

ÇOK DEĞİŞKENLİ OLASILIK YOĞUNLUK FONKSİYONUNUN ÇEKİRDEK KESTİRİMİNİN TÜRKİYE'DEKİ DEPREM VERİLERİNE UYGULANMASI

Serpil CULA*

*Başkent Üniversitesi, Uygulamalı Bilimler Yüksekokulu, Ankara, TÜRKİYE,
scula@baskent.edu.tr*

Öniz TOKTAMIŞ, Aytekin DÖNMEZ

Hacettepe Üniversitesi, Fen Fakültesi, İstatistik Bölümü, Ankara, TÜRKİYE

ÖZET

Bu çalışmada, 1900-2001 yılları arasında tüm Türkiye'de gerçekleşen büyüklüğü Richter ölçeğine göre 5 ve üzeri olan enlem ve boylamı verilen deprem verileri için çok değişkenli olasılık yoğunluk fonksiyonunun çekirdek kestirimi elde edilmiştir.

Anahtar Kelimeler: Olasılık yoğunluk fonksiyonu kestirimi, bant genişliği seçimi, çapraz geçerlilik.

THE APPLICATION OF THE MULTIVARIATE KERNEL DENSITY ESTIMATION IN EARTHQUAKES DATA IN TURKEY

ABSTRACT

In this study, it was obtained kernel estimation of probability density function for the earthquakes data with magnitude at least 5 on Richter scale occurring between 1990-2001 years in Turkey.

Key Words: Density estimation of probability density function, bandwidth selection, cross-validation

1. GİRİŞ

Bu çalışmada, çok değişkenli olasılık yoğunluk fonksiyonunun çekirdek kestirimi tanıtılmış ve her bir yönde eşit ve farklı düzleştirme miktarları olması durumları için, ayrı ayrı 1900-2001 döneminde Türkiye'de 32,00-44,00 Enlem ve 24,00-46,00 Boylam koordinatlarında, Richter ölçeğine göre büyüklüğü 5 ve üzeri olan deprem verileri üzerinde bir uygulama yapılmıştır.

1.1. Tek Değişkenli Olasılık Yoğunluk Fonksiyonunun Çekirdek Kestirim Yöntemi

$X_1, X_2, \dots, X_n$, sürekli, tek değişkenli bir dağılımdan alınan bağımsız özdeş dağılımlı gözlemlerin bir örnekleme olmak üzere, olasılık yoğunluk fonksiyonunun çekirdek kestiricisi,

$$\hat{f}(x, h) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right) \quad [1]$$

biçimindedir. Burada K , çekirdek fonksiyonu; h ise pencere genişliği, düzleştirme parametresi ya da bant genişliği olarak adlandırılır. K çekirdek fonksiyonu, h bant genişliği ile birlikte kestirimin ağırlıklarını belirlemektedir. Uygulayıcı tarafından seçilen K çekirdek fonksiyonu, genellikle sıfır etrafında simetrik tek tepeli olasılık yoğunluk fonksiyonlarından herhangi biri olarak alınabilir. Gözlemler merkez olmak üzere yerleştirilen çekirdek fonksiyonuna ve bant genişliğine göre elde edilen ağırlıklardan yararlanılarak bulunan, ağırlıklı ortalama olarak hesaplanan çekirdek kestirimlerinde, K çekirdek fonksiyonlarının kestirim üzerinde büyük değişiklik yaratmadığı ancak bant genişliği seçiminin çekirdek kestiriminin performansını önemli ölçüde etkilediği bilinmektedir. Çekirdek kestiriminin ana problemi bant genişliğinin seçimidir. h bant genişliğini elde etmek için birçok yöntem önerilmiştir. Genel olarak benimsenen bir yöntem halen mevcut olmayıp, önerilen yöntemlerden bazıları; subjektif seçim, standart bir dağılım kullanarak seçim, en küçük kareler çapraz geçerlilik yöntemi (ÇG), yanlış çapraz geçerlilik yöntemi (YÇG), bootstrap yöntemi (B), plug-in yöntemi, düzleştirilmiş çapraz geçerlilik yöntemidir (5, 11).

