GAZİ ÜNİVERSİTESİ

1926

ÇEKİRDEK DÜZGÜNLEŞTİRMESİYLE YOĞUNLUK FONKSİYONU TAHMİNİNDE BANT GENİŞLİĞİ SEÇİM YÖNTEMLERİNİN KARŞILAŞTIRILMASI

Şule KARAKOÇ

YÜKSEK LİSANS TEZİ

İSTATİSTİK ANA BİLİM DALI

GAZİ ÜNİVERSİTESİ

FEN BİLİMLERİ ENSTİTÜSÜ

OCAK 2023

ETİK BEYAN

Gazi Üniversitesi Fen Bilimleri Enstitüsü Tez Yazım Kurallarına uygun olarak hazırladığım bu tez çalışmasında;

- Tez içinde sunduğum verileri, bilgileri ve dokümanları akademik ve etik kurallar çerçevesinde elde ettiğimi,
- Tüm bilgi, belge, değerlendirme ve sonuçları bilimsel etik ve ahlak kurallarına uygun olarak sunduğumu,
- Tez çalışmasında yararlandığım eserlerin tümüne uygun atıfta bulunarak kaynak gösterdiğimi,
- Kullanılan verilerde herhangi bir değişiklik yapmadığımı,
- Bu tezde sunduğum çalışmanın özgün olduğunu,

bildirir, aksi bir durumda aleyhime doğabilecek tüm hak kayıplarını kabullendiğimi beyan ederim.

Şule KARAKOÇ
25/01/2023

ÇEKİRDEK DÜZGÜNLEŞTİRMESİYLE YOĞUNLUK FONKSİYONU TAHMİNİNDE BANT GENİŞLİĞİ SEÇİM YÖNTEMLERİNİN KARŞILAŞTIRILMASI

(Yüksek Lisans Tezi)

Şule KARAKOÇ

GAZİ ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

Ocak 2023

ÖZET

Parametrik olmayan yoğunluk tahminlerinde yaygın olarak kullanılan yöntemlerden biri çekirdek yoğunluk tahminidir. Bilinmeyen olasılık yoğunluk fonksiyonu için çekirdek tahmin edicisi kullanılarak ilgili rastgele değişkene ilişkin olasılık yoğunluk ya da dağılım fonksiyonu tahmininde bulunulur. Veriye dayalı olarak elde edilen çekirdek tahmin edicilerinin performansı düzgünleştirme parametresi olarak da bilinen bant genişliklerine bağlıdır. Bu nedenle, literatürde çeşitli bant genişliği seçim yöntemleri yer almaktadır. Bu çalışmanın amacı, bant genişliği seçim yöntemleri içinde yaygın olarak kullanılan normal ölçek, Silverman'ın pratik yaklaşımı, yerine koyma, denklem çözme, en küçük kareler çapraz doğrulama, yanlı çapraz doğrulama ve düzleştirilmiş çapraz doğrulama yöntemlerinin performanslarını çeşitli normal karışım yoğunluklarına göre karşılaştırmaktır. Yapılan simülasyon çalışması sonuçlarına göre, kurtotic tek modlu ve ayrılmış iki modlu yoğunlukların dışında ele alınan normal karışım yoğunlukları için yan ve ortalama hata kare (mean squared error) değerleri bakımından yanlı çapraz doğrulama yöntemi ile elde edilen bant genişliklerinin optimal bant genişliğine en yakın sonucu verdiği gözlemlenmiştir.

Bilim Kodu	:	20513
Anahtar Kelimeler	:	Çekirdek yoğunluk tahmini, bant genişliği seçim yöntemleri, normal ölçek, en küçük kareler çapraz doğrulama, yerine koyma
Sayfa Adedi	:	65
Danışman	:	Doç. Dr. Necla GÜNDÜZ

A COMPARISON OF BANDWIDTH SELECTION METHODS FOR ESTIMATING DENSITY FUNCTIONS WITH KERNEL SMOOTHING

(M. Sc. Thesis)

Şule KARAKOÇ

GAZİ UNIVERSITY

GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES

January 2023

ABSTRACT

One of the commonly used methods for non-parametric density estimation is kernel density estimation. For the unknown probability density function, the probability density or distribution function of the related random variable is estimated by using the kernel estimator. The performance of kernel estimators derived from the data depends on the bandwidth, also known as the smoothing parameter. Therefore, there are various bandwidth selection methods in the literature. The aim of this study is to compare the performances of normal scale, Silverman's rule of thumb, direct plug-in, solve-the equation, least squares cross-validation, biased cross-validation and smoothed cross-validation methods, which are widely used in bandwidth selection methods, according to various normal mixture densities. According to the results of the simulation study, it was observed that the bandwidths obtained by the biased cross-validation method in terms of side and mean squared error values for the normal mixture densities considered outside the kurtotic unimodal and separated bimodal densities gave the closest result to the optimal bandwidth.

Science Code : 20513
Key Words : Kernel density estimation, bandwidth selection methods, normal scale rules, least squares cross validation, plug-in
Page Number : 65
Supervisor : Assoc. Prof. Dr. Necla GÜNDÜZ

TEŞEKKÜR

Bu çalışmanın gerçekleşmesinde, bilgi ve tecrübelerini benden esirgemeyen her daim yol göstericiliği ile yolumu bulmamı sağlayan danışman hocam sayın Doç. Dr. Necla GÜNDÜZ'e sonsuz saygılarımı ve teşekkürlerimi sunarım. Bugünlere gelmemde maddi manevi desteklerini esirgemeyen ve her zaman yanımda olan annem Hülya KARAKOÇ ve babam Erol KARAKOÇ'a, çalışmalarım sırasında beni yüreklendiren kardeşlerim Sedanur KARAKOÇ ve Efe Erdem KARAKOÇ'a sonsuz teşekkürlerimi sunarım.

İÇİNDEKİLER

Sayfa

ÖZET iv
ABSTRACT v
TEŞEKKÜR vi
İÇİNDEKİLER vii
ÇİZELGELERİN LİSTESİ ix
ŞEKİLLERİN LİSTESİ xi
SİMGELER VE KISALTMALAR xiv
1. GİRİŞ 1
2. YOĞUNLUK TAHMİNİ 3
3. ÇEKİRDEK YOĞUNLUK TAHMİNİ 5
3.1. Çekirdek Yoğunluk Fonksiyonlarının Başarı Ölçütleri 9
3.2. Uygun Çekirdek Fonksiyonunun Seçimi 16
4. BANT GENİŞLİĞİ SEÇİM YÖNTEMLERİ 19
4.1. Normal Ölçek Bant Genişliği Seçim Yöntemi 20
4.2. Silverman'ın Pratik Yaklaşımı Bant Genişliği Seçim Yöntemi 21
4.3. Yerine Koyma Bant Genişliği Seçim Yöntemi 22
4.4. Denklem Çözme Bant Genişliği Seçim Yöntemi 23
4.5. En Küçük Kareler Çapraz Doğrulama Bant Genişliği Seçim Yöntemi 24
4.6. Yanlı Çapraz Doğrulama Bant Genişliği Seçim Yöntemi 25
4.7. Düzleştirilmiş Çapraz Doğrulama Bant Genişliği Seçim Yöntemi 26
5. SİMÜLASYON ÇALIŞMASI 27
6. UYGULAMA 57
7. SONUÇ VE ÖNERİLER 61
KAYNAKLAR 63

Sayfa

ÖZGEÇMİŞ 65

ÇİZELGELERİN LİSTESİ

Çizelge **Sayfa**

Çizelge 3.1. Çekirdek düzgünleştirmesinde yaygın olarak kullanılan fonksiyonlar....... 17

Çizelge 5.1. Normal karışım yoğunluk parametreleri.. 27

Çizelge 5.2. $N(0,1)$ modeli ile tanımlanan standart normal yoğunluk ve bu dağılımdan alınan çeşitli örnek çapları için elde edilen optimal bant genişliği ($h_{\text{İ}OHK}$) ile 1000 tekrarla üretilen 7 farklı bant genişliği seçim yöntemlerinin ortalamaları (h_{ORT}) standart sapmaları (S), yanları (YAN) ve ortalama hata kareleri (OHK).. 30

Çizelge 5.3. $\frac{1}{5}\mathrm{N}(0,1)+\frac{1}{5}\mathrm{N}\left(\frac{1}{2},(\frac{2}{3})^2\right)+\frac{3}{5}\mathrm{N}\left(\frac{13}{12},(\frac{5}{9})^2\right)$ modeli ile tanımlanan çarpık tek modlu yoğunluk olarak adlandırılan normal karışım yoğunluğundan alınan çeşitli örnek çapları için elde edilen optimal bant genişliği ($h_{\text{İ}OHK}$) ile 1000 tekrarla üretilen 7 farklı bant genişliği seçim yöntemlerinin ortalamaları (h_{ORT}) standart sapmaları (S), yanları (YAN) ve ortalama hata kareleri (OHK).. 31

Çizelge 5.4. $\frac{2}{3}\mathrm{N}(0,1)+\frac{1}{3}\mathrm{N}\left(0,(\frac{1}{10})^2\right)$ modeli ile tanımlanan kurtotic tek modlu yoğunluk olarak adlandırılan normal karışım yoğunluğundan alınan çeşitli çapları için elde edilen optimal bant genişliği ($h_{\text{İ}OHK}$) ile 1000 tekrarla üretilen 7 farklı bant genişliği seçim yöntemlerinin ortalamaları (h_{ORT}) standart sapmaları (S), yanları (YAN) ve ortalama hata kareleri (OHK).. 32

Çizelge 5.5. $\frac{1}{10}\mathrm{N}(0,1)+\frac{9}{10}\mathrm{N}\left(0,(\frac{1}{10})^2\right)$ modeli ile tanımlanan aykırı yoğunluk olarak adlandırılan normal karışım yoğunluğundan alınan çeşitli örnek çapları için elde edilen optimal bant genişliği ($h_{\text{İ}OHK}$) ile 1000 tekrarla üretilen 7 farklı bant genişliği seçim yöntemlerinin ortalamaları (h_{ORT}) standart sapmaları (S), yanları (YAN) ve ortalama hata kareleri (OHK)............... 33

Çizelge 5.6. $\frac{1}{2}\mathrm{N}\left(-1,(\frac{2}{3})^2\right)+\frac{1}{2}\mathrm{N}\left(1,(\frac{2}{3})^2\right)$ modeli ile tanımlanan iki modlu yoğunluk olarak adlandırılan normal karışım yoğunluğundan alınan çeşitli çapları için elde edilen optimal bant genişliği ($h_{\text{İ}OHK}$) ile 1000 tekrarla üretilen 7 farklı bant genişliği seçim yöntemlerinin ortalamaları (h_{ORT}) standart sapmaları (S), yanları (YAN) ve ortalama hata kareleri (OHK)............... 34

Çizelge 5.7. $\frac{1}{2}\mathrm{N}\left(-\frac{3}{2},(\frac{1}{2})^2\right)+\frac{1}{2}\mathrm{N}\left(\frac{3}{2},(\frac{1}{2})^2\right)$ modeli ile tanımlanan ayrılmış iki modlu yoğunluk olarak adlandırılan normal karışım yoğunluğundan alınan çeşitli çapları için elde edilen optimal bant genişliği ($h_{\text{İ}OHK}$) ile 1000 tekrarla üretilen 7 farklı bant genişliği seçim yöntemlerinin ortalamaları (h_{ORT}) standart sapmaları (S), yanları (YAN) ve ortalama hata kareleri (OHK). 35

Çizelge **Sayfa**

Çizelge 5.8. $\frac{3}{4}\mathrm{N}(0,1)+\frac{1}{4}\mathrm{N}\left(\frac{3}{2},(\frac{1}{3})^2\right)$ modeli ile tanımlanan çarpık iki modlu yoğunluk olarak adlandırılan normal karışım yoğunluğundan alınan çeşitli çapları için elde edilen optimal bant genişliği ($h_{\dot{I}OHK}$) ile 1000 tekrarla üretilen 7 farklı bant genişliği seçim yöntemlerinin ortalamaları (h_{ORT}) standart sapmaları (S), yanları (YAN) ve ortalama hata kareleri (OHK)................ 36

Çizelge 5.9. $\frac{9}{20}\mathrm{N}\left(-\frac{6}{5},(\frac{3}{5})^2\right)+\frac{9}{20}\mathrm{N}\left(\frac{6}{5},(\frac{3}{5})^2\right)+\frac{1}{10}\mathrm{N}\left(0,(\frac{1}{4})^2\right)$ modeli ile tanımlanan üç modlu yoğunluk olarak adlandırılan normal karışım yoğunluğundan alınan çeşitli çapları için elde edilen optimal bant genişliği ($h_{\dot{I}OHK}$) ile 1000 tekrarla üretilen 7 farklı bant genişliği seçim yöntemlerinin ortalamaları (h_{ORT}) standart sapmaları (S), yanları (YAN) ve ortalama hata kareleri (OHK)... 37

Çizelge 6.1. Boston veri seti değişkenleri için 7 farklı yönteme göre hesaplanan bant genişlikleri.. 57

ŞEKİLLERİN LİSTESİ

Şekil **Sayfa**

Şekil 3.1. Farklı bant genişlikleri için çekirdek düzgünleştirmesi. ... 8

Şekil 3.2. Çekirdek düzgünleştirmesinde yaygın olarak kullanılan çekirdek fonksiyon grafikleri. ... 18

Şekil 5.1. Normal karışım yoğunluk grafikleri ... 28

Şekil 5.2. $\mathrm{N}(0,1)$ modeli ile tanımlanan standart normal yoğunluğun olasılık yoğunluk fonksiyonu ve bu dağılımdan alınan çeşitli örnek çapları için elde edilen optimal bant genişliği ile 7 farklı bant genişliği seçim yöntemlerine göre elde edilen çekirdek yoğunluk tahminleri. ... 38

Şekil 5.3. $\frac{1}{5}\mathrm{N}(0,1)+\frac{1}{5}\mathrm{N}\left(\frac{1}{2},(\frac{2}{3})^2\right)+\frac{3}{5}\mathrm{N}\left(\frac{13}{12},(\frac{5}{9})^2\right)$ modeli ile tanımlanan çarpık tek modlu yoğunluk olarak adlandırılan normal karışım dağılımının olasılık yoğunluk fonksiyonu ve bu dağılımdan alınan çeşitli örnek çapları için elde edilen optimal bant genişliği ile 7 farklı bant genişliği seçim yöntemlerine göre elde edilen çekirdek yoğunluk tahminleri. ... 39

Şekil 5.4. $\frac{2}{3}\mathrm{N}(0,1)+\frac{1}{3}\mathrm{N}\left(0,(\frac{1}{10})^2\right)$ modeli ile tanımlanan kurtotic tek modlu yoğunluk olarak adlandırılan normal karışım dağılımının olasılık yoğunluk fonksiyonu ve bu dağılımdan alınan çeşitli örnek çapları için elde edilen optimal bant genişliği ile 7 farklı bant genişliği seçim yöntemlerine göre elde edilen çekirdek yoğunluk tahminleri. ... 40

Şekil 5.5. $\frac{1}{10}\mathrm{N}(0,1)+\frac{9}{10}\mathrm{N}\left(0,(\frac{1}{10})^2\right)$ modeli ile tanımlanan aykırı yoğunluk olarak adlandırılan normal karışım dağılımının olasılık yoğunluk fonksiyonu ve bu dağılımdan alınan çeşitli örnek çapları için elde edilen optimal bant genişliği ile 7 farklı bant genişliği seçim yöntemlerine göre elde edilen çekirdek yoğunluk tahminleri. ... 41

Şekil 5.6. $\frac{1}{2}\mathrm{N}\left(-1,(\frac{2}{3})^2\right)+\frac{1}{2}\mathrm{N}\left(1,(\frac{2}{3})^2\right)$ modeli ile tanımlanan iki modlu yoğunluk olarak adlandırılan normal karışım dağılımının olasılık yoğunluk fonksiyonu ve bu dağılımdan alınan çeşitli örnek çapları için elde edilen optimal bant genişliği ile 7 farklı bant genişliği seçim yöntemlerine göre elde edilen çekirdek yoğunluk tahminleri. ... 42

Şekil 5.7. $\frac{1}{2}\mathrm{N}\left(-\frac{3}{2},(\frac{1}{2})^2\right)+\frac{1}{2}\mathrm{N}\left(\frac{3}{2},(\frac{1}{2})^2\right)$ modeli ile tanımlanan ayrılmış iki modlu yoğunluk olarak adlandırılan normal karışım dağılımının olasılık yoğunluk fonksiyonu ve bu dağılımdan alınan çeşitli örnek çapları için elde edilen optimal bant genişliği ile 7 farklı bant genişliği seçim yöntemlerine göre elde edilen çekirdek yoğunluk tahminleri. ... 43

Şekil **Sayfa**

Şekil 5.8. $\frac{3}{4}\mathrm{N}(0,1)+\frac{1}{4}\mathrm{N}\left(\frac{3}{2},(\frac{1}{3})^2\right)$ modeli ile tanımlanan çarpık iki modlu yoğunluk olarak adlandırılan normal karışım dağılımının olasılık yoğunluk fonksiyonu tahmini ve bu dağılımdan alınan çeşitli örnek çapları için elde edilen optimal bant genişliği ile 7 farklı bant genişliği seçim yöntemlerine göre elde edilen çekirdek yoğunluk tahminleri.. 44

Şekil 5.9. $\frac{9}{20}\mathrm{N}\left(-\frac{6}{5},(\frac{3}{5})^2\right)+\frac{9}{20}\mathrm{N}\left(\frac{6}{5},(\frac{3}{5})^2\right)+\frac{1}{10}\mathrm{N}\left(0,(\frac{1}{4})^2\right)$ modeli ile tanımlanan üç modlu yoğunluk olarak adlandırılan normal karışım dağılımının olasılık yoğunluk fonksiyonu ve bu dağılımdan alınan çeşitli örnek çapları için elde edilen optimal bant genişliği ile 7 farklı bant genişliği seçim yöntemlerine göre elde edilen çekirdek yoğunluk tahminleri.. 45

Şekil 5.10. $\mathrm{N}(0,1)$ modeli ile tanımlanan standart normal yoğunluğun çeşitli örnek çapları için elde edilen 7 farklı bant genişliği seçim yöntemlerine ilişkin kutu grafikleri.. 46

Şekil 5.11. $\frac{1}{5}\mathrm{N}(0,1)+\frac{1}{5}\mathrm{N}\left(\frac{1}{2},(\frac{2}{3})^2\right)+\frac{3}{5}\mathrm{N}\left(\frac{13}{12},(\frac{5}{9})^2\right)$ modeli ile tanımlanan çarpık tek modlu yoğunluk olarak adlandırılan normal karışım dağılımından çeşitli örnek çapları için elde edilen 7 farklı bant genişliği seçim yöntemlerine ilişkin kutu grafikleri.. 47

Şekil 5.12. $\frac{2}{3}\mathrm{N}(0,1)+\frac{1}{3}\mathrm{N}\left(0,(\frac{1}{10})^2\right)$ modeli ile tanımlanan kurtotic tek modlu yoğunluk olarak adlandırılan normal karışım dağılımından çeşitli örnek çapları için elde edilen 7 farklı bant genişliği seçim yöntemlerine ilişkin kutu grafikleri.. 48

Şekil 5.13. $\frac{1}{10}\mathrm{N}(0,1)+\frac{9}{10}\mathrm{N}\left(0,(\frac{1}{10})^2\right)$ modeli ile tanımlanan aykırı yoğunluk olarak adlandırılan normal karışım dağılımından çeşitli örnek çapları için elde edilen 7 farklı bant genişliği seçim yöntemlerine ilişkin kutu grafikleri. 49

Şekil 5.14. $\frac{1}{2}\mathrm{N}\left(-1,(\frac{2}{3})^2\right)+\frac{1}{2}\mathrm{N}\left(1,(\frac{2}{3})^2\right)$ modeli ile tanımlanan iki modlu yoğunluk olarak adlandırılan normal karışım dağılımından çeşitli örnek çapları için elde edilen 7 farklı bant genişliği seçim yöntemlerine ilişkin kutu grafikleri.. 50

Şekil 5.15. $\frac{1}{2}\mathrm{N}\left(-\frac{3}{2},(\frac{1}{2})^2\right)+\frac{1}{2}\mathrm{N}\left(\frac{3}{2},(\frac{1}{2})^2\right)$ modeli ile tanımlanan ayrılmış iki modlu yoğunluk olarak adlandırılan normal karışım dağılımından çeşitli örnek çapları için elde edilen 7 farklı bant genişliği seçim yöntemlerine ilişkin kutu grafikleri.. 51

Şekil 5.16. $\frac{3}{4}\mathrm{N}(0,1)+\frac{1}{4}\mathrm{N}\left(\frac{3}{2},(\frac{1}{3})^2\right)$ modeli ile tanımlanan çarpık iki modlu yoğunluk olarak adlandırılan normal karışım dağılımından çeşitli örnek çapları için elde edilen 7 farklı bant genişliği seçim yöntemlerine ilişkin kutu grafikleri.. 52

Şekil **Sayfa**

Şekil 5.17. $\frac{9}{20}\mathrm{N}\left(-\frac{6}{5},(\frac{3}{5})^2\right)+\frac{9}{20}\mathrm{N}\left(\frac{6}{5},(\frac{3}{5})^2\right)+\frac{1}{10}\mathrm{N}\left(0,(\frac{1}{4})^2\right)$ modeli ile tanımlanan üç modlu yoğunluk olarak adlandırılan normal karışım dağılımından çeşitli örnek çapları için elde edilen 7 farklı bant genişliği seçim yöntemlerine ilişkin kutu grafikleri... 53

Şekil 6.1. Boston veri seti değişkenlerinin 7 farklı yönteme göre hesaplanan bant genişlikleri ile elde edilen çekirdek yoğunluk tahminleri............................. 59

SİMGELER VE KISALTMALAR

Bu çalışmada kullanılmış simgeler ve kısaltmalar, açıklamaları ile birlikte aşağıda sunulmuştur.

Simgeler	**Açıklamalar**
$\boldsymbol{n}$	Örnek çapı
$\boldsymbol{K}$	Çekirdek fonksiyonu
$\boldsymbol{h}$	Bant genişliği
$\boldsymbol{K_h}$	h'a bağlı çekirdek fonksiyonu
$\boldsymbol{h_n}$	n'ye bağlı bant genişliği
$(\boldsymbol{K_h} * \boldsymbol{f})(\boldsymbol{x})$	K_h ile f'nin konvülasyonu
$\sim$	Asimptotik denklik
$\boldsymbol{O}$	Büyük "O"
$\boldsymbol{o}$	Küçük "o"

Kısaltmalar	**Açıklamalar**
AİOHK	Asimptotik İntegrallenmiş Ortalama Hata Kare (Asymptotic Mean Integrated Squared Error (AMISE))
DÇD	Düzleştirilmiş Çapraz Doğrulama Bant Genişliği Seçim Yöntemi
EKKÇD	En Küçük Kareler Çapraz Doğrulama Bant Genişliği Seçim Yöntemi
İHK	İntegrallenmiş Hata Kare (Integrated Squared Error (ISE))
İOHK	İntegrallenmiş Ortalama Hata Kare (Mean Integrated Squared Error (MISE))
İOMH	İntegrallenmiş Ortalama Mutlak Hata (Mean Integrated Absolute Error (MIAE))
İYK	İntegrallenmiş Yan Kare (Integrated Squared Bias (ISB))
NÖ	Normal Ölçek Bant Genişliği Seçim Yöntemi
OHK	Ortalama Hata Kare (Mean Squared Error (MSE))
SJDÇ	Denklem Çözme Bant Genişliği Seçim Yöntemi
SJYK	Yerine Koyma Bant Genişliği Seçim Yöntemi

Kısaltmalar	Açıklamalar
SPY	Silverman'ın Pratik Yaklaşım Bant Genişliği Seçim Yöntemi
YÇD	Yanlı Çapraz Doğrulama Bant Genişliği Seçim Yöntemi

1. GİRİŞ

Verinin bütün özelliklerinin tanımlanmasını sağlayan olasılık yoğunluk fonksiyonu, istatistiğin temelini oluşturur. Bir rasgele değişkenin olasılık yoğunluk fonksiyonu, o rasgele değişkenin dağılımının belirlenmesine, konum ve ölçek parametrelerinin bulunmasına yardımcı olur. Konum parametresi (μ) verinin nerede yoğunlaştığını, ölçek parametresi (σ^2) ise verinin konumu etrafındaki değişkenliğinin ölçüsüdür. Yani bir problemle ilgili olarak olasılık yoğunluk fonksiyonu biliniyorsa her türlü olasılık sorusuna cevap verebiliriz. Bu nedenle pek çok analizde değişkenlerin olasılık yoğunluk fonksiyonları tahmin edilir. Olasılık yoğunluk tahminlerinde parametrik ve parametrik olmayan yöntemeler kullanılır. Parametrik yöntemler ilgili olasılık yoğunluk fonksiyonunun doğru olasılık yoğunluk fonksiyonu olduğu varsayımını yapar. Ancak parametrik olmayan yöntemler böyle bir varsayım yapsa bile, rasgele değişkenin kendini daha iyi temsil etmesini sağlayarak olasılık yoğunluk fonksiyonunu tahmin eder (Silverman, 1986: 1). Parametrik olmayan yoğunluk tahminleri yoğunluğa dair mümkün olduğunca az varsayım yaparak örneklem üzerinden "düzgünleştirme" işlemi yapar (Terrell, 1990).

Parametrik olmayan yoğunluk tahmin edicilerinden en iyisi "çekirdek yoğunluk" tahmin edicileridir (Rosenblatt, 1956; Parzen, 1962). Çekirdek yoğunluk tahmin edicilerin performansı, önemli ölçüde "bant genişliği" seçimine bağlıdır. Örneğin; küçük bant genişlikleri seçildiğinde dağılımın kuyruklarında gerçek olmayan pürüzler olurken, büyük bant genişlikleri seçildiğinde dağılımın kuyruklarında aşırı düzleşme olacağından dağılımın önemli özellikleri kaybolabilir. Bant genişliği seçiminde tek bir en iyi seçim yöntemini belirlemek mümkün olmasa da yoğunluk tahmini yapılmak istenen verinin gözlendiği yığın yapısının seçim yöntemleri üzerindeki etkisi büyüktür (Gramacki, 2018: 63; Gündüz, Aydın ve Balibeyoğlu, 2020).

Çekirdek yoğunluk tahmininde bant genişliği seçim yöntemleri Rudemo (1982), Bowman (1984), Silverman (1986), Scott ve Terrel (1987), Sheather ve Jones (1991) gibi birçok araştırmacı tarafından ele alınmıştır. Cao, Cuevas ve Gonzalez-Manteiga (1994) ve Harpole (2013) farklı bant genişliklerinin performanslarını karşılaştırmak için her bant genişliğinin İntegrallenmiş Ortalama Hata Kare (İOHK), Ortalama Hata Kare (OHK), ve yanlarını (bias) hesaplamışlar ve bu ölçütleri puanlayarak bant genişliklerinin performanslarını karşılaştırmışlardır.

Bu tez çalışmasında çekirdek yoğunluk tahminini etkileyen bant genişliği seçim yöntemlerinden normal ölçek, Silverman'ın pratik yaklaşımı, yerine koyma, denklem çözme, en küçük kareler çapraz doğrulama, yanlı çapraz doğrulama ve düzleştirilmiş çapraz doğrulama gibi yaygın olarak kullanılan yöntemler incelenmektedir. Bu bant genişliklerinin performansları karşılaştırılırken çeşitli yoğunluklar için basit ve hesaplamaya elverişli normal karışım yoğunlukları kullanılmıştır.

