İstatistik Araştırma Dergisi Cilt 01 No 02 Özel Sayı Sayfa 55-63 Ağustos 2002

Hata Terimlerinin Genelleştirilmiş Lojistik Dağılıma Sahip Olması Durumunda Parametrelerin Uyarlanmış En Çok Olabilirlik Tahmin Edicilerinin Bulunması

Birdal ŞENOĞLU* Hülya ŞEN**

ÖZET

Hata terimlerinin Genelleştirilmiş Lojistik dağılıma sahip olması durumunda;

$y_i = \mu + e_i \qquad (1 \le i \le n)$

modelindeki parametrelerin tahmin edicileri bulunacak ve bu aşamada izlenecek metodoloji bir uygulama ile desteklenecektir (Şenoğlu, 2000).

Söz konusu metodolojiye dayanan ilerleyen çalışmalarımızda aşağıda detayları verilen metodoloji kullanılarak Dengeli Eksik Bölük Tasarımları, Latin Kareleri ve Faktöriyel Tasarımlar vb. Deney Tasarımında yaygın olarak kullanılan tasarımlardaki parametrelerin tahmin edicileri bulunacak ve elde edilen sonuçlar ayrıca değerlendirilecektir. Sözü edilen metodolojinin burada detaylı olarak incelenmesinin başlıca amacı, ilerideki çalışmalarımızın temelini oluşturmasıdır.

Anahtar Kelimeler: *Uyarlanmış En Çok Olabilirlik tahmin edicileri, Genelleştirilmiş Lojistik Dağılımı, Mimimum Varyans Sınırı, Robust (sağlam) tahmin ediciler.*

1.GİRİŞ

μ ve σ konum ve ölçek parametrelerini göstermek üzere; $f(y|\mu,\sigma)$ dağılımından $y_1, y_2, \ldots\ldots, y_n$ rassal örnekleminin alındığını düşünelim. μ *ve* σ 'nın En Çok Olabilirlik (EÇO) tahmin edicileri, $\partial \ln L/\partial \mu = 0$ ve $\partial \ln L/\partial \sigma = 0$ olabilirlik denklemlerinin çözümünden bulunur. Söz konusu denklemler doğrusal fonksiyon ise, denklemlerin çözümü tektir. Genel olarak, $\partial \ln L/\partial \mu = 0$ ve $\partial \ln L/\partial \sigma = 0$ olabilirlik denklemleri doğrusal olmayan fonksiyon yapısına sahiptirler.

* Yrd.Doç.Dr., Osmangazi Üniversitesi, Fen Edeb Fak., İstatistik Bölümü, Eskişehir
** Yrd.Doç.Dr., Osmangazi Üniversitesi, Fen Edeb Fak., İstatistik Bölümü, Eskişehir

Cebirsel çözümleri bulunamadığından iteratif yöntemlerle çözülürler. Bununla beraber aşağıdaki nedenlerden dolayı sözkonusu denklemlerin çözümünde çeşitli sorunlarla karşılaşılabilir.

(i) Birden fazla kök bulunması

(ii) İterasyonların yakınsamaması

(iii) Yanlış değerlere yakınsaması (Barnett, 1966; Vaughan, 1992).

Yukarıda bahsedilen problemlerden dolayı EÇO tahmin edicilerini kullanmak her zaman uygun olmayabilir.

Bu çalışmada konum ve ölçek parametrelerinin Uyarlanmış En Çok Olabilirlik (UEÇO) tahmin edicilerinin özellikleri (Tiku, 1967, 1968) ve detayları tanımlanarak, Genelleştirilmiş Logistik dağılımının konum ve ölçek parametreleri UEÇO tahmin edicileri kullanılarak bulunacak ve bir uygulama ile desteklenecektir.