1.2. Çok Değişkenli Olasılık Yoğunluk Fonksiyonunun Çekirdek Kestirimi

Çok değişkenli çekirdek kestirimleri, tek değişkenli durumun bir genellemesidir. $X_1 = (X_{11}, X_{12}, \dots, X_{1d})'$, $x \in \mathcal{R}$ olmak üzere, $X_1, X_2, \dots, X_n$, d değişkenli f dağılımından alınan rasgele örneklem ve d boyutlu genel vektör $x = (x_1, x_2, \dots, x_n)'$ olsun. Bu gösterim altında d boyutlu olasılık yoğunluk fonksiyonunun çekirdek kestiricisi aşağıda verilmiştir:

$$\hat{f}(x; H) = \frac{1}{n|H|^{1/2}} \sum_{i=1}^n K\left(\frac{x - X_i}{H^{1/2}}\right) \quad [2]$$

Burada H , $d \times d$ boyutlu simetrik pozitif tanımlı bant genişliği matrisi; K ise, $\int_{-\infty}^{\infty} K(x) dx = 1$ koşulunu sağlayan d boyutlu çekirdek fonksiyonudur. $\int$ işareti, d katlı $\int \dots \int_{\mathcal{R}^d}$ integralinin, $dx, dx_1, dx_2, \dots, dx_d$ diferansiyeller çarpımının kısaltılmış biçimidir. K çekirdek fonksiyonu genellikle çok değişkenli olasılık yoğunluk fonksiyonu olarak seçilir. K çekirdek fonksiyonu için yaygın bir seçim, simetrik tek tepeli olan çok değişkenli standart normal olasılık yoğunluk fonksiyonudur. Ayrıca simetrik tek değişkenli çekirdek fonksiyonları kullanılarak çok değişkenli çekirdek fonksiyonu türetilir. Bunlardan biri $K(x) = \prod_{i=1}^d K(x_i)$ olan çarpım çekirdek, diğeri $K(x) = \pi \{(\mathbf{x}'\mathbf{x})^{1/2}\} / \int \pi \{(\mathbf{x}'\mathbf{x})^{1/2}\} d\mathbf{x}$ olan küresel çekirdek fonksiyonlarıdır.

(2) eşitliği ile verilen çekirdek kestiricisi, $\frac{d(d+1)}{2}$ farklı öğeye sahip bant genişliği matrisi H 'nın belirlenmesini gerektirir. Boyut sayısı arttıkça H matrisi ile yapılan işlemler kontrol edilemez olur. Çoğu kez boyut sayısı $d \geq 5$ için uygulanmaz (4). Bu durumda H 'nın daha basit bir biçim alması için H matrisi üzerinde bazı kısıtlar önerilir. H bant genişliği matrisi için en basit kısıt $H \in S$ almaktır ki bu $H = h^2 I (h > 0)$ seçimine karşılık gelmektedir.

I , $d \times d$ boyutlu birim matrisdir. Sabit tek bir h bant genişliği alınması durumu, her bir yöndeki düzleştirme miktarının aynı olacağını ifade eder. Bu durum tüm değişkenlerin ölçekleri yaklaşık olarak aynı ise uygundur. Ancak her bir değişken ortak bir ölçeğe göre standartlaştırılmış ise uygulanabilir (12). $H \in S$ durumu için bir çalışma Cula ve Toktamış tarafından yapılmıştır (2). Diğer bir durum, D , $d \times d$ boyutlu pozitif tanımlı köşegen matrisler sınıfı olmak üzere, bant genişliği matrisi $H \in D$, başka bir ifadeyle $H = \text{ks}g(h_1^2, h_2^2, \dots, h_d^2)$ alınmasıdır. Bu yaklaşım her bir d koordinat yönünde farklı miktarlarda düzleştirme esnekliğine sahiptir. Bu çalışmada da amaç, bu durum göz önüne alınarak olasılık yoğunluk fonksiyonunun çekirdek kestirim değerlerini elde etmektir. Bu durum altında, çok değişkenli olasılık yoğunluk fonksiyonunun çekirdek kestiricisi aşağıdaki biçimde yazılabilir (3):

$$\hat{f}(\mathbf{x}; \mathbf{h}) = \frac{1}{n} \left(\prod_{i=1}^d h_i \right)^{-1} \sum_{i=1}^n K \left(\frac{x_1 - X_{i1}}{h_1}, \frac{x_2 - X_{i2}}{h_2}, \dots, \frac{x_d - X_{id}}{h_d} \right)$$

Koordinat eksenlerinin yönlerinden farklı yönlerde düzleştirme istendiğinde ise tam (full) bir bant genişliği matrisi, başka bir ifadeyle, H uygun olacaktır. Burada F pozitif tanımlı, d boyutlu simetrik matrisler sınıfıdır. H durumunda çok fazla parametre söz konusu olacağı için işlemler zorlaşmaktadır (10).