İkinci bölümde yoğunluk tahmini açıklanmıştır. Üçüncü bölümde çekirdek yoğunluk tahmini, olasılık yoğunluk fonksiyonlarının başarı ölçütleri ve çekirdek tahmini yapılırken kullanılacak olan çekirdek fonksiyonları anlatılmıştır. Dördüncü bölümde normal ölçek, Silverman'ın pratik yaklaşımı, yerine koyma, denklem çözme, en küçük kareler çapraz doğrulama, yanlı çapraz doğrulama ve düzleştirilmiş çapraz doğrulama bant genişliği seçim yöntemleri anlatılmıştır. Beşinci bölümde bant genişliklerinin performanslarının karşılaştırılması amacıyla normal karışım yoğunlukları kullanılarak R paket programı (versiyon 4.2.2) desteğiyle bir simülasyon çalışmasına yer verilmiştir. Altıncı bölümde gerçek veri seti üzerinde bant genişliklerinin R paket programı (versiyon 4.2.2) uygulaması yapılmıştır. Yedinci bölümde ise elde edilen bulgular açıklanmıştır.

2. YOĞUNLUK TAHMİNİ

Yoğunluk tahmini, gözlem verilerinden olasılık yoğunluk fonksiyonunun bir tahmininin oluşturulmasıdır (Silverman, 1986: 1). Başka bir deyişle yoğunluk tahmini bir rastgele değişkenin örnek için aldığı değerlerle nasıl dağıldığını gösterir. Olasılık yoğunluk fonksiyonundan rastgele değişkenin sadece ortalama ve varyans gibi tanımlayıcı istatistikleri değil, aynı zamanda bu değişkenin belirli bir aralıkta değerler alma olasılıkları da hesaplanabilir (Hardle, Müller, Sperlich, 2004: 1).

Genel olarak yoğunluk tahmin etme ihtiyacı iki şekilde ortaya çıkar. İhtiyaçlardan birisi eldeki verinin potansiyel bir yoğunluğa ne kadar uyduğuna karar vermektir. Diğeri ise parametrik tahmin edicilerin asimptotik bir dağılımını araştırmaktır (Yolsal, 2017: 19).

Yoğunluk tahmininde parametrik ve parametrik olmayan yöntemeler kullanılır. Parametrik yöntemlerin en büyük dezavantajı gerçek olasılık yoğunluk fonksiyonu f'nin bilindiğinin varsayılmasıdır (Pagan ve Ullah, 1999: 6). Parametrik yöntemler olasılık yoğunluk fonksiyonu f için varsayımlar yapar. Örneğin, ortalaması μ ve varyansı σ^2 olan normal dağılımdan çekilen bir verinin olasılık yoğunluk fonksiyonu

$$f(x) = \begin{cases} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} & -\infty < x < \infty \\ 0 & d.h. \end{cases} \tag{2.1}$$

olsun. Burada μ ve σ^2 tahminleri bulunarak

$$\hat{\mu} = \frac{1}{n}\sum_1^n x_i \qquad \hat{\sigma}^2 = \frac{1}{n-1}\sum_1^n (x_i - \bar{x})^2 \tag{2.2}$$

ve bu tahminler Eş. 1.1'de yerine konularak

$$\hat{f}(x) = \begin{cases} \frac{1}{\hat{\sigma}\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\hat{\mu}}{\hat{\sigma}}\right)^2} & -\infty < x < \infty \\ 0 & d.h. \end{cases} \tag{2.3}$$

tahmini elde edilebilir. Parametrik olmayan yöntemlerde ise olasılık yoğunluk fonksiyonu *f* için varsayımlar yapmadan doğrudan gözlenen veriden tahmin edilir (Pagan ve Ullah, 1999: 6).

Parametrik yoğunlukların bilinen ve sıkça kullanılan pek çoğu tek modludur (yani, tek bir yerel maksimuma sahiptir), oysa bir dizi pratik problem çok modlu yoğunlukları içerir, bu da parametrik yoğunluk tahminini birçok uygulama için uygun olmayan bir hale getirir. Bu aynı zamanda olasılık yoğunluk fonksiyonunun parametrik formunun seçimini öneren hiçbir ipucu olmadığında da geçerlidir. Bu nedenle, parametrik olmayan yoğunluk tahminini kullanmak genellikle daha iyidir. Kısacası, tüm parametrik olmayan yöntemler, altta yatan dağılımın formu hakkında herhangi bir parametrik varsayım yapmadan, doğrudan verilerden yoğunluk tahminini yapmaya çalışır. Parametrik olmayan yoğunluk tahmini için Histogram, Parzen penceresi ve k-en yakın komşu gibi birçok yöntem mevcuttur. Çekirdek düzgünleştirme olarak da adlandırılan çekirdek yoğunluk tahmini de bu yöntemlerden en önemli ve yaygın olarak kullanılanlardan birisidir (Gramacki, 2018: 3, 4).

3. ÇEKİRDEK YOĞUNLUK TAHMİNİ

Parametrik olmayan olasılık yoğunluk tahmin yöntemleri ile yığın hakkında mümkün olduğunca az varsayım yapılarak ve örnek verileri kullanılarak, verinin gözlemlendiği yığının olasılık yoğunluk fonksiyonu yeniden tahmin edilmeye çalışılır. Veriye dayalı olan bu yöntemler örneklem üzerinden düzgünleştirme işlemi yapar (Terrell, 1990). Düzgünleştirme rasgele değişkenin yapısını modelleyerek aynı zamanda olasılık yoğunluk fonksiyonunu tahmin eder (Yolsal, 2017: 24).

Parametrik olmayan yoğunluk tahmininin en basit ve yaygın olarak kullanılan yöntemi histogramdır. Histogram, örnek verilerinin kutu adı verilen eşit boyutlu aralıklara bölünmesiyle oluşturulur. Kutuların genişliği b ile gösterilirse, x noktasındaki histogramın olasılık yoğunluk tahmini Eş. 3.1'deki gibi tanımlanmaktadır (Wand ve Jones, 1995: 5):

$$\hat{f}_b(x) = \frac{x \text{ içeren kutulardaki gözlem sayısı}}{nb}. \tag{3.1}$$

Histogramı oluştururken kutu genişliği (b) ve sınıfların başlangıç noktası seçimi önemlidir (Racine, 2008: 7). Histogram, parametrik olmayan yoğunluk tahmin edicisinin çok basit bir formu olsa da bazı ciddi dezavantajları vardır. Bunlardan biri duyarlılığının kutu genişliğine bağlı olmasıdır. Bir diğeri ise birçok durumda yoğunluk fonksiyonunun düzgün bir fonksiyon olduğu varsayılır, ancak yoğunluk fonksiyonu tahmin edicisinde histogram kullanılırsa, histogramı oluşturan kutuların keskin sınırları düzgün bir yoğunluk fonksiyonu oluşturmaz (Bowman ve Azzalini, 1997: 2). Bu nedenle yoğunluk tahmin edicisi olarak histogramdan ziyade, "çekirdek" tahmin edicileri tercih edilir (Rosenblatt, 1956; Parzen, 1962). Olasılık yoğunluk fonksiyonunun tahmini çekirdek fonksiyonu kullanılarak elde edilmişse bu tahmine kısaca çekirdek yoğunluk tahmini denilmektedir. Bir başka ifade ile, istatistikte, çekirdek yoğunluk tahmini, olasılık yoğunluk tahmini için çekirdek düzgünleştirme uygulamasıdır.

İlk kez Fix ve Hodges (1951) tarafından tanıtılan çekirdek tahmin edicileri histogram ile tahminin genelleştirilmiş ve düzgünleştirilmiş bir halidir. Çekirdek tahmin edicilerinin sadeliği matematiksel izlenebilirliği beraberinde getirir, böylece araştırmacılar çok kapsamlı matematik olmadan da bu tahmin edicilerin özellikleri hakkında derinlemesine araştırma yapabilir. Özetle, çekirdek düzgünleştirme çok çeşitli önemli sorunlara basit, güvenilir ve kullanışlı yanıtlar sağlar (Wand ve Jones, 1995: 7).

Çekirdek yoğunluk tahmini, gözlem yapılan deney noktalarından yararlanılarak, rasgele değişkenin tanım bölgesinin tamamı için yoğunluk tahmininde bulunmaktır, yani gözlem bulunmayan noktalar da dahil olmak üzere, yoğunluk tahmin edilmeye çalışılır. Kısaca, çekirdek düzgünleştirmesi istatistiksel bir model oluşturmak için varsayım ileri sürmeden sadece gözlem değerlerini kullanan parametrik olmayan bir istatistiksel modelleme yöntemidir.

Başka bir ifade, çekirdek yoğunluğu tahmini olasılık dağılımını tahmin etmek için istatistiksel momentler veya spesifik olasılık yoğunluk fonksiyonları gerektirmez. Her değer için elde edilen çekirdek fonksiyonlarının birleştirilmesi ile çekirdek yoğunluğu fonksiyonu elde edilir. Yani, eldeki mevcut verinin parametrik bir olasılık yoğunluk fonksiyonunu temsil etmediği düşünüldüğünde, sadece veriyi kullanan çekirdek yoğunluğu tahmininden yararlanılır.

$X_1, X_2, \dots, X_n$ rasgele örneği X rasgele değişkeninin dağılımından birbirinden bağımsız ve aynı dağılımlı olarak alınan rasgele değişkenlerdir. X rasgele değişkeninin olasılık yoğunluk fonksiyonu $f(x)$ bilinmemektedir ve bu olasılık yoğunluk fonksiyonunun çekirdek yoğunluk tahmini Eş. 3.2'deki gibi tanımlanmaktadır (Silverman, 1986: 15; Wand ve Jones, 1995: 11):

$$\tilde{f}_{n,h}(x) = \frac{1}{n \cdot h} \sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right) = \frac{1}{n \cdot h} \sum_{i=1}^{n} K(t). \qquad (3.2)$$

Çekirdek yoğunluk tahmini fonksiyonunda, *K(t)*, $\int_{-\infty}^{+\infty} K(t)\,dt = 1$'i sağlayan bir çekirdek fonksiyonudur. *K(t)* genellikle negatif olmayan, tek modlu ve sıfır etrafında simetriktir. Bu $\tilde{f}_{n,h}(x)$'in olasılık yoğunluk fonksiyonu olmasını sağlar. Yani *K(t)*;

- $\int K(t)dt = 1$
- $K(t) = K(-t)$ olduğundan $\int tK(t)dt = 0$
- $\mu_2 = \int t^2 K(t)dt > 0$

bu koşullar sağladığında çekirdek yoğunluk tahmin edicisi, $f(x)$'in tutarlı tahmin edicisidir.

h_n, n'ye bağlı olarak değişen pozitif değerli sayı dizisini göstersin. h_n değerleri çekirdek fonksiyonunun bant genişliği, pencere genişliği veya düzgünleştirme parametresi olarak adlandırılır. Fakat literatürde n indisi kaldırılarak yok edilir ve h olarak kullanılır. Bu çalışmada da aynı yol izlenecektir. Yani aşağıdaki $\lim_{n\to\infty} h_n = 0$ ifadesi yerine $\lim_{n\to\infty} h = 0$ ifadesi kullanılmıştır. Bu diziye ait istenen özellikler,

- $\lim_{n\to\infty} h = 0,$
- $\lim_{n\to\infty} nh = \infty$

dir (Wand ve Jones, 1995: 20).

Çekirdek yoğunluk tahmini için, Eş. 3.2 yerine, $K_h(t) = (1/h)\, K(t/h)$ yeniden ölçeklendirme uygulanarak daha basit bir formül verilebilir (Wand ve Jones, 1995: 12):

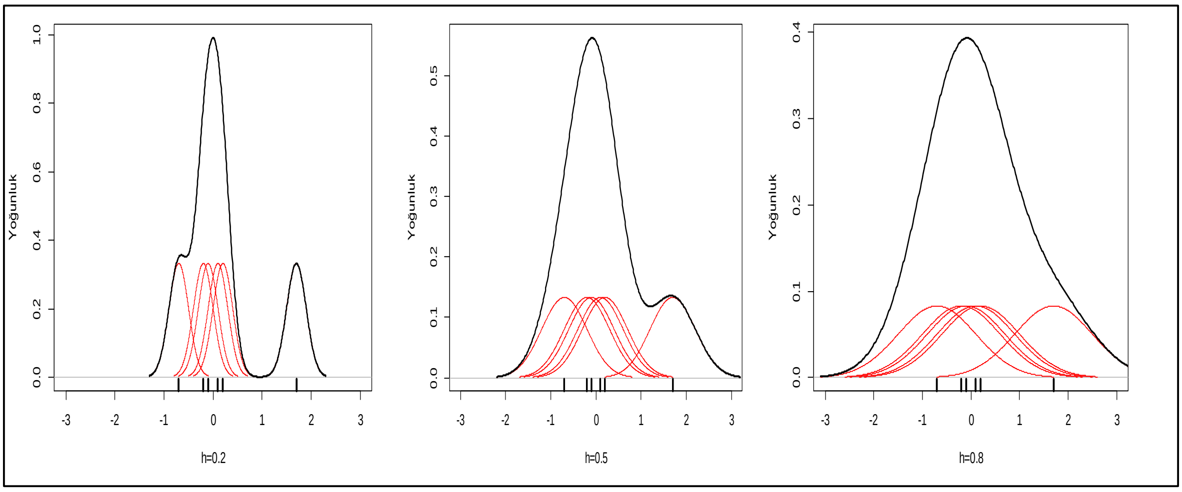
$$\tilde{f}_{n,h}(x) = \frac{1}{n}\sum_{i=1}^{n} K_h(x - X_i). \tag{3.3}$$

K çekirdek fonksiyonu düzgünleştirme işlevi gören bir fonksiyondur ve bu düzgünleştirme daha sonra da bahsedildiği gibi K ile f'nin bir konvülasyonu şeklinde ifade edilerek sağlanır.

Eş. 3.2, bir x noktasında X_i'ler verildiğinde, $\tilde{f}_{n,h}(x)$ tahmin edilmiş bir çekirdek yoğunluk fonksiyonudur. Bu eşitlikte tanımlanan çekirdek tahmin edicisi negatif değer alabilir. Bundan dolayı çekirdek yoğunluk tahmini de negatif değer alabilir. Negatif yoğunluk tahminleri istenmeyen bir etki yaratacağından, bu sorunun bir çözümü olarak daha yüksek derecelerden çekirdek fonksiyonları kullanılabilir. K fonksiyonunun pozitiflik varsayımının daha etkin tahmin ediciler bulmak için esnetilmesi fikri ilk olarak Parzen (1962) tarafından önerilmiştir. K çekirdek fonksiyonunun sürekli ve türevi alınabilir olması $\tilde{f}_{n,h}(x)$'in de sürekli ve türevi alınabilir olmasını sağlar. Bant genişliği h, n örneklem büyüklüğünün $n \to \infty$ iken, $h \to 0$ olması $x \neq X_i$ değerleri için $|t| \to \infty$ olmasını gerektir, bu da $K(-\infty) = K(\infty) = 0$ olmasına yol açar (Yolsal, 2017: 33). K için yapılan varsayımlar altında

$\int_{-\infty}^{+\infty} K(t)\, dt = 1$ olması sağlanmış olur. Bu varsayımlarla tanımlanan Eş. 3.2'deki $\tilde{f}_{n,h}(x)$, X_i'in olasılığının "Rosenblatt – Parzen çekirdek yoğunluk tahmin edicisi"dir (Hardle, 1994: 32). Burada $|t|$'nin büyük değeri için, X_i değerleri x'ten uzağa düşeceğinden, $K(t)$ küçük değerlerde olmalıdır (Pagan ve Ulah, 1999: 10).

Şekil 3.1'de çekirdek yoğunluğu tahmininin nasıl çalıştığı R paket programı (versiyon 4.2.2) ile çizilen yoğunluk grafikleri ile gösterilmeye çalışılmıştır. n=6 çaplı -0.7, -0.2, -0.1, 0.1, 0.2, 1.7 rasgele örneği için 3 ayrı bant genişliği ($h = 0.2,\ h = 0.5,\ h = 0.8$) kullanılarak çekirdek fonksiyonlarından standart normal çekirdek fonksiyonu ile çekirdek yoğunluk tahmini yapılarak yoğunluk grafiği çizilmiştir. Böylece farklı bant genişlikleri ile elde edilen çekirdek yoğunluk tahmininin düzgünleştirme dereceleri görülmektedir. Yatay eksende görülen dik koyu çizgiler 6 gözlem değeridir. Bu çizgileri merkez alan tümsekler (bump) her biri sırasıyla ilgili gözlem etrafındaki çekirdek fonksiyonlarını gösterir. Burada h bant genişliği tümseklerin genişliğini belirlerken, K çekirdek fonksiyonu tümseklerin şeklini belirler.



Şekil 3.1. Farklı bant genişlikleri için çekirdek düzgünleştirmesi.

Şekil 3.1'de görüldüğü gibi bant genişliği h küçük değerler aldığında yoğunluk tahmininin çok modlu olduğu ve h'ın değeri arttıkça yoğunluk tahminin giderek düzleştiği görülür. Bant genişliği h sıfıra yakın değerler aldığında gözlemlerde ani yükselişler olurken, h büyük değerler almaya başladığında bütün ayrıntılar belirsizleşmektedir. Bir başka ifade ile, küçük bant genişliği "eksik düzgünleştirme" (undersmoothing) yapmaktadır. Bu nedenle verinin geldiği yığının bir özetini sunmaktan uzak kalmakta ve belirgin bir form ortaya koyamamaktadır. Diğer taraftan bant genişliği büyüdükçe “aşırı düzgünleştirme” (oversmoothing) meydana gelir. Bundan dolayı verinin geldiği yığına ait olasılık yoğunluk fonksiyonunun birtakım karakteristikleri ortadan kalkar. Ayırt edici özellikleri yok edilmiş olur. Özetle, bant genişliği h’ın farklı değerleri için tümseklerin görünümü değişir ve bunun sonucu olarak, tümseklerin toplamından oluşan çekirdek yoğunluk tahmini $\tilde{f}_{n,h}(x)$’in görünümü değişir.

Eş. 3.3 ve Şekil 3.1’de görüldüğü gibi çekirdek tahmin edicileri araştırmacı tarafından belirlenen,

- Uygun çekirdek fonksiyonu *K(t)* ve
- Bant genişliği h

büyüklüklerine bağlıdır (Yolsal, 2017: 35).

3.1. Çekirdek Yoğunluk Fonksiyonlarının Başarı Ölçütleri

Elde edilen çekirdek yoğunluğu tahmininin, X rasgele değişkenin gerçekteki dağılımına ne kadar yakın ya da X rasgele değişkenin gerçekteki dağılımından ne kadar farklı olduğunu ölçmek için ortalama hata kareden (OHK) yararlanılabilir. Klasik istatistiksel yaklaşımda bir parametre tahmin edicisinin, gerçek değerine ne kadar yakın ya da uzak olduğunu OHK’ya dayalı olarak ($\hat{\theta}$, θ’nın verilen bir tahmin edicisi olmak üzere), Eş. 3.4’deki gibi ifade edilir:

$$OHK\left(\hat{\theta}\right) = E\,(\hat{\theta} - \theta)^2. \tag{3.4}$$

OHK ölçüsünün önemli bir özelliği, varyans ve yanın karesine ayrışabilmesidir:

$$OHK\,(\hat{\theta}) = Var\,(\hat{\theta}) + \left(E(\hat{\theta}) - \theta\right)^2. \tag{3.5}$$

Eğer, tahmin edici $\hat{\theta}$, θ’nın yansız tahmin edicisi ise Eş. 3.5’in sağındaki terim 0 olur ve $OHK\,(\hat{\theta})$, $\hat{\theta}$’nın varyansına eşit olur. OHK çekirdek yoğunluk tahmin edicisi için de yazılabilir.

$\tilde{f}_{n,h}(x)$'i, $x \in$R noktasında $f(x)$ olasılık yoğunluk fonksiyonunun bir çekirdek fonksiyonu kullanılarak elde edilen tahmin edicisini göstersin. OHK $(\tilde{f}_{n,h}(x))$'i hesaplamak için, $\tilde{f}_{n,h}(x)$'in ortalama ve varyansı için ifadeleri, Eş. 3.3'de verilen çekirdek yoğunluk tahmini fonksiyonu kullanılarak Eş. 3.6'daki gibi ifade edilir:

$$E\left(\tilde{f}_{n,h}(x)\right) = E\big(K_h(x - X)\big) = \int K_h(x - y) f(y)dy. \tag{3.6}$$

Burada konvülasyon tanıtılacak olursa, f ile g fonksiyonlarının konvülasyonu Eş. 3.7'deki gibi ifade edilir (Wand ve Jones, 1995: 15):

$$(f * g)(x) = \int f(x - y)g(y)dy. \tag{3.7}$$

Çekirdek fonksiyonu K_h ile bilinmeyen f olasılık yoğunluk fonksiyonunun konvülasyonunun düzgünleştirmeyi nasıl sağladığı ve işleyişi Şekil 3.1'de görülmektedir. Bu düzgünleştirmede önemli etkenlerden birisi örneklem çapı, kullanılan çekirdek fonksiyonu K ve bant genişliği h'dir. Düzgünleştirmenin verinin gözlemlendiği yığını iyi yansıtması için çok fazla düzgünleştirme yapması da gereğinden az düzgünleştirme yapması da istenmez. Optimal bir bant genişliği istenir. Optimallik tanımı tahmin edicinin varyansı ve yanlılığının dengelenerek ikisinin de toplamının en küçük yapılması anlamında OHK ile işlevsel hale getirilir. Bu çalışma bant genişliği seçiminin öneminin ortaya çıkarıldığı bir çalışmadır.

X, yoğunluğu $f(x)$ olan rastgele bir değişken olsun. $\tilde{f}_{n,h}(x)$ çekirdek yoğunluk tahmin edicisinin sırasıyla yan ve varyansı Eş. 3.7'de tanıtılan konvülasyon notasyonu ile Eş. 3.8 ve Eş. 3.9'daki gibi elde edilir:

$$E\left(\tilde{f}_{n,h}(x)\right) - f(x) = (K_h * f)(x) - f(x) \tag{3.8}$$

$$Var\left(\tilde{f}_{n,h}(x)\right) = n^{-1}\Big((K_h^2 * f)(x) - (K_h * f)^2(x)\Big). \tag{3.9}$$

Eş. 3.8 ve Eş 3.9 birleştirildiğinde çekirdek yoğunluk tahminine ilişkin OHK ölçüsü elde edilir:

$$OHK\left(\tilde{f}_{n,h}(x)\right) = n^{-1}\left((K_h^2 * f)(x) - (K_h * f)^2(x)\right) + ((K_h * f)(x) - f(x))^2. \quad (3.10)$$

Sabit bir noktada f'i tahmin etmek yerine, f'i veriyi de dikkate alarak reel sayıların tamamı üzerinden tahmin etmek istenir. Bu durumda çekirdek yoğunluk tahmin fonksiyonu $\tilde{f}_{n,h}(x)$ ve $f(x)$ fonksiyonu arasındaki uzaklık olarak ölçülen bir hata kriterinin dikkate alınması gerekir. Bu hata kriteri İntegrallenmiş Hata Kare (İHK) olarak Eş. 3.11'de tanımlanmaktadır (Wand ve Jones, 1995: 15):

$$\text{İ}HK\left(\tilde{f}_{n,h}(\cdot)\right) = \int_{-\infty}^{+\infty}\left(\tilde{f}_{n,h}(x) - f(x)\right)^2 dx. \quad (3.11)$$

Bu ölçü yalnızca eldeki veri kümesi ile ilgileniliyorsa uygundur. Olasılık yoğunluk fonksiyonu $f(x)$'ten türetilebilecek diğer olası veri kümelerini dikkate almaz. Bu nedenle, analizlerde genellikle İHK'nın beklenen değeri İntegrallenmiş Ortalama Hata Kare (İOHK) kullanılmaktadır ve İOHK,

$$\text{İ}OHK\left(\tilde{f}_{n,h}(\cdot)\right) = E\left(\text{İ}HK\left(\tilde{f}_{n,h}(\cdot)\right)\right) = E(\int_{-\infty}^{+\infty}(\tilde{f}_{n,h}(x) - f(x))^2)dx \quad (3.12)$$

Eş. 3.12'deki gibi ifade edilmektedir. Eş. 3.12'de integral ve beklenen değer operatörlerinin sırası değiştirilerek Eş. 3.13 elde edilir (Gramacki, 2018: 44):

$$\begin{aligned}\text{İ}OHK\left(\tilde{f}_{n,h}(\cdot)\right) &= \int_{-\infty}^{+\infty} E\left(\left((\tilde{f}_{n,h}(x) - f(x)\right)^2\right) dx = \int_{-\infty}^{+\infty} OHK\left(\tilde{f}_{n,h}(x)\right) dx \\ &= \int_{-\infty}^{+\infty} Var\left(\tilde{f}_{n,h}(x)\right) d\,x + \int_{-\infty}^{+\infty} Bias^2\left(\tilde{f}_{n,h}(x)\right) dx. \quad (3.13)\end{aligned}$$

İlerideki eşitlikler ve işlemlere geçilmeden önce karşımıza çıkacak bazı gösterimlerden söz edilecek olursa, a_n ve b_n gerçek sayı dizisi olsun. Sayı dizilerinin $n \to \infty$ iken birinin diğerine göre ilerleme hızını göstermek için "$O(.)$" yani büyük "O" işareti, "$o(.)$" yani küçük "o" işareti ve "$\sim$" işareti standart olarak kullanılır. Bu işaretlerle çalışmak matematiksel işlemlerin analitik olarak izlenemediği durumlarda araştırmacıya kolaylık sağlar. Özellikle açılımlarda serinin belli bir sayıdaki terimden sonra kalan terimlerin büyüklüğünü göstermek için kullanılır:

- $n \to \infty$ iken $a_n = O(b_n)$ olması demek a_n'nin b_n'nin hızının sabit belirli bir oranı içerisinde sabit bir sayıya yakınsadığını gösterir. Bu da $n \to \infty$ iken $a_n = O(b_n)$ ancak ve ancak $\lim_{n\to\infty} |a_n/b_n| < \infty$ olmasının bir ifadesidir.
- $a_n \sim b_n$, ancak ve ancak $\lim_{n\to\infty} (a_n/b_n) = 1$ olmasının ifadesidir. a_n, b_n'ye asimptotik olarak denk yani, yukarıdaki O notasyonundaki oranın 1 olması demektir. Eğer b_n terimleri işlemlere karmaşıklık getiriyorsa onun yerine daha az karmaşık olan a_n terimleri kullanılabilir.
- $n \to \infty$ iken $a_n = o(b_n)$, ancak ve ancak $\lim_{n\to\infty} |a_n/b_n| = 0$. b_n'den küçük ifadelerin sıfıra yakınsamasının ifadesidir. Bu a_n, b_n dizisinden daha hızlı bir şekilde 0'a yakınsadığını gösterir. Örneğin; a_n bir dizi olsun. b_n dizisinin terimlerinide n ifadesi ile gösterelim. $a_n = O\left(\frac{1}{n}\right) = O(n^{-1})$ a_n'nin terimleri ise 1/n'nin 0'a yakınsama hızının belirli bir oranında benzer yakınsamaya sahip olduklarını gösterir. $a_n = o(n^{-1})$ ise a_n terimlerinin 1/n'nin 0'a yakınsama hızından daha hızlı 0'a yakınsadığını gösterir. (Serfting, 1980: 1).

