İstatistik Araştırma Dergisi Cilt 01 No 02 Özel Sayı Sayfa 95-101 Ağustos2002

Kesilmiş Veri İle Tehlike Fonksiyonlarında Değişim Noktası Tahmini

Ülkü GÜRLER* Deniz YENIGÜN**

ÖZET

Bu çalışmada rastgele kesilmiş veriler kullanılarak tek bir değişim noktası olan iki parçalı sabit tehlike fonksiyonlarındaki değişimin yerleşimi ve büyüklüğü tahmin edilecektir. Tahminler soldan kesilmiş veriler için yapılan bir simulasyon calışmasıyla özetlenecektir.

Anahtar Kelimeler: *Tehlike fonksiyonu, kesme, değişim noktası modeli, yaşam analizi.*

1. GİRİŞ

Tehlike fonksiyonlarının güvenilirlik ve yaşam analizi çalışmalarında önemli bir rolü vardır. Bazı uygulamalarda bakım çalışmalarına, önemli işlemlere ya da yeniden gözden geçirmeye bağlı olarak tehlike fonksiyonlarında ani değişiklikler gözlenebilir. Bu tür durumlarda değişikliğin büyüklüğünü ve meydana geldiği anı tahmin etmek önemlidir. Bu konuyu ilk olarak Matthews ve Farewell (1982) incelemiş ve sıfır hipotezinin değişim noktası içermeyen sabit tehlike fonksiyonu olduğu durumda olabilirlik oran testi türetmişlerdir. Worsley (1988) olabilirlik oranı istatistiğinin kesin dağılımını türetmiştir. Loader (1991) en çok olabilirlik prensibine dayanan tahmin yöntemi önermiş ve değişim noktasının yeri ve büyüklüğü için güven aralıkları oluşturmuştur. Loader (1996) parametrik olmayan regresyon kullanarak bu konuda çalışmış, Antoniadis, Gijbels ve MacGibbon (1998) değişim noktası modelini parametrik olmayan yöntemler kullanarak ele almışlardır.

Yaşam analizi çalışmalarında kesilmiş veri ile karşılaşılabilir. X ilgi duyulan ve yaşam süresini temsil eden restgele değişken olsun. X'in gözlenmesi bir başka bağımsız rastgele değişken olan Y tarafından engellenebilir. Örneğin soldan kesilme modellerinde gözlemci (X,Y) ikililerini sadece $Y \leq X$ iken gözleyebilir. Kesilmiş veriler araştırmacıların ilgisini daha yakın bir geçmişte, bir ölçüde AIDS verilerine uygulanabilirlikleri açısından çekmişlerdir. Woodroofe (1985) X ve Y'nin dağılımı için parametrik olmayan en çok olabilirlik tahminleri önermiştir. Wang (1989) kesme mekanizmasını parametrik olarak düşünmüş ve yarı parametrik model için en çok olabilirlik tahminleri önermiştir. Kalbfleisch ve Lawless (1991) kesilmiş veri için regresyon modelleri üzerine çalışmıştır. Wanderlaan (1996), Grigoletto ve Akritas (1999) ve diğer birçok araştırmacı yakın geçmişte bu konuyu ele almış ve genellikle parametrik olmayan ve yarı parametrik yöntemler üzerine çalışmışlardır.

* Bilkent Üniversitesi Endüstri Mühendisliği Bölümü
** ODTÜ İstatistik Bölümü

Bu çalışmada, değişkenlerin kesilmiş olması durumunda tehlike fonksiyonundaki değişim noktasının tahmini düşünülmüştür. Tek bir değişim noktası olan iki parçalı sabit tehlike fonksiyonları için en çok olabilirlik ilkesine dayanan bir tahmin yöntemi önerilmiştir.

