ANKARA ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

YÜKSEK LİSANS TEZİ

BAYESCİ SEMİPARAMETRİK REGRESYON

Selma ÇETİN

İSTATİSTİK ANABİLİM DALI

ANKARA
2023

ÖZET

Yüksek Lisans Tezi

BAYESCİ SEMİPARAMETRİK REGRESYON

SELMA ÇETİN

Ankara Üniversitesi
Fen Bilimleri Enstitüsü
İstatistik Anabilim Dalı

Danışman: Prof. Dr. Olçay ARSLAN

İstatistik biliminde, tahmin yapılırken iki farklı felsefe kullanılır. Bunlar Klasik yaklaşım ve Bayesci yaklaşımdır. Klasik yaklaşımda parametreler sabit değişken olarak ele alınırken, Bayesci yaklaşımda parametreler rastgele değişkenler olarak ele alınır ve temeli Bayes teoremine dayanır. Bu yaklaşımla elde edilecek olan bilgi, önsel dağılım olarak bilinen daha önce var olan bilginin güncellenmesiyle bulunur. Bu noktada önsel dağılım geçmiş veriye ve/veya karar vericinin süreçle ilgili parametre hakkındaki kişisel görüşüne karşılık gelmektedir. Buna bağlı sonsal dağılım, ilgilenilen belirsiz parametre hakkındaki tüm bilginin toplam sonucudur. Bu çalışmada semiparametrik regresyon analizi Bayesci yaklaşımla ele alınmıştır. Regresyon analizi, aralarında sebep-sonuç ilişkisi bulunan iki veya daha fazla değişken arasındaki ilişkiyi belirlemek ve bu ilişkiyi kullanarak o konu ile ilgili tahminler yapabilmek amacıyla yapılır ancak bu tahminler parametrik regresyon analizinin varsayımlarının sağlanmadığı durumlarda iyi bir tahmin olma niteliğini kaybederler. Bu durumda daha iyi tahmin yapabilmek amacıyla bu varsayımların esnetilmesine olanak sağlayan regresyon yöntemleri olan nonparametrik ve semiparametrik regresyona ihtiyaç duyulur. Parametrik regresyon analizinden farklı olarak bu yöntemlerde doğrusallık aranmaz düzgünlük aranır. Bu çalışmada parametrik olmayan bu regresyon modelleri için tahmin yöntemlerinden, düzgünlüğün sağlanması için kullanılan düzleştirme yöntemlerinden, düzleştirme parametresinin nasıl hangi yöntemlerle seçilmesi gerektiğinden bahsedilmiştir. Bayesci yaklaşım ile semiparametrik regresyon tahmini anlatılmış ve bir simülasyon çalışmasıyla Bayesci semiparametrik regresyon için sonuçlar verilmiş olup yorumlanmıştır.

Ocak 2023, 79 sayfa

Anahtar kelimeler: Nonparametrik modeller, nonparametrik regresyon, semiparametrik regresyon, düzleştirme, bayes, bayesci semiparametrik regresyon

ABSTRACT

Master of Science

BAYESIAN SEMİPARAMETRİC REGRESSION

SELMA ÇETİN

Ankara University
Graduate School of Natural and Applied Sciences
Department of Statistics

Supervisor: Prof. Dr. Olçay ARSLAN

In statistics, two different philosophies are used when making predictions. These are the Classical approach and the Bayesian approach.While parameters are treated as constant variables in the classical approach, in the Bayesian approach, parameters are treated as random variables and are based on Bayes theorem. The knowledge to be obtained with this approach is found by updating previously existing knowledge, known as the a priori distribution. At this point, the a priori distribution corresponds to the historical data and/or the decision maker's personal opinion about the parameter related to the process. The resulting posterior distribution is the sum of all information about the uncertain parameter of interest. In this study, semiparametric regression analysis was handled with a Bayesian approach. Regression analysis is performed to determine the relationship between two or more variables that have a cause-effect relationship and to make predictions about that subject using this relationship. These estimates lose their quality of being good estimates when the assumptions of the parametric regression analysis are not met. In this case, nonparametric and semiparametric regression, which are regression methods that allow these assumptions to be stretched, are needed in order to make better predictions. Unlike parametric regression analysis, these methods do not seek linearity, but smoothness. In this study, the estimation methods, the smoothing methods used to ensure smoothness, and how the smoothing parameter should be selected are mentioned. Semiparametric regression estimation with Bayesian approach is explained and results for Bayesian semiparametric regression are interpreted with a simulation study.

January 2023, 79 pages

Key Words: Nonparametric models, nonparametric regression, semiparametric regression, smoothing, bayesian, bayesian semiparametric regression

TEŞEKKÜR

Yüksek lisans eğitimim boyunca bana her türlü konuda destek olan, bilgisini, anlayışını ve hoşgörüsünü esirgemeyen, değerli vaktini ayıran tez danışmanım Prof. Dr. Olçay Arslan'a teşekkürlerimi sunarım.

Sevgileriyle, varlıklarıyla bana güç veren, her zaman bana inanan, her durumda yanımda olan ve beni destekleyen anneme, babama ve kardeşlerime teşekkür ederim.

Selma ÇETİN
Ankara, Ocak 2023

İÇİNDEKİLER

SİMGELER ve KISALTMALAR DİZİNİ

B_j	j. kutu
m_j	j. kutunun merkezi
h	kutu genişliği ya da bant genişliği
f_j	Kutulara ait yükseklik
n_j	j. kutuya düşen gözlem sayısı
$\mathrm{I}(\cdot)$	Gösterge fonksiyonu, mantık argümanı doğru ise 1, değilse 0
$\|\cdot\|_2^2$	L_2 büyüklüğü (Norm)
h_o	Optimum kutu genişliği
K	Tek değişkenli Kernel fonksiyonu
$\mathcal{K}$	Çok değişkenli Kernel fonksiyonu
K_h	Ölçeklendirilmiş Kernel fonksiyonu
h_{opt}	Optimal bant genişliği
h_{pm}	Park & Marron tarafından bulunan bant genişliği
sd	Serbestlik derecesi
p	Paramatre sayısı
$\mu_2(K)$	K'nın 2. momenti
$f(\cdot)$	Bilinmeyen fonksiyon
H	Bant genişliği matrisi ya da şapka matrisi
I	Birim matris
∇_f	Gradyant Matrisi
$\mathcal{H}_f$	Hessian Matrisi
Diag	Köşegen matrisi
$w_i(\cdot)$	Ağırlıklandırılmış fonksiyon
S_λ	Düzleştirme matrisi
λ	Düzleştirme parametresi
$o(\cdot)$	$a = o(b)\ ancak\ ve\ ancak\ a/b \to 0,\ n \to \infty\ \ veya\ \ h \to \infty$
$O(\cdot)$	$a = O(b)\ ancak\ ve\ ancak\ \ a/b\ \ n \to \infty\ \ veya\ \ h \to \infty$
L	nxn boyutlu düzleştirme matrisi
τ	Skaler çarpan

Kısaltmalar

HKT	Hata kareler toplamı
HKO	Hata kareler ortalaması
BHKO	Bütünleşik hata kareler ortalaması
ABHKO	Asimptotik bütünleşik hata kareler ortalaması
OTKH	Ortalama toplam kare hatası
BKH	Bütünleşik kare hatası
BIAS	Yanlılık

ŞEKİLLER DİZİNİ

ÇİZELGELER DİZİNİ

1. GİRİŞ

İstatistik biliminde, tahmin yapılırken yaygın olarak kullanılan klasik yaklaşıma alternatif olarak Bayesci yaklaşım kullanılır. Bayesci yaklaşımda parametreler rastgele değişkenler olarak ele alınır ve temeli Bayes teoremine dayanmaktadır. Bu çalışmada semiparametrik regresyon analizi Bayesci yaklaşımla ele alınmıştır. Regresyon analizi, aralarında sebep-sonuç ilişkisi bulunan iki veya daha fazla değişken arasındaki ilişkiyi belirlemek ve bu ilişkiyi kullanarak o konu ile ilgili tahminler yapabilmek amacıyla yapılır ancak bu tahminler doğrusal regresyon analizinin varsayımlarının sağlanmadığı durumlarda iyi bir tahmin olma niteliğini kaybederler. Bu durumda daha iyi tahmin yapabilmek amacıyla bu varsayımların esnetilmesine olanak sağlayan regresyon yöntemleri olan nonparametrik ve semiparametrik regresyona ihtiyaç duyulur. Semiparametrik regresyon modelleri birçok araştırmacı tarafından incelenmiştir. Lenk (1999), bir Fourier gösterimi kullanarak, Natio (2002) çarpımsal ayarlama ile, Ruppert ve ark. (2003), cezalı regresyon spline'larına ve karma modellere dayalı semiparametrik regresyon modellerini çalışmışlardır. 2007'de ise Jensen ve Maheu Bayesci semiparametrik stokastik oynaklık modellemesi üzerinde çalıştılar. Choi, Lee ve Roy (2008), parametrik boş modeli semiparametrik alternatif modele karşı test etmek için Bayes faktörünün geniş örneklem özelliğini araştırdılar. Pelenis (2012), belirli bir örnek olarak doğrusal regresyon ile sınırlı koşullu moment modelinin Bayesci bir tahminini dikkate alan Bayesci semiparametrik regresyonu incelemiştir.

Çalışmanın ikinci bölümünde kısaca regresyon yöntemlerinden bahsedilmiş, tanımları yapılmıştır. Üçüncü bölümde Nonparametrik regresyon ayrıntılı bir şekilde ele alınmıştır. Öncelikle bir sürekli değişkenin nasıl dağıldığı hakkında fikir veren yoğunluk tahmin yöntemlerinden Histogram ve Kernel yöntemi anlatılmış, bant genişliği seçimi için uygulanan yöntemler incelenmiştir. Nonparametrik regresyon tahmini için Kernel regresyon tahmin yöntemlerinden Nadaraya-Watson tahmini ve Yerel polinom regresyon anlatılmıştır. Sonrasında nonparametrik regresyonda önemli bir kavram olan düzleştirme yöntemleri ele alınmıştır. Düzleştirme; bir ya da birden fazla bağımsız değişkenin fonksiyonu olan bağımlı değişkeninin sahip olduğu trendi ifade etmek için kullanılır. Düzleştiriciler değişkenler arasındaki ilişkinin biçimini kesin

bir şekilde varsaymadığı için parametrik olmayan regresyonda sık kullanılan bir araçtır ve düzleştirme parametresinin seçiminin de doğru bir şekilde yapılması gereklidir. Üçüncü bölümün sonunda nonparametrik regresyon Kernel tahmini ile bir örnekle gösterilmiştir.

Dördüncü bölümde semiparametrik regresyon anlatılmıştır. Semiparametrik regresyon, bağımlı değişkenin bazı bağımsız değişkenlerle parametrik bazılarıyla parametrik olmayan bir ilişki olduğu durumlarda kullanılan bir yöntemdir. Toplamsal modellerin özel bir durumudur. Cezalandırılmış en küçük kareler yöntemi ve kısmi spline yaklaşımı ile tahminin nasıl yapıldığından bahsedilip bir örnekle açıklanmaya çalışılmıştır.

Tezin son bölümünde, Bayesci yaklaşım ve tahmin yöntemleri anlatıldıktan sonra bayesci yaklaşımla semiparametrik regresyon ele alınmıştır. Bayesci yaklaşımda önemli bir nokta olan önsel dağılımı belirlenmesinden, Bayesci parametre tahmininin nasıl yapıldığı, hangi metodların kullanıldığından bahsedilmiştir. Son olarak bir simülasyon çalışması ile farklı yineleme sayılarıyla elde edilen sonsal olasılık sonuçlarına göre seçilebilecek regresyon modelleri verilmiştir.

2. REGRESYON ANALİZİ

Regresyon analizi, aralarında sebep sonuç ilişkisi bulunan değişkenler arasındaki ilişkinin matematiksel modelle ifade edilerek incelenen konuyla ilgili tahminler yapılmasıdır. Parametrik regresyon analizinde bağımlı ve bağımsız değişkenler arasındaki ortalama ilişki matematiksel bir modelle ifade edilirken fonksiyondaki parametreler açıkça gösterilmektedir. Parametrik regresyonda açıklayıcı değişkenlerin arasındaki ilişki doğrusal olarak varsayılmaktadır (Akman, 2018).

Basit doğrusal regresyon modelinde bir bağımlı (Y) ve bir bağımsız değişken (X) vardır. Amaç bağımlı değişkeni en iyi şekilde tahmin edecek modelin kurulmasıdır. X ve Y arasındaki gerçek ilişki;

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i \qquad i = 1,2,\dots,n$$

biçiminde gösterilir. Bu modeli tahmin etmek için kullanılan regresyon modeli ise;

$$Y_i = \hat{\beta}_0 + \hat{\beta}_1 X_i + e_i \qquad i = 1,2,\dots,n$$

Basit doğrusal regresyon modelinde katsayılar tahmini için en küçük kareler yöntemiyle hata kareler toplamı minumum yapılarak parametreler tahmin edilebilmektedir. Minumum katsayıları elde etmek için, hata kareler toplamının β_0 ve β_1'e göre türevleri alınarak 0'a eşitlenip katsayılar tahmin edilir.

$Y_i = \hat{\beta}_0 + \hat{\beta}_1 X_i + e_i$ denklemi için hata kareler toplamı:

$$HKT = \sum \varepsilon_i^2 = \sum \left[Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i\right]^2$$

Bu denklemden β_0 ve β_1 katsayılarının tahmin:

$$\hat{\beta}_1 = \frac{\sum X_i Y_i - \frac{\sum X_i \sum Y_i}{n}}{\sum X_i^2 - \frac{(X_i)^2}{n}}, \qquad \hat{\beta}_0 = \bar{Y} - \beta_1 \bar{X}$$

Çoklu doğrusal regresyonda ise birden çok açıklayıcı değişken ile bir yanıt değişkeni arasındaki doğrusal ilişki araştırılır. Genellikle Y yanıt değişkeni (bağımlı değişken) ve k tane X_1, X_2,…, X_k açıklayıcı değişkenler arasındaki ilişki,

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} +, \dots, +\beta_k X_{ik} + \varepsilon_i \qquad i = 1,2, \dots, n$$

şeklinde doğrusal bir modelle gösterilir. Tahmin edilen regresyon modeli,

$$Y_i = \hat{\beta}_0 + \hat{\beta}_1 X_1 +, \dots, +\hat{\beta}_k X_k + e_i$$ şeklindedir.

Modelin matris formunda gösterimi:

$Y = X\beta + \varepsilon$ şeklinde veya

$$Y = \begin{bmatrix} Y_1 \\ Y_2 \\ . \\ . \\ . \\ Y_n \end{bmatrix},\ X = \begin{bmatrix} 1 & X_{11} & . & . & X_{1k} \\ 1 & X_{21} & . & . & X_{2k} \\ . & . & . & . & . \\ . & . & . & . & . \\ . & . & . & . & . \\ 1 & X_{n1} & . & . & X_{nk} \end{bmatrix},\ \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ . \\ . \\ . \\ \beta_k \end{bmatrix} \text{ ve } \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ . \\ . \\ . \\ \varepsilon_n \end{bmatrix}$$

şeklinde gösterilir. $\hat{\beta}$ tahmini ise aşağıda gösterilen formülle elde edilir.

$$\hat{\beta} = (X'X)^{-1}X'Y$$

Parametrik yaklaşım varsayımlara dayanmaktadır ve bu varsayımlar sağlanmadığı takdirde yapılan tahminler etkin ve doğru tahmin olma özelliklerini kaybederler. Bu varsayımların sağlanmadığı durumlarda nonparametrik ve semiparametrik regresyon yöntemleri kullanılır.

Nonparametrik regresyonu parametrik regresyondan ayıran en önemli özelliği modele ait bilinen fonksiyonel bir yapının olmamasıdır. Parametrik regresyondaki doğrusallık kavramının yerini burada düzgünlük alır. Bir y bağımlı değişkeni ile yalnız bir x bağımsız değişkenin olduğu basit parametrik olmayan regresyon modeli,

$$y_i = f(x_i) + \varepsilon_i$$

şeklinde gösterilir. Amaç bilinmeyen ve fonksiyonel ilişkiye gösteren $f(x_i)$ fonksiyonunu düzleştirme tekniklerini kullanarak en uygun düzgün tahmini üretmektir. Bu modeller f fonksiyonuna pürüzsüzlük gibi kısıtlamalar yükleyerek bilinen parametrik regresyon modellerini desteklerler. Nonparametrik basit regresyon modeli, genel olarak saçılım grafiği düzleştiricisi (scatterplot smoothing) olarak da bilinir çünkü parametrik olmayan regresyon yöntemi ile saçılım grafiğindeki noktalar düzgün bir eğri ile temsil edilmeye çalışılır. Nonparametrik regresyon tahminlerinde Kernel regresyon yöntemlerinden Nadaraya-Watson tahmin edicisi ve yerel polinom regresyon yöntemleri kullanılarak tahminler yapılabilir.

Semiparametrik regresyon modelleri ise bağımlı değişkenin bağımsız değişkenlerin bir kısmıyla parametrik bir kısmıyla parametrik olmayan bir ilişki içinde olduğu durumlarda kullanılır. Bağımlı değişken ile her bir bağımsız değişkenin ayrı ayrı çizilecek grafikleri incelenerek, bağımlı değişken ile bağımsız değişkenler arasındaki ilişkinin yapısı hakkında fikir sahibi olunabilir. Her bir değişkenin ilişkisine tek tek bakıldıktan sonra bağımsız değişkenlerin bir veya birkaçı için parametrik, diğerleriyle parametrik olmayan bir ilişki varsa semiparametrik model tercih edilir. Semiparametrik modeller, parametrik ve parametrik olmayan bileşenlerin her ikisini içerdiğinden, bu modellere kısmi doğrusal modeller (partially linear model) de denir. Bu modeller standart regresyon modellerine göre çok daha esnektir. Semiparametrik regresyon modellerinde parametrik kısım, doğrusal olabileceği gibi dönüşüm yöntemleri (Logaritmik, karesel dönüştürme vs.), uygulanarak doğrusallaştırılabilen yapıda da olabilir (Aydın, 2005). Semiparametrik regresyon modeli:

$$\boldsymbol{y_i = \alpha + z_i^T \beta_i + m(x_i) + \varepsilon_i,\ \ \varepsilon_i \sim N(0, \sigma^2),\ \ \ 1 \leq i \leq n}$$

biçiminde ifade edilir. Semiparametrik regresyon tahmininde cezalı en küçük kareler yöntemi, kısmi spline yaklaşımı gibi yöntemler kullanılarak tahminler yapılabilir.