Burada aşağıda olasılık yoğunluk fonksiyonu verilen Genelleştirilmiş Lojistik dağılım üzerinde durulacaktır. Çünkü bu dağılım şekil parametresi b'nin aldığı değere göre negatif çarpık (b<1), pozitif çarpık (b>1) veya simetrik (b=1) olabilir. Dolayısıyla uygulamada yaygın olarak karşılaşılabilecek bir dağılımdır.

$$f(y|\mu,\sigma)=\frac{b}{\sigma}\frac{e^{-(y-\mu)/\sigma}}{(1+e^{-(y-\mu)/\sigma})^{b+1}} \quad , -\infty < y < \infty \, , \; -\infty < \mu < \infty \, , \; \sigma > 0 \tag{1}$$

Burada μ ve σ parametrelerinin tahmin edicileri bulunmak isteniyorsa, EÇO tahmin edicileri aşağıdaki denklemlerin çözümünden elde edilir.

$$\frac{\partial \ln L}{\partial \mu}=\frac{n}{\sigma}-\frac{b+1}{\sigma}\sum_{i=1}^{n} e^{-z_i}\left\{1+e^{-z_i}\right\}^{-1}=0 \tag{2}$$

ve

$$\frac{\partial \ln L}{\partial \sigma}=-\frac{n}{\sigma}+\frac{1}{\sigma}\sum_{i=1}^{n} z_i-\frac{b+1}{\sigma}\sum_{i=1}^{n} z_i e^{-z_i}\left\{1+e^{-z_i}\right\}^{-1}=0 \tag{3}$$

(2) ve (1.3) denklemlerinin cebirsel çözümleri aşağıda verilen $g\{z_i\}$ fonksiyonu yüzünden bulunamaz. Bir başka deyişle bu denklemlerin açık bir çözümü yoktur.

$$g\{z_i\}=e^{-z_i}\left\{1+e^{-z_i}\right\}^{-1} \tag{4}$$

2. UYARLANMIŞ EN ÇOK OLABİLİRLİK TAHMİN EDİCİLERİNİN BULUNMASI

UEÇO tahmin edicileri üç aşamada bulunur:

a) (2) ve (1.3) denklemleri sıralı istatistikler cinsinden $\left(z_{(i)} = \frac{y_{(i)} - \mu}{\sigma}, \quad 1 \le i \le n \right)$ aşağıdaki gibi ifade edilir.

$$\frac{\partial \ln L}{\partial \mu} = \frac{n}{\sigma} - \frac{b+1}{\sigma} \sum_{i=1}^{n} e^{-z_{(i)}} \left\{ 1 + e^{-z_{(i)}} \right\}^{-1} = 0 \tag{5}$$

ve

$$\frac{\partial \ln L}{\partial \sigma} = -\frac{n}{\sigma} + \frac{1}{\sigma} \sum_{i=1}^{n} z_{(i)} - \frac{b+1}{\sigma} \sum_{i=1}^{n} z_{(i)} e^{-z_{(i)}} \left\{ 1 + e^{-z_{(i)}} \right\}^{-1} = 0 \tag{6}$$

Çünkü terimler toplamı sıralı istatistiklerin kullanılmasıyla değişmez.

b) Problem yaratan $g\{z_{(i)}\}$ fonksiyonu Taylor Serisinin ilk iki terimi kullanılarak $E(z_{(i)}) = t_{(i)}$ etrafında açılarak doğrusal hale getirilir.

$$g\{z_{(i)}\} \cong g\{t_{(i)}\} + [z_{(i)} - t_{(i)}] \left\{ \frac{dg(z)}{dz} \right\}_{z=t(i)} = \alpha_i - \beta_i z_{(i)}, \quad 1 \le i \le n \tag{7}$$

$$t_i = E(z_{(i)}), \quad \alpha_i = g(t_{(i)}) - t_{(i)}\beta_i, \quad \beta_i = \left\{ \frac{dg(z)}{dz} \right\}_{z=t(i)} \tag{8}$$

Buradan,

$$\alpha_i = \frac{(1 + e^t + te^t)}{(1 + e^t)^2} \quad ve \quad \beta_i = \frac{e^t}{(1 + e^t)^2}$$
$$t = t_{(i)} = -\ln(q_i^{-1/b} - 1), \quad q_i = \frac{i}{n+1} \tag{9}$$

elde edilir.