2. MODEL

X ve Y, bilinmeyen F ve G dağılımlarına sahip, bağımsız pozitif restgele değişkenler olsun. Burada X'in ilgi duyduğumuz değişken, Y'nin de X'i soldan kesen değişken olduğunu düşünüyoruz. X değişkenine ait tek bir değişim noktası olan iki parçalı sabit tehlike fonksiyonu aşağıdaki gibi modellenebilir:

$$\lambda(t)=\begin{cases}\beta, & 0<t<\tau\\ \beta+\theta, & \tau<t\end{cases} \tag{1}$$

Burada τ, değişim noktası; β, değişimden önceki tehlike; θ ise değişimin miktarıdır.

X'e ait olasılık yoğunluk fonksiyonu (2) ilişkisi kullanılarak bulunabilir:

$$f(x)=\lambda(x)\exp\left\{-\int_0^x \lambda(t)dt\right\}, \tag{2}$$

$$f(x)=\begin{cases}0, & x<0\\ \beta\exp\{-\beta x\}, & 0\le x\le\tau,\ f_1(x)\\ (\beta+\theta)\exp\{-\beta x-\theta(x-\tau)\ \}, & x>\tau,\ f_2(x)\end{cases} \tag{3}$$

H^*, gözlenen (X,Y) ikililerinin bileşik dağılım fonksiyonu olsun. Bu durumda,

$$H^*(x,y)=P(X\le x, Y\le y|Y\le X)=\alpha^{-1}\int_0^x G(u\wedge y)dF(u) \tag{4}$$

$$,\ \alpha=P(Y\le X)=\int_0^\infty G(x)dF(x)$$

ve (X,Y) ikililerinin bileşik olasılık yoğunluk fonksiyonu h^*,

$$h^*(x,y)=\begin{cases}\alpha^{-1}f(x)g(y), & y\le x\\ 0, & \textit{öteki durumlarda}\end{cases} \tag{5}$$

olarak tanımlanır ve En çok olabilirlik fonksiyonu aşağıdaki gibidir:

$$L=\prod_{i=1}^{n} h^{*}(x_i, y_i) \tag{6}$$

İlgi duyduğumuz X rastgele değişkenini soldan kesen Y rastgele değişkeninin r parametresiyle üstel dağılıma sahip olduğunu düşünelim. Bu çalışmada β, θ, r ve τ parametrelerinin tümünü bilmediğimizi kabul ederek, en çok olabilirlik ilkesine dayanan bir yöntemlele bu parametrelerin tahmin edilmesini inceleyeceğiz.

En çok olabilirlik fonksiyonu $L(\beta,\theta,r,\tau)$ ve kesme oranı $\alpha = P(Y \le X)$ aşağıdaki gibidir:

$$L(\beta,\theta,r,\tau)=\prod_{A} f_1(x_i)\prod_{B} f_2(x_i)\prod_{i=1}^{n} g(y_i)\alpha^{-n}, \quad A=\{x_i : x_i<\tau\} \text{ve } B=\{x_i : x_i>\tau\}$$

$$L(\beta,\theta,r,\tau)=\beta^{n_\tau} e^{-\beta\sum_A x_i} (\beta+\theta)^{n-n_\tau} e^{\tau\theta(n-n_\tau)} e^{-(\beta+\theta)\sum_B x_i} r^n e^{-r\sum_{i=1}^{n} y_i} \alpha^{-n}, \tag{7}$$

$$\beta>0,\ \beta+\theta>0,\ r>0,$$

$$\alpha = P(Y \le X) = \frac{r\left[(\beta+\theta+r)-\theta e^{-\tau(\beta+r)}\right]}{(\beta+r)(\beta+\theta+r)}, \tag{8}$$