3. NONPARAMETRİK REGRESYON

Parametrik regresyon modellerinde fonksiyonel yapının bilindiği varsayılırken parametrik olmayan regresyon modellerinde uygun fonksiyonel yapı verilerden kestirilir. Bağımlı değişken ile bağımsız değişken arasındaki parametrik olmayan ilişki fonksiyonel yapı olmadan yerel kestirimlerin bir sonucu olarak elde edilir. Bir y bağımlı değişkeni ile yalnız bir x bağımsız değişkenin olduğu basit parametrik olmayan regresyon modeli,

$$y_i = f(x_i) + \varepsilon_i$$

şeklinde gösterilir. Bu eşitlikte $f(\cdot)$ bağımlı değişken ve bağımsız değişken arasındaki fonksiyonel ilişkiyi gösteren bilinmeyen ve belirli bir şekle sahip olmayan bir fonksiyondur. Nonparametrik regresyon yönteminin amacı $f(\cdot)$ fonksiyonunun düzgün ve sürekli olduğu, hata terimleri ε_i'lerin ortalaması 0 ve varyansı σ^2 ile özdeş bir dağılıma sahip olduğu varsayımı altında $f(\cdot)$ fonksiyonunu verilerden tahmin etmektir. Nonparametrik basit regresyon modeli, genel olarak saçılım grafiği düzleştiricisi (scatterplot smoothing) olarak da bilinir çünkü parametrik olmayan regresyon yöntemi ile saçılım grafiğindeki noktalar düzgün bir eğri ile temsil edilmeye çalışılır. İki ya da daha fazla bağımsız değişken olduğunda ise parametrik olmayan regresyon modelinin tahminini yapmak daha zordur ve elde edilen grafikler karmaşık bir yapıda olmaktadır. Bu durum 'boyutluluk sorunu' olarak adlandırılır. Nonparametrik regresyon modelini geometrik olarak göstermek zor olduğu için daha kısıtlayıcı modeller geliştirilmiştir. Bu modellerden biri toplamsal regresyon modelidir ve bu model:

$$y_i = \alpha + f_1(x_{i1}) + f_2(x_{i2}) + , \dots , + f_k(x_{ik}) + \varepsilon_i, \qquad i = 1, \dots, n$$

şeklinde gösterilir. Burada $f_j(\cdot)$ fonksiyonlarına kısmi regresyon fonksiyonları denir ve düzgün oldukları varsayılır (Fox, 2002). Nonparametrik regresyon modeliyle doğrusallık sağlanmaz, düzgünlük oluşturulur böylece doğrusallık varsayımı esnetilmiş olur. Doğrusallık varsayımının esnetilmesi daha karmaşık hesaplamalarla birlikte yorumlanması daha zor sonuçlar elde edilmesine neden olur. Ancak regresyon fonksiyonu daha doğru bir şekilde tahmin edilir. (Keele, 2008; Fox, 2002).

3.1 Nonparametrik Modellerde Yoğunluk Tahmini

Olasılık yoğunluk fonksiyonu, sürekli rasgele değişkenlerin nasıl dağıldığı hakkında fikir verirler. Olasılık yoğunluk fonksiyonu bilindiği zaman, ortalama ve varyans gibi parametreler hakkında fikir sahibi olunmasının yanında, bu değişkenin belli bir olasılıkla hangi aralıkta değerler alacağı hesaplanabilmektedir. Bu bölümde nonparametrik modeller için tahmin yöntemleri anlatılmıştır.

3.1.1 Histogram yöntemi

Histogram yöntemi parametrik olmayan yoğunluk tahmininde kullanılan en eski yaklaşımlardan biridir. Kullanışlı bir yöntemdir fakat gerçeği çok iyi yansıtamadığı söylenebilir. Ayrıca iki ya da daha fazla değişken için uyarlamak oldukça zordur. Bu dezavantajları sebebiyle histogram yöntemine alternatif olarak birçok parametrik olmayan tahminciler geliştirilmiştir. Histogram yöntemi ya da kutulama yöntemi için, $(X_1, X_2, \dots, X_n)$ bilinmeyen ve sürekli bir dağılımdan gelen rastgele değişkenler olsun (Hardle vd. 2004).

$$B_j = [x_0 + (j-1)h, x_0 + jh] \qquad j \epsilon Z$$

Burada B_j: j. kutu, x_0: başlangıç noktası, h: kutu genişliğini gösterir. Her kutuya ait yükseklik ise:

$$f_j = \frac{n_j}{nh} \text{ şeklinde ifade edilir.}$$

Burada, n_j, j. kutuya düşen gözlem sayısını, n: örneklem büyüklüğünü göstermektedir. Histogramın başka bir gösterim şekli ise:

$$\hat{f}_h(x) = \frac{1}{nh} \sum_{i=1}^{n} \sum_{j} I(X_i \epsilon B_j) I(x \epsilon B_j) \tag{3.1}$$

$$I(X_i \epsilon B_j) = \begin{cases} 1, & X_i \epsilon B_j \\ 0, & diğer\ durumlar \end{cases}$$

(3.1)'deki formül her x değeri için f'nin tahminini verir. m_j, B_j kutusunun merkezi olmak üzere $B_j = \left[m_j - \frac{h}{2}, m_j + \frac{h}{2}\right]$ aralığı (3.1) eşitliğinden çıkarımda bulunarak yazılabilir. Bu aralık her x değeri için aynı f değerini atar. Bu durum da kısıtlayıcı ve esnek olmayan bir durum olduğu için alternatif çözümler türetilmiştir. Histogramın şekli kutu genişliğine ve başlangıç noktasına göre değişir. Aynı kutu genişliği farklı başlangıç noktalarıyla çizilen histogramlar çok modlu ya da simetrik olacak şekilde sonuçlar verebilir. Gözlem sayısının fazla olduğu durumlarda histogram yönteminin kullanılması sorun yaratmazken, küçük örneklemlerde dar aralıklı kutu oluşturması sebebiyle histogramların kullanımı pratik değildir. Bunun nedeni her bir kutuda az sayıda gözlem yer alacağı için, i. kutuya ait örnek kutu ortalaması dengeli olmayacaktır. Dengeli bir ortalama için geniş aralıklı ve az sayıda kutuya ihtiyaç vardır. Bunu sağlamak için x değişkenine ait değerler eşit genişliğe sahip kutulara bölünebilir ya da hemen hemen eşit sayıda gözlem içeren kutular oluşturulabilir. Eğer x tekdüze dağılıma sahipse x değişkenine ait gözlemler eşit sayıda gözlem içeren kutulara bölünebilir. Fakat bu yöntem küçük örnekler için uygulanırsa, uygulamada içi boş kutularla karşılaşılabilir. Eğer dağılım sağa ya da sola çarpıksa yine eşit aralıklı kutuları kullanmak doğru olmayacaktır. Bu durumda da gözlemlerin çoğu aynı kutu içerisine düşecektir (Hardle vd. 2004).

Histogram yöntemi yansız bir tahmin edici değildir. Yanlılığın kesin değeri, gerçek yoğunluğun şekline bağlıdır. Bununla birlikte, küçük veya büyük bir sapmaya yol açan durumlar hakkında bazı genel açıklamalar yapmamızı sağlayan yaklaşık bir sapma formülü türetilebilir. Başlangıç noktasını $x_0 = 0$ olarak alınırsa yanlılık:

$$Bias\{\hat{f}_h(x)\} = E\{\hat{f}_h(x) - f(x)\} \approx f'(m_j)(m_j - x)$$

$m_j = \left(j - \frac{1}{2}\right)h$'dir. $x = m_j$ olduğunda yaklaşık yanlılık sıfır olur. Histogramın yaklaşık varyansı ise:

$Var\{\hat{f}_h(x)\} \approx \frac{1}{nh} f(x)$ şeklindedir.

Bu formüle göre nh artarken histogramın varyansı azalır. Artan h değeri varyansı azaltırken, yanlılığı arttırır. Histogram yönteminde geniş kutular, küçük varyans büyük yanlılıkken, dar kutular, büyük varyans, küçük yanlılık demektir. Örneklem sayısı arttıkça bu sorunun üstesinden gelmek daha kolaydır. Varyans ve yanlılık h değeri ile ters yönde değiştiği için, varyans ve yanlılık arasında optimal bir denge kuracak h değerlerinin bulunması gerekir. Bu amaçla kullanılan yöntemlerden biri hata kareler ortalamasıdır Histogram için HKO:

$$HKO\{\hat{f}_h(x)\} = var\left(\hat{f}_h(x)\right) + \left[Bias\{\hat{f}_h(x)\}\right]^2$$

Eğer n gözlem sayısı, çok fazla arttırılırsa ($n \to \infty$), h kutu genişliği çok küçük bir değer olarak alınırsa ($h \to 0$) ve bu küçülme çok yavaş bir şekilde gerçekleşirse ($nh \to \infty$), $\hat{f}_h(x)$ fonksiyonun HKO'ı sıfıra gider. Kare ortalamasındaki yakınsama, olasılıkta yakınsamayı belirttiği için, $\hat{f}_h(x)$, $f(x)$'e yakınsar yani $\hat{f}_h(x)$, bilinmeyen $f(x)$ fonksiyonunun tutarlı bir tahminidir. Hata kareler ortalaması, bilinmeyen f fonksiyonunun hem varyansına hem de yanlılığın karesine bağlı olduğu için uygulanması pratikte zordur. Tek bir nokta için f olasılık yoğunluk fonksiyonunun tahmincisi olarak $\hat{f}_h(x)$ yerine daha genel kabul görmüş bir ölçü olan bütünleşik hata kare ortalaması (BHKO) kullanılır. Herhangi bir x noktası için:

$$BHKO(\hat{f}_h(x) \approx \frac{1}{nh} + \frac{h^2}{12}\int\{f'(x)\}^2 dx = \frac{1}{nh} + \frac{h^2}{12}\|f'\|_2^2, \;\; h \to 0 \; için$$

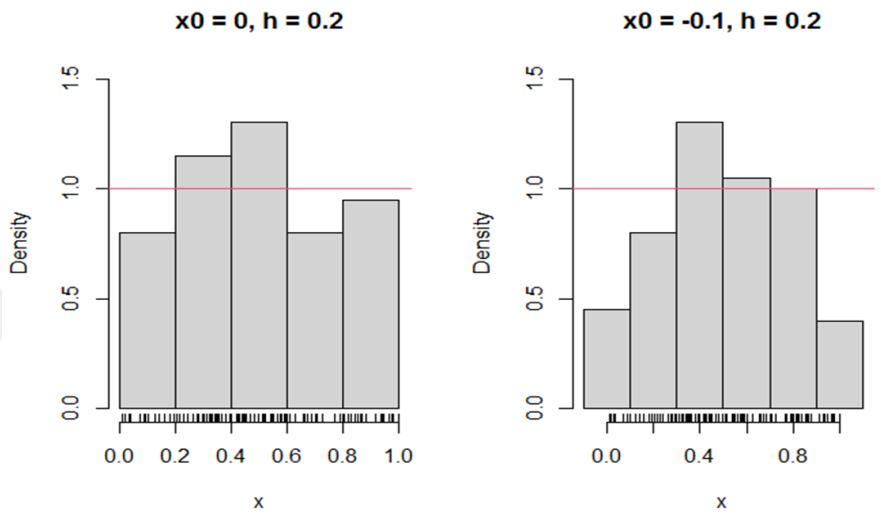
Varyans ve yanlılığın azaltılmasında önemi olan h'nin optimum değeri asimptotik BHKO'nun (ABHKO) minimize edilmesiyle bulunur. ABHKO değeri:

$$ABHKO\left(\hat{f}_h\right) = \frac{1}{nh} + \frac{h^2}{12}\|f'\|_2^2 \text{ şeklindedir.}$$

Bu değerin h'ye göre türevi alınırsa, optimum kutu genişliği yani h_o aşağıdaki şekilde bulunur:

$$h_o = \left(\frac{6}{n\|f'\|_2^2}\right)^{1/3} \sim n^{-1/3}$$

Histogram yönteminin en büyük dezavantajlarından biri başlangıç noktasına göre şeklinin değişmesi durumudur. Başlangıç noktasına göre aynı veri, farklı histogramla, farklı biçimde ifade edilmiş olur. Aşağıdaki gösterimlerde n örneklem hacmi 100 olmak üzere tekdüze (uniform) dağılımdan rastgele üretilen verilere dayanarak başlangıç noktaları farklı, kutu genişlikleri aynı olan histogram grafikleri gösterilmiştir.



Şekil 3.1 Başlangıç noktaları farklı kutu genişlikleri aynı iken histogramlar

Aynı kutu genişliği kullanılmasına rağmen histogramlar, verilerin bazı temel özelliklerine ilişkin farklı açıklamalar vermektedir. Bu sorunun üstesinden gelebilmek için aynı kutu genişliği fakat farklı başlangıç noktalarıyla çizilen histogramlar üzerinden bir ortalama hesaplanarak bir çözüm üretilebilir. Buna değiştirilmiş histogram yöntemi denir.

3.1.2 Kernel yöntemi

Histogram yaklaşımında $f(x)$'i tahmin etmenin yollarından biri şu fikre dayanmaktadır. $\frac{1}{n.(\text{kutu genişliği})}$ burada gözlemler x'i içeren küçük bir aralığa düşer. Kernel yoğunluk tahmin edicisinde ise histogramdaki gibi x_0 başlangıç noktası seçim problemi yoktur. $f(x)$'in tahmini $\frac{1}{n.(\text{kutu genişliği})}$'da artık gözlemler, x'i içeren küçük bir aralığa değil

x'in etrafında yer alan bir aralığa dayanır. Başka bir türetme ise, histogramdan farklı olarak aralık uzunluğunu 2h olarak seçmektir yani (x-h, x+h) aralığı dikkate alınmaktadır.$\hat{f}_h(x)$, tahmini şu şekilde yazılabilir:

$$\hat{f}_h(x) = \frac{1}{2hn}, \{X_i \epsilon (x-h, x+h)\} \tag{3.2}$$

Ağırlıklandırılmış bir tekdüze kernel fonksiyonu kullanılırsa,

$$K(u) = \frac{1}{2} I(|u| \leq 1) \tag{3.3}$$

Bu formülde $u = (x - X_i)/h$'dir. Tekdüze kernel fonksiyonu, x'e olan uzaklığı h'den büyük olmayan her bir X_i gözlemine $1/2$ ağırlık atar. x'ten daha uzaktaki noktalar sıfır ağırlık alır çünkü gösterge işlevi $I(|u| \leq 1)$, tanım gereği $u = (x - X_i)/h$ ölçeklenmiş mesafesinin 1'den büyük tüm değerleri için 0'a eşittir. Amaç $\frac{1}{n.(\text{kutu genişliği})}$ ifadesini formülize etmektir. Buradan yola çıkarak $\hat{f}_h(x)$ tahmini tekrar yazılırsa:

$$\begin{aligned} \hat{f}_h(x) &= \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right) \\ &= \frac{1}{nh} \sum_{i=1}^{n} \frac{1}{2} I\left(\left|\frac{x - X_i}{h}\right| \leq 1\right) \end{aligned} \tag{3.4}$$

(3.4)'ten, (x-h, x+h) aralığına düşen her gözlem için gösterge fonksiyonu1 değerini alır ve frekansa bir katkı sağlar. X_i gözlemi x'e ne kadar yakın olursa olsun (x'in h içinde olması koşuluyla) her katkı eşit olarak (yani bir çarpanıyla) ağırlıklandırılır.

Çizelge 3. 1 Çeşitli Kernel fonksiyonları

Kernel	K(u)
Uniform (Düzgün)	$\frac{1}{2}\,\mathrm{I}(\lvert u\rvert \le 1)$
Üçgen (Triangle)	$(1-\lvert u\rvert)\,\mathrm{I}(\lvert u\rvert \le 1)$
Epanechnikov	$\frac{3}{4}(1-u^2)\,\mathrm{I}(\lvert u\rvert \le 1)$
4.dereceden (Quartic,Biweight)	$\frac{15}{16}(1-u^2)^2\,\mathrm{I}(\lvert u\rvert \le 1)$
6.dereceden (Triweight)	$\frac{35}{32}(1-u^2)^3\,\mathrm{I}(\lvert u\rvert \le 1)$
Normal (Gaussian)	$\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}u^2\right)$
Cosinüs	$\frac{\pi}{4}\cos\left(\frac{\pi}{2}u\right)\mathrm{I}(\lvert u \le 1\rvert)$

f'den $X_1, X_2, \dots, X_n$ rastgele örneklemine dayalı olan f olasılık yoğunluk tahmin edicisinin kernel olasılık yoğunluk tahmin edicisi genel formu şu şekildedir.

$$\hat{f}_h(x) = \frac{1}{n}\sum_{i=1}^{n} K_h(x - X_i), \tag{3.5}$$

$K_h(\cdot) = \frac{1}{h}K(\cdot/h)$ olmak üzere $K(\cdot)$ tabloda verilen kernel fonksiyonlarından herhangi birini temsil eder. h ise bant genişliğini gösterir. Kernel fonksiyonunun ağırlıklandırma işlevi K ile temsil edilir. Kernel tahmini ise $\hat{f}_h(x)$'dir. Histograma benzer şekilde h, tahminin düzgünlüğünü kontrol eder ve h seçimi çok önemli bir problemdir. Eldeki bazı resmi kriterler olmaksızın, h'nin hangi değerinin optimal düzgünlük derecesini sağladığını belirlemek zordur. Bu sorun bant genişliği seçim yöntemleriyle çözülmeye çalışılır.