Çekirdek yoğunluk tahmininde asimptotik yaklaşımlar elde etmek için önemli bir matematiksel araç Taylor açılımıdır. Farz edelim ki f, R üzerinde tanımlanmış gerçek değerli bir fonksiyon ve $x \in R$ olsun. f 'nin yaklaşık $\delta > 0$ için bir $(x-\delta, x+\delta)$ aralığında p sürekli türevi olduğunu varsayalım. O zaman sıfıra yakınsayan herhangi bir α_n dizisi için Eş 3.14 elde edilir:

$$f(x+\alpha_n) = \sum_{j=0}^{p} \left(\frac{\alpha_n^j}{j!}\right) f^{(j)}(x) + o(\alpha_n^p). \tag{3.14}$$

Taylor teoremi, fonksiyonun yeterince düzgün (istenen türevlerinin var olması ve bunlara ilişkin sürekliliklere sahip olması) olması belirli bir noktaya yakın değerlerde fonksiyon değerlerine o noktada yüksek mertebeden türevleri kullanılarak doğrusal olarak yaklaşmamızı sağlar. α_n, sıfıra yakınsayan bir diziyse, $x > 0$ için Taylor teoremi şunu verir:

$$(x+\alpha_n)^{-\frac{1}{2}} = (x)^{-\frac{1}{2}} - \frac{1}{2}\alpha_n x^{-\frac{3}{2}} + \frac{3}{8}\alpha_n^2 x^{-\frac{5}{2}} + o(\alpha_n^2). \tag{3.15}$$

Bant genişliği parametresi h, yoğunluk tahmininin düzgünlüğünü kontrol eder. Bu nedenle $\dot{I}OHK\ \tilde{f}_{n,h}(\cdot)$'yi en küçük yapacak h,

$$h_{\dot{I}OHK} = \underset{h \in R^+}{\operatorname{argmin}}\, \dot{I}OHK\ \tilde{f}_{n,h}(\cdot) \tag{3.16}$$

olarak seçilmelidir. Burada amaç İOHK'yi en küçük yapacak h değerinin bulunmasıdır. Ancak $f(x)$ genel olarak bilinmediğinden İOHK kapalı bir forma sahip değildir. Bu nedenle asimptotik İOHK (AİOHK) elde edilmelidir (Gramacki, 2018: 44). Öncelikle $f(x)$'in $x \in R$ tahminini ele aldığımızda Eş. 3.6'da bir değişken değişikliği yapılırsa,

$$E\ (\tilde{f}_{n,h}(x)) = \int K(z) f(x - hz) dz \tag{3.17}$$

dır. $f(x - hz)$'yi x etrafında bir Taylor serisinde genişleterek elde ederiz:

$$f(x - hz) = f(x) - hzf'(x)\frac{1}{2}h^2 z^2 f''(x) + o(h^2). \tag{3.18}$$

Eş. 3.17 ve Eş. 3.18'den

$$E\ (\tilde{f}_{n,h}(x)) = f(x) + \frac{1}{2}h^2 f''(x) \int z^2 K(z) dz + o(h^2) \tag{3.19}$$

olur. Buradan K çekirdek fonksiyonu varsayımlarından yana yol açan $\mu_2(K) = \int z^2 K(z) dz$ notasyonu ile yanın ifadesi,

$$E\ (\tilde{f}_{n,h}(x)) - f(x) = \frac{1}{2}h^2 \mu_2(K) f''(x) + o(h^2) \tag{3.20}$$

ve varyans ifadesi,

$$\begin{aligned} Var\left(\tilde{f}_{n,h}(x)\right) &= (nh)^{-1} \int K(z)^2 f(x - hz) dz - n^{-1}\left(E\ \left(\tilde{f}_{n,h}(x)\right)\right)^2 \\ &= (nh)^{-1} \int K(z)^2 \big(f(x) + o(1)\big) dz - n^{-1}\big(f(x) + o(1)\big)^2 \\ &= (nh)^{-1} \int K(z)^2 dz\, f(x) + o((nh)^{-1}) \end{aligned} \tag{3.21}$$

dir. Herhangi bir integrallenebilir kare fonksiyon g için $R(g) = \int g(x)^2$ 'dir. Bu varyansı şu şekilde yazmamızı sağlar:

$$Var\left(\tilde{f}_{n,h}(x)\right) = (nh)^{-1}h^2R(K)f(x) + o((nh)^{-1}). \tag{3.22}$$

Eş. 3.22'deki ifadeler

$$R(K) = \int K(x)^2 dx$$
$$\mu_2(K) = \int x^2 K(x) dx$$
$$R(f'') = \int f''(x)^2 dx \tag{3.23}$$

dir. Standart normal çekirdek ve tek değişkenli durumlar için, Eş. 3.23'teki ifadeler,

$$R(K) = (2\pi^{1/2})^{-1}$$
$$\mu_2(K) = 1$$
$$R(f'') = 3(8\pi^{1/2})^{-1} \tag{3.24}$$

olur. Eş. 3.20'nin karesi ve Eş. 3.21 toplanarak Eş. 3.25 elde edilir (Gramacki, 2018: 45).

$$OHK\left(\tilde{f}_{n,h}(x)\right) = (nh)^{-1}R(K)f(x) + \frac{1}{4}h^4\mu_2(K)^2f''(x)^2 + o((nh)^{-1} + h^4) \tag{3.25}$$

ve buradan

$$\dot{I}OHK\left(\tilde{f}_{n,h}(\cdot)\right) = A\dot{I}OHK(\tilde{f}_{n,h}(\cdot)) + o((nh)^{-1} + h^4) \tag{3.26}$$

olarak yazılabilir. Buradan

$$A\dot{I}OHK\left(\tilde{f}_{n,h}(\cdot)\right) = (nh)^{-1}R(K) + \frac{1}{4}h^4\mu_2(K)^2R(f'') \tag{3.27}$$

dir. Eş. 3.26'da asimptotik İOHK (AİOHK) adı verilir (Wand ve Jones, 1995: 21). AİOHK'yi en küçük yapacak değer, Eş. 3.26'dan elde edilir. Bu da AİOHK bant genişliğini bulmaya izin verir:

$$h_{İOHK} \sim h_{AİOHK} = \left[\frac{R(K)}{n\mu_2(K)^2 R(f'')}\right]^{1/5}. \tag{3.28}$$

Eş. 3.28, Eş. 3.27'de yazıldığında,

$$AİOHK\left(\tilde{f}_{n,h}(\cdot)\right) = \frac{5}{4}\left(\mu_2(K)^2 R(K)^4 R(f'')\right)^{1/5} n^{-4/5} \tag{3.29}$$

dir. Bu K çekirdeğini kullanarak f'yi tahmin etmek için mümkün olan en küçük AİOHK'dir. $n^{-4/5}$ terimi, n'ye göre sabittir. Başka bir deyişle, AİOHK $n^{-4/5}$ oranında yakınsamaktadır.

Yakınsama oranı kavramı, örnek çapı büyüdükçe bir tahmin edicinin hedefine ne kadar "hızlı" yaklaştığını gösteren basit bir yoruma sahiptir. Ayrıca farklı tahmin edicilerin karşılaştırılması için yararlı olabilir (Wand ve Jones, 1995: 18). Ayrıca, histogramda AİOHK daha yavaş yani $n^{-2/3}$ oranında bir yakınsama yapar. Bu çekirdek yoğunluğu tahmin edicisinin histogramdan üstün olmasının bir başka nedenidir.

Eş. 3.28, bilinmeyen $f(x)$ yoğunluğuna bağlı olduğu için öncelikle $f(x)$ tahmin edilmelidir. $f(x)$'in çeşitli yollarla tahmin edildiği farklı bant genişliği seçim yöntemleri vardır.

Eş. 3.27'de, integrallenmiş yan karenin asimptotik olarak h^4 ile orantılı olduğundan bu kısmı azaltmak için h'yi olabildiğince küçük seçilmesi gerekir. Diğer taraftan integrallenmiş varyans asimptotik olarak $(nh)^{-1}$ ile orantılıdır, bu nedenle bu bölümü azaltmak için h'nin olabildiğince büyük seçilmesi gerekir. Optimum bant genişliği, AİOHK'nin minimuma indirileceği ve varyans ve yanlılık arasında en iyi dengeyi kuracak şekilde seçilmelidir (Gramacki, 2018: 46).

Eş. 3.27, Eş. 3.30'daki gibi de ifade edilebilir,

$$AİOHK\left(\tilde{f}_{n,h}(\cdot)\right) = (nh)^{-1}R(K) + \frac{1}{4}h^4\mu_2(K)^2\Psi_4 \tag{3.30}$$

ve

$$h_{İOHK} \sim h_{AİOHK} = \left[\frac{R(K)}{n\mu_2(K)^2\Psi_4}\right]^{1/5} \quad (3.31)$$

burada

$$\Psi_4 = \int_{-\infty}^{\infty} f^{(4)}(x)f(x)dx \quad (3.32)$$

veya genel olarak,

$$\Psi_r = \int_{-\infty}^{+\infty} f^{(r)}(x)f(x)dx = E(f^r(X)) \quad (3.33)$$

dir. Literatürde yalnızca İHK ve İOHK hata kriterlerinin kullanılmadığını belirtmekte fayda var. Bazen parametrik olmayan yoğunluk tahmini, İntegrallenmiş Ortalama Mutlak Hata (İOMH) kullanarak analiz edilebilir, burada İOHK kriterindeki karenin yerini mutlak değer alır, yani

$$İOMH\left(\tilde{f}_{n,h}(\cdot)\right) = \int_{-\infty}^{+\infty} \left|\tilde{f}_{n,h}(x) - f(x)\right| dx \quad (3.34)$$

ifade edilir (Gramacki, 2018: 47).

3.2. Uygun Çekirdek Fonksiyonunun Seçimi

Bir rasgele değişkenin çekirdek yoğunluk tahminini yaparken rasgele değişkenin yapısını doğru bir şekilde ortaya çıkarmak için kullanılan çekirdek fonksiyonunu seçmek, yoğunluk tahmininde daha az etkili olsa da önem taşımaktadır.

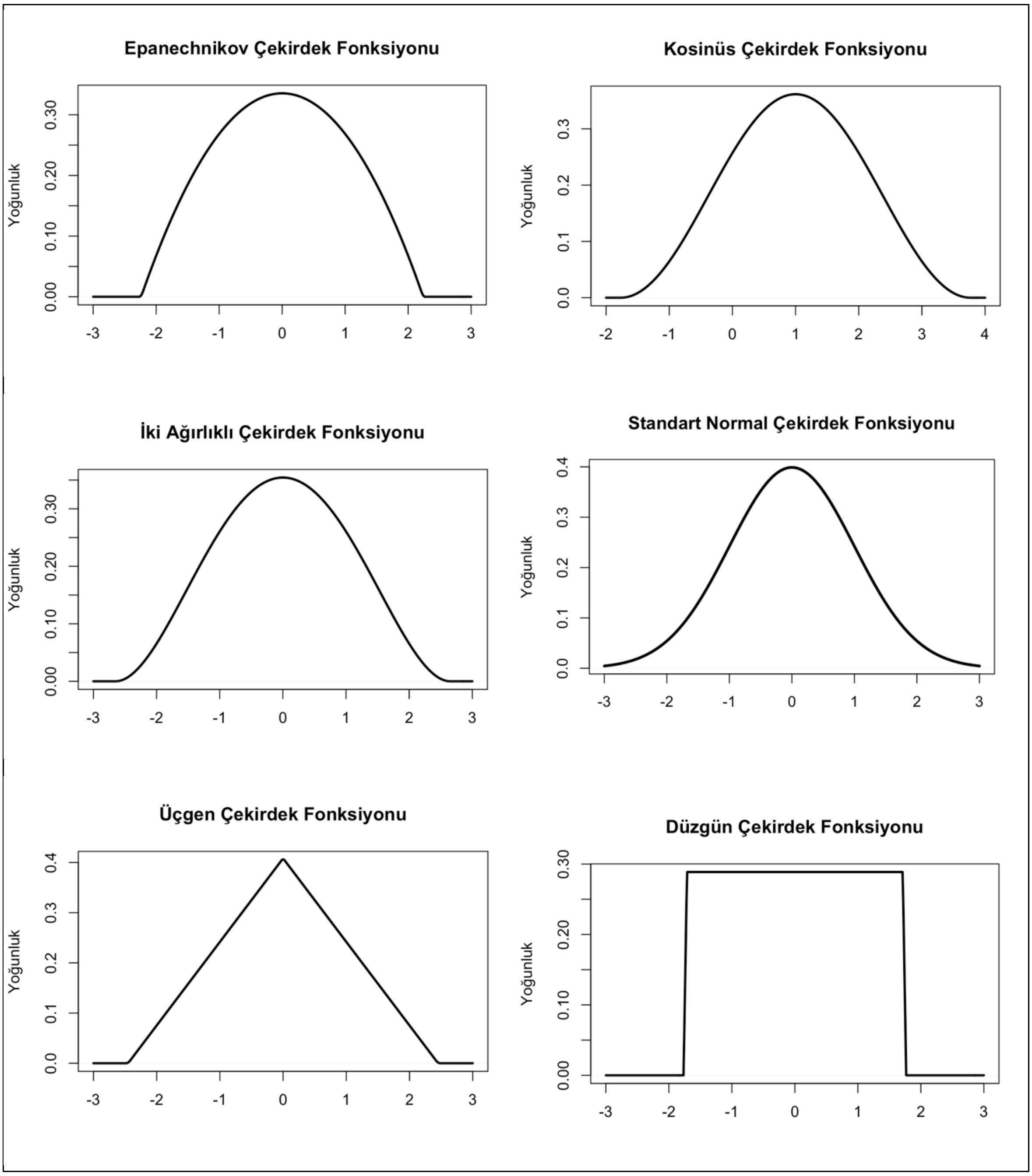
Çekirdek fonksiyonlarından en sık kullanılanları Çizelge 3.1'de etkinlikleri ile birlikte verilmektedir. Bu etkinlikler Epanechnikov çekirdek fonksiyonuna (K_e) göre hesaplanmıştır ve etkinlik Eş. 3.35'te tanımlanmıştır (Silverman 1986: 42):

$$etkinlik(K) = \left(\frac{C(K_e)}{C(K)}\right)^{5/4} = \frac{3}{5\sqrt{5}}\mu_2{}^{-1/2}\left(\int K(u)^2du\right)^{-1}. \quad (3.35)$$

Çizelge 3.1. Çekirdek düzgünleştirmesinde yaygın olarak kullanılan fonksiyonlar (Silverman, 1986: 43; Gramacki, 2018: 27).

Çekirdek Fonksiyonu	*K(x)*	Etkinlik
Epanechnikov	$K(x) = \begin{cases} \left(\frac{3}{4}\right)(1-x^2) & \lvert x\rvert \leq 1 \\ 0 & d.h. \end{cases}$	1.000
Kosinüs (Cosinüs)	$K(x) = \begin{cases} ((\frac{\pi}{4})(\cos(\frac{\pi}{2}x)) & \lvert x\rvert \leq 1 \\ 0 & d.h. \end{cases}$	0.999
İki Ağırlıklı (Biweight)	$K(x) = \begin{cases} \left(\frac{15}{16}\right)(1-x^2)^2 & \lvert x\rvert \leq 1 \\ 0 & d.h. \end{cases}$	0.994
Standart Normal (Gaussian)	$K(x) = 2\pi^{-1/2} \exp\left(\frac{1}{2}x^2\right)$	0.951
Üçgen (Triangular)	$K(x) = \begin{cases} (1-\lvert x\rvert) & \lvert x\rvert \leq 1 \\ 0 & d.h. \end{cases}$	0.986
Düzgün (Uniform)	$K(x) = \begin{cases} \left(\frac{1}{2}\right) & \lvert x\rvert \leq 1 \\ 0 & d.h. \end{cases}$	0.930

Çizelge 3.1'de verilen çekirdek fonksiyonları grafiksel olarak Şekil 3.2'de verilmektedir.



Şekil 3.2. Çekirdek düzgünleştirmesinde yaygın olarak kullanılan çekirdek fonksiyon grafikleri.

4. BANT GENİŞLİĞİ SEÇİM YÖNTEMLERİ

Bir rasgele değişkenin olasılık yoğunluk fonksiyonunun çekirdek yoğunluk tahmin edicilerinin performansı, büyük ölçüde düzgünleştirme parametresi olarak da bilinen bant genişliği seçimine bağlıdır. Bir başka ifade ile bant genişliği düzgünleştirme miktarını kontrol eden bir değerdir. Bant genişliği pek çok durumda öznel olarak seçilebilir. Öznel seçim rasgele değişkene farklı bant genişlikleri uygulanarak oluşturulan tahmin, olasılık yoğunluk fonksiyonlarının grafikleri karşılaştırılarak yapılır. Farklı bant genişlikleri denemeleri sonucunda olasılık yoğunluk fonksiyonlarının görünümüne bakarak en uygun bant genişliğine karar verilir. Ancak böyle bir yaklaşım tekrar edilebilir sonuçlar verirken iyi tahminler vermeyecektir. Ayrıca deneme yanılma yoluyla zaman alıcı ve yanılma olasılığının yüksek olduğu bir yöntemdir. Diğer bir dezavantajı ise genellikle veri yapısı hakkında önsel bilginin olmaması ve hangi bant genişliğinin gerçek yoğunluğa en yakın tahmini verdiği konusunda herhangi bir ipucunun olmamasıdır. Öznel seçimin kullanışlı olduğu tek durum, araştırmacının rasgele değişkenin modunun konumu hakkında belli bir yapının varlığına inandığı durumdur. Bu nedenle çekirdek tahmin edicilerinin bant genişliğini belirlerken genellikle otomatik seçim gereklidir (Wand ve Jones, 1995: 58).

Çok büyük bant genişliklerinin az miktarda “rastlantısallıkların” olduğu, aşırı düzleşmiş yoğunluk tahminleri vermesi beklenir. Bu durum olasılık yoğunluk fonksiyonunu gerçek şeklinden uzaklaşmasına ve tahminlerin yanlı olmasına yol açar. Küçük bant genişliklerinin ise rastlantısallıkların çok olduğu dalgalı (kıvrımlı) yoğunluk tahminleri vermesi beklenir. Bu durumda tahminin yanlılığı azalırken, varyansı artar. Bu iki durum dağılımın özelliklerini görmemizi engeller. Bu nedenle bant genişliğinin en doğru şekilde seçilmesine dikkat edilmelidir. Bant genişliği seçim yöntemleri üç sınıfa ayrılır.

İlk sınıftaki seçimler çok basit ve hesaplaması kolay matematiksel formülle geniş bir aralıktan başlanarak ve gittikçe daraltarak, optimum bant genişliğini bulmaya dayanır. Çok çeşitli durumları kapsayacak şekilde geliştirilmişlerdir, ancak sonucun optimum bant genişliğine yeterince yakın olduğunu garanti etmezler. Bu tür seçimler öznel seçimle yapılan bant genişlikleri için makul bir başlangıç noktası sağlar.

İkinci sınıftaki seçimler daha kesin bir matematiksel temele dayanan ve veri kümesinin altında yatan gerçek fonksiyona ulaşmak için çok daha fazla hesaplama gücü gerektiren seçim süreçleridir. Bu seçim türlerinden her biri İOHK'yi en küçük yapmayı amaçlar ve asimptotik olarak bu amaca ulaşırlar. Bundan dolayı İOHK'yi en küçük yaparak elde edilen bant genişliği seçim yöntemlerinin tutarlı olduğu söylenir (Yolsal, 2018: 54). Bu bant genişliği seçim yöntemleri: en küçük kareler çapraz doğrulama bazen tarafsız çapraz doğrulama olarak da adlandırılır, yanlı çapraz doğrulama ve düzleştirilmiş çapraz doğrulama yöntemleridir (Wand ve Jones, 1995: 59).

Üçüncü sınıftaki seçimler ise asimptotik olarak optimal bant genişliği için formüllerde görünen bazı bilinmeyen parametrelerin tahminlerini yerine koymaya (eklemeye) dayalı yöntemlerdir. Bu bant genişliği seçim yöntemleri ise yerine koyma ve denklem çözme olarak adlandırılırlar (Sheather, 2004; Gramacki, 2018: 65).

Aşağıda bu seçim yöntemleri kısaca tanıtılacaktır.

4.1. Normal Ölçek Bant Genişliği Seçim Yöntemi

Verinin olasılık yoğunluk fonksiyonu $f(x)$'i tahmin ederken İOHK'yi asimptotik olarak en aza indiren bant genişliği seçilmelidir. Normal Ölçek (NÖ) bant genişliği seçim yöntemi bilinmeyen olasılık yoğunluk fonksiyonun yerine bilinen bir dağılım fonksiyonu konularak geliştirilir.

Genellikle de bilinmeyen $f(x)$'in ortalaması μ ve varyansı σ^2 olan normal dağılımdan geldiği varsayılır. Bu varsayımla Eş. 3.28'de bulunan f'in ikinci türevinin karesinin integrali;

$$\int \left(f''(x)\right)^2 dx = \frac{3}{8\sqrt{\pi}\sigma^5} \approx 0.212\sigma^{-5} \tag{4.1}$$

olacaktır. Burada normal çekirdek fonksiyonu kullanıldığında,

$$\int K^2(u)du = \frac{1}{\sqrt{4\pi}} \qquad \text{ve} \int u^2 K(u)du \tag{4.2}$$

olur. Böylece bant genişliği,

$$h_{opt} = (4\pi)^{-1/10} \left(\frac{3}{8}\right)^{-1/5} \pi^{1/10} \sigma n^{-1/5} = 1.059\sigma n^{-1/5} \tag{4.3}$$

olarak elde edilir. Yaklaşık olarak ise,

$$h_{opt} = 1.06\sigma n^{-1/5} \tag{4.4}$$

olur. Burada σ yerine $\hat{\sigma} = s$ örneklem standart sapması konulursa optimal bant genişliği

$$h_{opt} = 1.06\hat{\sigma} n^{-1/5} \tag{4.5}$$

şeklinde hesaplanır (Silverman, 1986: 45; Wand ve Jones, 1995: 60).

$f(x)$'in normal dağılımdan geldiği varsayımı parametrik olmayan yoğunluk tahmini anlayışına aykırı olduğu unutulmamalıdır. X rasgele değişkeninin normal bir dağılıma sahip olduğunu bilindiğinde basitçe örnek ortalaması μ ve örnek varyansı σ^2'yi tahmin edip bu tahminleri normal yoğunluk formülüne koyarak yoğunluk çok daha kolay ve daha verimli bir şekilde tahmin edebilir. X rasgele değişkeninin dağılımı genellikle bilinmez ve böyle bir durumda, X'in dağılımı normal dağılımdan önemli ölçüde farklı değilse optimale yakın bir bant genişliği verir (Gramacki, 2018: 65).

4.2. Silverman'ın Pratik Yaklaşımı Bant Genişliği Seçim Yöntemi

Eş. 4.5, X'in dağılımı normal dağılımdan çok farklı değilse optimale yakın bir bant genişliği verir. Bu nedenle Eş. 4.5 tek modlu ve oldukça simetrik olan tüm dağılımlar için iyi sonuçlar verecek olan pratik bir bant genişliği olur. Özellikle aykırı gözlemler içeren veri kümlerinde dayanıklı (robust) tahmin edici kullanılmıştır. Yine normal dağılım varsayımı altında σ yerine örneklemden tahmin edilen standart sapması $\hat{\sigma}$ ve örneklem çeyrekler arası açıklık R üzerinden,

$$A = \min\left(\hat{\sigma}, \frac{R}{1.34}\right) \tag{4.6}$$

değerleri ile en uygun bant genişliği,

$$h_{opt} = 1.79Rn^{-1/5} \quad \text{veya} \qquad h_{opt} = 0.9An^{-1/5} \tag{4.7}$$

olarak hesaplanır (Silverman, 1986: 47; Yolsal, 2018: 57). İlk olarak B.W.Silverman tarafından önerilen bu yöntem bilinmeyen fonksiyonun yerine bilinen bir dağılımın konulması ile gerçekleştirildiğinden Silverman'ın Pratik Yaklaşımı (SPY) bant genişliği seçim yöntemi olarak adlandırılır (Silverman, 1986: 48).

4.3. Yerine Koyma Bant Genişliği Seçim Yöntemi

Sheather ve Jones (1991) tarafından önerilen Yerine Koyma (SJYK) bant genişliği seçim yöntemi özellikle hızlı yakınsama oranı ve düşük örnekleme değişkenliği bakımından hem teori hem de uygulama açısından istenen birçok özelliğe sahiptir. Bu özellikler, yerine koyma bant genişliği seçim yöntemini pratik uygulamalarda bir ilk seçenek haline getirir. Bu seçim yöntemi, AİOHK kriterine dayanmaktadır. Eş. 3.31'de AİOHK'yi en aza indiren bant genişliğinde tek bilinmeyen Ψ_4 değeri yerine bir tahmin edici yerleştirilmesine dayanır. Eşitlik 4.8'de Ψ_4 değeri $\Psi_4(g_4)$ çekirdek tahmin edici ile değiştirildiğinde,

$$\hat{h}_{DPI} = \left[\frac{\int K(u)^2 du}{n\mu_2(K)^2 \Psi_4(g_4)}\right]^{1/5} \tag{4.8}$$

elde edilir. Bu tahmin edici, değeri bilinmeyen g_4'e bağlı olduğu için doğrudan kullanılamaz. Wand ve Jones (1995)'de g_r'nin AİOHK optimal bant genişliği formülü kullanılarak hesaplanabileceği gösterilmiştir. g_r,

$$g_{r,\text{AİOHK}} = \left[\frac{2K^{(r)}(0)}{-n\mu_2(K)\Psi_{r+2}(g_{r+2})}\right]^{1/(r+3)} \tag{4.9}$$

şeklindedir. Bu Eş 4.9'da $r = 4$ yerine konulursa g_4,

$$g_{4,\text{AİOHK}} = \left[\frac{2K^{(4)}(0)}{-n\mu_2(K)\Psi_6(g_6)}\right]^{1/7} \tag{4.10}$$

elde edilir. Burada da $\Psi_6(g_6)$ bilinmeyen g_6 değerine bağlıdır. Bu durumda, g_6 benzer şekilde,

$$g_{6,AOHK} = \left[\frac{2K^{(6)}(0)}{-n\mu_2(K)\Psi_8(g_8)}\right]^{1/9} \qquad (4.11)$$

elde edilir. $\hat{h}_{DPI}$ pilot bant genişliği g'ye bağlı olduğundan tam otomatik değildir. Eş. 4.10 ve Eş. 4.11'de görüldüğü gibi dördüncü dereceden türevi tahmin etmek altıncı dereceden türeve, altıncı dereceden türevi tahmin etmek içinse sekizinci dereceden türeve ihtiyaç duyulur. Burada $\Psi_8(g_8)$ aşağıdaki normal ölçek formülüne göre hesaplandığını varsayar:

$$\Psi_r = \frac{(-1)^{r/2} r!}{(2\sigma)^{r+1}(r/2)!\pi^{1/2}}\,. \qquad (4.12)$$

Süreç bu şekilde devam ederken, yerine koymaların kaç aşamada sonlanacağı konusu da tartışmalıdır. Genellikle bant genişliğinin en az iki aşamada seçilmesi önerilir. Tüm bu aşamalarda bant genişliğini tahmin eden yerine koyma bant genişliği yöntemi oldukça maliyetli hesaplar gerektirir (Wand ve Jones, 1995: 71).

4.4. Denklem Çözme Bant Genişliği Seçim Yöntemi

Yerine koyma bant genişliği seçim yöntemine farklı bir bakış açısı getiren Denklem Çözme (SJDÇ) bant genişliği seçim yöntemi Eş. 4.8'de Ψ_4'ün tahmini için pilot bant genişliği h'nin bir γ fonksiyonu olması gibi ek bir gereklilik vardır, yani

$$\hat{h}_{DPI} = \left[\frac{\int K(u)^2 du}{n\mu_2(K)^2\Psi_4(\gamma(h))}\right]^{1/5} \qquad (4.13)$$

olur. $\gamma(h)$,

$$\gamma(h) = \frac{2K^{(4)}(0)\mu_2(K)\hat{\Psi}_4(g_4)}{\hat{\Psi}_4(g_4)R(K)} h^{5/7} \qquad (4.14)$$

dır (Gramacki, 2018: 67).