$t_{(i)}, (1 \le i \le n)$nin değerleri aşağıdaki denklemden bulunur.

$$\int_{-\infty}^{t_{(i)}} \frac{be^{-z}}{(1 + e^{-z})^{b+1}} dz = q_i, \qquad 1 \le i \le n$$

c) (7) denklemi (5) ve (2.2) daki denklemlerde yerine konularak uyarlanmış olabilirlik denklemleri bulunur.

$$\frac{\partial \ln L}{\partial \mu} \cong \frac{\partial \ln L^*}{\partial \mu} = \frac{n}{\sigma} - \frac{b+1}{\sigma}\sum_{i=1}^{n}\{\alpha_i - \beta_i z_{(i)}\} = 0 \tag{10}$$

ve

$$\frac{\partial \ln L}{\partial \sigma} \cong \frac{\partial \ln L^*}{\partial \sigma} = -\frac{n}{\sigma} + \frac{1}{\sigma}\sum_{i=1}^{n} z_{(i)} - \frac{b+1}{\sigma}\sum_{i=1}^{n} z_{(i)}\{\alpha_i - \beta_i z_{(i)}\} = 0 \tag{11}$$

(10) ve (11) denklemlerinin çözümleri UEÇO tahmin edicileri olarak adlandırılır.

$$\hat{\mu}_n = \frac{\sum_{i=1}^{n}\beta_i y_{(i)}}{m}, \quad m = \sum_{i=1}^{n}\beta_i$$

$$A = n, \quad B = (b+1)\sum_{i=1}^{n}\Delta_i(y_{(i)} - \hat{\mu}_n), \quad C = (b+1)\sum_{i=1}^{n}\beta_i(y_{(i)} - \hat{\mu}_n)^2,$$

$$\Delta_i = (b+1)^{-1} - \alpha_i, \quad \Delta = \sum_{i=1}^{n}\Delta_i$$

olmak üzere;

$$\hat{\mu} = \hat{\mu}_n + (\Delta / m)\hat{\sigma} \quad \text{ve} \quad \hat{\sigma} = \frac{B + \sqrt{B^2 + 4nC}}{2\sqrt{n(n-1)}} \quad \text{elde edilir.}$$

σ'nın yansız tahmin edicisini elde etmek için n yerine $n(n-1)$ kullanılmıştır (Şenoğlu, 2000).

Not: $\lim_{n\to\infty} \frac{1}{n}\left|\frac{\partial \ln L}{\partial \mu} - \frac{\partial \ln L^*}{\partial \mu}\right| = 0$ ve $\lim_{n\to\infty} \frac{1}{n}\left|\frac{\partial \ln L}{\partial \sigma} - \frac{\partial \ln L^*}{\partial \sigma}\right| = 0$

olduğundan dolayı UEÇO Tahmin edicileri olan $\hat{\mu}$ ve $\hat{\sigma}$ asimtotik olarak EÇO tahmin edicilerine eşittir.

Teorem : UEÇO tahmin edicisi olan $\hat{\mu}$ ve $\hat{\sigma}$ asimtotik olarak tam etkindir.

Tanım: μ parametresi için Minimum Varyans Sınırı (MVS)

$$MVS(\mu) = \frac{1}{-E\left(\frac{\partial^2 \ln L}{\partial \mu^2}\right)}$$

olarak tanımlanır. Burada $L=\prod_{i=1}^{n} f(y_i,\mu)$ olur. Eğer $\hat{\mu}$, μ parametresinin yansız bir tahmin

edicisi ve $V(\hat{\mu})=MVS(\mu)$ ise $\hat{\mu}$, MVS tahmin edicisi olarak adlandırılır.