Kernel Fonksiyonunun Seçimi: Kernel fonksiyonları genellikle olasılık yoğunluk fonksiyonlarıdır, yani K'nın tanım kümesindeki tüm u'lar için $K(u) \geq 0$'dır. $\int K(u)du = 1$ ifadesinin bir sonucu olarak $\int \hat{f}_h(x)dx = 1$'dir yani kernel yoğunluk tahmincisi de bir olasılık yoğunluk fonksiyonudur. Ayrıca, $\hat{f}_h$, K 'nın tüm süreklilik ve

türevlenebilme özelliklerini taşımaktadır. Eğer K'nın aralıksız k kez türevi alınırsa, $\hat{f}_h$ içinde bu durum geçerli olur. $\hat{f}_h$'nın bu özelliği $\hat{f}_h$ grafiğinin düzgünlüğüne de yansır. h bant genişliği aynı olsa bile sürekli kernel fonksiyonlarına dayalı tahminler arasında bile düzgünlükte önemli farklılıklar vardır. Optimal bant genişliği h_{opt}, bazı yöntemlerle bulunabilir ama belli bir h değeri farklı kernel fonksiyonlarıyla kullanıldığında aynı derecede düzgünlüğü garanti etmez (Wakefield Jon, 2013). Yeniden ölçeklendirilmiş kernel fonksiyonlarıyla kernel yoğunluğu tahmin etmek de farklı bir bakış açısıdır. Yeniden ölçeklendirilmiş kernel fonksiyonu şu şekilde ifade edilir:

$$\frac{1}{nh} K\left(\frac{x - X_i}{h}\right) = \frac{1}{n} K_h(x - X_i).$$

Yoğunluk tahmininin altındaki alan bire eşitken, yeniden ölçeklenen her kernel fonksiyonunun altındaki alan ise aşağıda verilen eşitliğe eşittir.

$$\int \frac{1}{nh} K\left(\frac{x - X_i}{h}\right) dx = \frac{1}{nh} \int K(u) h du = \frac{1}{nh} h \int K(u) du = \frac{1}{n}$$

$\hat{f}_h(x)$ tahmini ise şu şekilde yazılır:

$$\hat{f}_h(x) = \frac{1}{n} \sum_{i=1}^{n} K_h(x - X_i) = \sum_{i=1}^{n} \frac{1}{nh} K\left(\frac{x - X_i}{h}\right)$$

$\hat{f}_h(x)$ yeniden ölçeklendirilmiş Kernel fonksiyonlarının toplamı olarak yazılabilir. Grafiksel olarak, yeniden ölçeklenen Kernellar, her bir tümsekliğin gözlemlerden birinde ortalandığı küçük "tümsekler"dir. h'nin farklı değerleri için öncelikle tümseklerin görünümü ve sonuç olarak, toplamlarının $\hat{f}_h$ görünümü değişir. Kernel yoğunluk tahminin istatistiksel özellikleri:

- **Yanlılık**

$$\begin{aligned}\text{Bias}\{\hat{f}_h(x)\} &= E\{\hat{f}_h(x)\} - f(x)\\ &= \frac{1}{n}\sum_{i=1}^{n} E\{K_h(x - X_i)\} - f(x)\\ &= E\{K_h(x - X_i)\} - f(x)\end{aligned}$$

Buradan yanlılık:

$$\text{Bias}\{\hat{f}_h(x)\} = \int \frac{1}{h} K\left(\frac{x-u}{h}\right) f(u)du - f(x).$$

$s = \frac{u-x}{h}$ olarak tanımlanırsa, kernelin simetriği K(-s)=K(s) olmak üzere $f(u)'nun\ x$ etrafında ikinci dereceden Taylor açılımı gösterilebilir:

$$Bias\{\hat{f}_h(x)\} = \frac{h^2}{2} f''(x)\mu_2(K) + o(h^2), \quad as\ h \to 0. \tag{3.6}$$

Bu formülde $\mu_2(K) = \int s^2 K(s)ds$ şeklinde hesaplanır. (3.6)'dan yanlılığın h ile orantılı olduğu görülür. Bu nedenle, yanlılığı azaltmak için küçük bir h seçilmelidir. Ayrıca $Bias\{\hat{f}_h(x)\}$, x'deki yoğunluğun eğriliği olan $f''(x)$'e bağlıdır. Sapmanın büyüklüğü, f'' mutlak değerine yansıyan f'nin eğriliğine de bağlıdır. Büyük f'' değerleri, $Bias\{\hat{f}_h(x)\}$'ın büyük değerlerini ifade eder (Wakefield Jon 2013).

- **Varyansı**

$$Var\{\hat{f}_h(x)\} = \frac{1}{nh}\|K\|_2^2 f(x) + o\left(\frac{1}{nh}\right), \quad as\ \ nh \to \infty.$$

$\|K\|_2^2$, $\int K^2(s)ds$'nin kısaltılmış gösterimidir. Kernel yoğunluk tahmin edicisinin varyansı neredeyse nh^{-1} ile orantılıdır. Bu nedenle, küçük bir varyans için oldukça büyük bir h seçilmelidir. Büyük h değerleri, daha büyük aralıklar (x-h, x+h), her aralıkta daha fazla gözlem ve dolayısıyla toplam $\sum_{i=1}^{n} K_h(x - X_i)$'de sıfırdan farklı ağırlık alan daha fazla gözlem anlamına gelir. Daha fazla gözlem kullanarak toplamda daha az değişkenliğe sahip toplamlar üretilecektir. Benzer şekilde, belirli bir h değeri için (büyük veya küçük), örnek boyutu n'nin arttırılması 1/nh'yi azaltacak ve dolayısıyla

varyans da küçülecektir. Daha fazla toplam gözlem sayısına sahip olmak, ortalama her [x-h, x+h] aralığında daha fazla gözlem olacağı anlamına gelir.$\|K\|_2^2$ bu terim, tekdüze kernel gibi düz kerneller için oldukça küçük olacaktır. Pürüzsüz ve yassı kernellerin, her örnekte tüm gözlemlere kabaca eşit ağırlık verdiğinden, tekrarlı örneklemelerde daha az uç tahminler ürettiği söylenebilir.

- **Hata Kareler Ortalaması**

Parametrik olmayan yoğunluk tahmininde h bant genişliğini seçmek histogram yönteminde olduğu gibi kernel yöntemi için de bir sorundur. Hem varyans hem de sapma en aza indirmek istenir ancak h'yi arttırmak varyansı düşürürken sapmayı arttırır tersi durumda h'yi azaltmak varyansı arttırır sapmayı ise düşürür. Hata kareler ortalamasını ise en aza indirmek yani varyans ve kare yanlılık arasındaki toplam, aşırı ve az düzleştirme arasında bir dengeyi temsil eder. Ayrıca, HKO'na bakmak, kernel yoğunluk tahmin edicisinin tutarlı olup olmadığını değerlendirmenin bir yolunu da sağlar. $\hat{f}_h(x)$ için hata kareler ortalaması:

$$HKO\{\hat{f}_h(x)\} = \frac{h^4}{4} f''(x)^2 \mu_2(K)^2 + \frac{1}{nh} \|K\|_2^2 f(x) + o(h^4) + o\left(\frac{1}{nh}\right). \qquad (3.7)$$

Ortalama karedeki yakınsama, olasılıkta yakınsamayı, yani tutarlılığı gösterir. (3.7)'ye bakarak, $h \to 0$ ve $nh \to \infty$ giderken, kernel yoğunluk tahmincisinin hata kareler ortalaması da sıfıra gider buna dayanarak, kernel yoğunluk tahmincisinin gerçekten tutarlı olduğu söylenir. Fakat, HKO f ve f'' ye bağlıdır ve her iki fonksiyon da pratikte bilinmez.

Sonuç olarak, $f(x)$ ve $f''(x)$ için uygun ikameleri elde etmenin bir yolu bulunmadıkça, $h_{opt}(x)$ pratikte uygulanamaz. Ayrıca $h_{opt}(x)$, x'e bağlıdır ve dolayısıyla yerel bir bant genişliği olarak adlandırılır. Histogram yönteminde hata kareler ortalaması yerine, bütünleşik hata kareler ortalaması (BHKO) kullanılarak yerel bir tahmin ölçüsünden çıkılır ve daha genel kullanılan bir ölçü bulunur. Ayrıca sorunun boyutsallığı azaltılmış olur. Bu nedenle, BHKO optimal bant genişliği sürecine de bakılırsa, kernel yoğunluk tahmin edicisi için şu şekilde verilir:

$$\text{BHKO} = \frac{1}{nh}\|K\|_2^2 + \frac{h^4}{4}\{\mu_2(K)\}^2\|f''\|_2^2 + o\left(\frac{1}{nh}\right) + o(h^4) \;\; as\; h \to 0 \;\; nh \to \infty.$$

Yüksek dereceli terimler göz ardı edilerek, bütünleşik hata kareler ortalaması için asimptotik bütünleşik hata kareler ortalaması adı verilen yaklaşık bir formül şu şekilde verilebilir:

$$ABHKO\left(\hat{f}_h\right) = \frac{1}{nh}\|K\|_2^2 + \frac{h^4}{4}\{\mu_2(K)\}^2\|f''\|_2^2.$$

$ABHKO$ optimal bant genişliği ise:

$$h_{opt} = \left(\frac{\|K\|_2^2}{\|f''\|_2^2\{\mu_2(K)\}^2 n}\right)^{1/5} \sim n^{-1/5}. \tag{3.8}$$

Sonuç olarak bakıldığında, h_{opt} hala $\|f''\|_2^2$'ye bağlı olduğu için bilinmeyen niceliklerle uğraşma sorunu tam olarak çözülememiştir. (2.8)'e h_{opt} eklenmesi aşağıdaki formülü vermektedir.

$$ABHKO\left(\hat{f}_{hopt}\right) = \frac{5}{4}(\|K\|_2^2)^{4/5}(\mu_2(K)\|f''\|_2)^{2/5} \tag{3.9}$$

burada $n^{-4/5}$'den önceki terimlerin n'ye göre sabit olduğu belirtilmişti. Eğer örneklem boyutu giderek arttırılırsa, $ABHKO$, $n^{-4/5}$ oranında yakınsar. Histogram için $ABHKO$, daha düşük $n^{-2/3}$ oranıyla bir yakınsama gösterdiği için kernel yoğunluk tahmin edicisinin daha üstün olduğu söylenebilir (Hardle vd. 2004).

Uygulamalarda genellikle tek boyutlu tahminleri elde etmekle değil, çok değişkenli yoğunlukları tahmin etmekle ilgilenilir. $X = (X_1, \dots, X_d)^T$ d-boyutlu rastgele vektör olmak üzere burada $X_1, \dots, X_d$ tek boyutlu rastgele değişkenlerdir. Bu şekilde n boyutlu rastgele bir örnek, d rastgele değişkenlerin her biri için n gözleme sahip olduğu anlama gelir.X_i vektörü için d rastgele değişkenlerinin her biri için i. gözlemin toplamı olduğu varsayılsın. X_i:

$$X_i = \begin{pmatrix} X_{i1} \\ \vdots \\ X_{id} \end{pmatrix}, \;\; i = 1, \dots, n$$

X_{ij}, X_j rastgele değişkeninin i. gözlem değeridir. Burada amaç, $X = (X_1, \dots, X_d)^T$ olasılık yoğunluğunu tahmin etmektir. $X_1, \dots, X_d$ rastgele değişkenlerinin ortak olasılık yoğunluk fonksiyonu f:

$$f(x) = f(x_1, \dots, x_d).$$

Tek boyut için hesaplanan kernel yoğunluk tahmini d-boyut için uyarlanırsa;

$$\hat{f}_h(x) = \frac{1}{n}\sum_{i=1}^{n} \frac{1}{h^d} \mathcal{K}\left(\frac{x - X_i}{h}\right)$$

$$= \frac{1}{n}\sum_{i=1}^{n} \frac{1}{h^d} \mathcal{K}\left(\frac{x_1 - X_{i1}}{h}, \dots, \frac{x_d - X_{id}}{h}\right) \qquad (3.10)$$

$\mathcal{K}$, çok değişkenli kernel fonksiyonunu belirtir. Eğer bant genişliğini $h = (h_1, \dots, h_d)^T$ çok değişkenli durum için genişletirsek,

$$\hat{f}_h(x) = \frac{1}{n}\sum_{i=1}^{n} \frac{1}{h_1, \dots h_d} \mathcal{K}\left(\frac{x_1 - X_{i1}}{h_1}, \dots, \frac{x_d - X_{id}}{h_d}\right) \qquad (3.11)$$

Çok boyutlu kernel $\mathcal{K}(u) = \mathcal{K}(u_1, u_2, \dots, u_d)$ için çarpımsal kernel şeklinde ifade edilebilir:

$$\mathcal{K}(u) = K(u_1) \cdot \cdots \cdot K(u_d), \qquad (3.12)$$

Burada K tek değişkenli kernel fonksiyonunu temsil eder. Bu durumda 3.11'deki $\hat{f}_h(x)$ aşağıdaki şekilde tahmin edilir:

$$\hat{f}_h(x) = \frac{1}{n}\sum_{i=1}^{n} \left\{ \prod_{j=1}^{d} h_j^{-1} K\left(\frac{x_j - X_{ij}}{h_j}\right) \right\} \qquad (3.13$$

3.1.3 Yanlılık ve varyans

Öncelikle standart varsayımın bir sonucu olarak $\int \mathcal{K}(u)du = 1$ olduğu belirtilsin. $\hat{f}_H$ tahmini bir yoğunluk fonksiyonudur, yani, $\int \hat{f}_H(x)dx = 1$ 'dir. Ayrıca tahmin herhangi bir x noktasında süreklidir:

$$\hat{f}_H(x) = \frac{1}{n}\sum_{i=1}^{n} \mathcal{K}_H(X_i - x) \xrightarrow{P} f(x) \tag{3.14}$$

Bu bölümde asimptotik yaklaşımlardan bahsedilecektir. Tek değişkenli durumda olduğu gibi, ikinci dereceden bir Taylor açılımı kullanılır. ∇_f, gradyant matrisi ve $\mathcal{H}_f$, Hessian matrisi olmak üzere x etrafındaki $f(\cdot)$ Taylor açılımı:

$$f(x+t) = f(x) + t^T\nabla_f(x) + \frac{1}{2}t\mathcal{H}_f(x)t + o(t^Tt)$$

Beklenen değer:

$$E(\hat{f}_H(x)) \approx \int \mathcal{K}(s)\left\{f(x) + s^TH^T\nabla_f(x) + \frac{1}{2}s^TH^T\mathcal{H}_f(x)H_s\right\}ds$$

Yanlılık:

$$Bias\{\hat{f}_H(x)\} \approx \frac{1}{4}\mu_2^2(\mathcal{K})\int \left[iz\{H^T\mathcal{H}_f(x)H\}\right]^2 dx$$

Varyans:

$$\begin{aligned} Var\{\hat{f}_H(x)\} &= \frac{1}{n}\int \{\mathcal{K}_H(u-x)\}^2du - \frac{1}{n}\{E(\hat{f}_H(x))\}^2 \\ &\approx \int \frac{1}{ndet(H)}\mathcal{K}^2(s)\,f(x+Hs)ds \\ &\approx \int \frac{1}{ndet(H)}\mathcal{K}^2(s)\{f(x) + s^TH^T\nabla_f(x)\}ds \\ &\approx \int \frac{1}{ndet(H)}\|\mathcal{K}\|_2^2\,f(x) \end{aligned}$$

$\mathcal{K}$'nın d-boyutlu kare L_2 normu $\|\mathcal{K}\|_2^2$ ile belirtilir. Sonuç olarak, çok değişkenli kernel yoğunluk tahmincisi için Asimptotik bütünleşik hata karesi aşağıdaki şekilde ifade edilir:

$$ABHK(H) = \frac{1}{4}\mu_2^2(\mathcal{K}) \int \left[iz\{H^T \mathcal{H}_f(x) H\}\right]^2 dx + \frac{1}{ndet(H)} \|\mathcal{K}\|_2^2$$

Buradaki sorun ise ABHK optimal bant genişliğini seçmektir. Yine aynı şekilde bant genişliği ABKH'deki yanlılık ve varyans arasındaki dengeye bağlıdır. $H = hH_0$ ve $\det(H_0) = 1$ olmak üzere h bir skaler büyüklüğü belirtsin. ABKH şu şekilde yazılır:

$$ABKH(H) = \frac{1}{4}h^4\mu_2^2(\mathcal{K}) \int \left[iz\{H_0^T H_f(x) H_0\}\right]^2 dx + \frac{1}{nh^d} \|K\|_2^2 \qquad (3.15)$$

Yalnızca h'deki değişikliklere izin verilirse, h bant genişliği ve $ABKH$ için en uygun (optimal) sonuçları sırasıyla aşağıdaki gibidir:

$$h_{opt} \sim n^{-1/(4+d)}, \quad ABKH(h_{opt} H_0) \sim n^{-4/(4+d)}.$$

Bu nedenle, çok değişkenli yoğunluk tahmincisi, özellikle d büyük olduğunda, tek değişkenli yönteme kıyasla daha yavaş bir yakınsama hızına sahiptir. $H = hI_d{}'yi$ (tüm d boyutlarında aynı bant genişliği) göz önünde bulundurulursa ve n örnek boyutu sabitlenirse, tahminin makul ölçüde düzgün olduğundan emin olmak için asimptotik bütünleşik hata karesi, optimal bant genişliği tek boyutlu durumda olduğundan çok daha büyük olmalıdır. Ayrıca, boyutların sayısı d arttıkça, hesaplama zorluğu da artar. Bu nedenle, çok boyutlu yoğunluk tahmini genellikle $d \geq 5$ ise uygulanmaz.

3.1.4 Bant genişliği seçimi

H bant genişliğinin otomatik, veri odaklı seçim sorunu, çok değişkenli durum için daha fazla öneme sahiptir. Bir veya iki boyutta, sadece farklı bant genişlikleri için yoğunluk tahminlerinin grafiğine bakarak etkileşimli olarak uygun bir bant genişliği seçmek

kolayken, üç, dört veya daha fazla boyutta grafiksel temsil sorunu ortaya çıkar. Tek boyutlu durumda olduğu gibi, eklenti bant genişlikleri, özellikle basit bant genişlikleri ve çapraz doğrulama bant genişlikleri kullanılır. Çok değişkenli durumda, tek boyutlu bant genişliği seçicileri ile Silverman'ın kuralı ve en küçük kareler çapraz doğrulaması için genellemeler yapılır.