3. TAHMİN YÖNTEMİ

Parametre uzayı $\Psi=\{\beta,\theta,r,\tau\}$ iken, sabit τ için eksiltilmiş parametre uzayı $\psi_\tau=\{\beta,\theta,r\}$ olarak düşünülebilir. Bu durumda derece vektörü aşağıdaki gibidir:

$$U(\Psi_\tau)=\begin{bmatrix} \dfrac{\partial(\ln L)}{\partial\beta} \\ \dfrac{\partial(\ln L)}{\partial\theta} \\ \dfrac{\partial(\ln L)}{\partial r} \end{bmatrix}, L=L(\beta,\theta,r,\tau)\ , \tag{9}$$

Sabit τ için, $U(\Psi_\tau)=0$ çözümü, Ψ_τ'nin tahmin edicisi $\hat{\Psi}_\tau=(\hat{\beta},\hat{\theta},\hat{r})_\tau$'yi verir.

Bu çalışmada düşünülen tahmin yöntemi en çok olabilirlik fonksiyonu (7)'yi en büyükleyen τ değerini bulmak üzerine kurulmuştur. $[\tau_0,\tau_1]$ aralığında çok sayıda noktada τ sabitlenerek öteki parametreler için en çok olabilirlik tahminleri bulunur. $\tau_i \in [\tau_0, \tau_1]$ iken sabit τ_i değerinin ve bu değere göre β, θ ve r için elde edilen tahminlerin (7) de yerine konulmasını (10) ile ifade edelim:

$$L_{\tau_i} = L((\hat{\beta}, \hat{\theta}, \hat{r})_{\tau_i}, \tau_i) \tag{10}$$

$[\tau_0, \tau_1]$ aralığında (10)' yi en büyükleyen τ_i değeri, τ 'nın bir tahmin edicisi,

$$\hat{\tau} = \arg\max_{\tau_i} L_{\tau_i}$$

ve $\hat{\tau}$ ile hesaplanmış $(\hat{\beta}, \hat{\theta}, \hat{r})_{\hat{\tau}}$, öteki parametrelerin tahminleridir:

$$\hat{\Psi} = \{(\hat{\beta}, \hat{\theta}, \hat{r})_{\hat{\tau}}, \hat{\tau}\}$$

4. TAHMİNLERİN ASİMPTOTİK VARYANSLARI

Tahmin sürecinin bir devamı olarak, yine sabit bir τ değeri için bilgi matrisi hesaplanabilir ve buradan β, θ ve r parametreleri için varyans tahminleri yapılıp güven aralıkları oluşturulabilir. Bilgi matrisi oldukça karmaşık ve büyük olduğu için burada sunulmayacaktır, ancak bir sonraki bölümde simülasyon çalışmalarına ek olarak yaratılan tek bir rastgele örneklem için tahminler yapıldıktan sonra β, θ ve r parametreleri için güven aralıkları oluşturulacaktır. Yaratılan rastgele örneklemlerden hesaplanan asimptotik varyansların, aynı koşullarda yapılan simulasyon çalışmasında hesaplanan varyanslarla tutarlı olduğu gözlenmiştir.

5. SİMÜLASYON ÇALIŞMASI

Yukarıdaki çalışmanın başarısını özetleyen bir simülasyon çalışması yapılmıştır. Dört ayrı örneklem büyüklüğü ve dört ayrı kesme oranı kullanılarak çeşitli parametrelere sahip değişim noktası durumları incelenmiştir. Bu durumlardan bir tanesi burada sunulacaktır. Kesme oranı α, X rastgele değişkenini soldan kesen rastgele değişken Y'nin parametresi değiştirilerek ayarlanmıştır. Ortalamalar, varyanslar, yanlar (bias) ve ortalama karesel hatalar (mean square error) (MSE) hesaplanmıştır. Herhangi bir kesilmenin olmadığı değişim noktası modeli de çalışılmış ve sonuçlar simülasyon çalışmasına eklenmiştir.