4.5. En Küçük Kareler Çapraz Doğrulama Bant Genişliği Seçim Yöntemi

En Küçük Kareler Çapraz Doğrulama (EKKÇD) bant genişliği seçim yöntemi bant genişliği seçiminde çok sık olarak kullanılan yöntemlerden biridir. Bilgisayar yoğun hesaplamalara dayalı bir teknik olup açık formülasyonu yoktur. Kabaca bir histogram çizerek başlangıç bir h değeri değerlendirilmesi yapılır. Buna göre sonraki algoritma işleyişinde bir aralıkta yer alan h değerleri altında İOHK fonksiyonunu en küçük yapan değer seçilir. Yoğunluk tahmini yapılırken $\tilde{f}_{n,h,-i}(x)$ bir X_i değeri dışarıda bıraktığı için genellikle "bir gözlem dışarı" yoğunluk tahmin edicisi olarak adlandırılır. Çapraz doğrulama denilmesinin sebebi örneklemin bir kısmını kullanarak diğer kısmı hakkında çıkarsama yapmasıdır. Bu yöntemin çıkış noktası Eş. 3.11'de tanımlanan $f(x)$ ve $\tilde{f}_{n,h}(x)$ arasında alternatif bir uzaklık ölçüsü olan İHK'dır. İHK yeniden yazılırsa,

$$\text{İ}HK\ \left(\tilde{f}_{n,h}(x)\right) = \int \left(\tilde{f}_{n,h}(x) - f(x)\right)^2 dx$$
$$= \int \tilde{f}_{n,h}(x)^2\, dx - 2\int f(x)\tilde{f}_{n,h}(x)dx + \int f(x)^2 dx \tag{4.15}$$

dır (Bowman, 1984). Amaç, İHK'yı olabildiğince küçük yapacak bir h değeri seçmektir. Ancak Eş. 4.15 h'ye bağlı değildir. Bu nedenle İHK'nin beklenen değeri alınarak,

$$\text{İ}OHK\ \left(\tilde{f}_{n,h}(x)\right) = E\left(\int \tilde{f}_{n,h}(x)^2 dx\right) - 2\left(E\left(\int \tilde{f}_{n,h}(x)f(x)dx\right)\right) + \int f(x)^2 dx \tag{4.16}$$

eşitliğinden, $\int f(x)^2 dx$ terimi h'ye bağlı olmadığından ve h'nin seçimi ve optimalitede etkili değildir. Bu nedenle ihmal edilebilir ve

$$\text{İ}OHK\ \tilde{f}_{n,h}(x) - \int f(x)^2 dx = E\int \tilde{f}_{n,h}(x)^2 dx - 2\left(E\left(\int \tilde{f}_{n,h}(x)f(x)dx\right)\right) \tag{4.17}$$

elde edilir. Eş. 4.17'nin sağ tarafı, $f(x)$'e bağlı olduğundan bilinememektedir. Ancak bunun için yansız bir *EKKÇD* tahmin edicisi olduğu gösterilebilir:

$$EKKÇD = \frac{1}{n}\sum_{i=1}^{n}\int \tilde{f}_{n,h,-i}(x)(x)^2 dx - 2n^{-1}\sum_{i=1}^{n}\tilde{f}_{n,h,-i}(x)(x_i). \tag{4.18}$$

Burada,

$$\tilde{f}_{n,h,-i}(x) = (n-1)^{-1} \sum_{j \neq i}^{n} K_h(x - X_j) \tag{4.19}$$

dir. EKKÇD değerini en aza indiren h değeri bant genişliği olarak seçilir (Wand ve Jones, 1995: 63). Hall (1983), EKKÇD bant genişliği seçim yöntemlerinin Eş. 4.20'de daha kolay hesaplanacağını ispatlamıştır:

$$EKKÇD = \int \tilde{f}_{n,h}(x)dx - 2n^{-1} \sum_{i=1}^{n} \tilde{f}_{n,h,-i}(x_i). \tag{4.20}$$

EKKÇD yöntemi sıklıkla kullanılsa da örneklem değişkenliğinin fazla olması bir dezavantajıdır. Bu nedenle EKKÇD yöntemi yoğunluk tahmininde eksik düzgünleştirmeler yapar (Terrell, 1990). Park ve Marron (1990) EKKÇD'nin yüksek miktardaki örneklem değişkenliği nedeniyle, diğer bant genişliği seçim yöntemleri ile karşılaştırıldığında neredeyse tüm yoğunluk tahminlerinde iyi bir performans vermediğini göstermişlerdir.

4.6. Yanlı Çapraz Doğrulama Bant Genişliği Seçim Yöntemi

En küçük kareler çapraz doğrulama bant genişliği seçim yönteminin örneklem değişkenliğini iyileştirmek için Scott ve Terrell (1987) Yanlı Çapraz Doğrulama (YÇD) bant genişliği seçim yöntemini öne sürmüşlerdir. Bu yöntemin EKKÇD ile arasındaki fark İHK yerine Eş. 3.27'de verilen AİOHK'ye dayalı olmasıdır. YÇD bant genişliği seçim yönteminde bilinmeyen $R(f'')$ bir tahmin edici olan $R(\tilde{f}'')$ ile değiştirilir. Bu da yanlı bir tahmin üretir (Scott ve Terrel, 1987):

$$YÇD = (nh)^{-1}R(K) + \frac{1}{4}h^4\mu_2(K)^2R(\tilde{f}''). \tag{4.21}$$

YÇD ile EKKÇD ve diğer bant genişliklerinde bahsedilen sorunların çoğuna sebep olan örneklem değişkenliği miktarını azaltır. Ayrıca asimptotik varyansı EKKÇD'nin varyansından oldukça düşüktür. Ancak varyanstaki bu düşüş yanlılıkta artışa sebep olur (Yolsal, 2017: 61). Cao ve diğerleri (1994), Jones, Marron ve Sheather (1996), Park ve Marron (1990) simülasyon çalışmaları ile YÇD'nin performansının EKKÇD'den iyi olduğunu göstermiştir.

4.7. Düzleştirilmiş Çapraz Doğrulama Bant Genişliği Seçim Yöntemi

Düzleştirilmiş Çapraz Doğrulama (DÇD) bant genişliği seçim yöntemi, İOHK'nin integrallenmiş yan kare bileşenini (İYK) tahmin etmek için g pilot bant genişliğine sahip bir çekirdek tahmin edicisi kullanması bakımından yerine koyma bant genişliği seçimine benzerdir. Bu nedenle, yöntemler benzer teorik özelliklere sahiptir. Aradaki fark, DÇD'nin asimptotik yaklaşımı yerine yanlılığın karesinin integraline dayanmasıdır. Bu nedenle asimptotik yaklaşıma daha az bağımlı olma gibi bir özelliğe sahiptir. Düzleştirilmiş çapraz doğrulama amaç fonksiyonu, $f(x)$'in bir pilot tahmin edici ile değiştirilmesiyle elde edilir:

$$\tilde{f}_{n,g}(x) = \frac{1}{n}\sum_{i=1}^{n} L_g(x - X_i). \qquad (4.22)$$

Burada $L_g(x)$'ler farklı bant genişlikleri g'lerle elde edilen çekirdek fonksiyonlarıdır. Aynı zamanda asimptotik integrallenmiş varyans $(nh)^{-1}R(K)$'yi içeren DÇD,

$$DÇD\ (h) = \frac{1}{n}\ R(K) + \dot{I}YK(h) \qquad (4.23)$$

ile İYK tahmin edilir:

$$\dot{I}YK(h)\ = n^{-2}\sum_{i=1}^{n}\sum_{j=1}^{n}\left(K_h * \tilde{f}_{n,g}(x) - \left(\tilde{f}_{n,g}(x)\right)\right)^2 dx. \qquad (4.24)$$

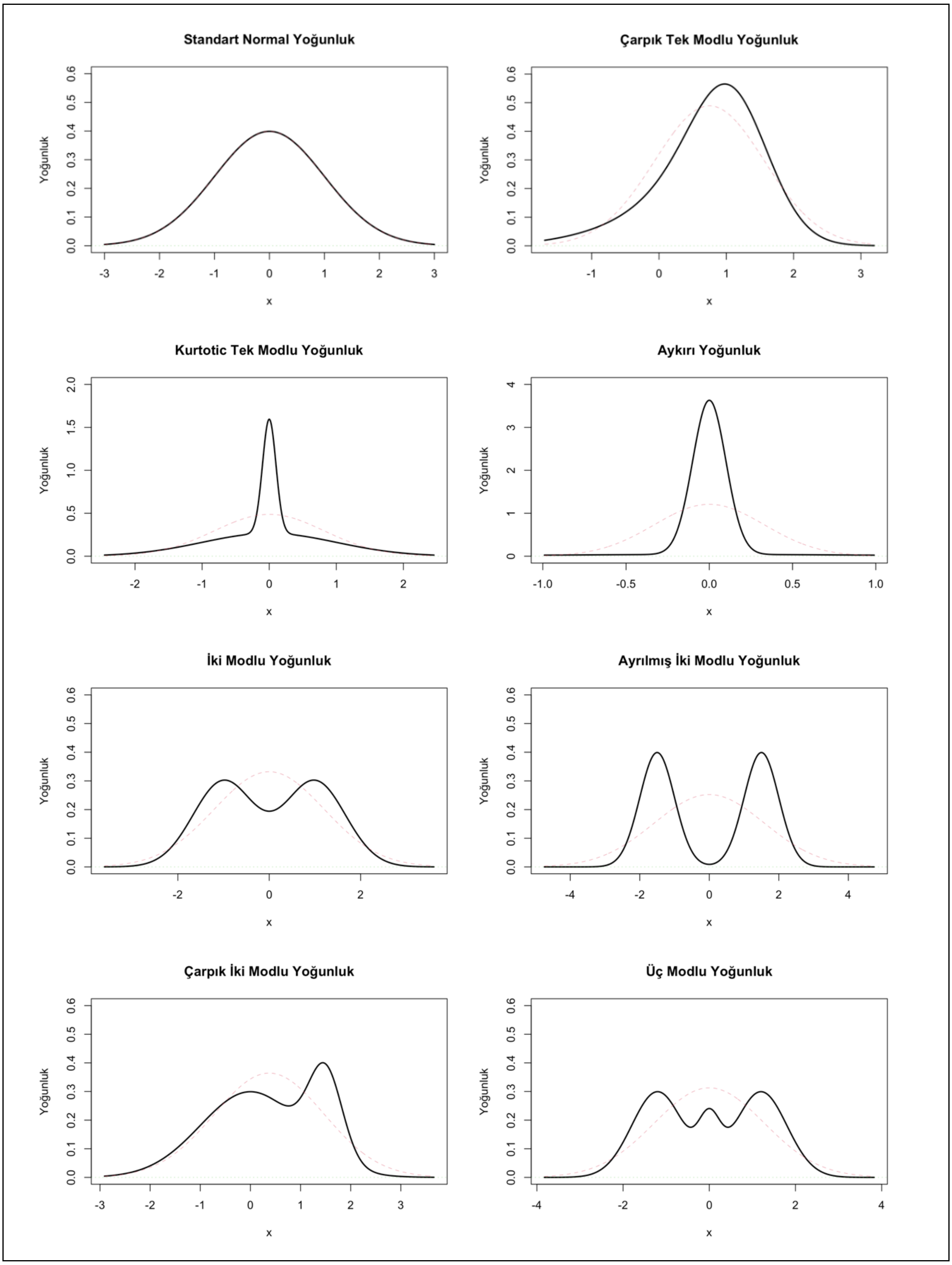
DÇD bant genişliği, $DÇD(h)$ değerinin en büyük minimumudur (Wand ve Jones, 1995: 75).

5. SİMÜLASYON ÇALIŞMASI

Bu bölümde NÖ, SPY, SJYK, SJDÇ, EKKÇD, YÇD ve DÇD bant genişliği seçim yöntemlerinin performanslarını OHK ve yan bakımından karşılaştırmak amacıyla simülasyon çalışması yapılmıştır. Bu amaçla Çizelge 5.1'de parametreleri verilen normal karışım yoğunlukları ile, R paket programı (versiyon 4.2.2) içerisinde bulunan, veriler için çekirdek düzgünleştirmede kullanılan "ks" paketi içindeki "rnorm.mixt" fonksiyonu kullanılarak 1000 tekrarlı rasgele veri üretilmiştir (Cran.r-project, 2022). Çeşitli çarpıklık, basıklık, simetri ve mod özelliklerine göre farklılık gösteren dağılımlarda bant genişliği seçim yöntemlerinin etkilerini görmek için normal karışım yoğunlukları kullanılmıştır. Bu dağılımlar gerçek verilerin farklı formlardaki dağılımlarının birer prototipleri olarak düşünülmektedir. Bu yoğunluklar Standart Normal (Gaussian), Çarpık Tek Modlu (Skewed Unimodal), Kurtotic Tek Modlu (Kurtotic Unimodal), Aykırı (Outlier), İki Modlu (Bimodal), Ayrılmış İki Modlu (Seperated Bimodal), Çarpık İki Modlu (Skewed Bimodal) ve Üç Modlu (Trimodal) dağılımı kapsamaktadır. Bu dağılımların grafikleri R paket programı (versiyon 4.2.2) içerisinde bulunan "nor1mix" paketi ile çizdirilerek Şekil 5.1'de verilmiştir (Cran.r-project, 2022; Gündüz ve Aydın, 2021).

Çizelge 5.1. Normal karışım yoğunluk parametreleri (Marron ve Wand, 1992).

Yoğunluk	$w_1 N(\mu_1 \sigma_1^2) + \cdots + w_k N(\mu_k \sigma_k^2)$
Standart Normal (Gaussian)	N(0,1)
Çarpık Tek Modlu (Skewed Unimodal)	$\frac{1}{5}\mathrm{N}(0,1) + \frac{1}{5}\mathrm{N}\left(\frac{1}{2}, (\frac{2}{3})^2\right) + \frac{3}{5}\mathrm{N}\left(\frac{13}{12}, (\frac{5}{9})^2\right)$
Kurtotic Tek Modlu (Kurtotic Unimodal)	$\frac{2}{3}\mathrm{N}(0,1) + \frac{1}{3}\mathrm{N}\left(0, (\frac{1}{10})^2\right)$
Aykırı (Outlier)	$\frac{1}{10}\mathrm{N}(0,1) + \frac{9}{10}\mathrm{N}\left(0, (\frac{1}{10})^2\right)$
İki Modlu (Bimodal)	$\frac{1}{2}\mathrm{N}\left(-1, (\frac{2}{3})^2\right) + \frac{1}{2}\mathrm{N}\left(1, (\frac{2}{3})^2\right)$
Ayrılmış İki Modlu (Separated Bimodal)	$\frac{1}{2}\mathrm{N}\left(-\frac{3}{2}, (\frac{1}{2})^2\right) + \frac{1}{2}\mathrm{N}\left(\frac{3}{2}, (\frac{1}{2})^2\right)$
Çarpık İki Modlu (Skewed Bimodal)	$\frac{3}{4}\mathrm{N}(0,1) + \frac{1}{4}\mathrm{N}\left(\frac{3}{2}, (\frac{1}{3})^2\right)$
Üç Modlu (Trimodal)	$\frac{9}{20}\mathrm{N}\left(-\frac{6}{5}, (\frac{3}{5})^2\right) + \frac{9}{20}\mathrm{N}\left(\frac{6}{5}, (\frac{3}{5})^2\right) + \frac{1}{10}\mathrm{N}\left(0, (\frac{1}{4})^2\right)$



Şekil 5.1. Normal karışım yoğunluk grafikleri.

Bant genişliği yöntemlerinin performanslarını karşılaştırmak amacıyla “ks” paketinde bulunan “Hmise.mixt” fonksiyonunu kullanılarak araştırmanın amacına uygun olarak Çizelge 5.1’de verilen çeşitli normal karışım yoğunlukları ve farklı örnek çapları için Eş. 3.27’de bulunan İOHK’yi en küçük yapan optimal bant genişlikleri ($h_{\text{İ}OHK}$) hesaplanmıştır.

Normal karışım yoğunluklarının her biri ile ilgili 7 farklı bant genişliklerini hesaplamak için R paket programı (versiyon 4.2.2) içerisinde bulunan, “stats” paketi içindeki “density” fonksiyonu kullanılmıştır (Stat.ethz, 2022). Ayrıca tüm bant genişliği hesaplamalarında standart normal çekirdek fonksiyonu kullanılmıştır.

Bant genişliklerinin farklı örnek çaplarında nasıl davrandığını görmek için Çizelge 5.1’de verilen normal karışım yoğunluklarından 5, 10, 25, 50, 100, 200 çaplı örnekler alınarak yoğunluk grafikleri oluşturulmuştur.

Optimal bant genişliğine en yakın bant genişliğini görmek için 7 bant genişliği seçim yönteminin her biri için üretilen bant genişliklerinin ortalamaları (h_{ORT}), standart sapmaları (S), yanı (YAN) ve OHK’sı hesaplanmıştır. Bu ölçütlere dair elde edilen sonuçlar Çizelge 5.2 ve 5.9 arasında sunulmuştur. Bu çizelgelerde optimal bant genişliği değerleri $h_{\text{İ}OHK}$, örnek çapı (n) ile birlikte verilmiştir. Çizelgeler sözü edilen normal karışım yoğunlukları için ayrı ayrı oluşturulmuş ve sonuçlar her örnek çapı için verilmiştir. Normal karışım yoğunlukları, standart normal yoğunluk, $h_{\text{İ}OHK}$ ve 7 farklı bant genişliği ile elde edilen çekirdek yoğunluk tahminlerini gösteren grafikler Şekil 5.2 ile 5.9 arasında gösterilmektedir. Ayrıca normal karışım yoğunlukları (standart normal ve çeşitli normal karışım yoğunlukları) için 7 farklı bant genişliği seçim yöntemine ilişkin kutu grafikleri Şekil 5.10 ile 5.17 arasında verilmektedir. Bu kutu grafiklerinin her birisinde ilgili $h_{\text{İ}OHK}$ değeri referans olarak işaretlenmiştir. Bu grafikler bize her bir yöntemin o dağılım türünde ne kadar etkili olduğunu gösterir.

Çizelge 5.2. $N(0,1)$ modeli ile tanımlanan standart normal yoğunluk ve bu dağılımdan alınan çeşitli örnek çapları için elde edilen optimal bant genişliği ($h_{\text{İ}OHK}$) ile 1000 tekrarla üretilen 7 farklı bant genişliği seçim yöntemlerinin ortalamaları (h_{ORT}), standart sapmaları (S), yanları (YAN) ve ortalama hata kareleri (OHK).

n=5 $h_{\text{İ}OHK}$=0,903					n=10 $h_{\text{İ}OHK}$=0,758				
YÖNTEM	h_{ORT}	S	YAN	OHK	YÖNTEM	h_{ORT}	S	YAN	OHK
NÖ	0,521	0,295	-0,382	0,233	NÖ	0,547	0,176	-0,212	0,076
SPY	0,442	0,250	-0,461	0,275	SPY	0,464	0,149	-0,294	0,109
SJYK	0,485	0,291	-0,418	0,259	SJYK	0,529	0,196	-0,230	0,091
SJDÇ	0,397	0,292	-0,506	0,341	SJDÇ	0,479	0,218	-0,279	0,125
EKKÇD	0,673	0,327	-0,230	0,160	EKKÇD	0,611	0,216	-0,147	0,069
YÇD	0,768	0,288	-0,135	0,101	YÇD	0,696	0,166	-0,062	0,031
DÇD	0,708	0,320	-0,195	0,140	DÇD	0,641	0,197	-0,118	0,053
n=25 $h_{\text{İ}OHK}$=0,609					n=50 $h_{\text{İ}OHK}$=0,520				
YÖNTEM	h_{ORT}	S	YAN	OHK	YÖNTEM	h_{ORT}	S	YAN	OHK
NÖ	0,497	0,098	-0,112	0,022	NÖ	0,453	0,061	-0,067	0,008
SPY	0,422	0,083	-0,187	0,042	SPY	0,384	0,052	-0,136	0,021
SJYK	0,491	0,120	-0,118	0,028	SJYK	0,447	0,082	-0,073	0,012
SJDÇ	0,465	0,142	-0,144	0,041	SJDÇ	0,431	0,098	-0,089	0,017
EKKÇD	0,514	0,157	-0,096	0,034	EKKÇD	0,444	0,128	-0,076	0,022
YÇD	0,596	0,085	-0,014	0,007	YÇD	0,517	0,052	-0,003	0,003
DÇD	0,547	0,112	-0,063	0,016	DÇD	0,477	0,078	-0,043	0,008
n=100 $h_{\text{İ}OHK}$=0,445					n=200 $h_{\text{İ}OHK}$=0,383				
YÖNTEM	h_{ORT}	S	YAN	OHK	YÖNTEM	h_{ORT}	S	YAN	OHK
NÖ	0,404	0,038	-0,042	0,003	NÖ	0,358	0,023	-0,025	0,001
SPY	0,343	0,032	-0,103	0,012	SPY	0,304	0,019	-0,079	0,007
SJYK	0,400	0,056	-0,045	0,005	SJYK	0,356	0,037	-0,027	0,002
SJDÇ	0,391	0,067	-0,054	0,007	SJDÇ	0,351	0,043	-0,032	0,003
EKKÇD	0,389	0,105	-0,056	0,014	EKKÇD	0,346	0,082	-0,037	0,008
YÇD	0,449	0,034	0,003	0,001	YÇD	0,391	0,022	0,008	0,001
DÇD	0,418	0,054	-0,027	0,004	DÇD	0,367	0,035	-0,016	0,001

Çizelge 5.3. $\frac{1}{5}\mathrm{N}(0,1)+\frac{1}{5}\mathrm{N}\left(\frac{1}{2},(\frac{2}{3})^2\right)+\frac{3}{5}\mathrm{N}\left(\frac{13}{12},(\frac{5}{9})^2\right)$ modeli ile tanımlanan çarpık tek modlu yoğunluk olarak adlandırılan normal karışım yoğunluğundan alınan çeşitli örnek çapları için elde edilen optimal bant genişliği ($h_{İOHK}$) ile 1000 tekrarla üretilen 7 farklı bant genişliği seçim yöntemlerinin ortalamaları (h_{ORT}), standart sapmaları (S), yanları (YAN) ve ortalama hata kareleri (OHK).

n=5 $h_{İOHK}$=0,898					n=10 $h_{İOHK}$=0,743				
YÖNTEM	h_{ORT}	S	YAN	OHK	YÖNTEM	h_{ORT}	S	YAN	OHK
NÖ	0,537	0,300	-0,362	0,220	NÖ	0,558	0,187	-0,185	0,069
SPY	0,456	0,254	-0,443	0,261	SPY	0,474	0,159	-0,269	0,098
SJYK	0,498	0,293	-0,401	0,246	SJYK	0,535	0,201	-0,208	0,084
SJDÇ	0,410	0,292	-0,488	0,324	SJDÇ	0,483	0,223	-0,260	0,117
EKKÇD	0,692	0,341	-0,206	0,159	EKKÇD	0,632	0,231	-0,111	0,066
YÇD	0,793	0,313	-0,105	0,109	YÇD	0,733	0,185	-0,009	0,034
DÇD	0,723	0,329	-0,175	0,139	DÇD	0,658	0,202	-0,084	0,048
n=25 $h_{İOHK}$=0,584					n=50 $h_{İOHK}$=0,492				
YÖNTEM	h_{ORT}	S	YAN	OHK	YÖNTEM	h_{ORT}	S	YAN	OHK
NÖ	0,506	0,109	-0,078	0,018	NÖ	0,464	0,069	-0,027	0,005
SPY	0,430	0,092	-0,154	0,032	SPY	0,394	0,058	-0,097	0,013
SJYK	0,489	0,125	-0,096	0,025	SJYK	0,443	0,085	-0,048	0,010
SJDÇ	0,460	0,143	-0,124	0,036	SJDÇ	0,424	0,100	-0,068	0,015
EKKÇD	0,528	0,162	-0,056	0,029	EKKÇD	0,453	0,133	-0,038	0,019
YÇD	0,622	0,094	0,038	0,010	YÇD	0,541	0,061	0,049	0,006
DÇD	0,550	0,117	-0,035	0,015	DÇD	0,476	0,080	-0,016	0,007
n=100 $h_{İOHK}$=0,417					n=200 $h_{İOHK}$=0,355				
YÖNTEM	h_{ORT}	S	YAN	OHK	YÖNTEM	h_{ORT}	S	YAN	OHK
NÖ	0,410	0,044	-0,007	0,002	NÖ	0,366	0,028	0,012	0,001
SPY	0,348	0,037	-0,069	0,006	SPY	0,311	0,023	-0,044	0,002
SJYK	0,390	0,057	-0,027	0,004	SJYK	0,344	0,038	-0,011	0,002
SJDÇ	0,379	0,066	-0,038	0,006	SJDÇ	0,336	0,045	-0019	0,002
EKKÇD	0,394	0,106	-0,022	0,012	EKKÇD	0,340	0,088	-0,014	0,008
YÇD	0,461	0,042	0,044	0,004	YÇD	0,393	0,033	0,038	0,003
DÇD	0,410	0,054	-0,006	0,003	DÇD	0,367	0,037	0,002	0,001

Çizelge 5.4. $\frac{2}{3}\mathrm{N}(0,1)+\frac{1}{3}\mathrm{N}\left(0,(\frac{1}{10})^2\right)$ modeli ile tanımlanan kurtotic tek modlu yoğunluk olarak adlandırılan normal karışım yoğunluğundan alınan çeşitli örnek çapları için elde edilen optimal bant genişliği ($h_{\text{İ}OHK}$) ile 1000 tekrarla üretilen 7 farklı bant genişliği seçim yöntemlerinin ortalamaları (h_{ORT}), standart sapmaları (S), yanları (YAN) ve ortalama hata kareleri (OHK).

n=5 $h_{\text{İ}OHK}$=0,374					n=10 $h_{\text{İ}OHK}$=0,198				
YÖNTEM	h_{ORT}	S	YAN	OHK	YÖNTEM	h_{ORT}	S	YAN	OHK
NÖ	0,335	0,248	-0,038	0,063	NÖ	0,336	0,175	0,138	0,050
SPY	0,285	0,211	-0,089	0,052	SPY	0,285	0,149	0,087	0,030
SJYK	0,305	0,241	-0,069	0,063	SJYK	0,297	0,177	0,099	0,041
SJDÇ	0,240	0,223	-0,134	0,068	SJDÇ	0,249	0,173	0,051	0,033
EKKÇD	0,467	0,310	0,093	0,104	EKKÇD	0,355	0,227	0,157	0,076
YÇD	0,591	0,294	0,217	0,133	YÇD	0,558	0,180	0,360	0,161
DÇD	0,513	0,304	0,139	0,112	DÇD	0,436	0,184	0,238	0,090
n=25 $h_{\text{İ}OHK}$=0,124					n=50 $h_{\text{İ}OHK}$=0,097				
YÖNTEM	h_{ORT}	S	YAN	OHK	YÖNTEM	h_{ORT}	S	YAN	OHK
NÖ	0,277	0,116	0,153	0,037	NÖ	0,241	0,078	0,144	0,027
SPY	0,235	0,099	0,111	0,022	SPY	0,204	0,066	0,108	0,016
SJYK	0,225	0,111	0,101	0,023	SJYK	0,172	0,066	0,075	0,010
SJDÇ	0,186	0,105	0,062	0,015	SJDÇ	0,134	0,054	0,037	0,004
EKKÇD	0,208	0,139	0,084	0,026	EKKÇD	0,116	0,059	0,020	0,004
YÇD	0,483	0,092	0,359	0,137	YÇD	0,421	0,061	0,324	0,109
DÇD	0,321	0,100	0,197	0,049	DÇD	0,243	0,059	0,146	0,025
n=100 $h_{\text{İ}OHK}$=0,079					n=200 $h_{\text{İ}OHK}$=0,065				
YÖNTEM	h_{ORT}	S	YAN	OHK	YÖNTEM	h_{ORT}	S	YAN	OHK
NÖ	0,202	0,050	0,123	0,018	NÖ	0,177	0,034	0,112	0,014
SPY	0,171	0,042	0,093	0,010	SPY	0,150	0,029	0,085	0,008
SJYK	0,125	0,033	0,047	0,003	SJYK	0,098	0,018	0,033	0,001
SJDÇ	0,098	0,024	0,020	0,001	SJDÇ	0,077	0,012	0,012	0,000
EKKÇD	0,081	0,026	0,003	0,001	EKKÇD	0,065	0,015	0,000	0,000
YÇD	0,357	0,062	0,279	0,081	YÇD	0,267	0,100	0,202	0,051
DÇD	0,178	0,029	0,100	0,011	DÇD	0,135	0,016	0,070	0,005