Lemma 2.1. : $\hat{\mu}$, μ nün $\dfrac{\sigma^2}{(b+1)m}$ varyansıyla MVS tahmin edicisidir ve normal dağılıma sahiptir.

İspat: Uyarlanmış olabilirlik denklemi asimtotik olarak olabilirlik denklemine eşittir ve

$$\frac{\partial \ln L^*}{\partial \mu}=\frac{(b+1)m}{\sigma^2}\{\hat{\mu}-\mu\}$$

olarak yazılabilir.

$\dfrac{\partial^r \ln L^*}{\partial \mu^r}=0$, $r\geq 3$ olmasından dolayı $\hat{\mu}$ normal dağılıma sahiptir (Bartlett, 1953).

Lemma 2.2. μ biliniyorken $\hat{\sigma}(\mu)$, σ'nın MVS tahmin edicisidir ve $n\hat{\sigma}^2(\mu)/\sigma^2$, n serbestlik derecesi ile ki-kare dağılır.

İspat: $\dfrac{\partial \ln L^*}{\partial \sigma}$ büyük n değerleri için aşağıdaki formda yazılır.

$$\frac{\partial \ln L^*}{\partial \sigma}\cong\frac{n}{\sigma^3}\left\{\left(\frac{C_0}{n}\right)-\sigma^2\right\}$$

Asimtotik olarak $\dfrac{C_0}{n}$, σ^2 nin MVS tahmin edicisidir.

$\dfrac{\partial \ln L^*}{\partial \sigma}$'nın kümülantlarının, $\dfrac{\partial \ln L^*}{\partial \sigma}$'nın türevlerinin beklenen değerleri cinsinden değerlendirilmesi $n\hat{\sigma}^2(\mu)/\sigma^2$ nin dağılımının n serbestlik derecesiyle ki-kare olduğu sonucuna götürür (Bartlett, 1953).

Not: $(n-1)\hat{\sigma}^2/\sigma^2$ nin dağılımı $(n-1)$ serbestlik derecesi ile ki-karedir.

μ ve σ'nın tahmin edicileri olan $\hat{\mu}$ ve $\hat{\sigma}$ varsayılan model altında tam etkin (veya yaklaşık olarak tam etkin) ve varsayılan modelin makul alternatifleri altında yüksek bir etkinliğe sahipse robust (sağlam) tahmin ediciler olarak adlandırılır (Tiku vd., 1986).

UEÇO tahmin edicileri olan $\hat{\mu}$ ve $\hat{\sigma}$ robust (sağlam) tahmin edicilerdir. Bunun sebebi;

$\frac{\partial \ln L^*}{\partial \mu}$ ve $\frac{\partial \ln L^*}{\partial \sigma}$ denklemlerinde $\beta_i\,(1 \le i \le n)$'lerin şemsiye sıralamasına sahip olmasıdır. Başka bir deyişle β_i'lerin pozitif sayıların önce artan sonra da azalan bir dizisi olmasıdır. (Burada dağılım simetrik olduğu için şemsiye sıralaması vardır. Eğer dağılım çarpık olsaydı yarı şemsiye sıralaması olacaktı. Yani β_i'ler çarpıklığın yönüne göre pozitif sayıların azalan veya artan bir dizisi olacaktı).

Örneğin $n = 10$ için aşağıdaki değerleri buluruz.

$\beta_i =$ 0.08264 0.14875 0.19833 0.23139 0.24793 0.24793 0.23139 0.19833 0.14875 0.08264

Buradan da görülebileceği gibi β_i'ler şemsiye sıralamasına sahiptir.