3.1.4.1 Silverman'ın temel kuralı

Bilinmeyen parametre içeren bir ifade varsa, bilinmeyen parametre, bir tahminle değiştirilir eklenti yöntemlerinin altında yatan bu kuraldır. Asimptotik bütünlesik hata kareleri ortalaması optimal bant genişliğini veren formülden yola çıkarak

$$h_{opt} = \left(\frac{\|K\|_2^2}{\|f''\|_2^2 \{\mu_2(K)\}^2 n} \right)^{1/5} \sim n^{-1/5}. \tag{3.16}$$

(3.16)'deki ifadede $\|f''\|_2^2$ değeri bilinmemektedir. Bilinmeyen f yoğunluk fonksiyonunun μ ortalama ve σ^2 varyans ile normal dağılım ailesine ait olduğu varsayılsın:

$$\begin{aligned} \|f''\|_2^2 &= \sigma^{-5} \int \{\varphi''(x)\}^2 dx \\ &= \sigma^{-5} \frac{3}{8\sqrt{\pi}} \approx 0.212\sigma^{-5}, \end{aligned} \tag{3.17}$$

burada $\varphi(\cdot)$, standart normal dağılımın olasılık yoğunluk fonksiyonunu gösterir. Bilinmeyen standart sapma σ yerine σ'nın tahmin edicisi, $\hat{\sigma}$ kullanılır,

$$\hat{\sigma} = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2} \tag{3.18}$$

(3.16)'deki formülü pratikte uygulamak için bir kernel fonksiyonu seçilmelidir. Standart normal dağılımı, olasılık yoğunluk fonksiyonuyla aynı olduğu için Gauss kerneli seçilir ve φ ile de gösterilirse aşağıdaki "Temel Kural" bant genişliği elde edilir:

$$\hat{h}_{rot} = \left(\frac{\|\varphi\|_2^2}{\left\|\widehat{f''}\right\|_2^2 \mu_2^2(\varphi) n} \right)^{1/5} \qquad \left\|\widehat{f''}\right\|_2^2 = \hat{\sigma}^{-5} \frac{3}{8\sqrt{\pi}} \quad \text{olmak üzere}$$

$$\hat{h}_{rot} = \left(\frac{4\hat{\sigma}^5}{3n} \right)^{1/5} \approx 1.06 \hat{\sigma} n^{-\frac{1}{5}} \tag{3.19}$$

Burada anlatılan uygulamada $f(x)$'in normal dağıldığı varsayılmıştır. Bu varsayım parametrik olmayan yoğunluk tahmin felsefesinin tam tersidir. Eğer X'in normal dağılıma sahip olunduğu bilinseydi yoğunluk çok daha kolay ve verimli bir şekilde tahmin edilebilirdi. Normallik varsayımı altında anlatılan temel kural bant genişliği kullanışlı ve uygulanabilir bir formüldür. Uygulamada X, normal dağılım gösteriyorsa $\hat{h}_{rot}$ optimal bant genişliğini verir. X, normal dağılımdan çok farklı değilse o zaman, $\hat{h}_{rot}$ optimumdan çok uzak olmayan bir bant genişliği verir. Sonuç olarak, tek modlu, neredeyse simetrik ve çok kalın kuyruklu olmayan tüm dağılımlar için "Temel Kural" bant genişliğinin makul sonuçlar verdiği söylenebilir.

Temel kural bant genişliği ile ilgili bir sorun, aykırı değerlere duyarlılığıdır. Tek bir aykırı değer olması, çok büyük bir σ tahminine neden olabilir bu da çok büyük bir bant genişliği verecektir. Bu noktada başka alternatif bir yöntem olarak, çeyrekler arası aralığından daha sağlam bir tahminci elde edilir:

$$R = X_{[0.75n]} - X_{[0.25n]}, \tag{3.20}$$

yani, sadece %75 üst çeyrek ve %25 alt çeyrek, arasındaki örnek çeyrekler arası aralığı hesaplanır. Yine de burada gerçek o.y.f'nun normal olduğu varsayılırsa yani, $X \sim N(\mu, \sigma^2)$ ve $Z = \frac{X-\mu}{\sigma} \sim N(0,1)$,

$$\begin{aligned} R &= X_{[0.75n]} - X_{[0.25n]} \\ &= \left(\mu + \sigma Z_{[0.75n]}\right) - \left(\mu + \sigma Z_{[0.25n]}\right) \\ &= \sigma(Z_{[0.75n]} - Z_{[0.25n]}) \\ &\approx \sigma\big(0.67 - (-0.67)\big) = 1.34\sigma, \end{aligned}$$

Buradan $\hat{\sigma} = \frac{R}{1.34}$ şeklinde tahmin edilir. (3.19)'de verilen $\hat{h}_{rot}$'da $\hat{\sigma}$ yerine konursa aralarındaki ilişki:

$$\hat{h}_{rot} = 1.06 \frac{R}{1.34} n^{-\frac{1}{5}} \approx 0.79 \hat{R} n^{-\frac{1}{5}} \tag{3.21}$$

(3.19) ve (3.20)'de verilen formüllerinden "daha iyi bir temel kural" yazılabilir:

$$\hat{h}_{rot} = 1.06 min\left\{\hat{\sigma}, \frac{R}{1.34}\right\} n^{-1/5} \tag{3.22}$$

Yine, hem (3.19) hem de (3.22) gerçek olasılık yoğunluğu normal dağılıma benziyorsa bu formül oldukça iyi sonuçlar verecektir, ancak gerçek yoğunluk normal dağılımın şeklinden önemli ölçüde saparsa (örneğin çok modlu olarak), temel bant genişliği kullanılarak elde edilen tahminlerde önemli ölçüde yanılma olacaktır (Wolfgang vd. 2004).

3.1.4.2 Yenilenmiş eklenti yöntemi

Normallik varsayımı altında hesaplanan Silverman'ın temel kuralı normallik sağlanmadığında büyük ölçüde başarısız olacaktır. Burada $\|f''\|_2^2$ için doğal bir iyileştirme, parametrik olmayan tahminleri kullanmaktan oluşur. Başka bir yöntem ise, bütünleşik hata kareler ortalaması (BHKO) için daha iyi bir yaklaşımın kullanılmasıdır. Çapraz geçerlilik yönteminin aksine eklenti bant genişliği seçicileri, bütünleşik hata kareler ortalamasını en aza indiren bir bant genişliği bulmaya çalışır. Seçilen bir $\hat{h}$ bant genişliğinin kalitesini değerlendirmenin yaygın bir yöntemi, onu nispi değerde BHKO'nı optimal bant genişliği olan h_{opt} ile karşılaştırmaktır. $\hat{h}$ 'nın, h_{opt}'a yakınsaklığı $O_p(n^{-\alpha})$ mertebesinde olduğu söylenir:

$$n^{\alpha} \frac{\hat{h} - h_{\,opt}}{h_{\,opt}} \xrightarrow{P} T,$$

Burada T, n'den bağımsız rastgele bir değişkendir. Eğer α=1/2 ise, buna genellikle $\sqrt{n}$ yakınsama denir ve bu yakınsama oranı da Hall & Marron'ın (1991) gösterdiği gibi elde edilebilecek en iyi orandır. Park & Marron (1990), f'nin parametrik olmayan bir tahminini kullanarak ve bu tahminin ikinci türevini alarak BHKO'da $\|f''\|_2^2$ tahminini önerdi (Wolfgang vd. 2004). Burada bir g bant genişliği kullanıldığı varsayılsın, o zaman f_g'nin ikinci türevi şu şekilde hesaplanabilir:

$$\hat{f}_g''(x) = \frac{1}{ng^3}\sum_{i=1}^{n} K''\left(\frac{x - X_i}{g}\right).$$

Burada da bant genişliği seçimi problemi ortaya çıkacaktır. İkinci türev için temel kural (Rule of thumb) ile bulunan bant genişliği ilk aşamada kullanılabilir. Başka bir problem ise, düzeltilmiş tahmin yanlılığını kullanarak üstesinden gelinebilecek olan $\left\|\hat{f}_g''\right\|_2^2$ sapması nedeniyle başka bir sorun ortaya çıkar.

$$\widehat{\|f''\|_2^2} = \left\|\hat{f}_g''\right\|_2^2 - \frac{1}{ng^5}\|K''\|_2^2. \tag{3.23}$$

$\|f''\|_2^2$ yerine Asimptotik bütünleşik hata kareler ortalamasında h'ye göre optimize etmek bant genişliği seçicisini verir.

$$\hat{h}_{pm} = \left(\frac{\|K\|_2^2}{\widehat{\|f''\|_2^2}\mu_2^2(K)n}\right)^{1/5} \tag{3.24}$$

Park & Marron (1990), $O_p(n^{-4/13})$ düzeyinde $\hat{h}_{pm}$'nin $h_{opt}{}'a$ göreceli bir yakınsama oranına sahip olduğunu gösterdiler yani optimal bant genişliği h_{opt}'a yakınsama oranı $\alpha = \frac{4}{13} < \frac{1}{2}$ anlamına gelmektedir. Dezavantajı, küçük bant genişlikleri sebebiyle $\widehat{\|f''\|_2^2}$ tahmin edicisinin negatif sonuçlar vermesidir (Wolfgang vd. 2004).

3.1.4.3 Çapraz geçerlilik kuralı

Çapraz geçerlilik yöntemi, parametre veya fonksiyon tahmininin özel yapısından bağımsızdır. Bant genişliği seçim problemi göz önüne alındığında, çapraz geçerlilik teknikleri, temel kural yaklaşımına (Rule of thumb) göre daha geniş bir f yoğunluk fonksiyonları sınıfına uyum sağlar. Temel kural bant genişliği referans dağılımın olasılık yoğunluk fonksiyonu için en uygun yöntem iken çok modlu yoğunluk fonksiyonları için başarısız olacaktır. Yoğunluk tahmini için en küçük kareler çapraz geçerlilik bütünleşik kare hatasıyla optimal bant genişliğini tahmin eder (Hardle vd. 2004). Çok değişkenli durum için,

$$BKH(H) = \int \{\hat{f}_H(x) - f(x)\}^2 dx$$

$$= \int \hat{f}_H^2(x)\, dx - 2\int \hat{f}_H(x)\, f(x) dx + \int f^2(x)\, dx.$$

Bu formülde bilinmeyen ve tahmin edilmesi gereken ikinci terimdir. İlk terim ise verilerden tahmin edilir. $\int \hat{f}_H(x)\, f(x) dx = E\hat{f}_H(X)$, bu terimi "leave-one-out" tahmin edicisi ile tahmin edersek,

$$\widehat{E\hat{f}_H(X)} = \frac{1}{n}\sum_{i=1}^{n} \hat{f}_{H,-i}(X_i)$$

$$\hat{f}_{H,-i(x)} = \frac{1}{n-1}\sum_{i\neq j, j=1}^{n} \mathcal{K}_H(X_j - x)$$

Tek değişkenli için verilen çapraz geçerlilik formülünü çok değişkenliye genellemesi ise aşağıdaki gibidir.

$$CV(H) = \frac{1}{n^2 \det(H)} \sum_{i=1}^{n}\sum_{j=1}^{n} \mathcal{K} * \mathcal{K}\{H^{-1}(X_j - X_i)\} - \frac{2}{n(n-1)} \sum_{i=1}^{n}\sum_{j=1, j\neq i}^{n} \mathcal{K}_H(X_j - X_i)$$

Buradaki zorluk ise bant genişliğinin d*d boyutlu H matrisine bağlı olmasıdır. H, bir köşegen matris olarak kabul edilirse, bu bir d-boyutlu optimizasyon problemi olarak kalır.

Bu, diğer çapraz geçerlilik yaklaşımları için de geçerlidir.

3.2 Nonparametrik Regresyon Tahmini

Nonparametrik regresyon tahmini için Kernel yöntemlerinden Nadaraya-Watson ve yerel polinom regresyonu kullanılabilir. Nadaraya-Watson ağırlıklandırılmış ortalama değerleri kullanırken yerel polinom regresyonu, ortalama değerleri yerine ağırlıklandırılmış en küçük kareler kestirim değerlerini kullanır. Yerel polinom regresyonu Nadaraya-Watson tahmin edicisinin genelleştirilmiş halidir.

3.2.1 Nadaraya-Watson Kernel tahmin edicisi

Nadaraya-Watson tahmin edicisi, herhangi bir x değerinin komşuluğunda bağımsız x_i değerlerine karşılık gelen bağımlı y_i değerlerinin ağırlıklı ortalamasıdır. Bu ağırlıklar x değerine daha uzak olan x_i değerleri için küçük olarak tanımlanırken, x değerine daha yakın olan x_i değerleri için daha büyük tanımlanmaktadır. Öncelikle $f(x)$'in tahmini aşağıdaki gibidir.

$$\begin{aligned} f(x) &= E[Y|x] \\ &= \int yp(y|x)dy \\ &= \frac{1}{p(x)} \int yp(x,y)dy. \end{aligned}$$

Kernel yöntemiyle $p(x,y)$'nin tahmini:

$$\hat{p}^{(\lambda_x,\lambda_y)}(x,y) = \frac{1}{n\lambda_x\lambda_y} \sum_{i=1}^{n} K_x\left(\frac{x-x_i}{\lambda_x}\right) K_y\left(\frac{y-y_i}{\lambda_y}\right) \tag{3.25}$$

$p(x)$'in tahmini: (Wakefield Jon 2013)

$$\hat{p}^{(\lambda_x)}(x) = \frac{1}{n\lambda_x} \sum_{i=1}^{n} K_x\left(\frac{x-x_i}{\lambda_x}\right)$$

(3.25)'de verilen tahminden yola çıkarak Nadaraya-Watson kernel regresyon tahmin edicisi:

$$\hat{f}(x) = \frac{\frac{1}{n\lambda_x\lambda_y}\sum_{i=1}^{n}\int y\, K_x\left(\frac{x-x_i}{\lambda_x}\right)K_y\left(\frac{y-y_i}{\lambda_y}\right)dy}{\frac{1}{n\lambda_x}\sum_{i=1}^{n}K_x\left(\frac{x-x_i}{\lambda_x}\right)}$$

$$= \frac{\sum_{i=1}^{n}K_x\left(\frac{x-x_i}{\lambda_x}\right)\int(y_i + u\lambda_y)K_y(u)du}{\sum_{i=1}^{n}K_x\left(\frac{x-x_i}{\lambda_x}\right)}$$

$$= \frac{\sum_{i=1}^{n}K\left(\frac{x-x_i}{\lambda}\right)y_i}{\sum_{i=1}^{n}K\left(\frac{x-x_i}{\lambda}\right)} \qquad (3.26)$$

Burada, $\int K_y(u)\,du = 1$ ve $\int uK_y(u)\,du = 0$. Ayrıca son satırda, $\lambda = \lambda_x$ ve $K = K_x$'e eşittir. Bu tahmin edici doğrusal düzleştirici olarak da yazılabilir:

$$\hat{f}^{(\lambda)}(x) = \sum_{i=1}^{n} S_i^{(\lambda)}(x)Y_i,$$

Formülde $S_i^{(\lambda)}(x)$, ağırlıklar olarak tanımlanır.

$$S_i^{(\lambda)}(x) = \frac{K\left(\frac{x-x_i}{\lambda}\right)}{\sum_{i=1}^{n}\left(\frac{x-x_i}{\lambda}\right)}$$

Tahmin edicinin makul davranışı için λ düzleştirme parametresinin seçimi çok önemlidir.

Her zaman olduğu gibi kare yanlılığı ve varyans nedeniyle bileşenlerine ayrılabilen asimptotik bütünleşik kare hatası incelenmelidir. $\lambda_n \to 0$ ve $n\lambda_n \to \infty$ iken x noktasında N-W tahmin edicisinin yanlılığı;

$$bias\left[\hat{f}^{(\lambda_n)}(x)\right] \approx \frac{\lambda_n^2\sigma_K^2}{2}\left(f''(x) + 2f'(x)\frac{p'(x)}{p(x)}\right)$$

Burada $p(x), x$'in bilinmeyen yoğunluğudur. λ_n arttıkça yanlılığın artması beklenen bir durumdur. Ayrıca yanlılık, kıvrımlarda $f(\cdot)$'nın arttığı noktalarda yani, büyük $f''(x)$ arttığı ve "tasarım yoğunluğunun" yani $p'(x)$ türevinin büyük olduğu noktalarda da artar. Sözde tasarım yanlılığı $2f'(x)p'(x)/p(x)$ olarak tanımlanır. x noktasında N-W tahmin edicisinin varyansı:

$$var\left[\hat{f}^{(\lambda)}(x)\right] \approx \frac{K_2\sigma^2}{n\lambda_n}\frac{1}{p(x)}$$

Daha kolay olması için varyans σ^2'nin, $var(Y|x)$' e eşit ve sabit olduğu kabul edilir. Tahmin edicinin varyansı, x değerlerinin yoğunluğunun artması ölçüm hatasının azalmasıyla ve yerel örnek hacmi $n\lambda_n$ ile azalır. Sonuç olarak, küçük λ 'nın yanlılığı azalttığı ancak varyansı arttırdığı görülür. Yanlılığın karesi ve varyansı birleştirerek ve x üzerinden integral alarak bütünleşik hata kareler ortalaması bulunur.

$$BHKO\left(\hat{f}^{(\lambda_n)}\right) \approx \frac{\lambda_n^4 \sigma_K^4}{4}\left(f''(x) + 2f'(x)\frac{p'(x)}{p(x)}\right)^2 dx + \frac{K_2\sigma^2}{n\lambda_n}\int \frac{1}{p(x)}dx. \tag{3.27}$$

Bu ifadenin türevi alınıp sıfıra eşitlenirse, optimal bant genişliği şu şekilde elde edilir:

$$\lambda_n^* = \left(\frac{1}{n}\right)^{1/5}\left(\frac{\sigma^2 K_2 \int p(x)^{-1}dx}{\sigma_K^4 \int (f''(x)p'(x)/p(x))^2 dx}\right)^{1/5} \tag{3.28}$$

Burada $\lambda^* = O\left(n^{-1/5}\right)$'dir. Bu ifadeyi (3.27)'de yerine koyarsak, BHKO'nın birçok parametrik olmayan tahmin edici için geçerli ve varyansı $O(n^{-1})$ olan çoğu parametrik tahmin edicinin aksine, $O(n^{-4/5})$ olduğunu gösterir. Verimlilikteki kayıp, parametrik olmayan yöntemlerin sağladığı esnekliğin bir dezavantajıdır. (3.28)'deki ifade birçok bilinmeyen niceliğe bağlıdır ve bu terimleri tahmin etmek için "eklenti" yöntemleri olsa da daha yaygın kullanılan bir yaklaşım çapraz geçerliliktir (Wakefield Jon 2013).