Optimal bant genişliğini elde etmek için, hata kareler ortalaması, toplanmış hata kareler ortalaması, asimtotik toplanmış hata kareler ortalaması, Hellinger uzaklık ölçüsü gibi hata ölçütleriyle çalışılmıştır. Rosenblatt, olasılık yoğunluk fonksiyonunun çekirdek kestiricisinin performans ölçütü olarak, matematiksel kolaylığı nedeniyle toplanmış hata kareler ortalamasının tercih edildiğini belirtmiştir (6).

Çok değişkenli olasılık yoğunluk fonksiyonunun çekirdek kestiricisine ilişkin toplanmış hata kareler ortalamasının yaklaşık bir ifadesini elde etmek için f olasılık yoğunluk fonksiyonu, H bant genişliği matrisi ve K çekirdek fonksiyonu ile ilgili aşağıdaki varsayımlar yapılmıştır:

i) $H_f(\mathbf{x})$, (i,j) ögesi $\frac{\partial^2}{\partial x_i \partial x_j} f(\mathbf{x})$ olan f 'nin $d \times d$ boyutlu Hessian matrisi olsun.

$H_f(\cdot)$ 'nin her bir ögesi parçalı sürekli ve karesi integrallenebilir olsun;

ii) $H = H_n$, bant genişlikleri matrisinin bir dizisi olmak üzere $n \rightarrow \infty$ iken $n^{-1} |H|^{-1/2}$ ve

H bant genişliği matrisinin tüm öğeleri sıfıra yaklaşsın;

iii) K çekirdek fonksiyonu sınırlı d değişkenli çekirdek fonksiyonu olup,

$$\int_{-\infty}^{\infty} K(\mathbf{z}) d\mathbf{z} = I, \quad \int_{-\infty}^{\infty} \mathbf{z} K(\mathbf{z}) d\mathbf{z} = 0, \quad \int_{-\infty}^{\infty} \mathbf{z} \mathbf{z}' K(\mathbf{z}) d\mathbf{z} = \mu_2(K) I$$

özelliklerini sağlasın; burada, $\mu_2(K) = \int_{-\infty}^{\infty} \mathbf{z}_1^2 K(\mathbf{z}) d\mathbf{z}$ 'dir.

Yukarıda verilen varsayımlar altında, çok değişkenli olasılık yoğunluk fonksiyonunun çekirdek kestiricisine ilişkin asimtotik toplanmış hata kareler ortalaması (ATHKO) aşağıdadır:

$$\text{ATHKO} \{ \hat{f}(\mathbf{x}; H) \} = \frac{1}{n} |H|^{-1/2} \int_{-\infty}^{\infty} K(\mathbf{z})^2 d\mathbf{z} + \frac{1}{4} \mu_2(K)^2 \int_{-\infty}^{\infty} i z^2 \{ H H_f(\mathbf{x}) \} d\mathbf{x} \quad [3]$$

Çok değişkenli olasılık yoğunluk fonksiyonu çekirdek kestiricisine ilişkin asimtotik toplanmış hata

kareler ortalamasından optimal bant genişliği matrisini elde etmek güçtür. Ancak, $\mathbf{H} \in \mathbf{D}$, ve $\mathbf{H} \in \mathbf{S}$, kısıtları altında asimtotik toplanmış hata kareler ortalaması daha basit olarak elde edilebilir. Örneğin $\mathbf{H} = h^2 \mathbf{I}$, kısıtı altında, bulunan optimal h bant genişliği aşağıda verilmiştir (14):

$$h_{ATHKO} = \left\{ \frac{\int_{-\infty}^{\infty} K(\mathbf{z})^2 d\mathbf{z}}{n\mu_2(K)^2 \int_{-\infty}^{\infty} \left\{ \nabla^2 f(\mathbf{x}) \right\}^2 d\mathbf{x}} \right\}^{1/(d+4)} \quad \nabla^2 f(\mathbf{x}) = \sum_{i=1}^d \left(\frac{\partial^2}{\partial x_i^2} \right) f(\mathbf{x}) \quad [4]$$