(10) ve (2.7) denklemlerinde aykırı hatalara ve karelerine verilen β_i değerleri kuyruk bölgelerinde küçük olur. Çünkü $z_{(i)} = e_{(i)}/\sigma.\,(1 \le i \le n)$ olur. Bu durum uzun kuyrukların ve aykırı değerlerin etkisini yok eder. Bu tür bir işlem istatistiksel sağlamlığı başarmada çok önemlidir (Huber 1981, Tiku vd. 1986, Tan ve Tiku 1999). (Bununla beraber EKK tahmin edicileri dağılım şekline ve aykırı değer olup olmamasına bakmaksızın bütün gözlemlere aynı ağırlığı verir. Bu da EKK tahmin edicilerinin aykırı değerlerden çok fazla etkilenmesini sağlayarak EKK tahmin edicilerinin robust (sağlam) istatistikler olmasını engeller. EKK tahmin edicileri sadece normal dağılım altında etkin tahmin edicilerdir (İslam vd, 1999).

Uygulama

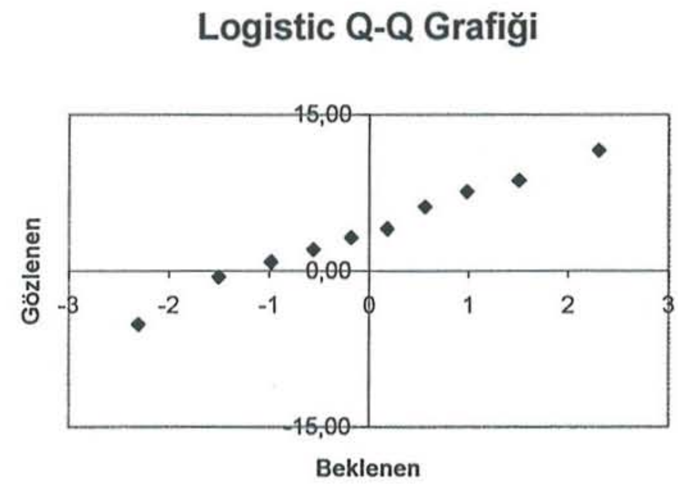
Burada yukarıda bahsedilen metot bir uygulama ile desteklenmiş ve 10,000 Monte Carlo simulasyonu yapılarak UEÇO tahmin edicileri uygulamada yaygın olarak kullanılan EKK tahmin edicileri ile karşılaştırılmıştır ($\mu = 0$, $\sigma = 1$ alınmıştır).

Bu örnekte kullanılan veri seti aşağıda verilmiştir.

y_i: -5.00 -0.61 0.80 2.02 3.10 3.99 6.12 7.60 8.65 11.50

Q-Q grafiğinden de görüleceği gibi Genelleştirilmiş Lojistik dağılımı (b=1) bu veri seti için uygun bir dağılımdır.*

* Genelleştirilmiş Lojistik dağılımının şekil parametresi 1 değerini aldığında Lojistik Dağılım olarak adlandırılır.



Genelleştirilmiş Lojistik dağılımının konum ve ölçek parametrelerinin tahmin değerleri $\hat{\mu} = 3.880$ ve $\hat{\sigma} = 2.894$ olarak bulunur.

$\hat{\mu}$ ve $\hat{\sigma}$ bulunurken kullanılan β_i değerleri aşağıda verilmiştir.

β_i = 0.08264 0.14875 0.19833 0.23139 0.24793 0.24793 0.23139 0.19833 0.14875 0.08264

Genelleştirilmiş Lojistik dağılımına sahip bir veri setinin konum ve ölçek parametrelerinin tahmininde kullanılan UEÇO ve EKK tahmin edicilerinin ortalaması, varyansı ve MSE (Hata Kareler Ortalaması)'leri Monte Carlo simulasyonu kullanılarak farklı b değerleri için hesaplanmıştır. Aşağıdaki tablodan da görüleceği gibi UEÇO tahmin edicileri ($\hat{\mu}$, $\hat{\sigma}$) hem yansızdırlar hem de EKK tahmin edicileri ($\bar{y}$, s)'ne göre çok daha küçük varyansa sahiptirler. Dolayısıyla etkinlikleri EKK tahmin edicileri ($\bar{y}$, s)'den daha fazladır.