3.2.2 Yerel polinom regresyonu

Yerel polinom regresyonu, Nadaraya-Watson kernel tahmin edicisinin genelleştirilmiş halidir. Kernel regresyonuna benzer fakat ağırlıklandırılmış ortalama değerleri yerine ağırlıklandırılmış en küçük kareler kestirim değerlerini kullanır. Herhangi bir x noktasındaki düzleştirme, ağırlıkların kernel fonksiyonunun yüksekliğine karşılık gelen ağırlıklandırılmış en küçük kareler doğrusuna uydurularak elde edilir. $w_i(x) = K[(x_i - x)/\lambda]$ ağırlıklandırılmış bir fonksiyon olmak üzere, ağırlıklandırılmış kareler toplamını minimize etmek için $\beta_{0x} = f(x)$ seçilmiş olsun. $\sum_{i=1}^{n} w_i(x)\,(Y_i - \beta_{0x})^2$ ifadesinin çözümü ile aşağıdaki tahmin elde edilir:

$$\hat{f}(x) = \hat{\beta}_{0x} = \frac{\sum_{i=1}^{n} w_i(x)\, Y_i}{\sum_{i=1}^{n} w_i(x)},$$

Bu da, (3.26)'daki Nadaraya-Watson kernel regresyon tahmin edicisinin ağırlıklı en küçük kareler kullanılarak tahmin edilen yerel sabit bir modele karşılık geldiğini göstermektedir. Notasyonal gösterimde kolaylık olması için, $w_i(x)$ ağırlığının λ yumuşatma parametresine bağlı olmadığı kabul edilmiştir. Elde edilmek istenen her tahmin için ayrı bir ağırlıklı en küçük kareler uygulanmaktadır. Bu formülasyon ile yerel sabit bir model olan Nadaraya-Watson tahmin edicisinin yerini yerel polinom alır. Sabit bir x komşuluğundaki u değerleri için polinom şu şekilde tanımlanır:

$$P_x(u; \beta_x) = \beta_{0x} + \beta_{1x}(u - x) + \frac{\beta_{2x}}{2!}(u - x)^2 + \cdots + \frac{\beta_{px}}{p!}(u - x)^p,$$

$\beta_x = [\beta_{0x}, \ldots, \beta_{px}]$'dir. Buradaki fikir, x komşuluğundaki f'yi polinomal $P_x(u; \beta_x)$ ile yaklaşık olarak bulmaktır. $\hat{\beta}_x$ parametresi yerel ağırlıklandırılmış kareler toplamını minimize etmek için seçilir:

$$\sum_{i=1}^{n} w_i(x)\,[Y_i - P_x(x_i; \beta_x)]^2.$$

u için f'nin yerel tahmini:

$$\hat{f}(u) = P_x(u; \hat{\beta}_x).$$

Bu tahmin x'in yerel bir komşuluğunda kullanabilir, ancak bunun yerine bulunmak istenen her x değeri için yeni bir yerel polinomal bulunabilir. $u = x$ değeri için,

$$\hat{f}(x) = P_x(x; \hat{\beta}_x) = \hat{\beta}_{0x}.$$

Ağırlık fonksiyonu $w(x_i) = K[(x_i - x)/\lambda]$'dır. Düzleştirme seviyesi düzleştirme parametresi λ tarafından kontrol edilir, $\lambda = 0$ ise $\hat{f}(x_i) = y_i$ ile sonuçlanır ve $\lambda = \infty$ ise lineer bir modele eşdeğerdir. $\hat{f}(x)$ sadece yerel polinomal modelin katsayısı $\hat{\beta}_{0x}$'e bağlıdır. Bu durum yerel sabit bir modelle karıştırılmamalıdır. x noktasında f

fonksiyonunun tahmini için, yerel regresyon, modele ağırlıklı en küçük kareler uygulanmasıyla eşdeğerdir. $E[\varepsilon_x] = 0$ ve $var(\varepsilon_x) = \sigma^2 W_x^{-1}$ ile,

$$Y = x_x \beta_x + \varepsilon_x$$

$$x_x = \begin{bmatrix} 1 & x_1 - x & \cdots & \frac{(x_1 - x)^p}{p!} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_n - x & \cdots & \frac{(x_n - x)^p}{p!} \end{bmatrix}$$

x_x, $n \times (p+1)$ tasarım matrisini ve W_x, $w_i(x)$ $i = 1, \ldots, n$ öğeleriyle $n \times n$ köşegen matrisini temsil eder. w_i'nin büyük değerleri $x - x_i$'nin küçük olmasına karşılık gelir, böylece x'e yakın olan x_i veri noktaları en etkilidir denir. W_x, kernel fonksiyonu K(·)'ya ve dolayısıyla bant genişliğine bağlı olduğuna dikkat edilmelidir.

$(Y - x_x\beta_x)^T W_x (Y - x_x\beta_x)$'in minimize edilmesiyle $\hat{\beta}_x$ aşağıdaki şekilde tahmin edilir:

$$\hat{\beta}_x = (x_x^T W_x x_x)^{-1} x_x^T W_x Y. \tag{3.29}$$

$(x_x^T W_x x_x)^{-1} x_x^T W_x$'nin ilk satırının Y ile iç çarpımını almak,$\hat{f}(x) = \hat{\beta}_{0x}$'i verir. (3.29)'daki tahmin edicinin verilerde doğrusal olduğu açıkça gözlemlenebilir.

$$\hat{f}(x) = \sum_{i=1}^{n} S_i^{(\lambda)}(x) Y_i.$$

Tahmin edicinin ortalaması,

$$E\left[\hat{f}(x)\right] = \sum_{i=1}^{n} S_i^{(\lambda)}(x) f(x_i)$$

Varyansı,

$$var\left[\hat{f}(x)\right] = \sigma^2 \sum_{i=1}^{n} S_i^{(\lambda)}(x)^2 = \sigma^2 \left\| S^{(\lambda)}(x) \right\|^2,$$

burada yine hata varyansının sabit ve gözlemlerin ilişkisiz olduğu varsayılmıştır. Etkin serbestlik derecesi $p^{(\lambda)} = iz(S^{(\lambda)})$ olarak tanımlanabilir. $S^{(\lambda)}$, $\hat{Y} = S^{(\lambda)} Y$'den elde

edilen şapka matrisidir. Genellikle $p = 1$, $f(\cdot)$'yi tahmin etmek için yeterli olacaktır. Bir doğrusal yerel polinomla aşağıda verilen sonuç elde edilir:

$$\hat{f}(x) = \frac{\sum_{i=1}^{n} w_i(x) Y_i}{\sum_{i=1}^{n} w_i(x)} + (x - \bar{x}_w) \frac{\sum_{i=1}^{n} w_i(x)(x - \bar{x}_w) Y_i}{\sum_{i=1}^{n} w_i(x)\,(x - \bar{x}_w)^2},$$

Formülde $\bar{x}_w = \sum_{i=1}^{n} w_i(x) x_i / \sum_{i=1}^{n} w_i(x)$ ve $w_i(x) = K((x - x_i)/\lambda)$. Sonuç olarak, yerel sabit (Nadaraya–Watson) tahmincisine ek olarak x_i'nin yerel eğimini ve çarpıklığını düzelten bir terim olduğu söylenebilir. Doğrusal yerel polinom model için beklenen değer ise şu şekilde tanımlanır:

$$E\left[\hat{f}^{(\lambda_n)}(x)\right] \approx f(x) + \frac{1}{2}\lambda_n^2 f''(x)\sigma_K^2$$

Varyansı,

$$var\left[\hat{f}^{(\lambda_n)}(x)\right] \approx \frac{1}{n\lambda_n} K_2 \sigma^2 \frac{1}{p(x)}.$$

Eğer, f, x'de doğrusal ise $\hat{f}$ kesinlikle yansızdır. Yerel doğrusal polinom tahmincisi,

$$\begin{aligned} BHKO\left(\hat{f}^{(\lambda_n)}\right) &= bias\left[\hat{f}^{(\lambda_n)}\right]^2 + var\left[\hat{f}^{(\lambda_n)}\right] \\ &\approx \frac{\lambda_n^4 \sigma_K^4}{4}\left[\int f''(x^2)dx\right] + \frac{K_2\sigma^2}{n\lambda_n}\int \frac{1}{p(x)}dx. \end{aligned}$$

Kernel regresyondaki bütünleşik hata kareler ortalaması ile karşılaştırıldığında, tasarım yanlılığı sıfırdır ve bu durum doğrusal polinomun Nadaraya–Watson tahmin edicisine göre avantajını gösterir. Optimal λ bu nedenle,

$$\lambda_n^* = \left(\frac{1}{n}\right)^{1/5} \left(\frac{\sigma^2 K_2 \int p(x)^{-1} dx}{\sigma_K^4 \int f''(x)^2 dx}\right)^{1/5} \qquad (3.30)$$

(3.30) ifadesindeki terimlerin her birinin, λ_n'nin bir "eklenti" tahmin edicisini vereceği tahmin edilebilir veya çapraz geçerlilik kullanılabilir (Wakefield Jon 2013).

3.3 Nonparametrik Regresyonda Düzleştirme Yöntemleri

Nonparametrik regresyon modelinde düzleştirme kavramı büyük önem taşımaktadır. Düzleştirme; bir ya da birden fazla bağımsız değişkenin fonksiyonu olan bağımlı değişkeninin sahip olduğu trendi ifade etmek için kullanılır. Düzleştiriciler değişkenler arasındaki ilişkinin biçimini kesin bir şekilde varsaymadığı için parametrik olmayan regresyonda sık kullanılan bir araçtır. Düzleştirme kavramının temeli ortalama alma fikrine dayanır. Bilinen en basit düzleştirici ise hareketli ortalamalardır. Bir düzleştirici tarafından tahmin edilen eğriye "düzgün" adı verilir. Tek bir bağımsız değişkenin olduğu durumlar ise "serpilme diyagramı düzleştirmesi" olarak adlandırılmaktadır. Serpilme diyagramındaki noktalar, herhangi bir olasılıksal model olmaksızın düzlem üzerindeki noktaların birikimi olarak kabul edilmektedir (Ruppert vd. 2004). Basit Hareketli ortalamalardan sonra en çok bilinen yöntem üstel düzleştirmedir. Basit üstel ve ikili üstel düzleştirme yöntemleri kernel regresyonun çeşitleridir. Bu bölümde, parametrik regresyon modellerinin kısıtlamalarına nasıl çözüm getirilebileceği anlatılacaktır.

3.3.1 Cezalandırılmış spline regresyonu

Spline yaklaşımında kestirimin pürüzlülüğü çok sayıda düğüm noktası olmasından kaynaklanır. Bu sorunun üstesinden gelmenin bir yolu, tüm düğümleri korumak, ancak etkilerini sınırlamaktır. Bu şekilde daha az değişkenle sonuçlanması beklenir. K'nin büyük olduğu K düğümlü genel bir spline modeli olsun. Sıradan en küçük kareler modeli şu şekilde yazılır,

$$\hat{y} = X\hat{\beta}, \qquad \hat{\beta}, \|y - X\beta\|^2 \text{ ifadesini minimize eder.}$$

Eşitlikte $\beta = [\beta_0, \beta_1, \beta_{11}, \dots, \beta_{1K}]^T$ olup β_{1K}, k. düğüme ilişkin parametreyi gösterir. β_{1K}'nın sınırlandırılmamış bir tahmini fazla kıvrımlı bir tahmine yol açar. Bu durumu düzeltebilecek β_{1K} üzerindeki kısıtlamalar şunlardır:

1) $max|\beta_{1K}| < C$
2) $\sum_{i=1}^{m}|\beta_{1K}| < C$

3) $\sum_{i=1}^{m} \beta_{1K}^2 < C$

C uygun bir şekilde seçildiğinde, kısıtların her birinden daha düz bir kestirim elde edilecektir. Diğer iki kısıta göre uygulanması en kolay kısıt üçüncü kısıttır. (K+2)×(K+2) boyutlu bir D matrisi aşağıdaki şekilde tanımlanırsa,

$$D = \begin{bmatrix} 0_{2x2} & 0_{2xK} \\ 0_{Kx2} & I_{KxK} \end{bmatrix}$$

minimizasyon problemi, $\beta^T D\beta \leq C$ kısıtı altında $\|y - X\beta\|^2$ olan hata kareler toplamının minimum yapılması problemine dönüşür. Bu fonksiyonla, Langrange çarpanları yöntemi kullanılarak yazılan aşağıdaki fonksiyonu minimize etmek eşdeğerdir. (Ruppert vd. 2004).

$$\|y - X\beta\|^2 + \lambda^2 \beta^T D\beta \quad \lambda \geq 0.$$

Bu durumda $\hat{\beta}_\lambda$ tahmini,

$$\hat{\beta}_\lambda = (X^T X + \lambda^2 D)^{-1} X^T y \quad \text{bulunur.}$$

$\lambda^2 \beta^T D\beta$ terimine pürüzlülük cezası denir, çünkü çok sert olan kıvrımları cezalandırır ve böylece daha düzgün bir sonuç verir. Düzleştirme miktarı, düzleştirme parametresi λ tarafından kontrol edilir. Cezalandırılmış bir spline regresyonu için uygun değerler daha sonra şu şekilde verilir:

$$\hat{y} = X(X^T X + \lambda^2 D)^{-1} X^T y. \qquad (3.31)$$

Cezalandırılmış tahmin, spline temel fonksiyonlarının tüm katsayılarını sıfıra doğru küçültür ve bazı katsayıları sıfıra kadar küçülten ve kalan katsayıları küçültmeden bırakan düğüm seçimi ile zıtlık oluşturabilir. (3.23)'deki eşitlik bir Ridge regresyonudur. (Ruppert vd. 2004).

3.3.2 Doğal kübik splinelar

Kübik spline modelinin yaygın olarak kullanılan hali doğal kübik spline temelidir. Doğal kübik splinelar kuyruklarında sınır düğümlerinin ötesinde doğrusal olma

kısıtlaması olan kübik spline'lardır. Doğrusallık, spline f'nin sınır düğümlerinde $f'' = f''' = 0$ sağladığı kısıtlamalar aracılığıyla sağlanmaya çalışılır. Kübik spline, $\hat{f}$, artık kareler toplamını ve ikinci dereceden türevinin karesi olan $(\hat{f}'')^2$ integrali üzerindeki cezayı minimize eder.

Bu durumda ortaya çıkan hata kareler toplamı HKT:

$$HKT = \sum_{i=1}^{n}\{y_i - \hat{f}(x_i)\}^2 + \lambda^3 \int \{\hat{f}^{(2)}(x)\}^2 dx \qquad (3.32)$$

burada $\lambda > 0$ düzleştirme parametresidir. $(\hat{f})$'nın bir spline olması için önceden herhangi bir kısıtlama getirilmemesine rağmen, bu cezalı kareler toplamının minimize edicisine, x_i'de (sınır ve iç) düğümleri olan doğal bir kübik spline'dır denir. Düzleştirici splinelar, parametrik olmayan uygulamalarda yaygın olarak kullanılmaktadır.

Doğal kübik splinelar "doğal" olarak adlandırılır çünkü bir optimizasyon probleminin çözümü olarak ortaya çıkarlar. Herhangi bir uygulamada sınır kısıtlamaları varsa spline doğal değildir. Dolayısıyla, bu kısıtlamalar olmadan doğal kübik spline modelini, kübik spline modele tercih etmek mantıklı bir seçim olmayacaktır. Cezalı splinelar düzleştirme splinelarından daha geneldir. Çünkü doğal kübik splinelara sınır kısıtlamalarının uygulanıp uygulanamama durumları söz konusudur. Cezalı splinelarda ise istenen sayıda ya da daha az sayıda düğüm kullanılabilir bu da cezalı splineları daha genel yapmaktadır. Bir doğal kübik spline her bir x_i düğüm noktasındaki değeri ve ikinci mertebeden türevinin verilmesi ile belirlenebilir. Bu gösterim ikinci türev-değer gösterimi olarak adlandırılır. f fonksiyonunun $x_1 < \cdots < x_n$ düğümlerinde doğal kübik spline fonksiyonu olduğu varsayılsın ve

$$f_i = f(x_i) \text{ ve } \gamma_i = f''(x_i), \quad i = 1, \dots, n$$

olarak tanımlansın. Doğal kübik spline tanımına göre, x_1 ve x_n noktalarında f fonksiyonun ikinci türevi sıfır olduğu için $\gamma_1 = \gamma_n = 0$ dır. $f = (f_1, \dots, f_n)'$ ve $\gamma = (\gamma_2, \dots, \gamma_{n-1})'$ olsun. f ve γ vektörleriyle f eğrisinin tam olarak belirlenmesi sağlanır ve f eğrisi, herhangi bir x noktasında f'nin değeri ve türevleri için, f ve γ vektörleri ile açık olarak formüle edilebilir. Fakat her f ve γ vektörü doğal kübik eğrisel çizgi belirtmez. f ve γ vektörlerinin kübik eğrisel çizgi belirtmesi için Q ve R

gibi iki bant matrisi tanımlanmalıdır. Bir matrisin sıfır olmayan elemanlarının hepsi az sayıda köşegen elemanları üzerinde toplanmışsa, bu matrise bant matrisi denir. Matrisin sıfır olmayan köşegen elemanlarının sayısı da matrisin bant genişliği olarak adlandırılır. $k_i = x_{i+1} - x_i$ olmak üzere Q ve R matrislerinin elemanları aşağıdaki gibi tanımlanır.

$$q_{ij} = \begin{cases} \frac{1}{k_{j-1}}, & i = j-1 \\ -\left(\frac{1}{k_{j-1}} + \frac{1}{k_j}\right), & i = j \\ 0, & |i-j| \geq 2 \\ \frac{1}{k_j}, & i = j+1 \end{cases} \qquad i = 1,2,\dots,n \; ve \; j = 2,\dots,n-1$$

$$r_{ij} = \begin{cases} \frac{1}{6}k_{j-1} & i = j-1 \\ \frac{1}{3}(k_{j-1} + k_j), & i = j \\ 0, & |i-j| \geq 2 \\ \frac{1}{6}k_j, & i = j+1 \end{cases} \qquad i = 1,2,\dots,n-1 \; ve \; j = 2,\dots,n-1$$

Bu eşitliklere göre R ve Q matrisleri aşağıdaki gibi gösterilebilir: R, (n-2)×(n-2) boyutlu simetrik bir matris, Q, n×(n-2) boyutlu matrisdir:

$$R: \begin{bmatrix} \frac{1}{3}(k_1+k_2) & \frac{1}{6}k_2 & 0 & . & . & . & 0 \\ \frac{1}{6}k_2 & \frac{1}{3}(k_2+k_3) & \frac{1}{6}k_3 & 0 & . & . & . \\ 0 & \frac{1}{6}k_3 & \frac{1}{3}(k_3+k_4) & \frac{1}{6}k_4 & 0 & . & . \\ . & . & . & . & . & . & . \\ . & . & . & . & . & . & 0 \\ . & . & . & 0 & \frac{1}{6}k_{n-3} & \frac{1}{3}(k_{n-3}+k_{n-2}) & \frac{1}{6}k_{n-2} \\ . & . & . & . & 0 & \frac{1}{6}k_{n-2} & \frac{1}{3}(k_{n-2}+k_{n-1}) \end{bmatrix}$$

$$Q: \begin{bmatrix} \frac{1}{k_1} & -\left(\frac{1}{k_1}+\frac{1}{k_2}\right) & \frac{1}{k_2} & 0 & 0 & . & . & . & 0 \\ 0 & \frac{1}{k_2} & -\left(\frac{1}{k_2}+\frac{1}{k_3}\right) & \frac{1}{k_3} & 0 & . & . & . & . \\ . & . & . & . & . & . & . & . & . \\ . & . & . & . & . & . & . & . & . \\ . & . & . & . & . & . & . & . & . \\ . & . & . & . & 0 & \frac{1}{k_{n-3}} & -\left(\frac{1}{k_{n-3}}+\frac{1}{k_{n-2}}\right) & \frac{1}{k_{n-2}} & 0 \\ 0 & . & . & . & 0 & 0 & \frac{1}{k_{n-2}} & -\left(\frac{1}{k_{n-2}}+\frac{1}{k_{n-1}}\right) & \frac{1}{k_{n-1}} \end{bmatrix}$$

R ve Q indisleri $|i-j| \geq 2$ koşulunu sağlayan elemanları sıfır olan matrislerdir bu matrislere üç köşegen matrisler denir. R matrisi her bir i elemanı için $|r_{ij}| > \sum_{i \neq j} |r_{ij}|$ ise köşegen dominanttır. Bu özellik R matrisinin pozitif tanımlı dolayısıyla R matrisinin tersi olduğunu gösterir ve

$$K = QR^{-1}Q^T$$

olacak şekilde bir K matrisi tanımlanabilir. Bu eşitlik sonucu iki önemli teorem tanımlanır.