(1) modelinde $\beta = 1$, $\theta = 5$ ve $\tau = 0.5$ iken rastgele örneklemler yaratılmış ve simülasyon çalışması Tablo 1'de sunulmuşur. Değişim noktasının yeri τ, [0.3,0.7] aralığında aranmış ve bu örneklemde tahminlerin oldukça başarılı olduğu gözlenmiştir. Bunda $\theta = 5$' in büyük bir değişiklik olmasının etkisi vardır. Örneklem büyüklüğü n arttıkça tahminlerin başarısında artış gözlenmiştir. Kesme oranı α arttıkça, yani örneklemlerde kesilme sonucu oluşan hata azaldıkça tahminlerde genel bir iyileşme söz konusudur.

α	n	$\beta=1$ $\hat{\beta}$	yan	varyans	MSE	$\theta=5.0$ $\hat{\theta}$	yan	varyans	MSE	$r=1.14$ $\hat{r}$	yan	varyans	MSE	$\tau=0.5$ $\hat{\tau}$	yan	varyans	MSE
0.4	100	0.9932	0.0068	0.0471	0.0471	5.0950	-0.0950	0.4655	0.4745	1.2320	-0.0920	0.2642	0.2726	0.5007	-7.669E-04	1.240E-04	1.246E-04
	300	1.0029	-0.0029	0.0169	0.0169	4.9804	0.0196	0.1559	0.1563	1.1589	-0.0189	0.0979	0.0982	0.4989	1.071E-03	2.319E-05	2.434E-05
	500	1.0054	-0.0054	0.0098	0.0098	4.9685	0.0315	0.0833	0.0843	1.1376	0.0024	0.0597	0.0597	0.4988	1.158E-03	1.398E-05	1.532E-05
	800	1.0040	-0.0040	0.0058	0.0058	4.9518	0.0482	0.0555	0.0578	1.1356	0.0044	0.0381	0.0381	0.4984	1.577E-03	1.168E-05	1.417E-05

α	n	$\beta=1$ $\hat{\beta}$	yan	varyans	MSE	$\theta=5.0$ $\hat{\theta}$	yan	varyans	MSE	$r=2.2$ $\hat{r}$	yan	varyans	MSE	$\tau=0.5$ $\hat{\tau}$	yan	varyans	MSE
0.6	100	1.0020	-0.0020	0.0422	0.0422	5.0866	-0.0866	0.5082	0.5157	2.2179	-0.0179	0.3131	0.3134	0.5002	-2.843E-04	1.023E-04	1.024E-04
	300	1.0046	-0.0046	0.0144	0.0144	4.9693	0.0307	0.1505	0.1515	2.2189	-0.0189	0.1024	0.1028	0.4989	1.088E-03	2.005E-05	2.124E-05
	500	1.0075	-0.0075	0.0087	0.0088	4.9543	0.0457	0.0919	0.0940	2.2063	-0.0063	0.0644	0.0644	0.4987	1.228E-03	1.418E-05	1.569E-05
	800	1.0062	-0.0062	0.0059	0.0059	4.9469	0.0531	0.0576	0.0604	2.2024	-0.0024	0.0404	0.0404	0.4985	1.417E-03	1.161E-05	1.362E-05

α	n	$\beta=1$ $\hat{\beta}$	yan	varyans	MSE	$\theta=5.0$ $\hat{\theta}$	yan	varyans	MSE	$r=4.65$ $\hat{r}$	yan	varyans	MSE	$\tau=0.5$ $\hat{\tau}$	yan	varyans	MSE
0.8	100	1.0012	-0.0012	0.0368	0.0368	5.0819	-0.0819	0.5913	0.5980	4.6961	-0.0461	0.5453	0.5474	0.5001	-1.019E-04	1.290E-04	1.290E-04
	300	1.0074	-0.0074	0.0126	0.0126	4.9683	0.0317	0.1900	0.1910	4.6489	0.0011	0.1745	0.1745	0.4990	9.342E-04	2.021E-05	2.108E-05
	500	1.0040	-0.0040	0.0074	0.0074	4.9455	0.0545	0.1036	0.1066	4.6559	-0.0059	0.0974	0.0975	0.4987	1.235E-03	1.319E-05	1.471E-05
	800	1.0040	-0.0040	0.0046	0.0047	4.9468	0.0532	0.0707	0.0736	4.6493	0.0007	0.0646	0.0646	0.4985	1.472E-03	1.151E-05	1.368E-05