Yukarıdaki ifadeden görüldüğü gibi toplanmış hata kareler ortalamasını en küçük yapan optimal bant genişliği değeri, bilinmeyen f fonksiyonunun ikinci mertebeden türevlerinin oluşturduğu Hessian matrisine bağlıdır. Bu nedenle h bant genişliği (4) eşitliğinden bulunamaz, ancak optimal bant genişliğini elde edebilmek için birçok yöntem önerilmiştir. Bu yöntemler daha önce tek değişkenli için verilmişti. Tek değişkenli olasılık yoğunluk fonksiyonunun çekirdek kestirim durumu için kullanılan yöntemler çok büyük değişime uğramadan çok değişkenli duruma genellenebilir. Örneğin Rudemo (7) ve Bowman (1)'ın ÇG yöntemi, $\mathbf{H}$ bant genişliği matris seçimine genellenebilir ve bu durumda ÇG fonksiyonu,

$$\text{ÇG}(\mathbf{H}) = \int_{-\infty}^{\infty} \hat{f}(\mathbf{x}; \mathbf{H})^2 d\mathbf{x} - 2n^{-1} \sum_{i=1}^n \hat{f}_{-i}(\mathbf{X}_i; \mathbf{H}) \quad [5]$$

olarak verilir. Burada, $\mathbf{X}_i$ 'nin veri kümesinden çıkarıldıktan sonra elde edilen çekirdek kestiricisidir. Buradaki temel amaç, (5) eşitliği ile verilen ifadeyi en küçük yapan bant genişliği ya da genişliklerini bulmaktır. $\mathbf{H}^* \mathbf{D}$ kısıtı altında K çekirdek fonksiyonu yerine, standart normal olasılık yoğunluk fonksiyonu kullanılarak, (5) eşitliği ile verilen ÇG fonksiyonu,

$$\begin{aligned} \text{ÇG}(h_1, \dots, h_d) = & \frac{1}{(2\sqrt{\pi})^d n h_1 \dots h_d} + \frac{1}{(2\sqrt{\pi})^d n^2 h_1 \dots h_d} \sum_{i=1}^n \sum_{j \neq i} \left\{ \exp \left[-\frac{1}{4} \sum_{k=1}^d \left(\frac{x_{ik} - x_{jk}}{h_k} \right)^2 \right] \right. \\ & \left. - \left(2 \times 2^{d/2} \right) \exp \left[-\frac{1}{2} \sum_{k=1}^d \left(\frac{x_{ik} - x_{jk}}{h_k} \right)^2 \right] \right\} \end{aligned} \quad [6]$$

biçiminde elde edilmiştir. Bu fonksiyonu en küçük yapan h bant genişliği bulunur. Çapraz geçerlilik fonksiyonu birden fazla yerel minimuma sahip olabilir. Yapılan çalışmalar bu durumda, en büyük yerel minimuma sahip olan bant genişliğinin alınmasının uygun olduğunu ortaya koymuştur. Bu değer elde edilerek olasılık yoğunluk fonksiyonunun kestiriminde kullanılır.

Scott ve Terrell (9)'ün tek değişkenli olasılık yoğunluk fonksiyonunun çekirdek kestiricisine ilişkin h bant genişliğinin elde edilmesi için geliştirdiği YÇG yöntemi, çok değişkenli durum için genelleştirilmiştir. Sain ve arkadaşları (4) eşitliği ile verilen çok değişkenli olasılık yoğunluk fonksiyonu optimal bant genişliği formülündeki bilinmeyen olasılık yoğunluk fonksiyonu f 'nin ikinci mertebeden

türevinin karesinin integrali yerine daha küçük yana sahip olan $\frac{1}{n} \sum_{i=1}^n \hat{f}_{-i}^{iv}(\mathbf{x}_i)$ kestirimini kullanmışlardır (8). $\mathbf{H} \in \mathbf{D}$ kısıtı altında, K çekirdek fonksiyonu yerine standart normal olasılık

yoğunluk fonksiyonu kullanılarak, elde edilen çok değişkenli olasılık yoğunluk fonksiyonunun çarpım çekirdek kestiricisine ilişkin YÇG fonksiyonu,

$$YÇG(h_1, \dots, h_d) = \frac{1}{(2\sqrt{\pi})^d n h_1 \dots h_d} + \frac{1}{4n(n-1)h_1 \dots h_d} \sum_{i=1}^n \sum_{j \neq i} \left\{ \left[\sum_{k=1}^d \left(\frac{x_{ik} - x_{jk}}{h_k} \right)^2 \right]^2 - (2d+4) \left[\sum_{k=1}^d \left(\frac{x_{ik} - x_{jk}}{h_k} \right)^2 \right] + (d^2 + 2d) \right\} \times \prod_{k=1}^d \Phi \left(\frac{x_{ik} - x_{jk}}{h_k} \right) \quad [7]$$

olarak elde edilmiştir. Burada standart normal olasılık yoğunluk fonksiyonudur. (7) eşitliği ile verilen ifadeyi en küçük yapan bant genişlikleri elde edilir. Çapraz geçerlilik yönteminde olduğu gibi yanlış çapraz geçerlilik yönteminden de birden çok yerel minimum elde edilebilir, bu durumda da yine en büyük minimuma sahip olan bant genişliği, optimal bant genişliği olarak alınır.

