...,x_n))-F(f_1(x_1,x_2,...,x_n))]dF(x_1)...dF(x_n)$$

$$=n!\int_c^d\int_c^{x_1}...\int_c^{x_{n-1}}\left[F\left(\sum_{i=1}^n b_i X_{[i]}\right)-F\left(\sum_{i=1}^n a_i X_{[i]}\right)\right]dF(x_1)...dF(x_n)$$

$$=n!\int_c^d\int_c^{x_1}...\int_c^{x_{n-1}}\left[\left(\sum_{i=1}^n b_i(\frac{x_i-c}{d-c})\right)-\left(\sum_{i=1}^n a_i(\frac{x_i-c}{d-c})\right)\right](\frac{1}{d-c})^n dx_n...dx_2dx_1$$

$$=\frac{1}{n+1}\sum_{i=1}^n (n-i+1)(b_i-a_i)$$

dır.

2.1. Düzgün Dağılım Fonksiyonlarına Sahip Dağılımlar Ailesi İçin α=0.90 Seviyesinde En Dar İnvaryant Güven Aralığının Elde Edilmesi

Bilindiği gibi büyük bir güven katsayısında güven aralığı ne kadar dar olursa tahmin o ölçüde güvenilir olur. Dolayısıyla α güven katsayısının 1 değerine çok yakın seçilmesi durumunda, $\left(\sum_{i=1}^n a_i X_{[i]}, \sum_{i=1}^n b_i X_{[i]}\right)$ rasgele aralığının olabildiğince dar olması için b_i ve a_i ’ ler arasındaki fark minimum olmalıdır.

$\alpha=0{,}90$ seçilmesi durumunda $\underline{a}$ ve $\underline{b}$ majorant vektörler olmak üzere (1.3) ’ deki koşul altında amaç fonksiyonu:

$$\min\ f=\sum_{i=1}^n |b_i-a_i|$$

ve kısıtları:

$$\begin{cases} \sum_{i=1}^k a_i \le \sum_{i=1}^k b_i \ ,\quad k\le n \\ \sum_{i=1}^n a_i = \sum_{i=1}^n b_i \\ \frac{1}{n+1}\sum_{i=1}^n (n-i+1)(b_i-a_i)=0.90 \\ a_i, b_i \in R \ ,\quad i=1,2,...,n \end{cases} \tag{2.2}$$

şeklinde olan bir optimizasyon problemi karşımıza çıkar. Bu problem tümevarımsal yöntemler kullanılarak aşağıdaki şekilde çözülmüştür.

$n = 2$ için model;

$$\begin{cases} \min \quad f = \sum_{i=1}^{2} \left| b_i - a_i \right| \\ b_1 \geq a_1 \\ b_1 + b_2 = a_1 + a_2 \\ \frac{1}{3} \sum_{i=1}^{2} (b_i - a_i)(3 - i) = 0.90 \\ a_i, b_i \in R \quad , \quad i = 1,2 \end{cases} \tag{2.3}$$

şeklindedir. Bu modelin çözümü Tablo 1' de verilmiştir.

$n = 3$ için model;

$$\begin{cases} \min \quad f = \sum_{i=1}^{3} \left| b_i - a_i \right| \\ b_1 \geq a_1 \\ b_1 + b_2 \geq a_1 + a_2 \\ b_1 + b_2 + b_3 = a_1 + a_2 + a_3 \\ \frac{1}{4} \sum_{i=1}^{3} (b_i - a_i)(4 - i) = 0.90 \\ a_i, b_i \in R \quad , \quad i = 1,2,3 \end{cases} \tag{2.4}$$

biçimindedir. Bu modelin çözümü Tablo 2' de verilmiştir.

$n = 4$ için model;

$$\begin{cases} \min \quad f = \sum_{i=1}^{3} \left| b_i - a_i \right| \\ b_1 \geq a_1 \\ b_1 + b_2 \geq a_1 + a_2 \\ b_1 + b_2 + b_3 \geq a_1 + a_2 + a_3 \\ b_1 + b_2 + b_3 + b_4 = a_1 + a_2 + a_3 + a_4 \\ \frac{1}{5} \sum_{i=1}^{4} (b_i - a_i)(5 - i) = 0.90 \\ a_i, b_i \in R \quad , \quad i = 1,2,3,4 \end{cases} \tag{2.5}$$

biçimindedir. Bu modelin çözümü de Tablo 3' te verilmiştir.

Çeşitli n değerleri verildiğinde (2.2) modelinin çözümü bilgisayar programı yardımıyla bulunmuştur. Sonuçlar Tablo 4 ve Tablo 5' deki gibi elde edilmiştir.