Çizelge 5.5. $\frac{1}{10}\mathrm{N}(0,1)+\frac{9}{10}\mathrm{N}\left(0,(\frac{1}{10})^2\right)$ modeli ile tanımlanan aykırı yoğunluk olarak adlandırılan normal karışım yoğunluğundan alınan çeşitli örnek çapları için elde edilen optimal bant genişliği ($h_{\dot{I}OHK}$) ile 1000 tekrarla üretilen 7 farklı bant genişliği seçim yöntemlerinin ortalamaları (h_{ORT}), standart sapmaları (S), yanları (YAN) ve ortalama hata kareleri (OHK).

n=5, $h_{\dot{I}OHK}$=0,804					n=10, $h_{\dot{I}OHK}$=0,654				
YÖNTEM	h_{ORT}	S	YAN	OHK	YÖNTEM	h_{ORT}	S	YAN	OHK
NÖ	0,464	0,281	-0,340	0,195	NÖ	0,482	0,180	-0,172	0,062
SPY	0,394	0,239	-0,410	0,225	SPY	0,410	0,153	-0,245	0,083
SJYK	0,433	0,280	-0,371	0,216	SJYK	0,465	0,197	-0,189	0,075
SJDÇ	0,354	0,277	-0,450	0,279	SJDÇ	0,419	0,214	-0,235	0,101
EKKÇD	0,621	0,325	-0,183	0,139	EKKÇD	0,551	0,219	-0,103	0,058
YÇD	0,718	0,292	-0,086	0,093	YÇD	0,655	0,167	0,001	0,028
DÇD	0,656	0,322	-0,148	0,125	DÇD	0,588	0,193	-0,066	0,042
n=25, $h_{\dot{I}OHK}$=0,483					n=50, $h_{\dot{I}OHK}$=0,351				
YÖNTEM	h_{ORT}	S	YAN	OHK	YÖNTEM	h_{ORT}	S	YAN	OHK
NÖ	0,437	0,105	-0,046	0,013	NÖ	0,402	0,071	0,051	0,008
SPY	0,371	0,089	-0,112	0,020	SPY	0,342	0,060	-0,010	0,004
SJYK	0,422	0,128	-0,061	0,020	SJYK	0,376	0,092	0,025	0,009
SJDÇ	0,394	0,147	-0,089	0,030	SJDÇ	0,350	0,108	-0,001	0,012
EKKÇD	0,443	0,172	-0,040	0,031	EKKÇD	0,354	0,148	0,003	0,022
YÇD	0,564	0,086	0,081	0,014	YÇD	0,488	0,055	0,137	0,022
DÇD	0,491	0,114	0,008	0,013	DÇD	0,415	0,082	0,064	0,011
n=100, $h_{\dot{I}OHK}$=0,218					n=200, $h_{\dot{I}OHK}$=0,146				
YÖNTEM	h_{ORT}	S	YAN	OHK	YÖNTEM	h_{ORT}	S	YAN	OHK
NÖ	0,358	0,046	0,139	0,022	NÖ	0,318	0,030	0,172	0,031
SPY	0,304	0,039	0,085	0,009	SPY	0,270	0,025	0,124	0,016
SJYK	0,319	0,067	0,101	0,015	SJYK	0,268	0,047	0,122	0,017
SJDÇ	0,294	0,082	0,076	0,012	SJDÇ	0,243	0,059	0,097	0,013
EKKÇD	0,259	0,123	0,041	0,017	EKKÇD	0,182	0,092	0,036	0,010
YÇD	0,422	0,040	0,203	0,043	YÇD	0,363	0,032	0,217	0,048
DÇD	0,346	0,061	0,127	0,020	DÇD	0,287	0,044	0,141	0,022

Çizelge 5.6. $\frac{1}{2}\mathrm{N}\left(-1,(\frac{2}{3})^2\right)+\frac{1}{2}\mathrm{N}\left(1,(\frac{2}{3})^2\right)$modeli ile tanımlanan iki modlu yoğunluk olarak adlandırılan normal karışım yoğunluğundan alınan çeşitli örnek çapları için elde edilen optimal bant genişliği ($h_{\dot{I}OHK}$) ile 1000 tekrarla üretilen 7 farklı bant genişliği seçim yöntemlerinin ortalamaları (h_{ORT}), standart sapmaları (S), yanları (YAN) ve ortalama hata kareleri (OHK).

n=5 $h_{\dot{I}OHK}$=1,149						n=10 $h_{\dot{I}OHK}$=0,899				
YÖNTEM	h_{ORT}	S	YAN	OHK		YÖNTEM	h_{ORT}	S	YAN	OHK
NÖ	0,698	0,333	-0,452	0,315		NÖ	0,742	0,183	-0,156	0,058
SPY	0,592	0,283	-0,557	0,390		SPY	0,630	0,156	-0,268	0,096
SJYK	0,624	0,302	-0,525	0,367		SJYK	0,670	0,195	-0,229	0,090
SJDÇ	0,492	0,302	-0,657	0,523		SJDÇ	0,584	0,231	-0,315	0,152
EKKÇD	0,802	0,340	-0,348	0,236		EKKÇD	0,724	0,245	-0,175	0,090
YÇD	0,945	0,279	-0,204	0,120		YÇD	0,858	0,153	-0,041	0,025
DÇD	0,848	0,308	-0,302	0,185		DÇD	0,754	0,200	-0,145	0,061
n=25 $h_{\dot{I}OHK}$=0,603						n=50 $h_{\dot{I}OHK}$=0,472				
YÖNTEM	h_{ORT}	S	YAN	OHK		YÖNTEM	h_{ORT}	S	YAN	OHK
NÖ	0,654	0,084	0,051	0,010		NÖ	0,581	0,045	0,109	0,014
SPY	0,555	0,072	-0,048	0,007		SPY	0,493	0,038	0,021	0,002
SJYK	0,580	0,115	-0,023	0,014		SJYK	0,500	0,080	0,027	0,007
SJDÇ	0,522	0,149	-0,081	0,029		SJDÇ	0,456	0,105	-0,016	0,011
EKKÇD	0,570	0,183	-0,032	0,034		EKKÇD	0,474	0,153	0,002	0,023
YÇD	0,713	0,078	0,111	0,018		YÇD	0,624	0,048	0,152	0,025
DÇD	0,606	0,125	0,003	0,016		DÇD	0,510	0,090	0,038	0,009
n=100 $h_{\dot{I}OHK}$=0,385						n=200 $h_{\dot{I}OHK}$=0,322				
YÖNTEM	h_{ORT}	S	YAN	OHK		YÖNTEM	h_{ORT}	S	YAN	OHK
NÖ	0,506	0,027	0,121	0,015		NÖ	0,441	0,016	0,120	0,015
SPY	0,430	0,023	0,044	0,002		SPY	0,375	0,014	0,053	0,003
SJYK	0,420	0,058	0,034	0,005		SJYK	0,344	0,038	-0,011	0,002
SJDÇ	0,388	0,073	0,002	0,005		SJDÇ	0,328	0,047	0,007	0,002
EKKÇD	0,395	0,118	0,010	0,014		EKKÇD	0,332	0,086	0,010	0,008
YÇD	0,535	0,044	0,149	0,024		YÇD	0,425	0,059	0,103	0,014
DÇD	0,424	0,064	0,038	0,006		DÇD	0,352	0,043	0,030	0,003

Çizelge 5.7. $\frac{1}{2}\mathrm{N}\left(-\frac{3}{2},(\frac{1}{2})^2\right)+\frac{1}{2}\mathrm{N}\left(\frac{3}{2},(\frac{1}{2})^2\right)$ modeli ile tanımlanan ayrılmış iki modlu yoğunluk olarak adlandırılan normal karışım yoğunluğundan alınan çeşitli örnek çapları için elde edilen optimal bant genişliği ($h_{\dot{I}OHK}$) ile 1000 tekrarla üretilen 7 farklı bant genişliği seçim yöntemlerinin ortalamaları (h_{ORT}), standart sapmaları (S), yanları (YAN) ve ortalama hata kareleri (OHK).

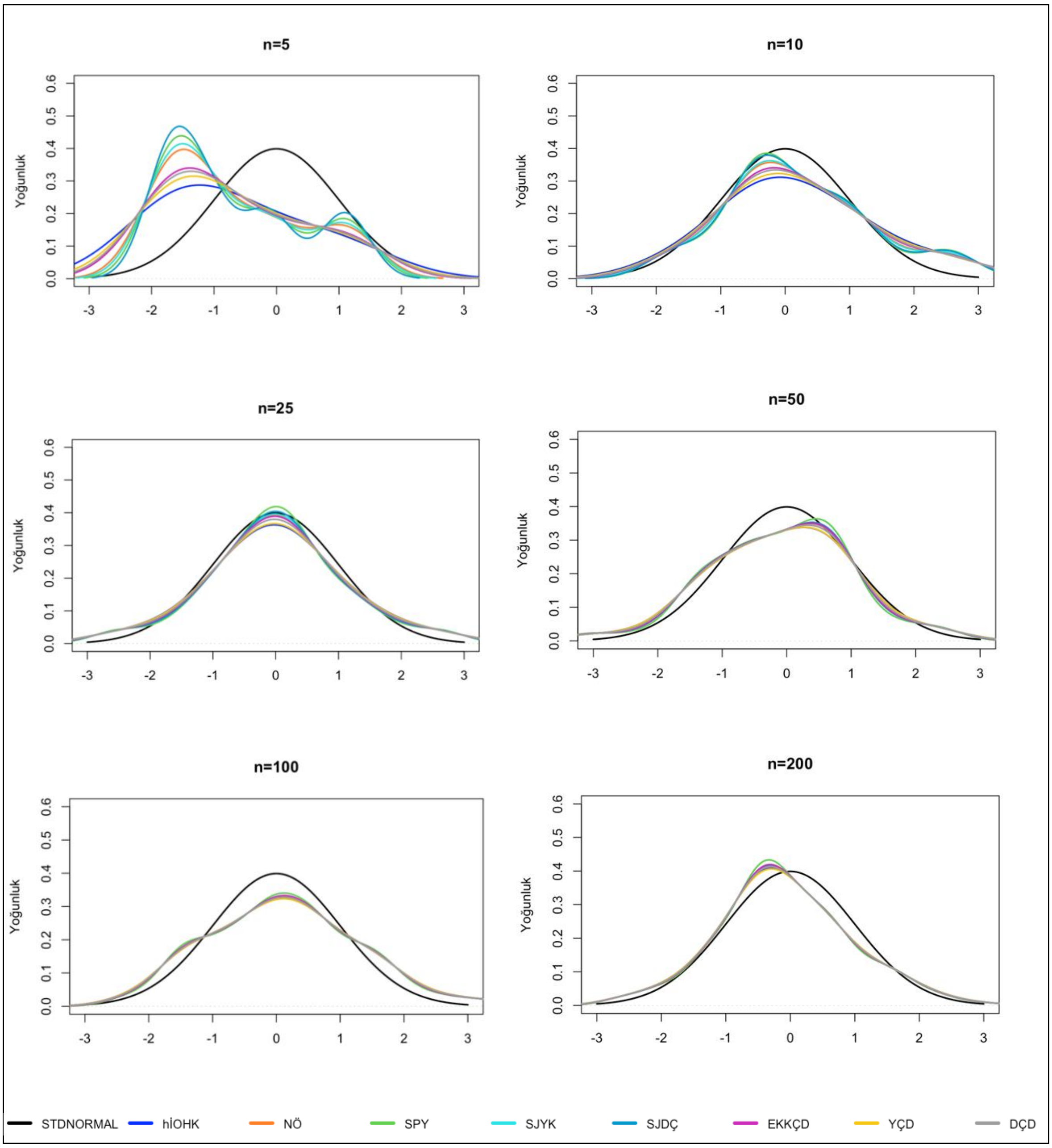
n=5 $h_{\dot{I}OHK}$=0,586					n=10 $h_{\dot{I}OHK}$=0,469				
YÖNTEM	h_{ORT}	S	YAN	OHK	YÖNTEM	h_{ORT}	S	YAN	OHK
NÖ	0,929	0,498	0,344	0,366	NÖ	1,005	0,235	0,535	0,341
SPY	0,789	0,423	0,203	0,220	SPY	0,853	0,199	0,383	0,187
SJYK	0,700	0,355	0,115	0,139	SJYK	0,658	0,153	0,189	0,059
SJDÇ	0,465	0,255	-0,120	0,080	SJDÇ	0,477	0,128	0,008	0,016
EKKÇD	0,804	0,423	0,218	0,226	EKKÇD	0,576	0,238	0,107	0,068
YÇD	1,261	0,315	0,675	0,555	YÇD	1,136	0,140	0,666	0,464
DÇD	0,899	0,282	0,314	0,178	DÇD	0,689	0,132	0,219	0,066
n=25 $h_{\dot{I}OHK}$=0,366					n=50 $h_{\dot{I}OHK}$=0,308				
YÖNTEM	h_{ORT}	S	YAN	OHK	YÖNTEM	h_{ORT}	S	YAN	OHK
NÖ	0,870	0,086	0,503	0,261	NÖ	0,766	0,036	0,458	0,211
SPY	0,738	0,073	0,372	0,144	SPY	0,651	0,030	0,342	0,118
SJYK	0,492	0,054	0,126	0,019	SJYK	0,394	0,028	0,086	0,008
SJDÇ	0,388	0,053	0,022	0,003	SJDÇ	0,330	0,032	0,021	0,001
EKKÇD	0,384	0,113	0,018	0,013	EKKÇD	0,312	0,082	0,004	0,007
YÇD	0,920	0,121	0,554	0,321	YÇD	0,436	0,141	0,128	0,036
DÇD	0,482	0,053	0,116	0,016	DÇD	0,381	0,028	0,073	0,006
n=100 $h_{\dot{I}OHK}$=0,262					n=200 $h_{\dot{I}OHK}$=0,224				
YÖNTEM	h_{ORT}	S	YAN	OHK	YÖNTEM	h_{ORT}	S	YAN	OHK
NÖ	0,667	0,022	0,405	0,164	NÖ	0,581	0,013	0,357	0,128
SPY	0,566	0,018	0,304	0,093	SPY	0,493	0,011	0,269	0,073
SJYK	0,319	0,017	0,057	0,004	SJYK	0,262	0,011	0,039	0,002
SJDÇ	0,278	0,021	0,017	0,001	SJDÇ	0,237	0,014	0,014	0,000
EKKÇD	0,261	0,060	-0,001	0,004	EKKÇD	0,223	0,046	-0,001	0,002
YÇD	0,304	0,028	0,043	0,003	YÇD	0,247	0,019	0,023	0,001
DÇD	0,307	0,017	0,046	0,002	DÇD	0,253	0,011	0,030	0,001

Çizelge 5.8. $\frac{3}{4}\mathrm{N}(0,1)+\frac{1}{4}\mathrm{N}\left(\frac{3}{2},(\frac{1}{3})^2\right)$modeli ile tanımlanan çarpık iki modlu yoğunluk olarak adlandırılan normal karışım yoğunluğundan alınan çeşitli örnek çapları için elde edilen optimal bant genişliği ($h_{İOHK}$) ile 1000 tekrarla üretilen 7 farklı bant genişliği seçim yöntemlerinin ortalamaları (h_{ORT}), standart sapmaları (S), yanları (YAN) ve ortalama hata kareleri (OHK).

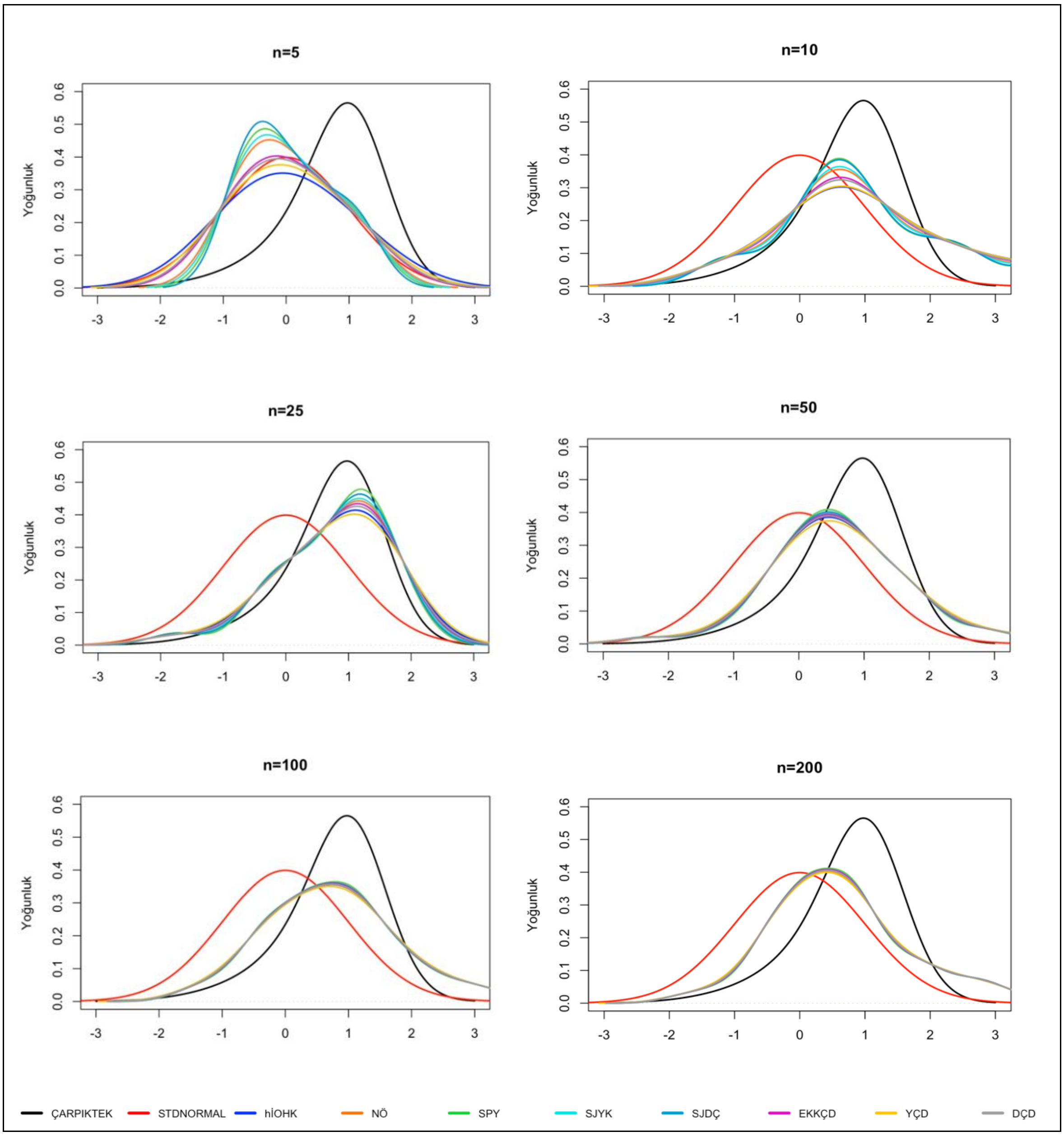
n=5 $h_{İOHK}$=1,017					n=10 $h_{İOHK}$=0,817				
YÖNTEM	h_{ORT}	S	YAN	OHK	YÖNTEM	h_{ORT}	S	YAN	OHK
NÖ	0,619	0,309	-0,398	0,254	NÖ	0,651	0,181	-0,166	0,061
SPY	0,525	0,262	-0,492	0,310	SPY	0,553	0,154	-0,265	0,094
SJYK	0,558	0,292	-0,459	0,296	SJYK	0,599	0,193	-0,218	0,085
SJDÇ	0,445	0,294	-0,571	0,413	SJDÇ	0,524	0,221	-0,293	0,135
EKKÇD	0,739	0,332	-0,278	0,188	EKKÇD	0,641	0,236	-0,176	0,087
YÇD	0,850	0,284	-0,167	0,108	YÇD	0,771	0,164	-0,046	0,029
DÇD	0,768	0,315	-0,249	0,161	DÇD	0,687	0,195	-0,130	0,055
n=25 $h_{İOHK}$=0,555					n=50 $h_{İOHK}$=0,408				
YÖNTEM	h_{ORT}	S	YAN	OHK	YÖNTEM	h_{ORT}	S	YAN	OHK
NÖ	0,587	0,089	0,032	0,009	NÖ	0,525	0,047	0,117	0,016
SPY	0,498	0,076	-0,057	0,009	SPY	0,446	0,040	0,038	0,003
SJYK	0,533	0,110	-0,022	0,013	SJYK	0,459	0,073	0,050	0,008
SJDÇ	0,482	0,142	-0,073	0,025	SJDÇ	0,415	0,099	0,007	0,010
EKKÇD	0,515	0,176	-0,040	0,033	EKKÇD	0,421	0,146	0,013	0,021
YÇD	0,653	0,082	0,098	0,016	YÇD	0,568	0,050	0,159	0,028
DÇD	0,565	0,116	0,011	0,014	DÇD	0,472	0,081	0,064	0,011
n=100 $h_{İOHK}$=0,318					n=200 $h_{İOHK}$=0,258				
YÖNTEM	h_{ORT}	S	YAN	OHK	YÖNTEM	h_{ORT}	S	YAN	OHK
NÖ	0,461	0,029	0,143	0,021	NÖ	0,403	0,017	0,146	0,022
SPY	0,391	0,024	0,073	0,006	SPY	0,342	0,014	0,085	0,007
SJYK	0,380	0,052	0,062	0,007	SJYK	0,314	0,036	0,056	0,005
SJDÇ	0,342	0,069	0,024	0,005	SJDÇ	0,283	0,046	0,026	0,003
EKKÇD	0,333	0,109	0,015	0,012	EKKÇD	0,266	0,081	0,009	0,007
YÇD	0,493	0,034	0,176	0,032	YÇD	0,416	0,046	0,159	0,027
DÇD	0,386	0,058	0,068	0,008	DÇD	0,316	0,041	0,059	0,005

Çizelge 5.9. $\frac{9}{20}\mathrm{N}\left(-\frac{6}{5},(\frac{3}{5})^2\right)+\frac{9}{20}\mathrm{N}\left(\frac{6}{5},(\frac{3}{5})^2\right)+\frac{1}{10}\mathrm{N}\left(0,(\frac{1}{4})^2\right)$ modeli ile tanımlanan üç modlu yoğunluk olarak adlandırılan normal karışım yoğunluğundan alınan çeşitli örnek çapları için elde edilen optimal bant genişliği ($h_{\dot{I}OHK}$) ile 1000 tekrarla üretilen 7 farklı bant genişliği seçim yöntemlerinin ortalamaları (h_{ORT}), standart sapmaları (S), yanları (YAN) ve ortalama hata kareleri (OHK).

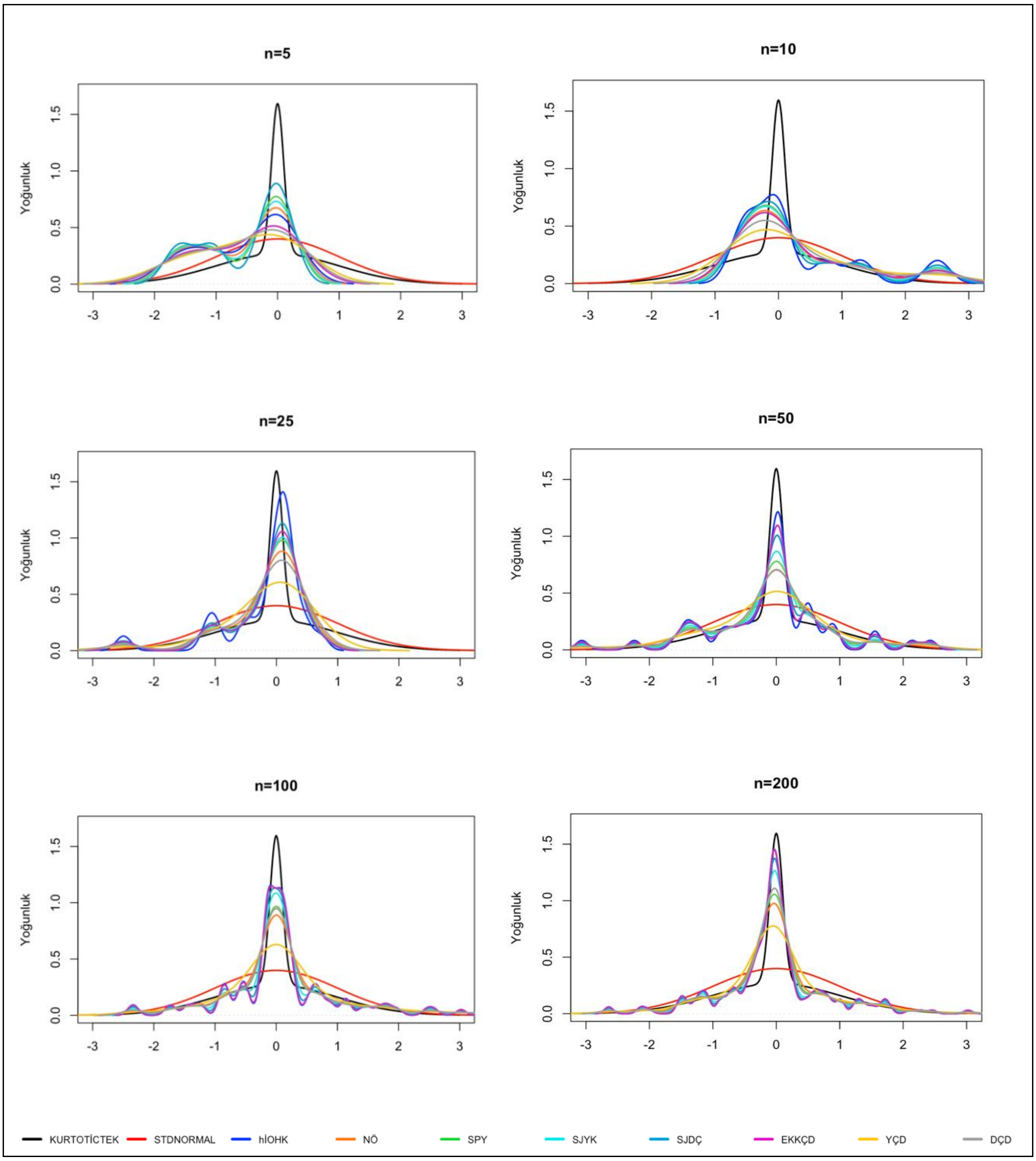
n=5 $h_{\dot{I}OHK}$=0,926					n=10 $h_{\dot{I}OHK}$=0,778				
YÖNTEM	h_{ORT}	S	YAN	OHK	YÖNTEM	h_{ORT}	S	YAN	OHK
NÖ	0,549	0,279	-0,378	0,220	NÖ	0,573	0,172	-0,205	0,071
SPY	0,466	0,237	-0,460	0,268	SPY	0,486	0,146	-0,291	0,106
SJYK	0,495	0,256	-0,431	0,251	SJYK	0,546	0,184	-0,231	0,087
SJDÇ	0,394	0,251	-0,533	0,347	SJDÇ	0,487	0,212	-0,291	0,129
EKKÇD	0,668	0,293	-0,258	0,152	EKKÇD	0,610	0,205	-0,167	0,070
YÇD	0,766	0,251	-0,161	0,089	YÇD	0,694	0,146	-0,084	0,028
DÇD	0,698	0,266	-0,229	0,123	DÇD	0,645	0,179	-0,133	0,050
n=25 $h_{\dot{I}OHK}$=0,617					n=50 $h_{\dot{I}OHK}$=0,512				
YÖNTEM	h_{ORT}	S	YAN	OHK	YÖNTEM	h_{ORT}	S	YAN	OHK
NÖ	0,511	0,089	-0,106	0,019	NÖ	0,459	0,048	-0,053	0,005
SPY	0,433	0,075	-0,183	0,039	SPY	0,390	0,041	-0,122	0,017
SJYK	0,496	0,108	-0,121	0,026	SJYK	0,449	0,071	-0,063	0,009
SJDÇ	0,463	0,133	-0,153	0,041	SJDÇ	0,426	0,093	-0,085	0,016
EKKÇD	0,495	0,151	-0,122	0,038	EKKÇD	0,422	0,126	-0,089	0,024
YÇD	0,580	0,071	-0,036	0,006	YÇD	0,505	0,043	-0,007	0,002
DÇD	0,537	0,105	-0,080	0,018	DÇD	0,471	0,071	-0,041	0,007
n=100 $h_{\dot{I}OHK}$=0,415					n=200 $h_{\dot{I}OHK}$=0,328				
YÖNTEM	h_{ORT}	S	YAN	OHK	YÖNTEM	h_{ORT}	S	YAN	OHK
NÖ	0,406	0,028	-0,009	0,001	NÖ	0,356	0,016	0,028	0,001
SPY	0,345	0,024	-0,070	0,006	SPY	0,302	0,013	-0,026	0,001
SJYK	0,393	0,050	-0,022	0,003	SJYK	0,341	0,034	0,013	0,001
SJDÇ	0,377	0,067	-0,038	0,006	SJDÇ	0,328	0,046	-0,000	0,002
EKKÇD	0,363	0,103	-0,052	0,013	EKKÇD	0,307	0,086	-0,021	0,008
YÇD	0,440	0,027	0,025	0,001	YÇD	0,382	0,018	0,054	0,003
DÇD	0,406	0,051	-0,010	0,003	DÇD	0,349	0,036	0,021	0,002