h bant genişliğini elde etmek için Taylor tarafından kullanılan bootstrap yöntemi de çok değişkenli durum için geliştirilmiştir (13). Bu duruma ilişkin THKO,

$$THKO(\mathbf{h}) = \int_{\mathbb{R}^d} E^* \left\{ \hat{f}^*(\mathbf{x}) - \hat{f}(\mathbf{x}) \right\}^2 d\mathbf{x} \quad [8]$$

dır. Burada $\hat{f}(\mathbf{x})$, çok değişkenli olasılık yoğunluk fonksiyonunun çarpım çekirdek kestiricisi, $\hat{f}^*(\mathbf{x})$, $\hat{f}(\mathbf{x})$ 'den seçilen veriler ile elde edilen çekirdek kestiricisi ve E^* , $(\mathbf{x})$ 'e göre alınan beklenen değerdir. (8) ile verilen eşitlik, çok değişkenli olasılık yoğunluk fonksiyonu çekirdek kestiricisinin THKO değerine çok yakın bir değere sahip olacağı düşünülmektedir. $\mathbf{H} \in D$ kısıtı altında, çekirdek fonksiyonu olarak standart normal olasılık yoğunluk fonksiyonu kullanılırsa, çok değişkenli bootstrap fonksiyonu,

$$B(h_1, \dots, h_d) = \frac{1}{(2\sqrt{\pi})^d n h_1 \dots h_d} + \frac{1}{(2\sqrt{\pi})^d n^2 h_1 \dots h_d} \sum_{i=1}^n \sum_{j \neq i} \left\{ \frac{n-1}{2^{d/2}} \exp \left[-\frac{1}{8} \sum_{k=1}^d \left(\frac{x_{ik} - x_{jk}}{h_k} \right)^2 \right] + \exp \left[-\frac{1}{4} \sum_{k=1}^d \left(\frac{x_{ik} - x_{jk}}{h_k} \right)^2 \right] - \frac{2 \times 2^{d/2}}{3^{d/2}} \exp \left[-\frac{1}{6} \sum_{k=1}^d \left(\frac{x_{ik} - x_{jk}}{h_k} \right)^2 \right] \right\} \quad [9]$$

olarak elde edilir. Bu fonksiyonu en küçük yapan h bant genişlikleri seçilir ve elde edilen değerler çok değişkenli olasılık yoğunluk fonksiyonunun çekirdek kestiriminde kullanılır.

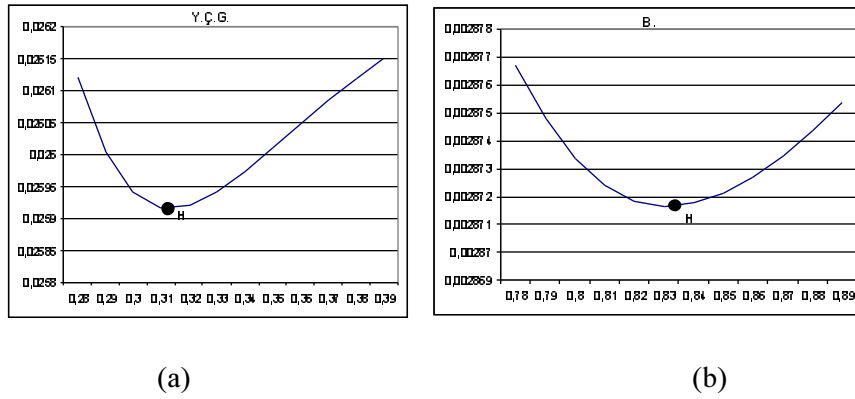
2. UYGULAMA

Bugüne kadar bant genişliği seçimi için bir çok çalışma yapılmış olmasına rağmen, halen genel anlamda kabul edilmiş bir yöntem bulunamamıştır. Çalışmaların çoğu bilgisayar benzetim çalışmalarının sonuçlarına dayanmaktadır. Ancak, h bant genişliğinin elde edilmesi için paket programlar içine alınmış, standart olarak kabul edilmiş yaygın bir düzeltme süreci olmadığından, araştırmacılar kendi programlarını kendileri geliştirmektedirler. Yapılan uygulamada, h bant genişliğinin elde edilmesi için geliştirilen yöntemlerden yansız çapraz geçerlilik, yanlış çapraz geçerlilik ve bootstrap yöntemleri kullanılmıştır. Yöntemlere ilişkin bilgisayar programları ve çok değişkenli olasılık yoğunluk fonksiyonunun çekirdek kestirim değerlerini veren program Delphi 6.0 programlama dili ile yazılmıştır.