Teorem 1: f ve γ vektörleri, yalnız ve yalnız aşağıdaki koşullar sağlanırsa doğal kübik eğrisel çizgi belirtir. $Q^T f = R\gamma$ bu eşitlik sağlanırsa,

$$\int_a^b f''(x)^2 dx = \gamma^T R\gamma = m^T Km$$

eşitliği elde edilir. (Green ve Silverman 2000, Rupert vd. 2003).

3.3.3 Diğer cezalar

Bir cezayı anlamanın en iyi yollardan biri limiti belirlemektir. Ceza parametresi λ sonsuza yakınsadıkça, düzgün uyum bir limite doğru yakınsar. Kesilmiş güç fonksiyonu temeli ile p. derece spline kullanırsak, bu durumda ceza;

$$\lambda^{2p} \sum_{k=1}^{K} \beta_{pk}^2, \tag{3.33}$$

Bu uyum, p dereceli bir polinoma uyan en küçük karelere doğru küçültülür. Böylece, spline derecesi ve küçülme hedefi birbirine bağlanır. (3.33)'de verilen ceza ile düz bir çizgiye doğru küçülmek için doğrusal spline'lar kullanılmalıdır. Bazı durumlarda, spline modelini ve küçültme limitini bağımsız olarak seçmede esneklik istenebilir. Bu esnekliği sağlamak için başka cezalar getirilebilir. Örneğin, eşit aralıklı düğümlere sahip B-spline'lar kullanılabilir ve B-spline katsayılarında (q+1). dereceden farkla cezalandırılır (Eilers ve Marx 1996, Rupert vd. 2003). Bu nedenle, örneğin q = 1 kullanıldığında, spline'ın derecesinden bağımsız olarak küçülme düz bir çizgiye doğrudur; oysa q = 2 için küçülme bir parabole doğrudur. Eilers ve Marx cezası, eşit

aralıklı düğümler uygun olduğunda kesinlikle yararlıdır (Rupert, 2008). Uygulamada ise düğümler eşit olmayan aralıklarda genellikle düzensiz bulunur. Geniş bir sınırlar aralığında küçülmeye izin veren ve herhangi bir düğüm aralığı için uygun olan (3.33)'e bir alternatif, formun ikinci dereceden integral cezasıdır;

$$\lambda^{2p} \int_a^b \left\{f^{(q+1)}(x)\right\}^2 dx, \qquad (3.34)$$

burada q+1 ≤ p'dir ve a ve b en küçük değerleri iken, x en büyük değeridir. (3.34)'deki ceza bir q. derece polinomuna doğru küçülür. p = 3 ve q = 1 kullanılması, kübik düzleştirme eğrisine çok benzer bir uyum sağlar. Aslında, her x'te düğümlü doğal kübik spline kullanmak, tam olarak kübik bir düzleştirme spline'ı verir.

$B(x) = [B_1(x), \dots, B_N(x)]^T$ spline tabanlı fonksiyonların bir vektörü olsun. Burada X'in i. satırı $B_{(xi)}^T$'dir ve,

$$\hat{f}(x) = \hat{\beta}^T B(x).$$

$\lambda^{2p} \int_a^b \left\{f^{(q+1)}(x)\right\}^2 dx$, bu eşitlik $\beta^T DB$'ye eşit olur. D, B'nin ikinci moment matrisidir:

$$D = \int_a^b B^{(q+1)}(x) \left\{B^{(q+1)}(x)\right\}^T dx.$$

$\hat{y} = X(X^T X + \lambda^{2p} D)^{-1} X^T y$ formülüyle tahmin edilmektedir (Ruppert vd. 2004).

3.3.4 Doğrusal düzleştiriciler

Cezalandırılmış spline uygulamaları, sıradan en küçük kareler uygulamalarından temelde farklı olsa da bazı ortak noktaları vardır. Özellikle her ikiside veri vektörü y'nin doğrusal fonksiyonlarıdır. Doğrusal bir regresyon modeli için:

$$y = X\beta + \varepsilon,$$

Sıradan en küçük kareler:

$$\hat{y} = Hy, \qquad \text{burada } H = X(X^T X)^{-1} X^T \qquad (3.35)$$

olmak üzere şapka matrisidir. Cezalandırılmış spline modeli bu formülü şu şekilde genelleştirir:

$$\hat{y} = S_\lambda y, \qquad S_\lambda = X(X^T X + \lambda^{2p} D)^{-1} X^T \tag{3.36}$$

X, p dereceli kesilmiş polinom tabanına karşılık gelir. Parametrik olmayan regresyonda şapka matrisi terimi kullanılsa da S_λ genellikle daha düzgün matris ya da düzleştirici matris olarak adlandırılır. (3.35) ve (3.36)'deki formüllerinin her ikisi de aşağıdaki şekilde yazılabilir.

$$\hat{y} = Ly$$

L, y'ye bağlı olmayan $n \times n$ boyutlu bir matristir. Buna doğrusal düzleştiriciler sınıfı denilmektedir. Çoğu zaman, L bir düzleştirme parametresiyle y'ye bağlıdır. Bu durumda, düzleştirici gerçekten doğrusal değildir yaklaşık olarak doğrusal olduğu kabul edilerek uygulamasının yapılması daha yaygındır (Ruppert vd. 2004).

3.4 Düzleştirme Hatası

Dağılım grafiği düzleştirme modelinde $\hat{f}$, f'nin bir tahminci edicisi olsun,

$$y_i = f(x_i) + \varepsilon_i,$$

$X \subseteq \mathbb{R}$, $f(x)$ tahmininin ilgilenildiği bir x değerleri kümesi olsun. Her $x \in X$ için, $\hat{f}(x)$, $f(x)$'in tahminidir. İstatistiksel tahmin teorisinin temel taşlarından biri tahminci tarafından yapılan hatadır. En yaygın hata ölçüsü, hata kareler ortalamasıdır (HKO). Hata kareler ortalaması, bir regresyon eğrisinin bir dizi noktaya ne kadar yakın olduğunu söyler ve her zaman pozitif değerlidir. Eğer HKO değeri sıfıra yakınsa tahmin edicinin iyi bir performans gösterdiği söylenebilir. Hata kareler ortalaması;

$$HKO\{\hat{f}(x)\} \equiv E\left[\{\hat{f}(x) - f(x)\}^2\right].$$

HKO aşağıdaki şekilde de yazılabilir:

$$HKO\{\hat{f}(x)\} = \left[E\{\hat{f}(x) - f(x)\}\right]^2 + var\{\hat{f}(x)\},$$

Bu karesel yanlılığı ve genel hataya varyans katkısını gösterir. Hatayı daha genel olarak X değerleri arasında ölçmek daha yaygın bir uygulamadır. Bunun için önerilen bir olasılık, bütünleşik hata kareler ortalamasıdır (BHKO).

$$BHKO\{\hat{f}(\cdot)\} \equiv \int HKO\{\hat{f}(x)\}dx.$$

X'e bağlı olmayan daha kolay bir alternatif ortalama toplam kare hatasıdır (OTKH).

$$OTKH\{\hat{f}(\cdot)\} \equiv E\sum_{i=1}^{n}\{\hat{f}(x_i) - f(x_i)\}^2,$$

Burada sadece gözlemlerdeki hata dikkate alınır. Ortalama toplam kare hatasının bir avantajı, daha basit matris cebirsel işlemlerle yapılıyor olmasıdır. $\hat{f} = [\hat{f}(x_1), \dots, \hat{f}(x_n)]^T$, y ortalama uygun değerlerinin bir vektörü olsun. f'de, $f(x_i)$ değerlerinin vektörünü benzer şekilde tanımlasın.

$$OTKH(\hat{f}) = E\|\hat{f} - f\|^2.$$

Doğrusal düzleştiricilerde tanımlanan $\hat{y}$=Ly yerine, $\hat{f}$ = Ly şeklinde yazılır. Ortalama toplam kare hatası:

$$\begin{aligned} OTKH(\hat{f}) &= \sum_{i=1}^{n}\left\{E\left(\hat{f}(x_i) - f(x_i)\right)\right\}^2 + var\{\hat{f}(x_i)\} \\ &= \sum_{i=1}^{n}\{E(Ly)_i - f_i\}^2 + var\{(Ly)_i\} \\ &= \sum_{i=1}^{n}\{E(Ly)_i - f_i\}^2 + \{Cov(Ly)\}_{ii} \\ &= \|(L-I)f\|^2 + tr\{Cov(Ly)\} \\ &= \|(L-I)f\|^2 + tr\{LCov(y)L^T\}. \end{aligned}$$

$Cov(y) = \sigma_\varepsilon^2 I$ ile sabit varyans varsayımı altında,

$$OTKH(\hat{f}) = \|(L-I)f\|^2 + \sigma_\varepsilon^2 tr(LL^T) \qquad (3.37)$$

(3.37)'de yazılan formülün ilk kısmı kare yanlılık katkısını temsil ederken, ikinci kısım varyansların toplamına karşılık gelir. Düzleştirmede, yanlılık ve varyans ikilemi birbirine bağlı olarak değişir. $L = S_\lambda$ olan cezalandırılmış spline düzleştiriciler için, daha büyük λ değerleri yanlılıkta bir artışa ancak varyansta bir azalmaya yol açar. Daha küçük λ değerleri içinse bu durumun tersi bir durum oluşmaktadır (Ruppert vd. 2004).

3.5 Düzleştirme Rankı

Klasik düzleştirme spline'ları n temel fonksiyon kullanırken, doğrusal cezalı spline'lar veya K düğümlü kübik radyal düzleştiriciler K + 2 temel fonksiyonları kullanır. "Azaltılmış düğümlü" düzleştiriciler "tam düğümlü" düzleştiricilerin bileşenlerini pürüzsüzlük için atma eğilimindedir. L'nin n × n boyutlu düzleştirici matrisiyle genel doğrusal bir düzleştirici olduğu varsayılsın. Burada L, simetrik ve yarı pozitif tanımlıdır. L matrisinin $v_1, \dots, v_n$ özvektörlerine karşılık gelen özdeğerleri $\lambda_1 \geq \cdots \geq \lambda_n$ şeklinde olsun. Tam düğüm düzleştirme eğrisine karşılık gelen özvektörlerin tahmin değerlerine karşı çizilen grafikleri daha kıvrımlıdır. Yani özdeğer değeri büyüdükçe grafik düzgünleşir ya da tam tersi özdeğer değeri küçüldükçe grafiğin kıvrımlılığı artar. Özdeğerler ve özvektörler arasındaki ilişki şu şekilde tanımlanır:

$$L_{vi} = \lambda_i v_i, \quad i = 1, \dots, n. \tag{3.38}$$

Özvektörler $\mathbb{R}^n$'de tanımlı olmak üzere, y yanıt değişkeni vektörü aşağıdaki şekilde yazılabilir:

$$y = \sum_{i=1}^{n} \alpha_i v_i.$$

(3.38)'deki eşitliği kullanarak, $\hat{y}$ değerleri aşağıdaki şekilde hesaplanır:

$$\hat{y} = \sum_{i=1}^{n} \alpha_i \lambda_i v_i.$$

Matris cebir sonucunun kullanılmasıyla,
Bir L matrisinin rankı ≡ L'nin bağımsız doğrusal sütun sayısı

= L'nin sıfır olmayan özdeğerlerinin sayısı şeklinde ifade edilir.

Hesaplama karmaşıklığının üstesinden gelmek için düzleştirici matrisinin rankı kullanılabilir. n temel fonksiyonu daha az kullanan düzleştiriciler, düşük ranklı olarak adlandırılırken, temel fonksiyonları örneklem boyutuyla yaklaşık olarak aynı olanlara tam ranklı denir. Örneklem boyutu büyük olduğunda düşük ranklı düzleştiricilerin kullanılmasının pek bir faydası olmadığı söylenebilir. Ancak daha büyük örnek boyutları ve birkaç düzleştirme içeren daha karmaşık modeller için, düşük ranklı düzleştiriciler hesaplamalarda kolaylık sağlayabilir (Ruppert vd. 2004).

3.6 Düzleştiricinin Serbestlik Derecesi

Düzleştirme parametresi λ'nın küçük bir değeriyle verilerin tahminine biraz yakın kıvrımlı bir dağılım grafiği elde edilirken, çok büyük bir λ değeri için spline temel fonksiyonların derecesine bağlı olarak parametrik bir uyum elde edilir. Buradaki sorun, λ'nın ne küçük ne de büyük olan ara değerleriyle elde edilen düzleştirme ve λ arasındaki ilişkinin net olmamasıdır yani düzleştirme parametresinin değeri, grafiğin şeklinin belirlenmesinde doğrudan etkili değildir. Bu sorunu çözmenin uygun bir yolu, λ'nın bir $t(\lambda)$ dönüşümü yani, $t(\lambda) = iz(S_\lambda)$'ya λ düzleştirme parametresine karşılık gelen uyumun serbestlik derecesi olarak adlandırılır. Parametrik regresyondaki şapka matrisi H'nin izi, iz(H) = uyum parametre sayısı = serbestlik derecesi şeklindedir. Dolayısıyla $iz(S_\lambda)$, bu tanımın saçılım grafiği düzleştiricilerine genelleştirilmiş halidir. Eşdeğer parametre sayılarına bağlı olarak genel bir yorum yapılırsa polinom uyumu ile ilişkilendirilebilir yani ν serbestlik derecesine sahip düz bir dağılım grafiği ile $(\nu - 1)$ dereceli bir polinom verileri yaklaşık olarak aynı ölçüde özetler. Kısaltma olarak $sd_{fit} \equiv iz(S_\lambda)$ gösterimi kullanılacaktır. Bu durumda $iz(AB) = iz(BA)$ matris özelliğine dayanarak cezalandırılmış spline için:

$$sd_{kestirim} = iz\{(X^T X + \lambda^2 D)^{-1} X^T X\}.$$

K düğüm ve p derecesi ile cezalandırılmış bir spline dağılım grafiği için,
$tr(S_0) = p + 1 + K$ olduğu söylenir. Diğer uçta, $tr(S_\lambda) \rightarrow p + 1$, λ→∞ iken yani λ'nın pozitif değerleri, $p + 1 < sd_{fit} < p + 1 + K$'ye karşılık gelir. Sonuç olarak

özetlenirse, λ ve sd_{fit} arasında monoton olarak azalan bir ilişki olduğu söylenir. Böylece, her $\lambda > 0$ değerine karşılık gelen benzersiz bir sd_{fit} vardır. λ değerleri için sd_{fit} değerleri birbirine yakınsarsa bu benzer görünümlerini açıklar, kabaca aynı sayıda eşdeğer parametreye sahip oldukları söylenir (Ruppert vd. 2004).

3.7 Düzleştirme Parametresi Seçimi

Nonparametrik regresyon modelinde düzleştirme kavramı büyük önem taşımaktadır. Düzleştirme; bir ya da birden fazla bağımsız değişkenin fonksiyonu olan bağımlı değişkeninin sahip olduğu trendi ifade etmek için kullanılır. Düzleştiriciler değişkenler arasındaki ilişkinin biçimini kesin bir şekilde varsaymadığı için parametrik olmayan regresyonda sık kullanılan bir araçtır. Düzleştirme parametresini seçmek de bu yüzden önemli bir konudur. "Çapraz geçerlilik", "Mallows'un C_p ölçütü" ve "Akaike bilgi kriteri" gibi yöntemler parametre seçiminde kullanılmaktadır.