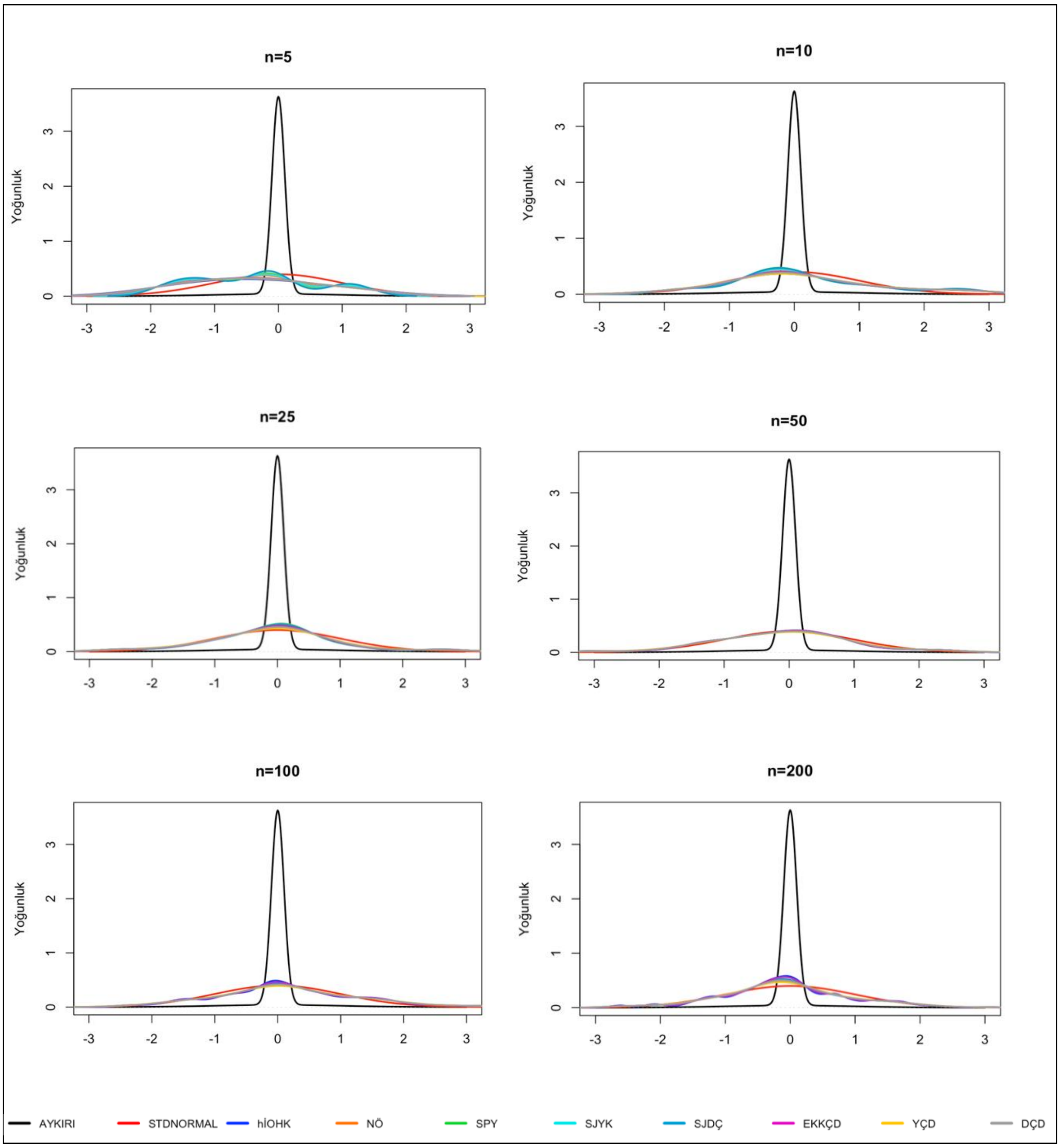
Şekil 5.2. N(0,1) modeli ile tanımlanan standart normal yoğunluğun olasılık yoğunluk fonksiyonu ve bu dağılımdan alınan çeşitli örnek çapları için elde edilen optimal bant genişliği ile 7 farklı bant genişliği seçim yöntemlerine göre elde edilen çekirdek yoğunluk tahminleri.



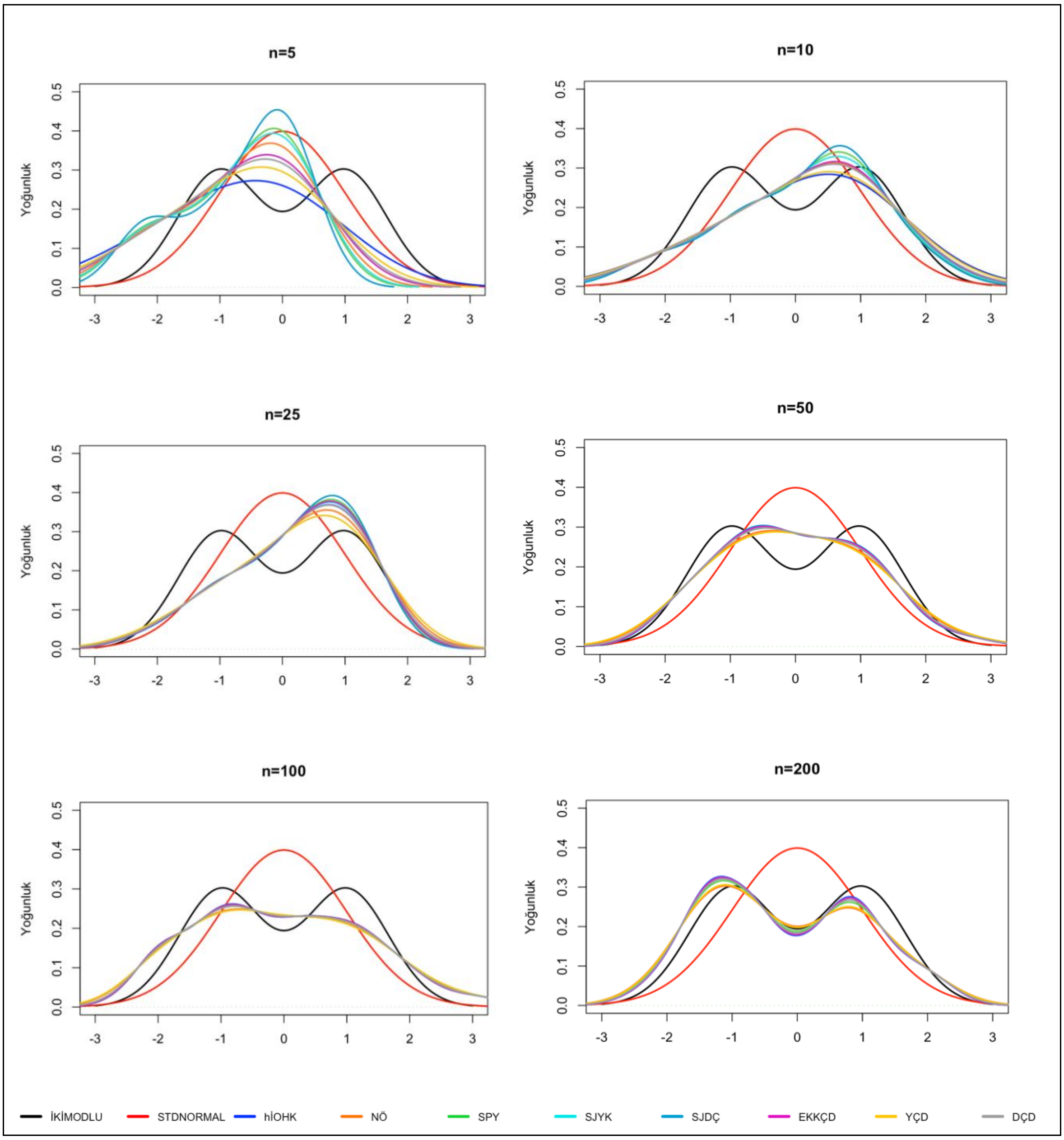
Şekil 5.3. $\frac{1}{5}\mathrm{N}(0,1)+\frac{1}{5}\mathrm{N}\left(\frac{1}{2},(\frac{2}{3})^2\right)+\frac{3}{5}\mathrm{N}\left(\frac{13}{12},(\frac{5}{9})^2\right)$ modeli ile tanımlanan çarpık tek modlu yoğunluk olarak adlandırılan normal karışım dağılımının olasılık yoğunluk fonksiyonu ve bu dağılımdan alınan çeşitli örnek çapları için elde edilen optimal bant genişliği ile 7 farklı bant genişliği seçim yöntemlerine göre elde edilen çekirdek yoğunluk tahminleri.



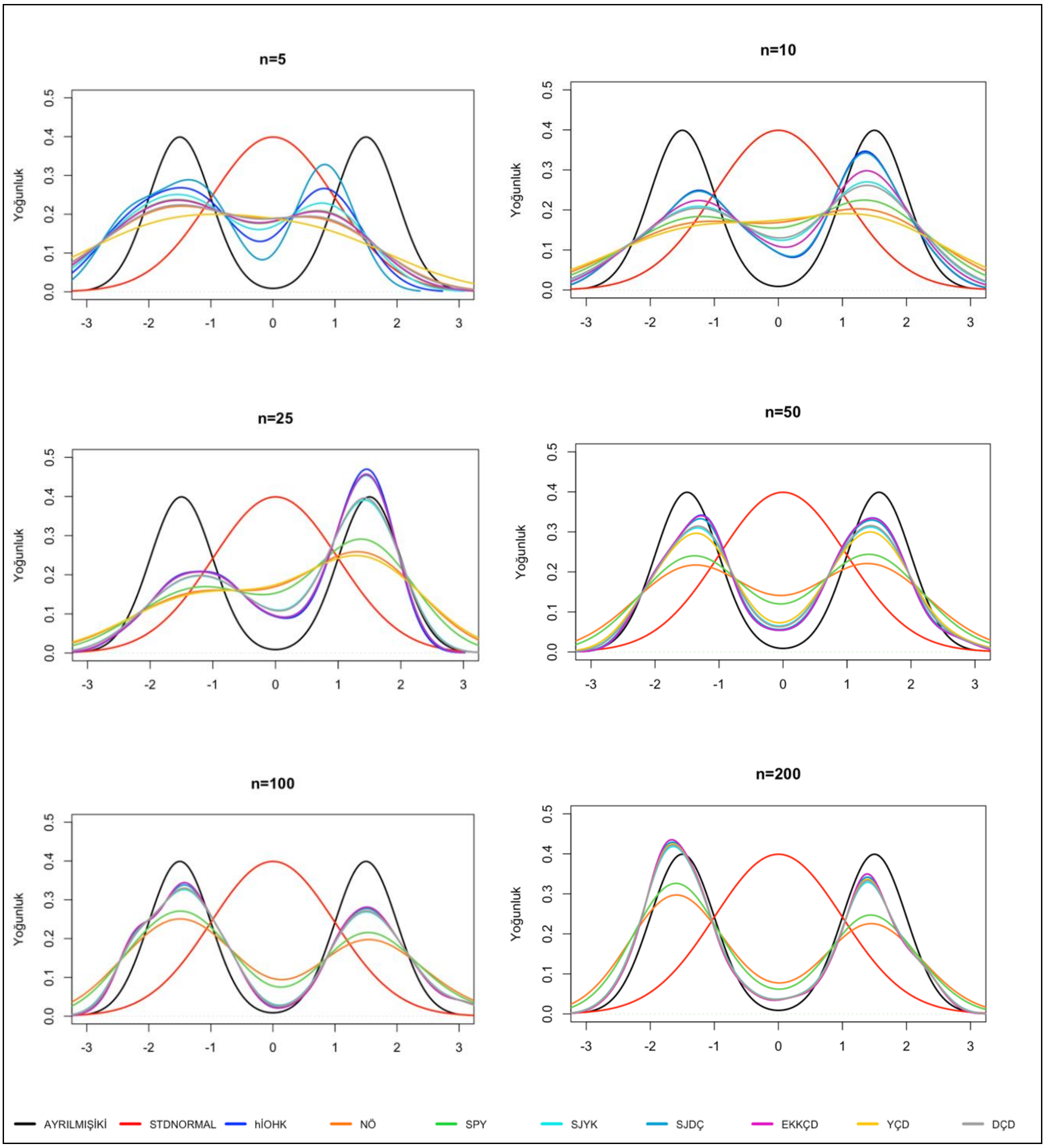
Şekil 5.4. $\frac{2}{3}\mathrm{N}(0,1)+\frac{1}{3}\mathrm{N}\left(0,(\frac{1}{10})^2\right)$ modeli ile tanımlanan kurtotic tek modlu yoğunluk olarak adlandırılan normal karışım dağılımının olasılık yoğunluk fonksiyonu ve bu dağılımdan alınan çeşitli örnek çapları için elde edilen optimal bant genişliği ile 7 farklı bant genişliği seçim yöntemlerine göre elde edilen çekirdek yoğunluk tahminleri.



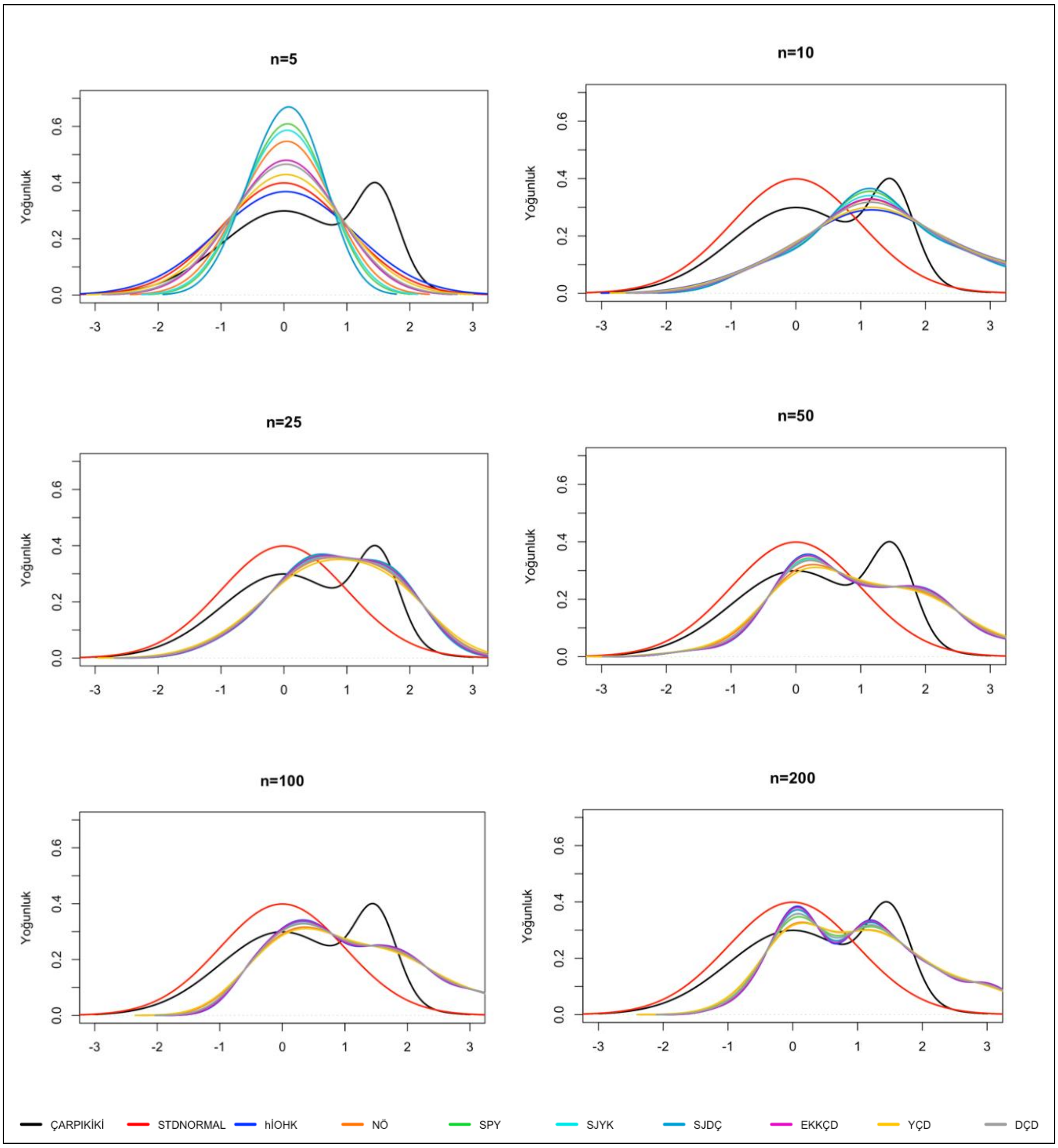
Şekil 5.5. $\frac{1}{10}\mathrm{N}(0,1) + \frac{9}{10}\mathrm{N}\left(0,(\frac{1}{10})^2\right)$ modeli ile tanımlanan aykırı yoğunluk olarak adlandırılan normal karışım dağılımının olasılık yoğunluk fonksiyonu ve bu dağılımdan alınan çeşitli örnek çapları için elde edilen optimal bant genişliği ile 7 farklı bant genişliği seçim yöntemlerine göre elde edilen çekirdek yoğunluk tahminleri.



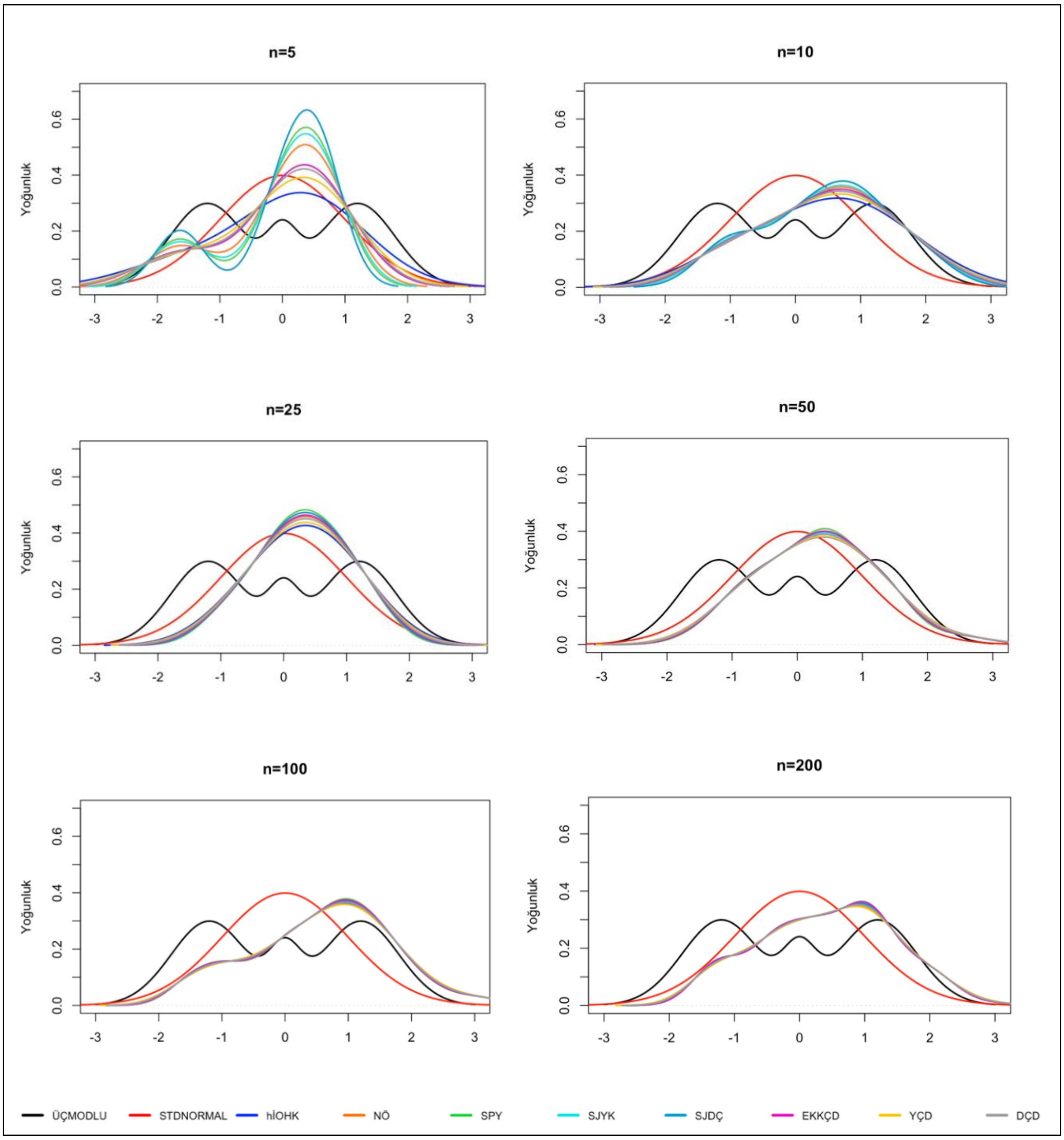
Şekil 5.6. $\frac{1}{2}\mathrm{N}\left(-1,(\frac{2}{3})^2\right)+\frac{1}{2}\mathrm{N}\left(1,(\frac{2}{3})^2\right)$ modeli ile tanımlanan iki modlu yoğunluk olarak adlandırılan normal karışım dağılımının olasılık yoğunluk fonksiyonu ve bu dağılımdan alınan çeşitli örnek çapları için elde edilen optimal bant genişliği ile 7 farklı bant genişliği seçim yöntemlerine göre elde edilen çekirdek yoğunluk tahminleri.



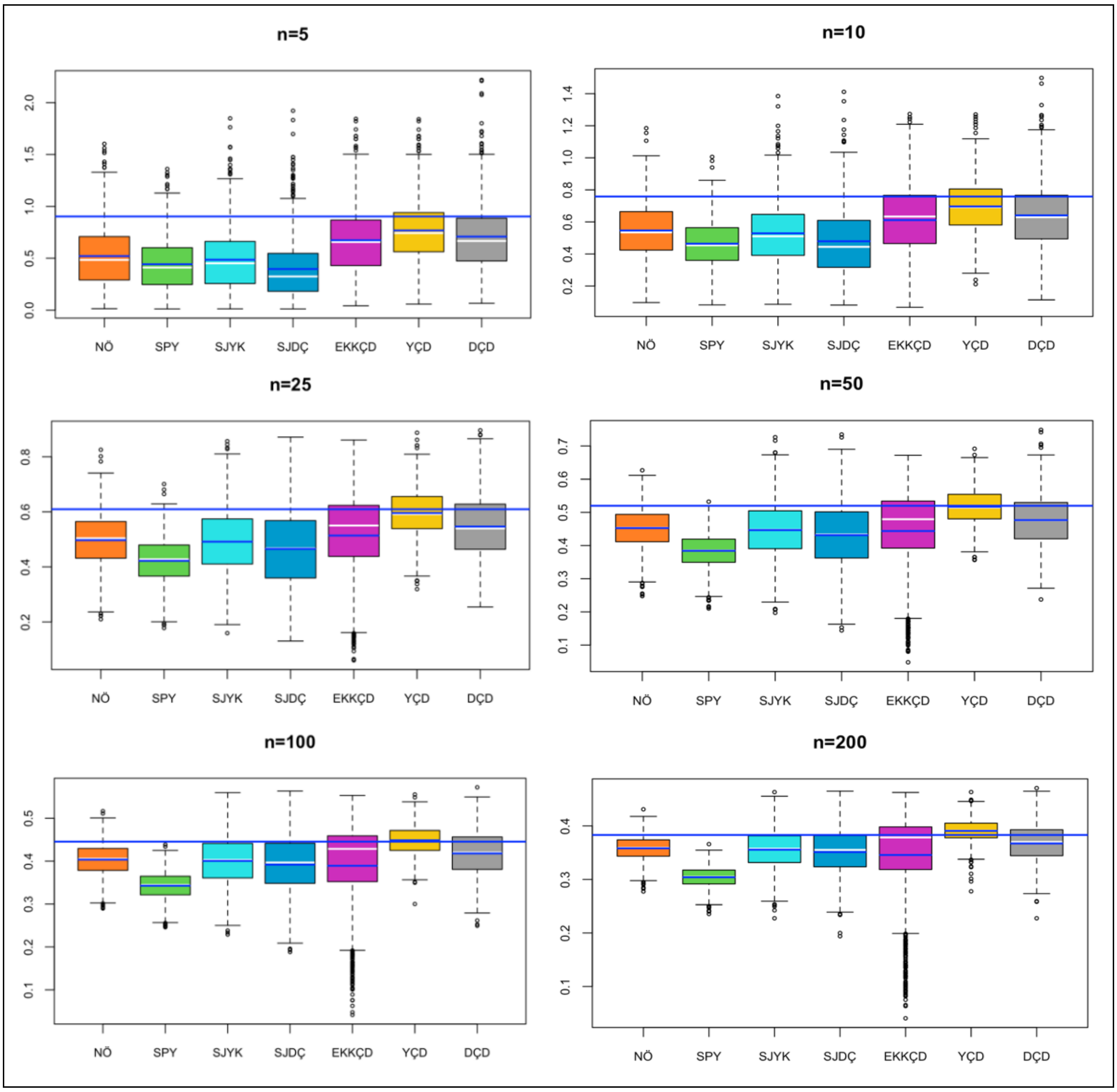
Şekil 5.7. $\frac{1}{2}\mathrm{N}\left(-\frac{3}{2},(\frac{1}{2})^2\right)+\frac{1}{2}\mathrm{N}\left(\frac{3}{2},(\frac{1}{2})^2\right)$ modeli ile tanımlanan ayrılmış iki modlu yoğunluk olarak adlandırılan normal karışım dağılımının olasılık yoğunluk fonksiyonu ve bu dağılımdan alınan çeşitli örnek çapları için elde edilen optimal bant genişliği ile 7 farklı bant genişliği seçim yöntemlerine göre elde edilen çekirdek yoğunluk tahminleri.