Elde edilen sonuçların grafik çizimleri için Excel 7.0 ve S-Plus 2000 paket programları kullanılmıştır. Daha önce Marmara ve Ege bölgeleri için tek bant genişliği kullanılarak ($H \in S$ durumu), deprem verilerine ilişkin çok değişkenli olasılık yoğunluk fonksiyonunun çekirdek kestirimleri elde edilmişti (2). Bu çalışmada ise, 1900-2001 yılları arasında tüm Türkiye'de gerçekleşen, büyüklüğü 5 ve üzeri olan, enlem ve boylamı verilen deprem verileri için iki değişkenli olasılık yoğunluk fonksiyonunun çekirdek kestirim değerlerinin farklı bant genişliklerine sahip durum için ($H \in D$ durumu) elde edilmesi ve $H \in S$ durumu için elde edilen kestirim değerleri ile karşılaştırılması amaçlanmıştır. Her iki durumda da K çekirdek fonksiyonu olarak, iki değişkenli standart normal olasılık yoğunluk fonksiyonu kullanılmıştır.

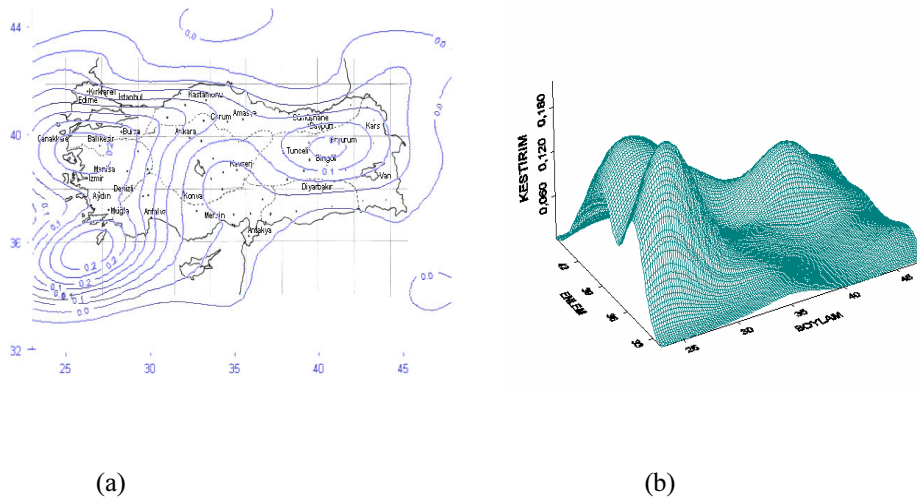
$H \in S$ durumunda, değişkenler arasındaki yayılım farkını önlemek için, veriler $\frac{x - \bar{x}}{s}$ dönüşümü

yapılarak standartlaştırılmış ve tek bir h bant genişliği kullanılmıştır. Benzetim çalışmasında, h bant genişliği 0'dan başlayıp 0,01 artırılarak ÇG, YÇG ve B fonksiyon değerleri elde edilmiş ve bunların grafikleri çizilerek, bu fonksiyonları minimum yapan bant genişlikleri bulunmuştur. ÇG fonksiyonundan yerel minimum elde edilememiştir. YÇG fonksiyonunu en küçük yapan h bant genişliği 0,32, B fonksiyonunu en küçük yapan h bant genişliği 0,84 olarak bulunmuştur. h bant genişliklerine karşılık gelen YÇG ve B fonksiyonlarının grafikleri Şekil 1a ve Şekil 1b'de verilmiştir:



Şekil 1. a)YÇG Fonksiyonunun Grafiği b) B Fonksiyonunun Grafiği

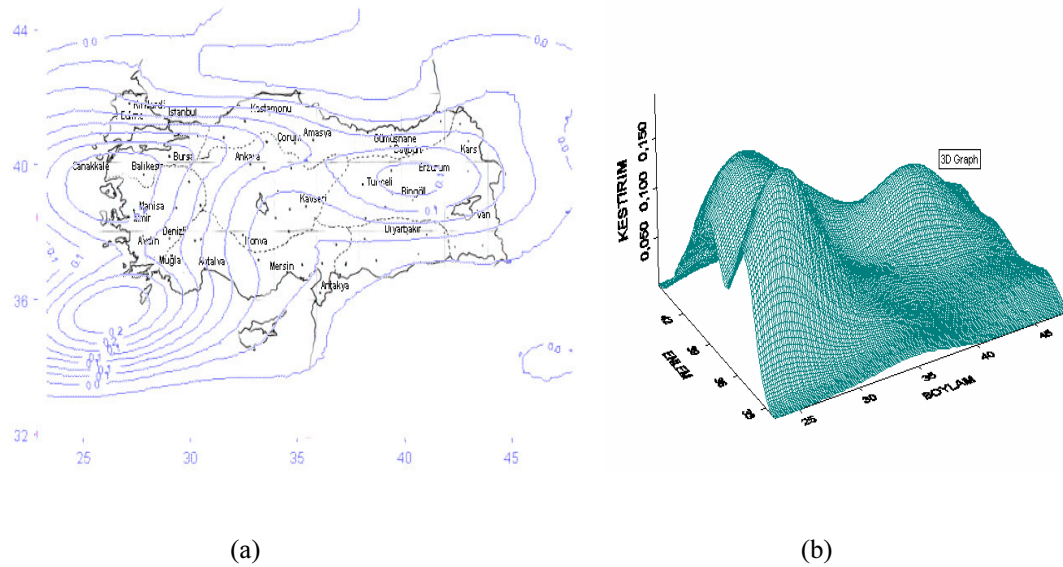
Daha önce yaptığımız bir çalışmada, YÇG yönteminden elde edilen bant genişliğinin optimal bant genişliğine daha yakın sonuçlar verdiği görülmüştür (14). Bu nedenle, bu çalışmada da YÇG yönteminden elde edilen ve değeri 0,32 olan bant genişliği, iki değişkenli olasılık yoğunluk fonksiyonunun çekirdek kestiricisinde yerine konularak kestirim değerleri elde edilmiştir. Elde edilen bu kestirim değerlerinin dağılımına ilişkin, Türkiye haritası üzerine yerleştirilen düzey eğrileri ve 3 boyutlu grafiği Şekil 2'de verilmiştir:



Şekil 2. h=0,32 İçin olasılık yoğunluk fonksiyonunun çekirdek kestirim değerlerinin dağılımına ilişkin
a) Düzey eğrileri b) 3 Boyutlu grafiği

$H \in D$ durumunda, her bir koordinat eksen yönünde farklı bant genişliğinin elde edilmesi amaçlanmıştır. h_1 ile h_2 bant genişliklerini elde etmek için yapılan benzetim çalışmasında, h_1 ile h_2 bant genişlikleri 0'dan başlayıp 0.01 artırılarak ÇG, YÇG ve B fonksiyon değerleri elde edilmiş ve bunların grafikleri çizilerek genel görünimleri izlenmiş; ÇG ve B fonksiyonlarını minimum yapan bant genişliği değeri bulunamıştır. YÇG fonksiyonunu minimum yapan, h_1 bant genişliği 0.3, h_2 bant genişliği 0.4 olarak bulunmuştur.

YÇG fonksiyonunu en küçük yapan h bant genişliği, iki değişkenli olasılık yoğunluk fonksiyonunun çekirdek kestiricisinde yerine konularak kestirim değerleri elde edilmiştir. Bu değerlerin dağılımına ilişkin, Türkiye haritası üzerine yerleştirilen düzey eğrileri ve 3 boyutlu grafiği Şekil 3'te verilmiştir:



Şekil 3. $h_1 = 0,3$, $h_2 = 0,4$ İçin Olasılık Yoğunluk Fonksiyonunun Çekirdek Kestirim Değerlerinin Dağılımına İlişkin a) Düzey Eğrileri b) 3 Boyutlu Grafiği