Tablo 1. (2.3) modelinin çözümü

$b_1 - a_1$	$b_2 - a_2$	$\min f$
2.7	-2.7	5.4

Tablo2. (2.4) modelinin çözümü

$b_1 - a_1$	$b_2 - a_2$	$b_3 - a_3$	$\min f$
1.8	0	-1.8	3.6

Tablo 3. (2.5) modelinin çözümü

$b_1 - a_1$	$b_2 - a_2$	$b_3 - a_3$	$b_4 - a_4$	$\min f$
1.5	0	0	-1.5	3

Tablo 4. Çeşitli *n=2,...,10* değerleri için (2.2) modelinin çözümü

n	$b_1 - a_1$	$b_2 - a_2$	$b_3 - a_3$	$b_4 - a_4$	$b_5 - a_5$	$b_6 - a_6$	$b_7 - a_7$	$b_8 - a_8$	$b_9 - a_9$	$b_{10} - a_{10}$	...	$\min f$
2	2.700	-2.700	-	-	-	-	-	-	-	-	-	5.400
3	1.800	0	-1.800	-	-	-	-	-	-	-	-	3.600
4	1.500	0	0	-1.500	-	-	-	-	-	-	-	3.000
5	1.350	0	0	0	-1.350	-	-	-	-	-	-	2.700
6	1.260	0	0	0	0	-1.260	-	-	-	-	-	2.520
7	1.200	0	0	0	0	0	-1.200	-	-	-	-	2.400
8	1.157	0	0	0	0	0	0	-1.157	-	-	-	2.314
9	1.125	0	0	0	0	0	0	0	-1.125	-	-	2.250
10	1.099	0	0	0	0	0	0	0	0	-1.099	-	2.198

Tablo 5. Çeşitli *n=15,20,25,50,100* değerleri için (2.2) modelinin çözümü

n	$b_1 - a_1$	$b_2 - a_2$	...	$b_{15} - a_{15}$	...	$b_{20} - a_{20}$	...	$b_{25} - a_{25}$	...	$b_{50} - a_{50}$	...	$b_{100} - a_{100}$	$\min f$
15	1.029	0	0	-1.029	-	-	-	-	-	-	-	-	2.058
20	0.995	0	0	0	0	-0.995	-	-	-	-	-	-	1.990
25	0.975	0	0	0	0	0	0	-0.975	-	-	-	-	1.950
50	0.937	0	0	0	0	0	0	0	0	-0.937	-	-	1.874
100	0.919	0	0	0	0	0	0	0	0	0	0	-0.919	1.838

3. Sonuç ve Öneriler

$\mathrm{P} = \left\{ F(x){:}F(x) = \dfrac{x-c}{d-c} \quad , \ c < x < d \right\}$ düzgün dağılım fonksiyonlarına sahip dağılımlar sınıfı olmak üzere majorant vektörler yardımıyla $\left(\sum_{i=1}^{n} a_i X_{[i]}, \sum_{i=1}^{n} b_i X_{[i]} \right)$ rasgele aralığının $\alpha = \dfrac{1}{n+1} \sum_{i=1}^{n} (n-i+1)(b_i - a_i)$ seviyeli ana kitleyi kapsayan invaryant güven aralığı olduğu [1] tarafından gösterilmiştir.

Bu güven aralığı kullanılarak yapılacak testlerin gücünün yüksek çıkması için bu aralığın olabildiğince dar olması gerekir. Bu da b_i ve a_i ’ ler arasındaki farkın en küçük olması ile sağlanır. Bu çalışmada $a_1 \geq ... \geq a_n$ ve $b_1 \geq ... \geq b_n$, $\underline{a} \, \pi \, \underline{b}$ majorant vektörler ve α =0.90 olmak üzere;

$$
\begin{cases}
\min \ f = \sum_{i=1}^{n} |b_i - a_i| \\
\sum_{i=1}^{k} a_i \leq \sum_{i=1}^{k} b_i \quad , \quad k \leq n \\
\sum_{i=1}^{n} a_i = \sum_{i=1}^{n} b_i \\
\dfrac{1}{n+1} \sum_{i=1}^{n} (n-i+1)(b_i - a_i) = 0.90 \\
a_i , b_i \in R \quad , \quad i = 1,2,...,n
\end{cases}
$$

şeklinde bir optimizasyon modeli ile karşılaşılmış ve bu modelin çözümü Tablo 4 ve Tablo 5 ’ deki gibi elde edilmiştir. Majorant vektörler arasındaki ilişki bilgisayar programı yardımıyla tespit edilmiştir. Bu ilişkiyi sağlayan her majorant vektörü kesin çözüm olarak kabul edebiliriz. Bu vektörler yardımıyla oluşturulan invaryant güven aralığı bir kitlenin dağılımının düzgün dağılımdan gelip gelmediğini test etmek amacıyla kullanılabilir. Bununla ilgili çalışmalar devam etmektedir.

Kaynaklar

[1] Kaya, M.F., Saraçoğlu, B., Kuş, C. **Düzgün Dağılım Fonksiyonları Ailesi İçin İnvaryant Güven Aralıkları,** İstatistik Araştırma Dergisi, 2, 47-55, Ağustos 2003, DİE, Ankara.

[2] Marshal, A.W., Olkin, I. **Inequalities. Theory of Majorizations and It’s Applications.** Academic Press. - (1979).