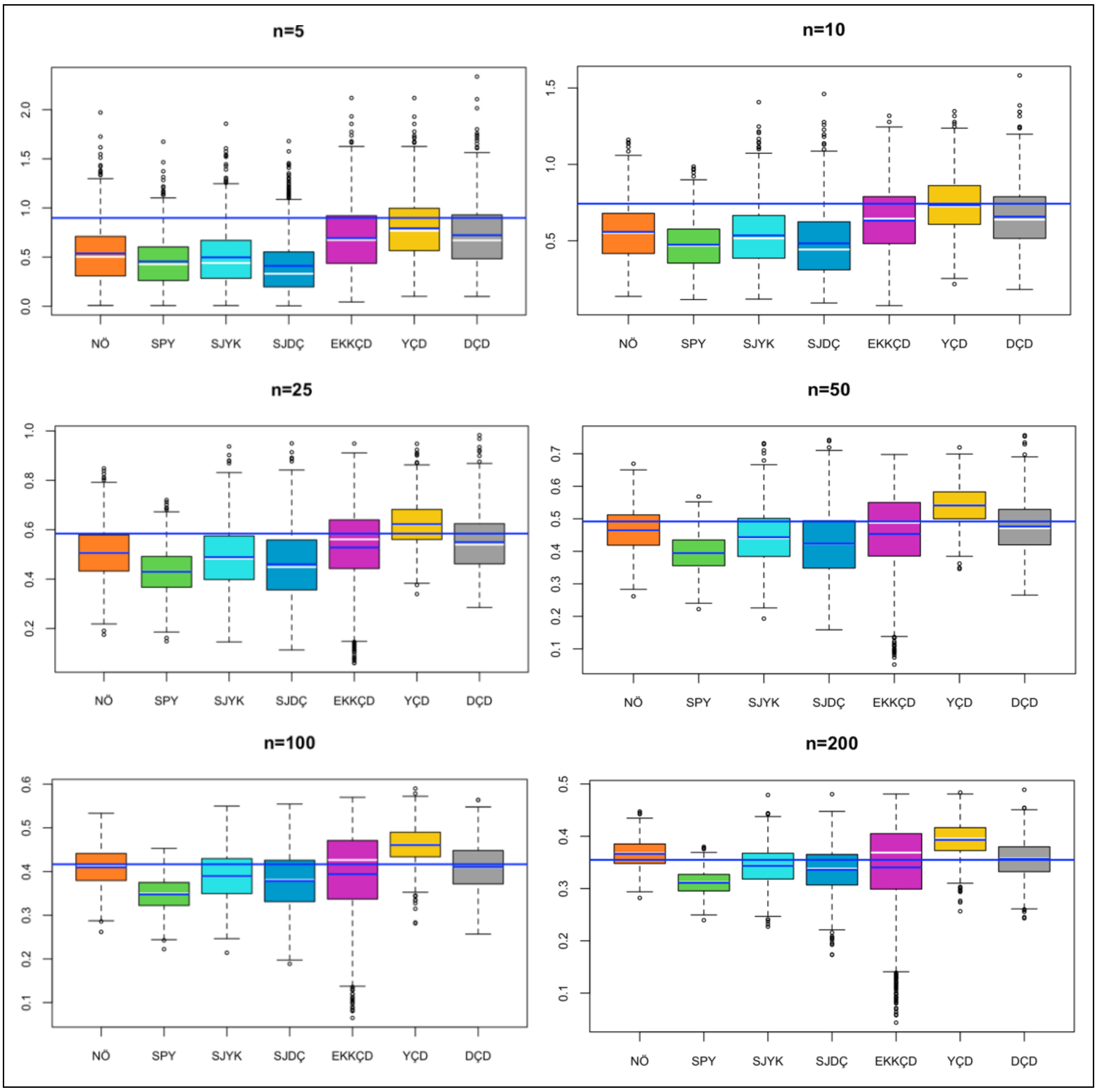
Şekil 5.8. $\frac{3}{4}\mathrm{N}(0,1)+\frac{1}{4}\mathrm{N}\left(\frac{3}{2},(\frac{1}{3})^2\right)$ modeli ile tanımlanan çarpık iki modlu yoğunluk olarak adlandırılan normal karışım dağılımının olasılık yoğunluk fonksiyonu tahmini ve bu dağılımdan alınan çeşitli örnek çapları için elde edilen optimal bant genişliği ile 7 farklı bant genişliği seçim yöntemlerine göre elde edilen çekirdek yoğunluk tahminleri.



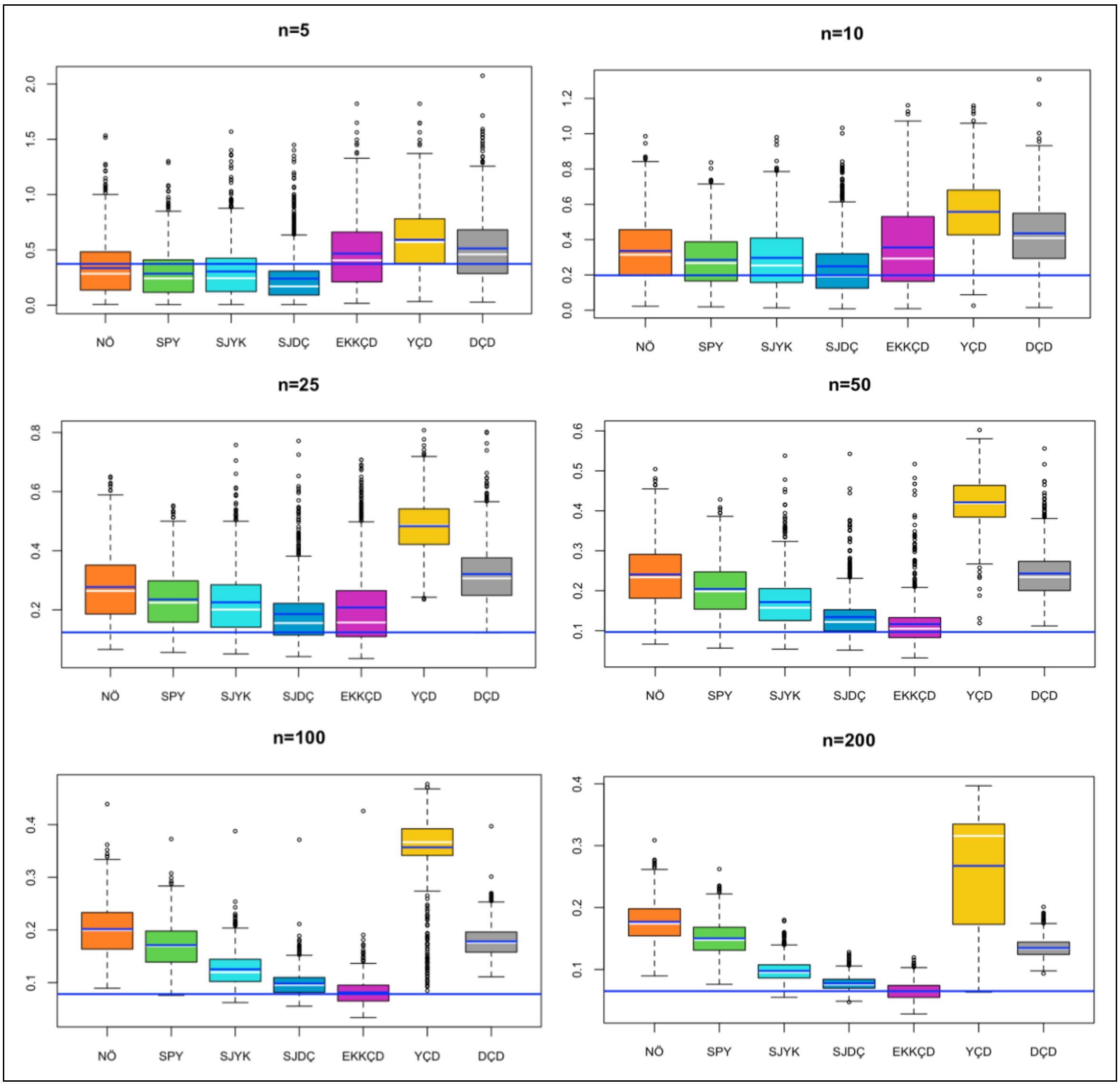
Şekil 5.9. $\frac{9}{20}N\left(-\frac{6}{5},(\frac{3}{5})^2\right)+\frac{9}{20}N\left(\frac{6}{5},(\frac{3}{5})^2\right)+\frac{1}{10}N\left(0,(\frac{1}{4})^2\right)$ modeli ile tanımlanan üç modlu yoğunluk olarak adlandırılan normal karışım dağılımının olasılık yoğunluk fonksiyonu ve bu dağılımdan alınan çeşitli örnek çapları için elde edilen optimal bant genişliği ile 7 farklı bant genişliği seçim yöntemlerine göre elde edilen çekirdek yoğunluk tahminleri.



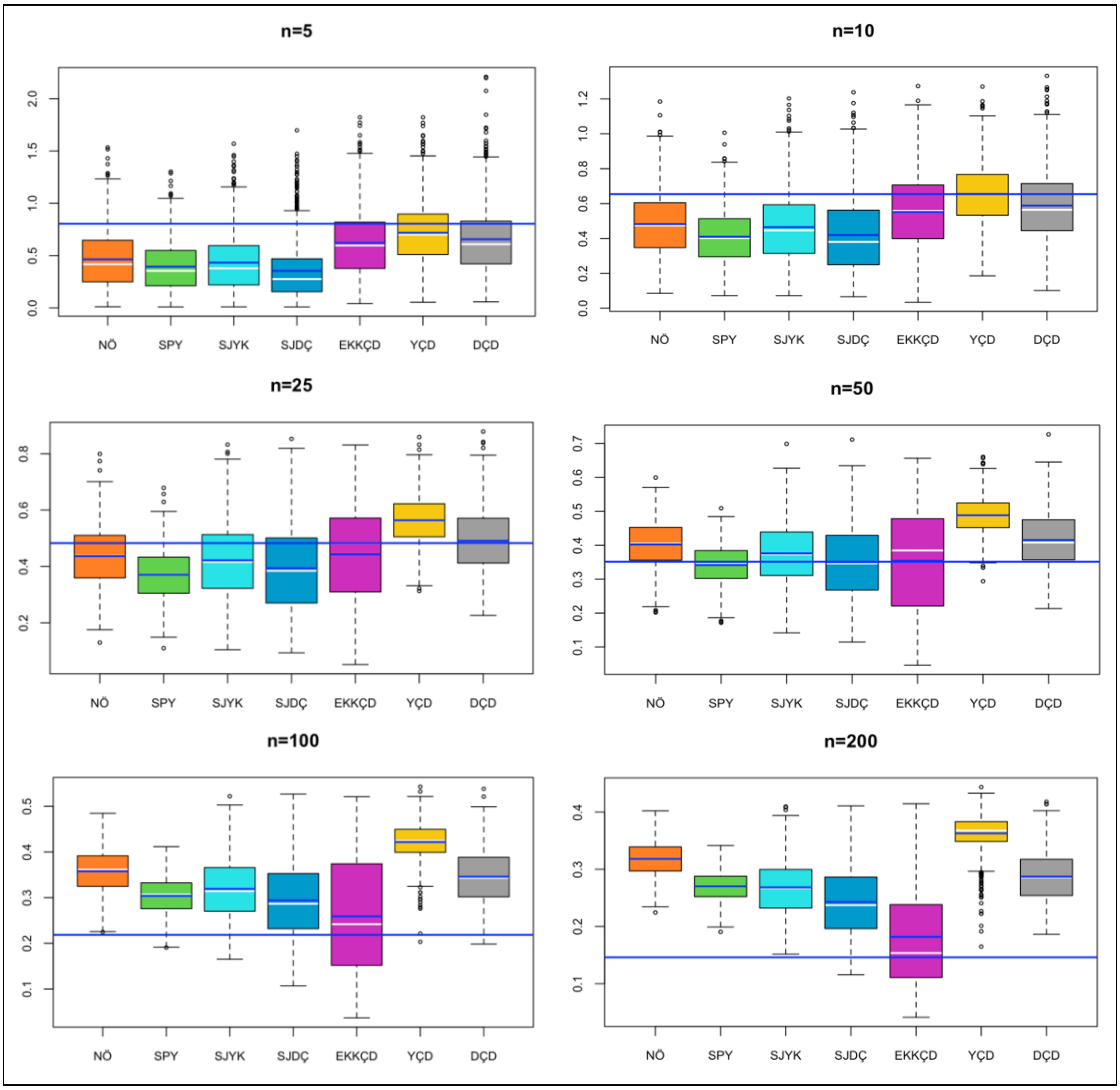
Şekil 5.10. N(0,1) modeli ile tanımlanan standart normal yoğunluğun çeşitli örnek çapları için elde edilen 7 farklı bant genişliği seçim yöntemlerine ilişkin kutu grafikleri. (Not: Her örnek çapı için grafik üzerinde referans değer olarak optimal bant genişliği doğrusu (mavi), her kutuda ise bant genişliklerinin ortalama değerleri (mavi) ve medyan değerleri (beyaz) işaretlenmiştir.)



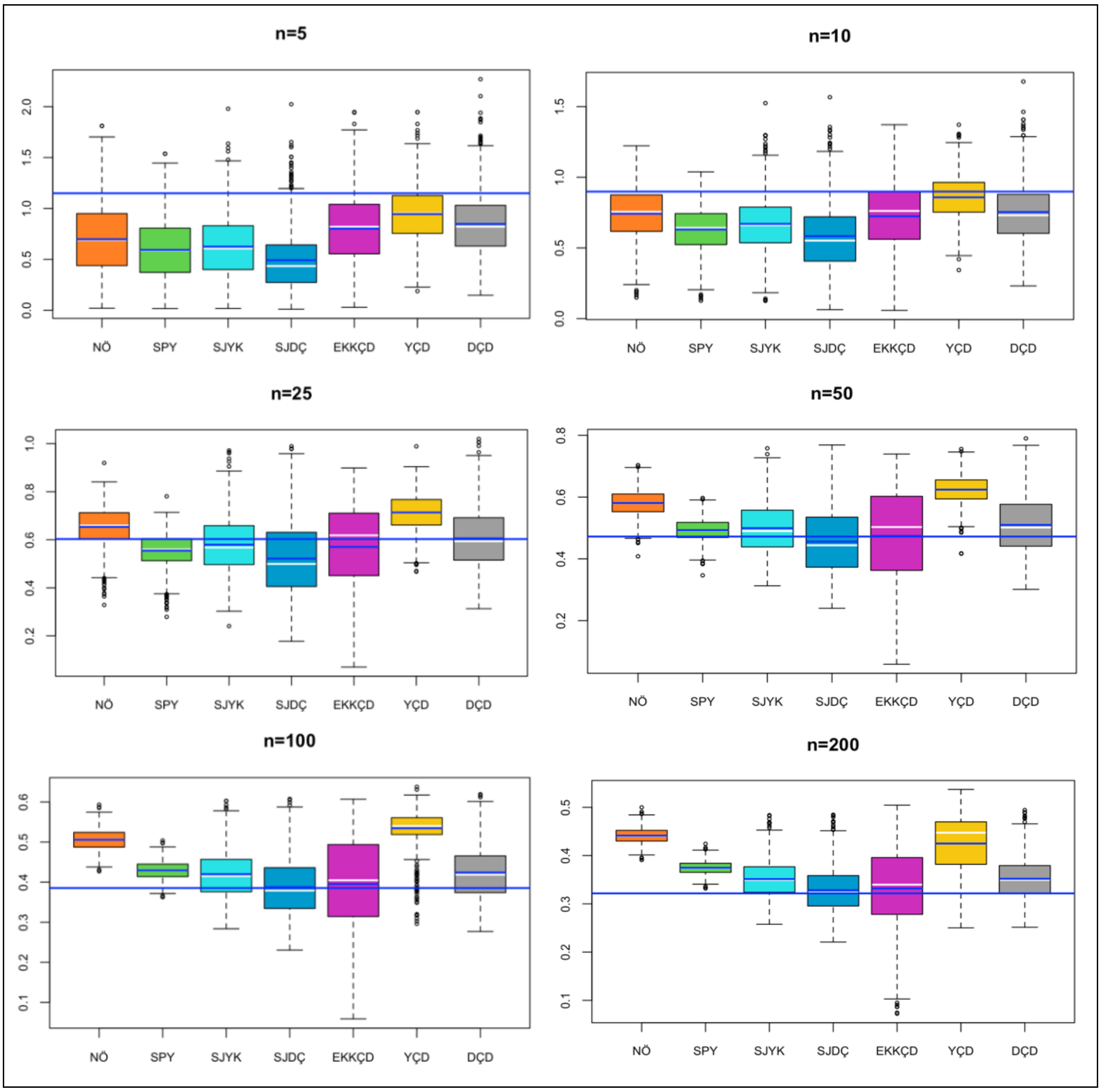
Şekil 5.11. $\frac{1}{5}\mathrm{N}(0,1)+\frac{1}{5}\mathrm{N}\left(\frac{1}{2},(\frac{2}{3})^2\right)+\frac{3}{5}\mathrm{N}\left(\frac{13}{12},(\frac{5}{9})^2\right)$ modeli ile tanımlanan çarpık tek modlu yoğunluk olarak adlandırılan normal karışım dağılımından çeşitli örnek çapları için elde edilen 7 farklı bant genişliği seçim yöntemlerine ilişkin kutu grafikleri. (Not: Her örnek çapı için grafik üzerinde referans değer olarak optimal bant genişliği doğrusu (mavi), her kutuda ise bant genişliklerinin ortalama değerleri (mavi) ve medyan değerleri (beyaz) işaretlenmiştir.)



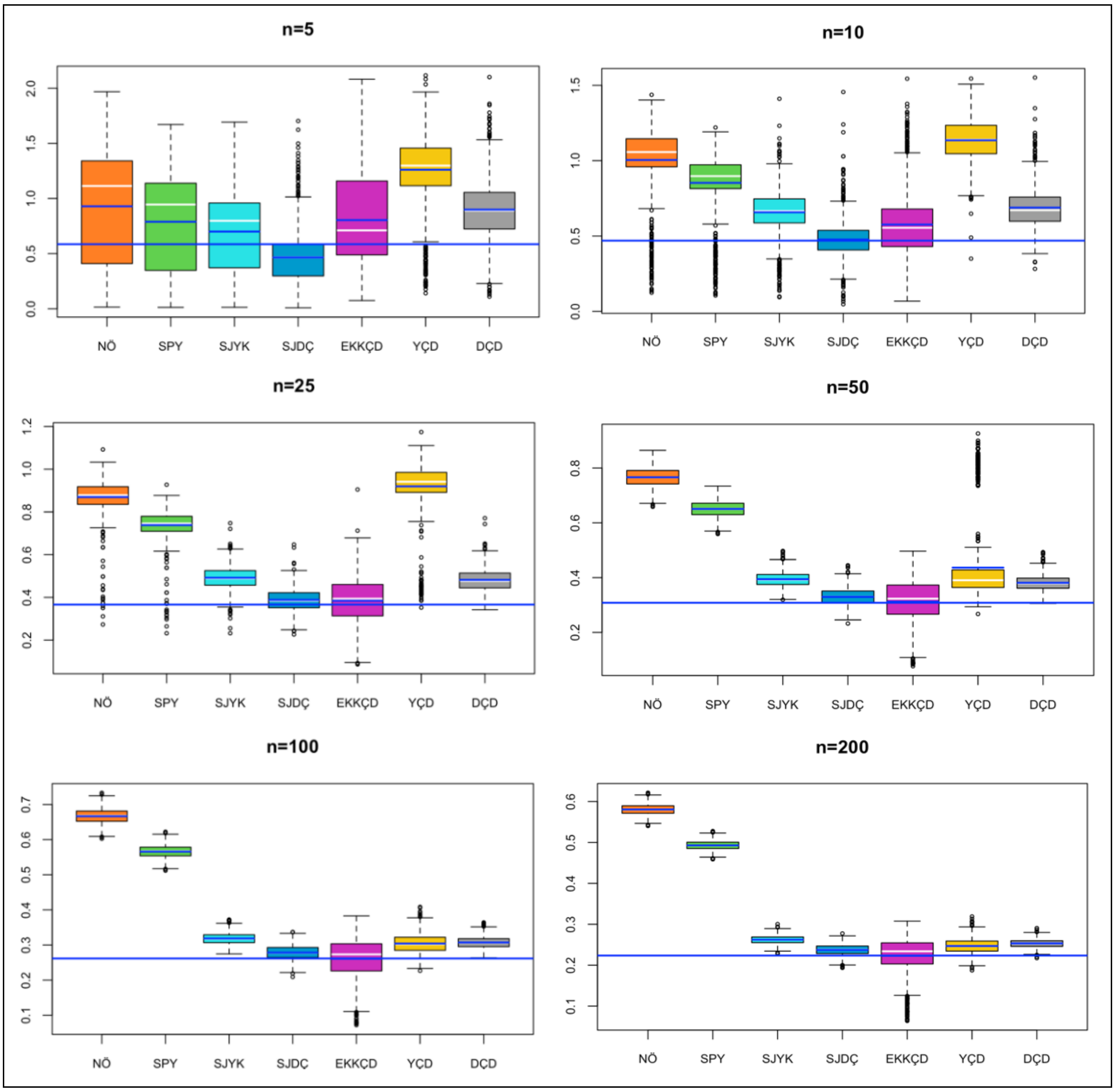
Şekil 5.12. $\frac{2}{3}\mathrm{N}(0,1) + \frac{1}{3}\mathrm{N}\left(0, (\frac{1}{10})^2\right)$ modeli ile tanımlanan kurtotic tek modlu yoğunluk olarak adlandırılan normal karışım dağılımından çeşitli örnek çapları için elde edilen 7 farklı bant genişliği seçim yöntemlerine ilişkin kutu grafikleri. (Not: Her örnek çapı için grafik üzerinde referans değer olarak optimal bant genişliği doğrusu (mavi), her kutuda ise bant genişliklerinin ortalama değerleri (mavi) ve medyan değerleri (beyaz) işaretlenmiştir.)



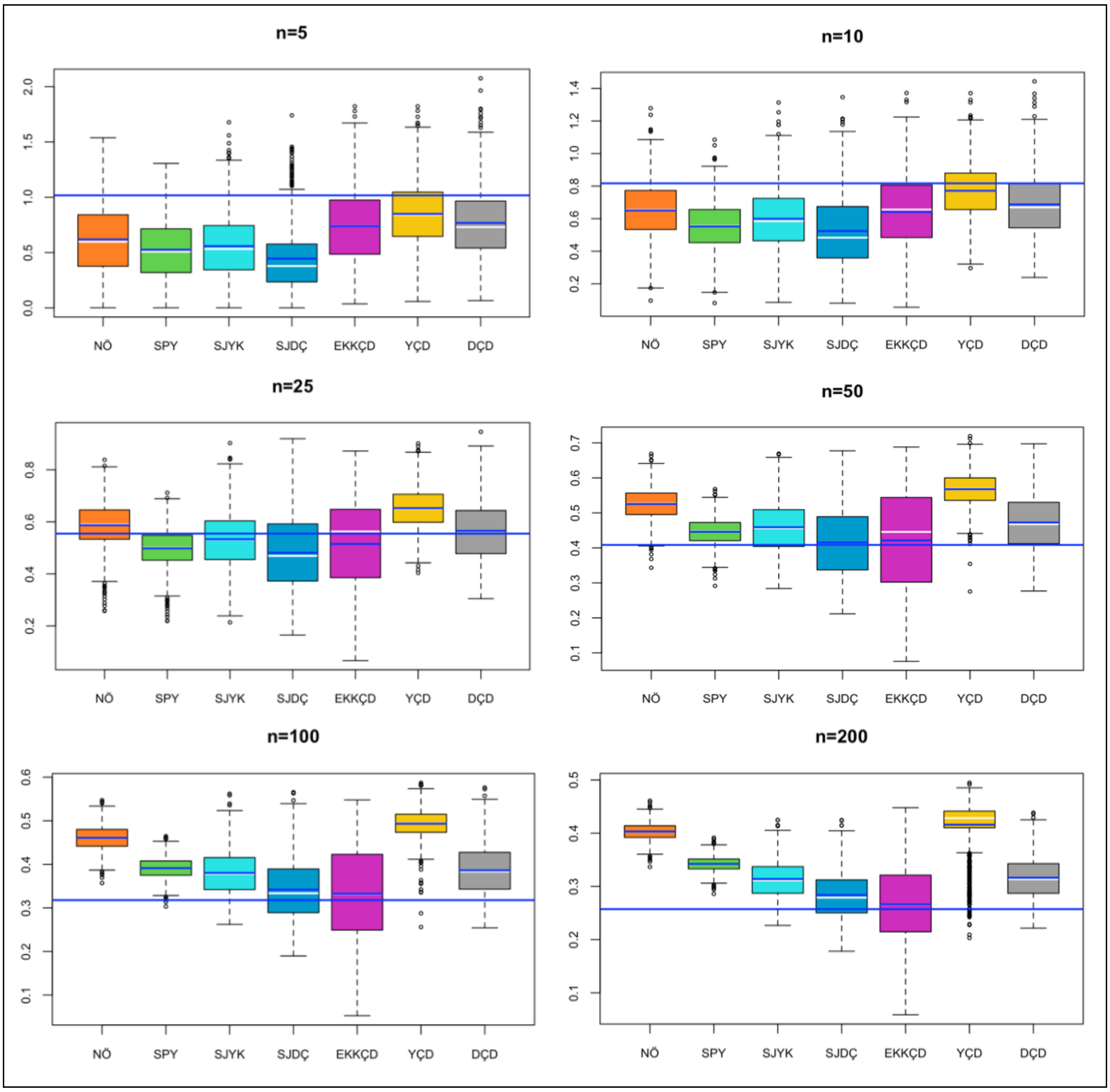
Şekil 5.13. $\frac{1}{10}N(0,1)+\frac{9}{10}N\left(0,(\frac{1}{10})^2\right)$ modeli ile tanımlanan aykırı yoğunluk olarak adlandırılan normal karışım dağılımından çeşitli örnek çapları için elde edilen 7 farklı bant genişliği seçim yöntemlerine ilişkin kutu grafikleri. (Not: Her örnek çapı için grafik üzerinde referans değer olarak optimal bant genişliği doğrusu (mavi), her kutuda ise bant genişliklerinin ortalama değerleri (mavi) ve medyan değerleri (beyaz) işaretlenmiştir.)



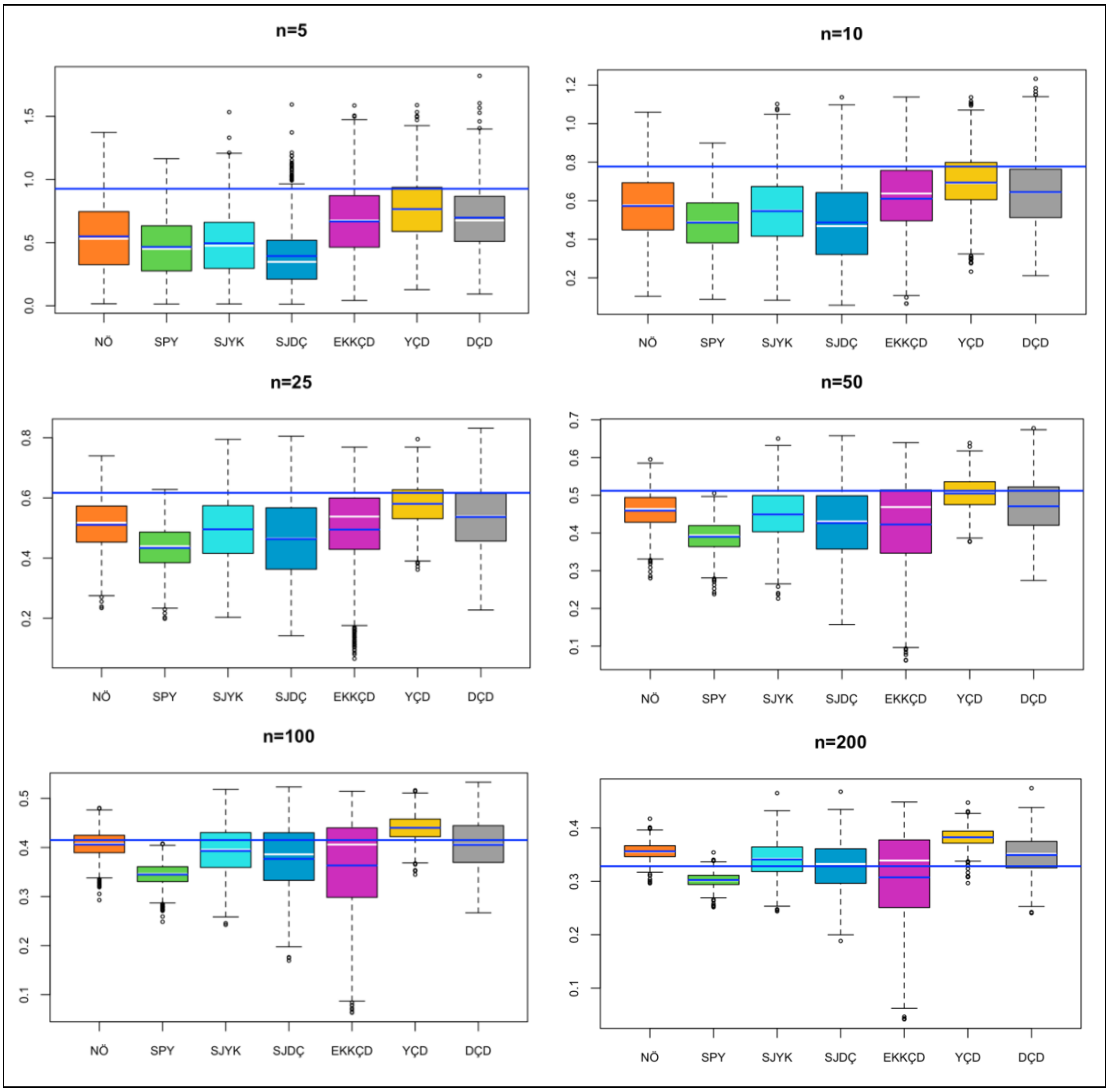
Şekil 5.14. $\frac{1}{2}\mathrm{N}\left(-1,(\frac{2}{3})^2\right)+\frac{1}{2}\mathrm{N}\left(1,(\frac{2}{3})^2\right)$ modeli ile tanımlanan iki modlu yoğunluk olarak adlandırılan normal karışım dağılımından çeşitli örnek çapları için elde edilen 7 farklı bant genişliği seçim yöntemlerine ilişkin kutu grafikleri. (Not: Her örnek çapı için grafik üzerinde referans değer olarak optimal bant genişliği doğrusu (mavi), her kutuda ise bant genişliklerinin ortalama değerleri (mavi) ve medyan değerleri (beyaz) işaretlenmiştir.)