3. SONUÇLAR

1900-2001 döneminde tüm Türkiye'de (32,00-44,00 Enlem ve 24,00-46,00 Boylam koordinatları) Richter ölçeğine göre büyüklüğü 5 ve üzeri olan deprem verileri ile yapılan uygulama sonucunda, $H \in S$ yaklaşımı ile, yanlış çapraz geçerlilik fonksiyonundan elde edilen h bant genişliği 0,32, bootstrap fonksiyonundan elde edilen h bant genişliği ise 0,84 olarak bulunmuştur (Şekil 1). Çapraz geçerlilik yönteminden h bant genişliği elde edilememiştir. $H \in D$ yaklaşımı ile, yanlış çapraz geçerlilik fonksiyonundan boylam için bant genişliği 0,3 (h_1), enlem için bant genişliği 0,4 (h_2) olarak bulunmuştur. Çapraz geçerlilik ve bootstrap yöntemlerinden bant genişliği elde edilememiştir. $H \in S$ durumu için yanlış çapraz geçerlilik yönteminden elde edilen bant genişliği değeri olasılık yoğunluk fonksiyonunun çekirdek kestiricisinde yerine konularak elde edilen ve Şekil 2'de verilen kestirim değerleri ile $H \in D$ durumu için elde edilen ve Şekil 3'te verilen kestirim değerleri karşılaştırıldığında, yaklaşık değerler elde edildiği görülmüştür. Grafiklerden, Marmara ve Ege Bölgelerinde Richter ölçeğine göre büyüklüğü 5 ve üzerinde deprem olma olasılığının diğer bölgelere göre daha yüksek olduğu görülmektedir. Bunun yanı sıra, Doğu Anadolu Bölgesi içinde yer alan Erzurum, Erzincan, Bingöl yerleşim alanlarında da deprem olma olasılığının Marmara ve Ege Bölgelerine göre daha düşük, ancak bu bölgedeki yerleşim alanlarına göre daha yüksek olduğu gözlenmektedir.

Bir depremin olma olasılığının bilinmesi bir çok bakımdan önemlidir. Örneğin, ekonomik olarak büyük maliyet gerektiren yatırımların yapılmasında bu bilginin bilinmesi, yatırımcının karar vermesi açısından yararlı olacaktır. Yatırımcı, yatırım yaparken deprem riski büyük olan bölgelerde bina gibi yapıların depreme dayanıklı olarak yapılmasını sağlayabilir ya da deprem riski büyük olan bölge yerine, yatırımların deprem riski daha az olan bölgelere kaydırılmasını tercih edebilir.

KAYNAKLAR

1. Bowman, A. W., "An alternative method of cross-validation for the smoothing of density estimates", *Biometrika*, 71(2): 353-360 (1984).
2. Cula, G. S., Toktamış, Ö., "Estimation of multivariate probability density function with kernel functions", *Journal of the Turkish Statistical Association*, 3: 29-39 (2000).
3. Epanechnikov, V. A., "Nonparametric estimation of a multivariate probability density", *Theory of Probability and Its Applications*, 14: 153-158 (1969).
4. Hardle, W., "Smoothing Techniques with implementation in S", *Springer-Verlag*, New-York Inc (1991).
5. Jones, M. C. and Marron, J.S., "A brief survey of bandwidth selection for density estimation", *Journal of the American Statistical Association*, 91(433): 401-407 (1996).
6. Rosenblatt, M., "Remarks on some nonparametric estimates a density function", *Annals Math. Statist.*, 27: 832-837 (1956).
7. Rudemo, M., 1982, "Empirical choice of histograms and kernel density estimator", *Scand. J. Statist.* 9: 65-78 (1982).
8. Sain, R. S., Scott, D. W. and Baggerly, K. A., "Cross-validation of multivariate densities", *Journal of the American Statistical Association*, 89(427): 807-817 (1994).
9. Scott, D. W. and Terrell, G. R., "Biased and unbiased cross-validation in density estimation", *Journal of the American Statistical Association*, 82(400): 1131-1146 (1987).
10. Scott, W. D. and Wand, M. P., "Feasibility of multivariate density estimates", *Biometrika*, 78(1): 197-205 (1991).
11. Silverman, B. W., "Density estimation for statistics and data analysis", *Chapman and Hall*, London 1-97 (1986).
12. Simonoff, J. S., "Smoothing Methods in Statistics", Verlag, New York, 96-130 (1996).
13. Taylor, C. C., "Bootstrap choice of the smoothing parameter in kernel density estimation", *Biometrika*, 76(4): 705-712 (1989).
14. Wand, M. P. and Jones, M. C., "Kernel smoothing", *Chapman and Hall*, New York, 1-113 (1995).

Geliş Tarihi:17.06.2002

Kabul Tarihi:14.03.2003