3.7.1 Çapraz geçerlilik (Cross validation)

Çapraz geçerlilik ölçütü nonparametrik regresyonda parametrik regresyona benzer olarak açıklanabilir. Parametrik regresyonda çapraz geçerlilik aşağıdaki şekilde ifade edilir.

$$CV = \sum_{i=1}^{n} \left(y_i - \hat{y}_{i,-i}\right)^2 = \sum_{i=1}^{n} \left(\frac{(y_i - \hat{y}_i)}{1 - h_{ii}}\right)^2$$

h_{ii}, H şapka matrisinini köşegen elemanlarını göstermektedir. Nonparametrik regresyonda da benzer olarak,

$$CV(\lambda) = \sum_{i=1}^{n} \left(y_i - \hat{f}_{-i}(x_i; \lambda)\right)^2$$

şeklinde yazılır. $\hat{f}(x_i; \lambda)$, x_i noktasında λ düzleştirme parametresiyle nonparametrik regresyon tahmini göstermek üzere nonparametrik regresyonda kestirim değerleri aşağıdaki gibidir:

$$S_\lambda y = \begin{pmatrix} \hat{f}(x_1;\lambda) \\ \vdots \\ \hat{f}(x_n;\lambda) \end{pmatrix}$$

$\hat{f}(x_i;\lambda) = \sum_{j=1}^{n} S_{\lambda,\mathrm{ij}} y_j$, $S_{\lambda,\mathrm{ij}}$, S_λ'nın (i,j). elemanıdır. Düzleştiricilerin birçoğu aşağıdaki şekilde ifade edilebilir. (Ruppert vd. 2003, Türkan 2009).

$$\hat{f}_{-i}(x_i;\lambda) = \frac{\sum_{\mathrm{i}\neq\mathrm{j}} S_{\lambda,\mathrm{ij}} y_j}{S_{\lambda,\mathrm{ij}}}$$

Veri kümesinden (x_i, y_i) gözlem değeri çıkarıldıktan sonra elde edilen tahmin değeridir ve genellikle bu eşitliğin yaklaşık olarak sağlandığı bilinmektedir. Çapraz geçerlilik ölçütü de bu gösterim ile bulunabilir. Düzleştiriciler her zaman $y_i \equiv 1$ ise $\hat{y}_i \equiv 1$ özelliğine sahiptir ve bu özellik tüm düzleştiriciler için,

$$\sum_{i=1}^{n} S_{\lambda,\mathrm{ij}} = 1, \quad \text{tüm i'ler için}$$

olduğunu gösterir. Buna göre $\sum_{i=1}^{n} S_{\lambda,\mathrm{ij}} = 1 - S_{\lambda,\mathrm{ii}}$'dir ve nonparametrik regresyonda çapraz geçerlilik ölçütü aşağıdaki gibi bulunur (Türkan, 2009).

$$\begin{aligned} CV(\lambda) &= \sum_{i=1}^{n} \left(y_i - \hat{f}_{-i}(x_i;\lambda)\right)^2 \\ &= \sum_{i=1}^{n} \left(y_i - \frac{\sum_{\mathrm{i}\neq\mathrm{j}} S_{\lambda,\mathrm{ij}} y_j}{1 - S_{\lambda,\mathrm{ii}}}\right)^2 \\ &= \sum_{i=1}^{n} \left(\frac{y_i - \hat{f}(x_i;\lambda)}{1 - S_{\lambda,\mathrm{ii}}}\right)^2 \\ &= \sum_{i=1}^{n} \left(\frac{\{(\mathrm{I} - S_\lambda)y\}_i}{1 - S_{\lambda,\mathrm{ii}}}\right)^2 \\ &= \sum_{i=1}^{n} \left(\frac{y_i - \hat{y}_i}{1 - S_{\lambda,\mathrm{ii}}}\right)^2 \end{aligned}$$

Çapraz geçerlilik ölçütüne göre en uygun düzleştirme parametresi (λ), $CV(\lambda)$ fonksiyonunu minumum yapan değerdir (Rupert vd. 2003, Türkan 2009). Genelleştirilmiş çapraz geçerlilik ölçütünde ise, $S_{\lambda,\mathrm{ii}}$'nin yerine $S_{\lambda,\mathrm{ii}}$'nin ortalaması kullanılır. Buna göre,

$$GCV(\lambda) = \sum_{i=1}^{n} \left\{ \frac{y_i - \hat{f}(x_i;\lambda)}{1 - iz(S_\lambda)/n} \right\}^2$$

$$= \sum_{i=1}^{n} \left\{ \frac{y_i - \hat{f}(x_i;\lambda)}{1 - iz(S_\lambda)/n} \right\}^2$$

$$= \frac{\|(\mathrm{I} - \mathrm{S}_\lambda)\|^2}{\{1 - n^{-1} iz(S_\lambda)\}^2}$$

şeklinde ifade edilir. Genelleştirilmiş çapraz geçerlilik ölçütüne göre en uygun düzleştirme parametresi λ, $GCV(\lambda)$ fonksiyonunu minumum yapan değerdir (Türkan, 2009).

3.7.2 Mallows'un C_p ölçütü

Mallows C_p ölçütü Mallows tarafından önerilen, Craven ve Wahba tarafından spline düzeltmeye uyarlanan bir ölçüttür (Rupert vd. 2003). Spline düzeltmede yansız risk yöntemi olarak bilinir. Çeşitli formlarda yazılabilir, bunlardan biri aşağıdaki gibidir:

$$C_p = HKT(p) = 2\hat{\sigma}_\varepsilon^2 p \qquad (3.39)$$

p modeldeki parametre sayısıdır, HKT(p) aynı model için hata kareler toplamıdır,

$$\hat{\sigma}_\varepsilon^2 = \frac{1}{n - p_{enb}} HKT_{(p_{enb})}$$

Burada $\hat{\sigma}_\varepsilon^2$, ilgilenilen en geniş modele ait hata varyansının tahminidir. Amaç, en küçük C_p değerine sahip modeli seçmektir. Parametrik olmayan regresyon fikrini kullanarak C_p değeri elde edilebilir. Aşağıdaki model ele alınsın:

$$y = f + \varepsilon, \quad Cov(\varepsilon) = \sigma_\varepsilon^2 I,$$

f'nin tahmin edicisi, doğrusal bir düzleştiricidir ve $\hat{f}$ ile gösterilir.

$$\hat{f} = Ly.$$

$\hat{f}$'nın ortalama hata kareler toplamı:

$$OHKT(\hat{f}) = \|(L - I)f\|^2 + \sigma_\varepsilon^2 iz(LL^T).$$

Hata kareler toplamının beklenen değeri,

$$E(HKT) = OHKT(\hat{f}) + \sigma_\varepsilon^2(n - 2sd_{kestirim})$$

$sd_{kestirim} = iz(L)$, serbestlik derecesini gösterir. $\hat{\sigma}_\varepsilon^2, \sigma_\varepsilon^2$'nun yansız tahmin edicisidir.

$HKT + 2\hat{\sigma}_\varepsilon^2 sd_{kestirim}$ ifadesi de $OHKT(\hat{f}) + n\sigma_\varepsilon^2$'nun yansız tahmin edicisidir. $n\sigma_\varepsilon^2$, L matrisine bağlı olmadığından, C_p'nin minimizasyonu yaklaşık olarak $OHKT(\hat{f})$'nın minimizasyonuyla aynıdır. $L = S_\lambda$ şeklinde alınırsa seçilen λ için C_p kriteri aşağıdaki gibi yazılır.

$$C_p(\lambda) = HKT(\lambda) + 2\hat{\sigma}_\varepsilon^2 sd_{kestirim}(\lambda).$$

$C_p(\lambda)$'yı minimize eden düzleştirici parametresini $\hat{\lambda}_{C_p}$ile gösterilsin. $\hat{\sigma}_\varepsilon^2$ tahmini için $sd_{kestirim}$'in seçilmesi gerekir. Artıklar gerçek hataları tahmin ettiğinden, σ_ε^2'nun bir tahmini olarak HKT/n kullanılabilir ama bu istatistik σ_ε^2'nu olduğundan az tahmin eder çünkü artıklar gerçek hatalara göre daha az değişkendir. Bu sorunla parametrik regresyonda da karşılaşılır ve sorun, HKT'nı n yerine (n – p)'ye bölünerek çözülür. Aynı sorunu burada da (n – p)'yi, $sd_{artık}(\lambda)$ ile değiştirerek çözmek mümkündür. Sonuç olarak $\hat{\sigma}_\varepsilon^2$'nu aşağıdaki şekilde yazabiliriz.

$$\hat{\sigma}_\varepsilon^2 = \frac{HKT(\lambda)}{sd_{artık}(\lambda)}$$

$\hat{\sigma}_\varepsilon^2$ tahmini için λ'nın seçilmesi gerekir Bunun için çapraz geçerlilik veya genelleştirilmiş çapraz geçerlilik kriterlerini minumum yapan λ'yı kullanmak uygundur ya da bunun yerine $\hat{f}$ varyansını ve yanlılığını dengelediği için bu kriterlerin küçültücülerinden biraz daha küçük λ seçilebilir. Bununla birlikte, $\hat{\sigma}_\varepsilon^2$ tahmini edilme amacı olsa bile, $\hat{f}$'daki yanlılıktan kaçınılmalıdır. Çünkü bu yanlılık $\hat{\sigma}_\varepsilon^2$'nu düzeltilemeyecek şekilde büyütür. $\hat{f}$'daki varyans ise $\hat{\sigma}_\varepsilon^2$ düşürür ancak bu durum HKT'nın n yerine $sd_{artık}$'a bölerek düzeltilebilecek etkidir. Sonuç olarak, $\hat{\sigma}_\varepsilon^2$ tahmin edilirken çok küçük bir düzeltme miktarı kabul edilebilir. Cezalandırılmış splinelar kullanılırken, EKK yöntemi çok fazla düğüm olması sebebiyle tutarlı sonuçlar vermeyeceği için $\hat{\sigma}_\varepsilon^2$'nu tahmin ederken λ = 0 bile kullanılabilir.

3.7.3 Akaike bilgi kriteri (AIC)

Akaike bilgi kriteri, ilk olarak parametrik modeller için geliştirilmiştir. Ancak doğrusal regresyon ve zaman serileri için Akaike bilgi kriterinin sapmasının oldukça büyük olduğu Hurvich, Simonoff ve Tsai (1998) tarafından gösterilmiştir ve daha küçük sapma içeren düzeltilmiş AIC_c geliştirmişlerdir. Düzeltilmiş akaike bilgi kriteri, parametrik olmayan düzelticilerde düzeltme parametresinin seçimi için kullanılmıştır. Öncelikle Akaike bilgi kriteri şu şekilde hesaplanır (Akaike 1973, Rupert vd. 2003):

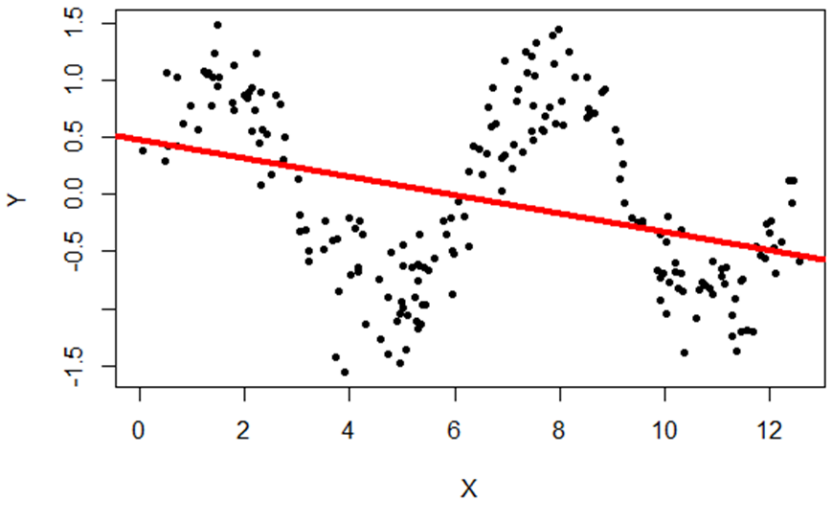
$$AIC(\lambda) = log\{HKT(\lambda)\} + 2sd_{kestirim}(\lambda)/n.$$

Düzeltilmiş Akaike bilgi kriteri ise aşağıdaki şekilde hesaplanır:

$$AIC_c(\lambda) = log\{HKT(\lambda)\} + \frac{2\{sd_{kestirim}(\lambda) + 1\}}{n - sd_{kestirim}(\lambda) - 2}$$

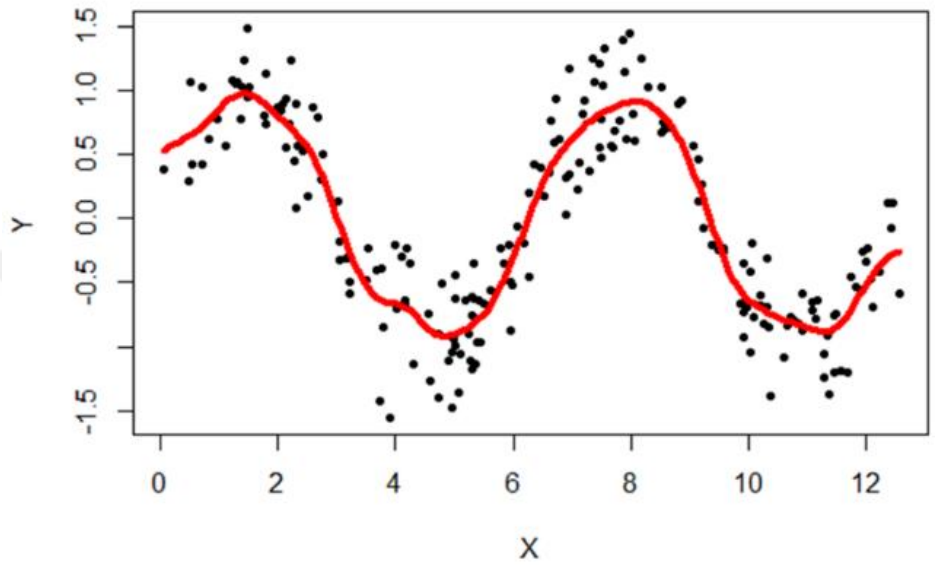
3.8 Nonparametrik Regresyonda Kernel Yöntemi ile Tahmin

Örneklem sayıları 200 olmak üzere tekdüze dağılımdan X değişkeni, normal dağılımdan Y değişkeni üretilsin. X_i ve Y_i değişkenleri arasındaki ilişki basit doğrusal regresyonla tahmin edildiğinde aşağıdaki şekil elde edilir.



Şekil 3.2 Basit doğrusal regresyon tahmini

Basit regresyon yöntemiyle yukarıdaki veride doğru bir tahmin yapılamamıştır. Bu iki değişken arasındaki tahmini yapabilmek için parametrik olmayan yöntemlerden Kernel regresyon tercih edilebilir. Kernel normal dağılım olarak ayarlandığında fonksiyon olarak Gauss kullanılır bant genişliği h ise 0.9 olarak alınırsa aşağıdaki Kernel tahmini elde edilecektir.



Şekil 3.3 Nonparametrik regresyonda Kernel tahmini

4. SEMİPARAMETRİK REGRESYON

Semiparametrik regresyon modeli, bağımlı değişkenin bağımsız değişkenlerin bir kısmıyla parametrik bir kısmıyla parametrik olmayan bir ilişki içinde olduğu durumlarda kullanılır. Semiparametrik modelin parametrik kısmının belirlenmesinde farklı modeller tahmin edilebilir. Tahmin edilen bu modellerden artık kareler toplamını minimum yapan model semiparametrik modelin parametrik kısmını oluşturur. Bağımlı değişken ile her bir bağımsız değişkenin ayrı ayrı çizilecek grafikleri incelenerek, bağımlı değişken ile bağımsız değişkenler arasındaki ilişkinin yapısı hakkında fikir sahibi olunabilir. Her bir değişkenin ilişkisine tek tek bakıldıktan sonra bağımsız değişkenlerin bir veya birkaçı için parametrik, diğerleriyle parametrik olmayan bir ilişki varsa semiparametrik model tercih edilir.

Semiparametrik modeller, parametrik ve parametrik olmayan bileşenlerin her ikisini içerdiğinden, bu modellere kısmi doğrusal modeller (partially linear model) de denir. Bu modeller standart regresyon modellerine göre çok daha esnektir. Semiparametrik regresyon modellerinde parametrik kısım, doğrusal olabileceği gibi dönüşüm yöntemleri (Logaritmik, karesel dönüştürme vs), uygulanarak doğrusallaştırılabilen yapıda da olabilir. Semiparametrik modelin parametrik kısmının belirlenmesinde farklı modeller tahmin edilebilir. Tahmin edilen bu modellerden artık kareler toplamını minimum yapan model semiparametrik modelin parametrik kısmı olarak tahmin edilebilir. Tahmin yöntemlerinden biri spline düzeltme yöntemidir ve bu yöntemin temeli cezalı en küçük kareler regresyon tahminine dayanır. Sıradan en küçük kareler yönteminden farklı olarak burada hata kareler toplamına, düzeltme parametresine sahip ceza fonksiyonu eklenir ve daha tutarlı tahminler elde edilebilir (Aydın, 2005)

Semiparametrik Regresyon Modeli: Semiparametrik regresyon modeli toplamsal modellerin (additive model) özel bir durumudur. Nonparametrik regresyonda çoklu regresyon modellerinin tahmini, çok boyutluluğun yarattığı sıkıntı nedeniyle zorlaşmaktadır. Bu zorluğun üstesinden gelebilmek için toplamsal modeller kullanılır.

Toplamsal model tanımı:

$$\boldsymbol{Y_i = \alpha + m_1(x_1) + m_2(x_2) \ldots + m_k(x_k) + \varepsilon}$$

Semiparametrik regresyon modeli:

$$\boldsymbol{y_i = \alpha + z_i^T \beta_i + m(x_i) + \varepsilon_i, \;\; \varepsilon_i \sim N(0, \sigma^2), \;\; 1 \leq i \leq n}$$

biçiminde ifade edilir.

Burada:

$\boldsymbol{z_i^T \beta}$ **;** (1 x p) boyutlu parametrik kısma karşılık gelen i.gözlem vektörünü

$\boldsymbol{m(x_i)}$**;** modelin parametrik olmayan kısmına karşılık gelen ikinci mertebeden türevi alınabilen fonksiyonlar uzayının bir elemanını

$\boldsymbol{x_i}$**:** i. gözlem değerini

$\boldsymbol{\varepsilon_i}$ **;** i. gözlem için hata terimini göstermektedir.