$\bar{y}$, s, $\hat{\mu}$, $\hat{\sigma}$ 'nın Ortalama, Varyans ve MSE (Hata Kareler Ortalaması)'leri (b=0.5, 1, 4).

		Ortalama				Varyans				MSE			
b	n	$\bar{y}$	s	$\hat{\mu}$	$\hat{\sigma}$	$\bar{y}$	s	$\hat{\mu}$	$\hat{\sigma}$	$\bar{y}$	s	$\hat{\mu}$	$\hat{\sigma}$
0.5	10	-1.3822	2.4492	-0.0299	1.0352	0.659	0.606	0.539	0.090	2.569	6.604	0.540	1.162
	20	-1.3832	2.5028	-0.0091	1.0177	0.323	0.319	0.263	0.041	2.237	6.583	0.263	1.076
1	10	-0.0047	1.7471	-0.0048	1.0348	0.330	0.248	0.309	0.084	0.330	3.300	0.309	1.154
	20	0.0024	1.7725	0.0014	1.0153	0.166	0.124	0.153	0.038	0.166	3.266	0.153	1.068
4	10	1.8401	1.3328	0.0707	1.0036	0.195	0.157	0.238	0.072	3.581	1.934	0.243	1.079
	20	1.8366	1.3618	0.0303	1.0049	0.097	0.086	0.114	0.035	3.470	1.941	0.115	1.044

3. SONUÇ

UEÇO tahmin edicileri normal dağılım varsayımının sağlanmadığı durumlarda regresyon, zaman serileri, deney tasarımı gibi alanlarda yaygın olarak kullanılmaktadır.

Çünkü bu tahmin ediciler asimtotik olarak tam etkin ve robust tahmin edicilerdir ve bu tahmin edicilere dayanan testlerin gücü daha fazladır.

KAYNAKLAR

BARNETT, V.D. (1966), *Evaluation of the maximum likelihood estimator where the likelihood equation has multiple roots.* Biometrika 53, 153-165.

BARTLETT, M.S. (1953). *Approximate confidence intervals. Biometrika* 40, 12-19.

HUBER, P.J. (1981). *Robust Statistics.* New York:John Wiley.

ISLAM, M.Q, TİKU, M.L. and YILDIRIM, F. (1999). *Nonnormal regression with skew distributions. Commun. Stat.-Theory Meth.* (submitted for publication).

ŞENOĞLU, B. (2000). *Experimental design under nonnormality. Ph.D Thesis, METU.*

TAN, W.Y. and TİKU, M.L. (1999). *Sampling Distibutions In Terms of Laguerre Polynomials and Aplications.* New Delhi: John Wiley Eastern.

TİKU. M.L. (1967). *Estimating the mean and standart deviation from censored normal samples. Biometrika* 54, 155-165.

TİKU. M.L. (1968). *Estimating the parameters of log-normal distribution from censored samples. J. Amer. Stat Assoc.* 63, 134-140.

TİKU. M.L., TAN, W.Y., and BALAKRİSHNAN, N. (1986). *Robust Inference.* New York: Marcel Dekker.

VAUGHAN, D.C. (1992). *On the Tiku-Suresh method of estimation. Commun. Stat.-Theory Meth.* 21, 391-404.

Derivation Of Modified Maximum Likelihood Estimators Of Parameters When The Error Distribution Has Generalized Logistic Distribution

ABSTRACT

The estimators of parameters in the model

$$y_i = \mu + e_i \qquad (1 \le i \le n)$$

will be derived when the error distribution has Generalized Logistic distribution and the methodology explained in this study will be supported with an example (Şenoğlu, 2000).

In our future studies, the estimators of the parameters in the models used commonly in statistical design of experiments such as Balanced Incomplete Block Design, Latin Square Design and

Factorial Design etc. will be derived by using the methodology given in the following sections. This methodology is the basic of our future studies that is why we investigate it in detail.

Key Words: *Modified Maximum Likelihood Estimators, Generalized Logistic Distribution, Minimum Variance Bound, Robust Estimators.*