Kesme Yok

n	$\beta=1$ $\hat{\beta}$	yan	varyans	MSE	$\theta=5.0$ $\hat{\theta}$	yan	varyans	MSE		$\tau=0.5$ $\hat{\tau}$	yan	varyans	MSE
100	1.00876	-0.00876	0.02707	0.02714	5.12738	-0.12738	0.64563	0.66186		0.50166	-0.00166	0.00018	0.00018
300	1.00322	-0.00322	0.00895	0.00896	4.97460	0.02540	0.21360	0.21425		0.49863	0.00137	0.00002	0.00002
500	1.00315	-0.00315	0.00500	0.00501	4.93405	0.06595	0.11734	0.12169		0.49851	0.00149	0.00001	0.00002
800	1.00388	-0.00388	0.00319	0.00321	4.92717	0.07283	0.07373	0.07903		0.49840	0.00160	0.00001	0.00001

$\beta = 1$, $\theta = 5$, r = 2.2 ve τ =0.5 durumunda yaratılan 800 büyüklüğünde tek bir örneklem kullanarak yapılan tahminler ve güven aralıkları aşağıdaki gibidir:

$\hat{\beta} = 1.01952$, $\hat{\theta} = 4.90936$, $\hat{r} = 1.95048$, $\hat{\tau} = 0.49999$,

$V(\hat{\beta}) = 0.00839$, $V(\hat{\theta}) = 0.04336$, $V(\hat{r}) = 0.05872$:

95% güven aralıkları , β için, (0.99595, 1.04909) ;θ için, (4.89720, 4.92153); r için ,(1.78548, 2.11549) olur.

Burada tahmin edilen varyanslar, simülasyon çalışmasıyla elde edilen varyanslarla tutarlı görünmektedir.

KAYNAKLAR

ANTONIADIS, A., GIJBELS, I. and MACGIBBON, B. (1998). *Technical Report,* University of Quebec at Montreal.

GRIGOLETTO, M. and AKRITAS, M.G. (1999). *Analysis of covariance with incomplete data via semiparametric model transformations. Biometrics,* **55,** 1177-1187.

KALBFLEISCH, J. D. and LAWLESS, J. F. (1991). *Regression Models for right truncated data with applications to AIDS incubation times and reporting lags.* Statistics Sinica, **1**, 19-32.

LOADER, C.R. (1991). *Inference for hazard rate change-point.* Biometrika, **78**, 835-843

LOADER, C.R. (1996). *Change Point Estimation Using Nonparametric Regression.* The Annals of Statistics, 24, No. 4, 1667-1678

MATTHEWS, D.E. and Farewell, V.T. (1982). *On testing for a constant hazard against a change-point alternative.* Biometrics, **38**, 463-468

WANDERLAAN, M. J. (1996). *Nonparametric estimation of the bivariate survival function with truncated data.* Journal of Multivariate Analysis. **58**, 107-131.

WANG, M.C.(1989). *A Semiparametric Model for Randomly Truncated Data. JASA,* **84**, 742-748

WOODROOFE, M. (1985). *Estimating a distribution function with truncated data.* Annals of Statistics, **13,** 163-177.

Hazard Change – Point Estimation With Truncadet Data

ABSTRACT

In this study we consider hazard models with a single change – point when the observations are subject to random truncation.

For a piecewise constant hazard function with a single change-point, we consider an estimation procedure based on the maximum likelihood ideas. The performance of the proposed estimators is illustrated by simulation results.

Key Words : *Hazard Function, Truncation, Change-point model, Life analysis*