Şekil 5.15. $\frac{1}{2}\mathrm{N}\left(-\frac{3}{2},(\frac{1}{2})^2\right)+\frac{1}{2}\mathrm{N}\left(\frac{3}{2},(\frac{1}{2})^2\right)$ modeli ile tanımlanan ayrılmış iki modlu yoğunluk olarak adlandırılan normal karışım dağılımından çeşitli örnek çapları için elde edilen 7 farklı bant genişliği seçim yöntemlerine ilişkin kutu grafikleri. (Not: Her örnek çapı için grafik üzerinde referans değer olarak optimal bant genişliği doğrusu (mavi), her kutuda ise bant genişliklerinin ortalama değerleri (mavi) ve medyan değerleri (beyaz) işaretlenmiştir.)



Şekil 5.16. $\frac{3}{4}N(0,1) + \frac{1}{4}N\left(\frac{3}{2}, (\frac{1}{3})^2\right)$ modeli ile tanımlanan çarpık iki modlu yoğunluk olarak adlandırılan normal karışım dağılımından çeşitli örnek çapları için elde edilen 7 farklı bant genişliği seçim yöntemlerine ilişkin kutu grafikleri. (Not: Her örnek çapı için grafik üzerinde referans değer olarak optimal bant genişliği doğrusu (mavi), her kutuda ise bant genişliklerinin ortalama değerleri (mavi) ve medyan değerleri (beyaz) işaretlenmiştir.)



Şekil 5.17. $\frac{9}{20}\mathrm{N}\left(-\frac{6}{5},(\frac{3}{5})^2\right)+\frac{9}{20}\mathrm{N}\left(\frac{6}{5},(\frac{3}{5})^2\right)+\frac{1}{10}\mathrm{N}\left(0,(\frac{1}{4})^2\right)$ modeli ile tanımlanan üç modlu yoğunluk olarak adlandırılan normal karışım dağılımından çeşitli örnek çapları için elde edilen 7 farklı bant genişliği seçim yöntemlerine ilişkin kutu grafikleri. (Not: Her örnek çapı için grafik üzerinde referans değer olarak optimal bant genişliği doğrusu (mavi), her kutuda ise bant genişliklerinin ortalama değerleri (mavi) ve medyan değerleri (beyaz) işaretlenmiştir.)

Örneğin çekildiği standart normal yoğunluk fonksiyonu, farklı örnek çapları için oluşturulan $h_{İOHK}$ ile elde edilen çekirdek yoğunluk tahmini ve 7 farklı bant genişliği seçim yöntemlerinin çekirdek yoğunluk tahminleri için elde edilen simülasyon sonuçlarında; tüm örnek çaplarında $h_{İOHK}$'ya en yakın sonucu veren bant genişliği seçim yönteminin YÇD olduğu, h_{ORT}, S, YAN ve OHK'nin verildiği çizelgelerde, çekirdek yoğunluk tahminlerinin grafiklerinde ve kutu grafiklerinde görülmektedir (Çizelge 5.2, Şekil 5.2, Şekil 5.10).

Örneğin çekildiği normal karışım yoğunluklarından çarpık tek modlu yoğunluğun olasılık yoğunluk fonksiyonu, standart normal yoğunluk fonksiyonu, farklı örnek çapları için oluşturulan $h_{İOHK}$ ile elde edilen çekirdek yoğunluk tahmini ve 7 farklı bant genişliği seçim yöntemlerinin çekirdek yoğunluk tahminleri için elde edilen simülasyon sonuçlarında; $h_{İOHK}$'ya en yakın sonucu veren bant genişliği seçim yönteminin $n \leq 10$ örnek çaplarında YÇD olduğu, $n \geq 25$ örnek çaplarında DÇD olduğu, h_{ORT}, S, YAN ve OHK'nin verildiği çizelgelerde, çekirdek yoğunluk tahminlerinin grafiklerinde ve kutu grafiklerinde görülmektedir (Çizelge 5.3, Şekil 5.3, Şekil 5.11).

Örneğin çekildiği normal karışım yoğunluklarından kurtotic tek modlu yoğunluğun olasılık yoğunluk fonksiyonu, standart normal yoğunluk fonksiyonu, farklı örnek çapları için oluşturulan $h_{İOHK}$ ile elde edilen çekirdek yoğunluk tahmini ve 7 farklı bant genişliği seçim yöntemlerinin çekirdek yoğunluk tahminleri için elde edilen simülasyon sonuçlarında; $h_{İOHK}$'ya en yakın sonucu veren bant genişliği seçim yönteminin $n = 5$ örnek çapında NÖ, $n = 10$ ile $n = 25$ örnek çaplarında SJDÇ ve $n \geq 50$ örnek çaplarında EKKÇD olduğu, h_{ORT}, S, YAN ve OHK'nin verildiği çizelgelerde, çekirdek yoğunluk tahminlerinin grafiklerinde ve kutu grafiklerinde görülmektedir (Çizelge 5.4, Şekil 5.4, Şekil 5.12).

Örneğin çekildiği normal karışım yoğunluklarından aykırı yoğunluğun olasılık yoğunluk fonksiyonu, standart normal yoğunluk fonksiyonu, farklı örnek çapları için oluşturulan $h_{İOHK}$ ile elde edilen çekirdek yoğunluk tahmini ve 7 farklı bant genişliği seçim yöntemlerinin çekirdek yoğunluk tahminleri için elde edilen simülasyon sonuçlarında; $h_{İOHK}$'ya en yakın sonucu veren bant genişliği seçim yönteminin $n = 5$ ile $n = 10$ örnek çapında YÇD, $n = 25$ örnek çapında DÇD, $n = 50$ örnek çapında SJDÇ ve $n \geq 100$ örnek çaplarında EKKÇD olduğu, h_{ORT}, S, YAN ve OHK'nin verildiği çizelgelerde, çekirdek yoğunluk tahminlerinin grafiklerinde ve kutu grafiklerinde görülmektedir (Çizelge 5.5, Şekil 5.5, Şekil 5.13).

Örneğin çekildiği normal karışım yoğunluklarından iki modlu yoğunluğun olasılık yoğunluk fonksiyonu, standart normal yoğunluk fonksiyonu, farklı örnek çapları için oluşturulan $h_{\text{İ}OHK}$ ile elde edilen çekirdek yoğunluk tahmini ve 7 farklı bant genişliği seçim yöntemlerinin çekirdek yoğunluk tahminleri için elde edilen simülasyon sonuçlarında; $h_{\text{İ}OHK}$'ya en yakın sonucu veren bant genişliği seçim yönteminin $n = 5$ ile $n = 10$ örnek çaplarında YÇD, $n = 25$ örnek çapında DÇD, $n = 50$ örnek çapında EKKÇD ve $n \geq 100$ örnek çaplarında SJDÇ olduğu, h_{ORT}, S, YAN ve OHK'nin verildiği çizelgelerde, çekirdek yoğunluk tahminlerinin grafiklerinde ve kutu grafiklerinde görülmektedir (Çizelge 5.6, Şekil 5.6, Şekil 5.14).

Örneğin çekildiği normal karışım yoğunluklarından ayrılmış iki modlu yoğunluğun olasılık yoğunluk fonksiyonu, standart normal yoğunluk fonksiyonu, farklı örnek çapları için oluşturulan $h_{\text{İ}OHK}$ ile elde edilen çekirdek yoğunluk tahmini ve 7 farklı bant genişliği seçim yöntemlerinin çekirdek yoğunluk tahminleri için elde edilen simülasyon sonuçlarında; $h_{\text{İ}OHK}$'ya en yakın sonucu veren bant genişliği seçim yönteminin $n \leq 10$ örnek çaplarında SJDÇ ve $n \geq 25$ örnek çaplarında EKKÇD olduğu, h_{ORT}, S, YAN ve OHK'nin verildiği çizelgelerde, çekirdek yoğunluk tahminlerinin grafiklerinde ve kutu grafiklerinde görülmektedir (Çizelge 5.7, Şekil 5.7, Şekil 5.15).

Örneğin çekildiği normal karışım yoğunluklarından çarpık iki modlu yoğunluğun olasılık yoğunluk fonksiyonu, standart normal yoğunluk fonksiyonu, farklı örnek çapları için oluşturulan $h_{\text{İ}OHK}$ ile elde edilen çekirdek yoğunluk tahmini ve 7 farklı bant genişliği seçim yöntemlerinin çekirdek yoğunluk tahminleri için elde edilen simülasyon sonuçlarında; $h_{\text{İ}OHK}$'ya en yakın sonucu veren bant genişliği seçim yönteminin $n = 5$ ile $n = 10$ örnek çaplarında YÇD, $n = 25$ örnek çapında DÇD, $n = 50$ örnek çapında SJDÇ ve $n \geq 100$ örnek çaplarında EKKÇD olduğu, h_{ORT}, S, YAN ve OHK'nin verildiği çizelgelerde, çekirdek yoğunluk tahminlerinin grafiklerinde ve kutu grafiklerinde görülmektedir (Çizelge 5.8, Şekil 5.8, Şekil 5.16).

Örneğin çekildiği normal karışım yoğunluklarından üç modlu yoğunluğun olasılık yoğunluk fonksiyonu, standart normal yoğunluk fonksiyonu, farklı örnek çapları için oluşturulan $h_{\text{İ}OHK}$ ile elde edilen çekirdek yoğunluk tahmini ve 7 farklı bant genişliği seçim yöntemlerinin çekirdek yoğunluk tahminleri için elde edilen simülasyon sonuçlarında; $h_{\text{İ}OHK}$'ya en yakın sonucu veren bant genişliği seçim yönteminin $n \leq 50$ örnek çaplarında YÇD, $n = 100$ örnek çapında NÖ ve $n = 200$ örnek çapında SJDÇ olduğu, h_{ORT}, S, YAN ve OHK'nin verildiği çizelgelerde, çekirdek yoğunluk tahminlerinin grafiklerinde ve kutu grafiklerinde görülmektedir (Çizelge 5.9, Şekil 5.9, Şekil 5.17).

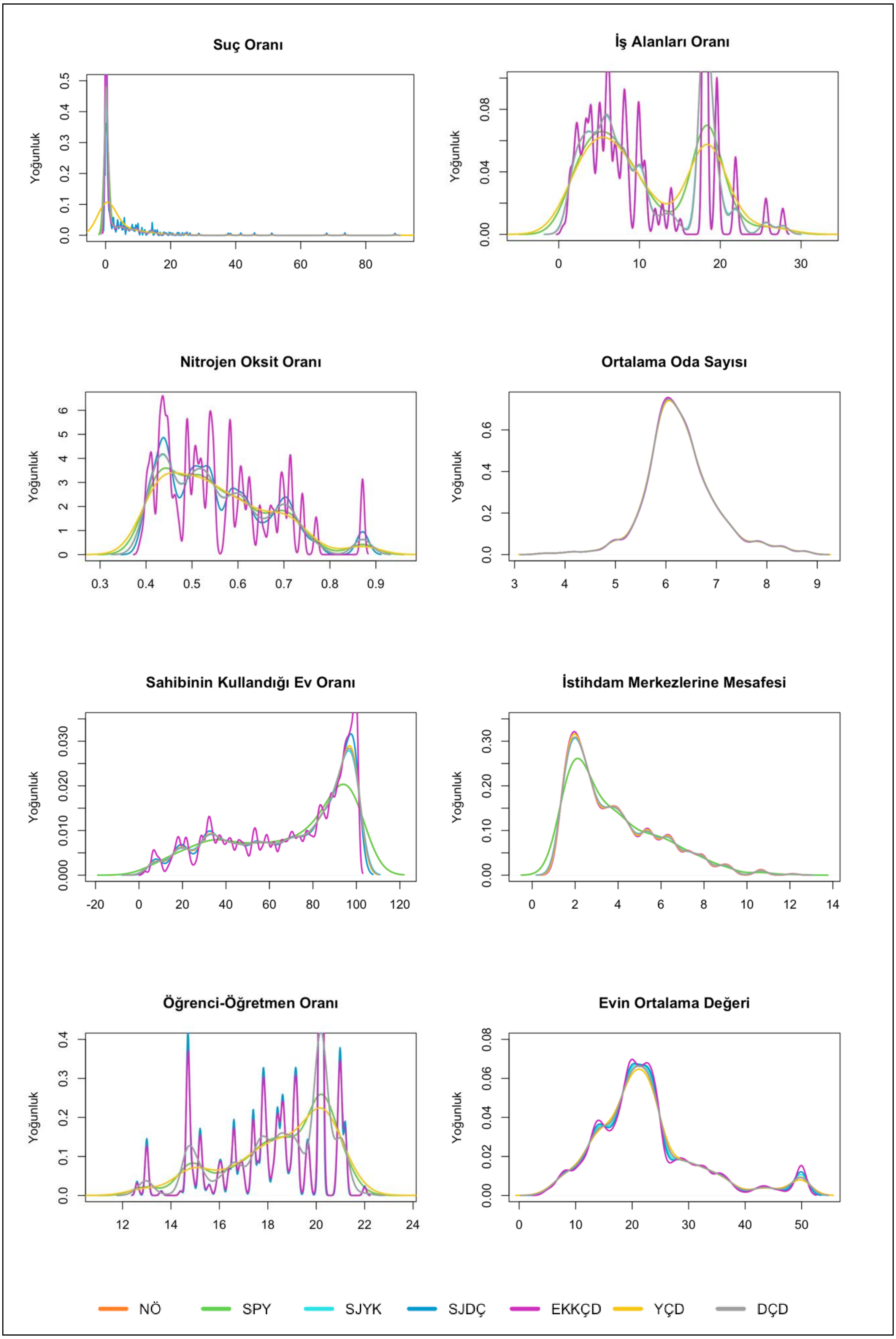
6. UYGULAMA

Bu bölümde dördüncü bölümde tanıtılan bant genişliklerinin gerçek veri seti üzerinde uygulaması sunulmuştur. Uygulamada R paket programı (versiyon 4.2.2) içerisinde bulunan "MASS" paketi içindeki "Boston" veri seti kullanılmıştır (Cran.r-project, 2023). 506 gözlemden oluşan veri setinde Boston banliyölerindeki konutlara ilişkin 14 değişkenin bulunmaktadır. Uygulamada bu değişkenlerden 8 tanesi kullanılmıştır. Bu değişkenler; konutun bulunduğu konuma göre kişi başına düşen suç oranı (Suç Oranı), konutun bulunduğu konuma göre ticari olmayan iş alanları oranı (İş Alanları Oranı), konutun bulunduğu konumun havadaki nitrojen oksit oranı (Nitrojen Oksit Oranı), konut başına düşen ortalama oda sayısı (Ortalama Oda Sayısı), sahibi tarafından kullanılan evlerin oranı (Sahibinin Kullandığı Ev Oranı), konutun Boston'daki 5 istihdam merkezine olan mesafesinin ağırlıklı ortalaması (İstihdam Merkezlerine Mesafesi), konutun bulunduğu konumun öğrenci-öğretmen oranı (Öğrenci-Öğretmen Oranı) ve ev sahiplerinin oturduğu evlerin ortalama değeri (Evin Ortalama Değeri)'dir. R paket programı (versiyon 4.2.2) içerisinde bulunan "stats" paketi ile her değişken için hesaplanan 7 farklı bant genişliği (NÖ, SPY, SJYK, SJDÇ, EKKÇD, YÇD, DÇD) sonuçları Çizelge 6.1'de verilmiştir. Her değişken için 7 bant genişliği seçim yönteminin çekirdek yoğunluk tahminleri aynı grafik üzerinde verilmiştir (Şekil 6.1). Böylece farklı bant genişliği seçim yöntemlerinin düzgünleştirme miktarları ve yoğunluklara etkileri Şekil 6.1'de görülmektedir.

Çizelge 6.1. Boston veri seti değişkenleri için 7 farklı yönteme göre hesaplanan bant genişlikleri.

Değişkenler	NÖ	SPY	SJYK	SJDÇ	EKKÇD	YÇD	DÇD
Suç Oranı	0,819	0,695	0,189	0,064	0,295	2,821	0,498
İş Alanları Oranı	2,093	1,777	0,753	0,249	0,235	2,248	0,710
Nitrojen Oksit Oranı	0,035	0,030	0,020	0,013	0,004	0,038	0,020
Ortalama Oda Sayısı	0,168	0,143	0,144	0,141	0,137	0,159	0,149
Sahibinin Kullandığı Ev Oranı	8,589	7,292	3,639	2,536	0,964	3,284	3,468
İstihdam Merkezlerine Mesafesi	0,643	0,546	0,317	0,269	0,244	0,268	0,305
Öğrenci-Öğretmen Oranı	0,638	0,541	0,278	0,064	0,074	0,710	0,275
Evin Ortalama Değeri	1,816	1,542	1,307	1,133	0,860	1,857	1,571

Uygulamada elde edilen sonuçlara göre Suç Oranı, İş Alanları Oranı, Nitrojen Oksit Oranı, Öğrenci-Öğretmen Oranı, Evin Ortalama Değeri değişkenlerinde diğer 6 yönteme göre YÇD bant genişliği seçim yöntemi ve Ortalama Oda Sayısı, Sahibinin Kullandığı Ev Oranı, İstihdam Merkezlerine Mesafesi değişkenlerinde diğer 6 yönteme göre NÖ bant genişliği seçim yöntemi yüksek bant genişliği sonucu vererek aşırı düzgünleştirme yapmıştır. Suç oranı, Öğrenci-Öğretmen Oranı değişkenlerinde diğer 6 yönteme göre SJDÇ bant genişliği seçim yöntemi ve İş Alanları Oranı, Nitrojen Oksit Oranı, Ortalama Oda Sayısı, Sahibinin Kullandığı Ev Oranı, İstihdam Merkezlerine Mesafesi, Evin Ortalama Değeri değişkenlerinde diğer 6 yönteme göre EKKÇD bant genişliği seçim yöntemi küçük bant genişliği sonucu vererek eksik düzgünleştirme yapmıştır (Çizelge 6.1, Şekil 6.1).



Şekil 6.1. Boston veri seti değişkenlerinin 7 farklı yönteme göre hesaplanan bant genişlikleri ile elde edilen çekirdek yoğunluk tahminleri.

7. SONUÇ VE ÖNERİLER

Bu çalışmada, simülasyon uygulaması ile normal karışım yoğunluklarından alınan farklı örnek çapları için elde edilen çekirdek yoğunluk tahminlerinde kullanılan optimum bant genişlikleri ile 7 bant genişliği seçim yöntemi bu bant genişlikleri seçim yöntemlerinin YAN ve OHK'sı kullanılarak karşılaştırılmıştır.

Simülasyon sonuçlarına göre;

- NÖ bant genişliği seçim yöntemi ile elde edilen bant genişlikleri kurtotic tek modlu yoğunlukta düşük örnek çapında ve üç modlu yoğunlukta ise yüksek örnek çapında optimal bant genişliğine en yakın sonucu vermektedir.
- SPY bant genişliği seçim yöntemi ile elde edilen bant genişlikleri normal karışım yoğunluklarının hiçbirisinde optimal bant genişliğine yaklaşamamaktadır.
- SJYK bant genişliği seçim yöntemi ile elde edilen bant genişlikleri normal karışım yoğunluklarının hiçbirisinde optimal bant genişliğine yaklaşamamaktadır.
- SJDÇ bant genişliği seçim yöntemi ile elde edilen bant genişlikleri kurtotic tek modlu ve ayrılmış iki modlu yoğunlukta düşük örnek çaplarında; iki modlu ve üç modlu yoğunlukta yüksek örnek çaplarında optimal bant genişliğine en yakın sonucu vermektedir.
- EKKÇD bant genişliği seçim yöntemi ile elde edilen bant genişlikleri ayrılmış kurtotic tek modlu, aykırı, ayrılmış iki modlu ve çarpık iki modlu yoğunlukta yüksek örnek çaplarında optimal bant genişliğine en yakın sonucu vermektedir.
- YÇD bant genişliği seçim yöntemi ile elde edilen bant genişlikleri standart normal yoğunlukta tüm örnek çaplarında; aykırı, iki modlu, çarpık iki modlu ve üç modlu yoğunlukta küçük örnek çaplarında optimal bant genişliğine en yakın sonucu vermektedir.
- DÇD bant genişliği seçim yöntemi ile elde edilen bant genişlikleri çarpık tek modlu yoğunlukta büyük örnek çaplarında optimal bant genişliğine en yakın sonucu vermektedir.

Gerçek veri uygulamasıyla da 7 bant genişliği seçim yöntemi Boston veri seti değişkenlerinin çekirdek yoğunluk tahminleri elde edilerek karşılaştırılmıştır.

Gerçek veri uygulaması sonuçlarına göre;

- NÖ ve YÇD bant genişliği seçim yöntemleri bazı değişkenlerde yüksek bant genişliği sonucu vererek aşırı düzgünleştirme yapmıştır.
- SJDÇ ve EKKÇD bant genişliği seçim yöntemleri bazı değişkenlerde küçük bant genişliği sonucu vererek eksik düzgünleştirme yapmıştır.

İncelenen pek çok simülasyon çalışmasında bant genişliği seçim yöntemlerinde son zamanlarda önemli ilerlemeler kaydedilmiş olsa da, bu bant genişliği seçim yöntemleri her durumda tatmin edici sonuçlar vermemektedir. Yani bütün çekirdek yoğunluk tahminleri için tek bir en iyi bant genişliği seçim yöntemi belirlemek mümkün değildir. Tahmin edilen yoğunluğun şeklinin ve yapısının da bant genişliği seçim yöntemlerinde etkisi büyüktür. Bu nedenle farklı bant genişliği seçim yöntemleri için tahminler elde edilmelidir.

KAYNAKLAR

Aydın C., Gündüz, N. ve Balibeyoğlu, J. (2020). Optimal bandwidth selection for a kernel density with a location-scale property. *Communications in Statistics - Theory and Methods*, 50(7), 1671-1684.

Bowman, A.W. (1984). An alternative method of cross-validation for the smoothing of density estimates. *Biometrika*, 71, 353-360.

Bowman, A. ve Azzalini, A. (1997). *Applied smoothing techniques for data analysis: the kernel approach with S-Plus ıllustrations* (First Edition). New York: Oxford University Press, 2.

Cao, R., Cuevas, A. ve Gonzalez Manteiga, W. (1994). A comparative study of several smoothing methods in density estimation. *Computational Statistics & Data Analysis*, 17(2), 153–176.

Fix, E. ve Hodges, J.L. (1951). Discriminatory analysis, nonparametric discrimination: consistency properties. *USAF School of Aviation Medicine*, 4(3), 1-24.

Gramacki, A. (2018). *Nonparametric kernel density estimation and ıts computational aspects* (First Edition). Berlin: Springer International Publishing AG, 26-42.

Gündüz, N. ve Aydın C. (2021). Optimal bandwidth estimators of kernel density functionals for contaminated data. *Journal of Applied Statistics*, 48(13-15), 2239-2258

Hall, P. (1983). Large sample optimality of least squares cross-validation in density estimation. *Annals Statistics*, 11, 1156-1174.

Hardle, W. (1994). *Applied nonparametric regression*. Londra: Cambridge University Press, 32.

Hardle, W., Müller, M., Sperlich, S. ve Werwatz, A. (2004). *Nonparametric and semiparametric models* (First Edition). Berlin: Springer-Verlag, 1.

Harpole, J.K. (2013). *How Bandwidth Selection Algorithms Impact Exploratory Data Analysis Using Kernel Density Estimation*. Master Thesis, University of Kansas, Kansas, 18.

Jones, C., Marron, J. ve Sheather, S. (1996). Progress in data-based bandwidth selection for kernel density estimation. *Computational Statistics*, 11, 337–381.

İnternet: Cran.r-project, (2022). Web: https://cran.rproject.org/web/packages/ks/ks.pdf, Son Erişim Tarihi: 10.01.2022.

İnternet: Cran.r-project, (2022). Web: https://cran.r-project.org/web/packages/nor1mix/nor1mix.pdf, Son Erişim Tarihi: 10.01.2022.

İnternet: Cran.r-project, (2023). Web: https://cran.r-project.org/web/packages/MASS/MASS.pdf, Son Erişim Tarihi: 10.01.2022.

İnternet: Stat.ethz, (2022). Web: https://stat.ethz.ch/R-manual/R-devel/library/stats/html/stats-package.html, Son Erişim Tarihi: 10.01.2022.

Marron, S. ve Wand, M. (1992). Exact mean integrated squared error. *Annals of Statistcs,* 20, 712–736.

Pagan, A. ve Ullah, A. (1999). *Nonparametric estimates* (First Edition). London: Cambrige University Press, 26-32.

Park, B. ve Marron, J. (1990). Comparison of data-driven bandwidth selectors. *Journal of the American Statistical Association*, 85(409), 66–72.

Parzen, E. (1962). On the estimation of a probability density function and the mode. *The Annals of Mathematical Statistics*, 33, 1065-1076.

Racine, J.S. (2008). *Nonparametric econometrics: A primer.* USA: Now Publishers Inc, 7.

Rosenblatt, M. (1956). Remarks on some nonparametric estimates of a density function. *The Annals of Mathematical Statistics*, 27, 832-837.

Rudemo, M. (1982). Empirical choice of histograms and kernel density estimators. *Scandinavian Journal of Statistics*, 9, 65–78.

Scott, D. ve Terrell, G. (1987). Biased and unbiased cross-validation in density estimation. *Journal of the American Statistical Association*, 82(400), 1131–1146.

Serfting, R.J. (1980). *Approximation theorems of mathematical statistics* (First Edition). New York: Wiley, 1.

Sheather, S.J. (2004). Density estimation. *Statistical Science*, 19(4), 558-597.

Sheather, S. ve Jones, M. (1991). A reliable data-based bandwidth selection method for kernel density estimation. *Journal of the Royal Statistical Society*. 53, 683– 690.

Silverman, B.W. (1986). *Density estimation for statistics and data analysis*. London: Chapman and Hall, 1-48.

Terrell, G.R. (1990). The maximal smoothing principle in density estimation. *Journal of the American Statistical Association* 85(410), 470-477.

Wand, M.P. ve Jones, M.C. (1995). *Kernel smoothing* (First Edition). New York: Chapman & Hall, 82-95.

Yolsal, H. (2017). *Parametrik olmayan yoğunluk tahmincileri ve regresyon analizi* (Birinci Baskı). Ankara: Detay Yayıncılık, 19-61.

Gazili olmak ayrıcalıktır…