Bu modelde amaç, parametrik kısımdaki parametreler vektörü β'yı ve parametrik olmayan $m(x_i)$ fonksiyonunu $\{x_i, y_i, z_i\}$ veri kümesinden tahmin etmektir.

Formülün açık hali de şu şekilde ifade edilir:

$$y_i = \alpha + m_1(x_1) + m_2(x_2) + \cdots + m_j(x_j) + \beta_1 z_1^T + \cdots + \beta_k z_k^T + \varepsilon \qquad (4.1)$$

Matris-vektör terimi ile ifade edersek;

$$\boldsymbol{y = Z\beta + m + \varepsilon}$$

$$y = \begin{bmatrix} y_1 \\ y_2 \\ . \\ . \\ . \\ y_n \end{bmatrix}, \; Z = \begin{bmatrix} z_{11} & . & . & . & z_{1k} \\ z_{22} & . & . & . & z_{2k} \\ . & . & . & . & . \\ . & . & . & . & . \\ . & . & . & . & . \\ z_{n1} & . & . & . & z_{nk} \end{bmatrix}, \; \beta = \begin{bmatrix} \beta_1 \\ \beta_2 \\ . \\ . \\ . \\ \beta_k \end{bmatrix}, \; m = \begin{bmatrix} m(x_1) \\ m(x_2) \\ . \\ . \\ . \\ m(x_n) \end{bmatrix} \text{ ve } \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ . \\ . \\ . \\ \varepsilon_n \end{bmatrix}$$

(4.1) modelinde yer alan tüm açıklayıcı değişkenlerin belirlenmiş olduğu varsayılır ve söz konusu modelde $m(x_i)$ sabit terimi kapsadığından parametrik kısım sabit terim içermez (Aydın 2005).

4.1 Semiparametrik Regresyon Modellerinin Tahmini

Bu tür regresyon modellerinin kestirimde kullanılan yöntemlerden biri spline düzeltme yöntemidir. Spline düzeltme yönteminin temeli, cezalı en küçük kareler regresyonuna dayanmaktadır. Sıradan en küçük karelerden farklı olarak, hata kareler toplamına bir düzeltme parametresi sahip ceza fonksiyonu eklenir. Bu ceza fonksiyonuyla birlikte esnek eğimli uyumlar ve sabit eğimli uyumlar arasında ortak bir nokta sağlanır ve böylece söz konusu modelle çok daha tutarlı tahminler yapılır (Aydın, 2005).

4.1.1 Cezalandırılmış en küçük kareler yöntemi

Klasik yaklaşımdan yola çıkarak artık kareler toplamını minumum yapan m fonksiyonu ve β parametreleri elde edilmeye çalışılır. Fakat m fonksiyonu üzerinde bir kısıt bulunmuyorsa bu yöntem başarısız olacaktır. Farklı x_i değerleri için, β'nın herhangi bir değerine karşılık m fonksiyonu $m(x_i) = y_i - z_i^T\beta$ şeklinde elde edilebilir. Ancak burada β'nın bilinmesi gerekir bilinmediği durumlarda ise bu yaklaşımla m fonksiyonu elde edilemez. Bu sorunun üstesinden gelebilmek için cezalandırılmış en küçük kareler yöntemi kullanılır. Semiparametrik regresyon için cezalandırılmış en küçük kareler toplamı aşağıdaki şekilde ifade edilir:

$$S(\beta, m) = \sum_{i=1}^{n} \{y_i - z_i^T\beta - m(x_i)\}^2 + \lambda \int_a^b \{m''(x)\}^2 dx \qquad (4.2)$$

Bu yöntem, kübik spline düzeltmeye dayanan spline düzleştirme yöntemini esas alan bir çözümdür. (4.2)'deki eşitlikteki ilk kısım hata kareler toplamını, ikinci kısım ise nonparametrik regresyonda olduğu gibi pürüzlülük cezasını gösterir. Burada λ skaler bir değerdir ve $\{x_i, y_i, z_i\}_{i=1}^n$ gözlem değerlerine bağlı olarak belirlenen bilinmeyen regresyon fonksiyonu ile sabit bir düzeltme parametresidir.

Parametre tahmini için, $x_1, \dots, x_n$ düğüm noktaları farklı ve sıralı olmak üzere yani, $x_1 < \cdots < x_n$ koşuluyla semiparametrik regresyon için cezalandırılmış en küçük kareler toplamı:

$$S(\beta, m) = \sum_{i=1}^{n} \{y_i - z_i^T \beta - m(x_i)\}^2 + \lambda \int_a^b \{m''(x)\}^2 dx$$

şeklinde yazılır. $m(\cdot)$ fonksiyonu $[x_i, x_{i+1}]$ aralığında parçalı kübik bir polinom ise o zaman $\int \{m''(x)\}^2 dx$ pürüzlülük cezası $K = QR^{-1}Q$ şeklinde ayrıştırılabilen karesel bir formdur bu durumda cezalandırılmış en küçük kareler toplamında bu formu yerine koyarsak formül şu şekilde ifade edilir:

$$S(\beta, m) = (y - Z\beta - m)^T (y - Z\beta - m) + \lambda m^T K m$$

Eğer x_i'ler farklı ve sıralı değilse, N tekrarlanma matrisi (incidence matrix) ile farklı ve sıralı hale getirilir. Burada ise $x_1, \dots, x_n$ düğüm noktalarının farklı ve sıralı değerleri $s_1, \dots, s_q$ ile gösterilsin. Böylece tekrarlanma matrisinin elemanları $x_i = s_j$ ise $N_{ij} = 1$ aksi durumda $N_{ij} = 0$ biçimindedir. Buna göre, cezalandırılmış en küçük kareler toplamı aşağıdaki şekilde ifade edilir:

$$S(\beta, m) = (y - Z\beta - Nm)^T (y - Z\beta - Nm) + \lambda m^T K m \qquad (4.3)$$

Her iki formül içinde β ve m vektörlerinin kestirimleri, bu eşitliklerin β ve m'ye göre türevi alınıp sıfıra eşitlenerek bulunur. (4.3) eşitliği için elde edilen denklemler matris formunda aşağıdaki şekilde yazılır:

$$\begin{bmatrix} Z^T Z & Z^T N \\ N^T Z & N^T N + \lambda K \end{bmatrix} \begin{bmatrix} \beta \\ m \end{bmatrix} = \begin{bmatrix} Z^T \\ N^T \end{bmatrix} y \qquad (4.4)$$

Bu modelin parametrik kısmı göz ardı edildiğinde yani β=0 olduğunda yukarıdaki formül

$(N'N + \lambda K)m = N'y$ formülüne indirgenir. Nm uyum değerlerini elde edebilmek için y vektörüne uygulanan ve λ>0 sabitine bağlı düzeltme matrisi,

$$S_\lambda = N(N'N + \lambda K)^{-1} N' \qquad (4.5)$$

biçiminde elde edilir.

Eğer x_i düğüm noktaları farklı ve önceden sıralı ise, N=I olmasından dolayı S_λ düzeltme matrisi aşağıdaki formüle indirgenir. I, burada bir birim matristir (Green ve Silverman, 1994; Tabakan, 2009).

$$S_\lambda = (I + \lambda K)^{-1} \qquad (4.6)$$

4.1.2 Kısmi spline yaklaşımı

Kısmi spline yaklaşımı ile önceden belirlenmiş λ değeri için $y = Z\beta + m + \varepsilon$ modelindeki m ve β vektörlerinin tahmini bir S_λ düzeltme matrisi yardımıyla elde edilir. Elde edilen tahmin bir kısmi spline olarak adlandırılır. Modele göre m, β ve μ'ye karşılık gelen tahmin ediciler:

$$\hat{m}_p = Nm = S_\lambda(y - Z\beta)$$
$$\hat{\beta}_p = [Z^T(I - S_\lambda)Z]^{-1}Z^T(I - S_\lambda)$$
$$= \left(\tilde{Z}^T Z\right)^{-1}\tilde{Z}^T y$$

Semiparametrik regresyon modelinin ortalama vektörü ise,

$$\mu_p = Z\hat{\beta}_p + \hat{m}_p = Z\hat{\beta}_p + S_\lambda\left(y - Z\hat{\beta}_p\right)$$
$$= S_\lambda y + (I - S_\lambda)Z\hat{\beta}_p$$
$$= S_\lambda y + (I - S_\lambda)Z[Z^T(I - S_\lambda)Z]Z^T(I - S_\lambda)y$$
$$= \left[S_\lambda + \tilde{Z}\left(\tilde{Z}^T Z\right)^{-1}\tilde{Z}^T\right]y$$
$$= H_p y \qquad (4.7)$$

$H_p = S_\lambda + \tilde{Z}\left(\tilde{Z}^T Z\right)^{-1}\tilde{Z}^T$ matrisi $\tilde{Z} = (I - S_\lambda)Z$ doğrusal regresyondaki şapka matrisine karşılık gelir. $\tilde{Z}$ matrisi tam ranklı değilse $\left(\tilde{Z}^T Z\right)^{-1}$ genelleştirilmiş ters olarak yorumlanır.

4.2 Düzeltme Parametresinin Seçimi

Spline düzeltme yöntemine dayalı semiparametrik modeli değerlendirmek için bir λ düzeltme parametresinin seçilmesi gerekir. İkinci bölümde ayrıntılı olarak incelenen ve düzeltme parametresinin seçiminde kullanılan Çapraz Geçerlilik, Geliştirilmiş Akaike Bilgi Kriteri ve Mallow'un C_p Kriteri olan klasik yöntemlerin yardımıyla λ düzeltme parametresinin seçimi yapılabilir. Semiparametrik model için λ düzeltme parametresinin belirlenmesinde kullanılan kriterlerin, parametrik olmayan bir model için kullanılan kriterlerden farkı S_λ matrisi yerine H matrisinin hesaplanmasıdır. Kısmi spline yaklaşımı kullanıldığında, (4.6)'de verilen ve simgesel olarak $tr(S_\lambda)$ ile gösterilen matris izi yerine, (4.7)'de ifade edilen matrisin izi hesaplanır. Diğer bir ifadeyle $tr(H_p)$

hesaplanır. Buna göre, λ düzeltme parametresi söz konusu bu matrisin boyutları ile orantılı bir zamanda elde edilir.

4.3 Varyans ve Kovaryans Tahmini

Semiparametrik regresyonda varyans ve kovaryans tahmin edilmesindeki amaçlar şu şekilde sıralanabilir:

i. Düzeltme parametresinin seçimindeki ölçütler olan yansız risk kriteri ve risk tahmin kriterlerinin hesaplanmasında
ii. Modelin parametrik bileşeni için çıkarımlar yapılmasında
iii. Modelin parametrik olmayan fonksiyonu hakkındaki çıkarımların yapılmasında gereklidir.

Semiparametrik modelin parametrik kısmını temsil eden $Z\beta$ için geliştirilen varyans tahmini:

$$\hat{\sigma}^2 = \frac{y'A'(I-P)A_y}{tr(A'(I-P)A)}$$

Bu formülde $P = Az(z'A'Az)^{-1}z'A'$, A matrisi $(n-2)xn$ boyutludur ve herhangi bir i. satırının i. elemanı $a_i c_i$, (i+1). elemanı $-c_i$ ve (i+2). elemanı $b_i c_i$ olmak üzere, bu satırın diğer tüm elemanları sıfırdır. A matrisi aşağıdaki elemanlardan oluşmaktadır.

$$a_i = \frac{x_{i+1} - x_i}{x_{i+1} - x_{i-1}}, \quad b_i = \frac{x_i - x_{i-1}}{x_{i+1} - x_{i-1}}$$

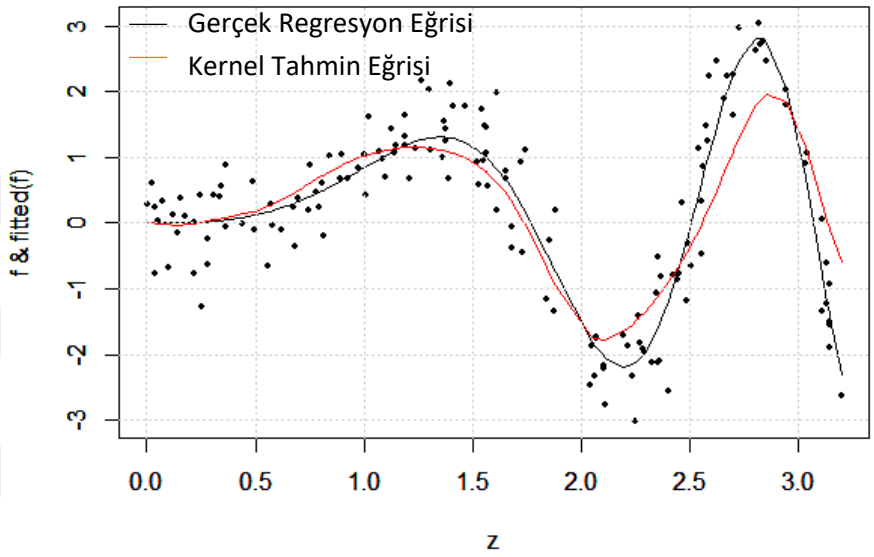
$$c_i = (a_i^2 + b_i^2 + 1)^{-1/2}, \quad i = 1, \dots, n-1$$

Semiparametrik regresyonun parametrik katsayılarının varyanslarının tahmininde kullanılan varyans kestiricileri için Schimek tarafından yapılan çalışmada iki yaklaşım ele alınmıştır. Bu yaklaşımlar: *Kısmi spline (partial spline)* ve *Speckman yaklaşımıdır*. Burada kısmi spline β_p için varyans-kovaryans matrisi verilmiştir (Aydın, 2005).

$$V_p = \sigma^2\left(\tilde{Z}'Z\right)^{-1}\tilde{Z}'Z\left(Z'\tilde{Z}\right)^{-1}$$

4.4 Semiparametrik Regresyonda Kernel Yöntemi ile Tahmin

$\beta_0 = 1$ ve $\beta_1 = 2$ gerçek değerler, h bant genişliği 0.5 olarak alınmış, düzleştirme parametresi genelleştirilmiş çapraz geçerlilik yöntemiyle seçilmiştir. Sonuç olarak Kernel düzleştiricisiyle tahmin aşağıdaki grafik de görüldüğü gibi elde edilmiştir.



Şekil 4.1 Semiparametrik regresyonda Kernel tahmini

Tahmin edilen beta değerleri ise sırasıyla $\beta_0 = 1.020775$ ve $\beta_1 = 1.981537$ şeklinde gerçek tahminlere çok yakındır.

5. BAYESCİ İSTATİSTİK

Bayesci istatistiğin temelini Thomas Bayes tarafından ortaya atılan Bayes teoremi oluşturmaktadır. Bu yaklaşımda parametreler klasik yaklaşımın aksine rastgele değişkenler olarak ele alınır. Klasik yaklaşımdan ayıran en önemli özellik gözlemleri ve öznel görüşleri kullanmasıdır. Veriden elde edilen olabilirlik fonksiyonu ve daha önceki araştırmalardan elde edilen ya da araştırmacının kendi görüşlerini yansıtan ve önsel bilgi adı verilen bu önsel bilgi ile olabilirlik fonksiyonuna dayanarak, sonsal dağılım elde edilir. Elde edilen bu sonsal dağılımdan parametreye ilişkin tüm bilgiye ulaşılabilir. Bayes teoremi, sonuçlardan hareket ederek nedenlerin olasılıklarını belirlemeye çalışan, alışılagelenin dışında bir bakış açısı sunmuştur. Koşullu olasılık tanımından yola çıkılarak Bayes teoremi açıklanabilir. Kesikli olasılık modeli için:

$$P(B|A) = \frac{P(A|B)P(B)}{P(A)}$$

$P(B)$ ön bilgiyi, $P(A|B)$ yeni bilgiyi içeren olabilirlik fonksiyonunu (likelihood) ve $P(B|A)$ sonsal olasılığı göstermektedir. Değişkenlerin sürekli olduğu durumlar için Bayes Teoremi:

$$p(\theta|Y = y) = \frac{p(\theta, y)}{p(y)}$$

$$p(\theta|y) = p(y|\theta)p(\theta)$$

$$p(y) = \int_{-\infty}^{+\infty} p(\theta)p(y|\theta)d\theta$$

$$p(\theta|Y = y) = \frac{p(\theta)p(y|\theta)}{\int_{-\infty}^{+\infty} p(\theta)p(y|\theta)d\theta}$$

Yukarıdaki formülün paydada yer alan integral sebebiyle hesaplanması oldukça güçtür. Hesaplamayı kolaylaştırmak için önsel dağılımlar kullanılır. Önsel ve olabilirlik fonksiyonları aynı aileye mensup ise, sonsal dağılımlar analitik olarak elde edilebilir.

5.1 Önsel Dağılımın Belirlenmesi

Önsel dağılım, daha önce yapılmış araştırma ve denemelerle elde edilmiş bilgiye ait dağılımdır. Bayesci analizinin en önemli noktalarından biri önsel dağılımın seçimidir. Tahmini yapılmak istenen belirsiz nicelikler veya tesadüfi değişkenlerle ilgili bilgi olasılıksal olarak ifade edilmeli ve ardından Bayes Teoremi uygulanmalıdır. Çoğu zaman tesadüfi değişkenin alabileceği değerler belliyken, belli olmadığı durumlarda değerler sübjektif olarak belirlenmektedir. Çok sayıda mümkün değer olması durumunda hesaplamaları kolaylaştırmak için daha az değer alınabilir ya da sürekli olarak ele alınabilecek bir değişken kesikli varsayılabilir. Kesikli olması durumunda, önsel olasılık dağılımı tesadüfi değişkenin değerler kümesinden ve olasılıklarından oluşur. Burada olasılıkların pozitif olması ve toplamlarının 1 olması şartı aranır (Winkler, 2003). Bayesci yaklaşımda önsel dağılım seçiminde farklı görüşler vardır. Klasik Bayesciler, bilgi vermeyen önsel dağılımları örneğin dikdörtgen önsel dağılım gibi dağılımların (non-informative prior) seçilmesini uygun görürler. Modern parametrik Bayesciler tasarlanmış özelliklere sahip eşlenik önsel dağılımları (conjugate prior) seçmeyi tercih ederler. Sübjektif Bayesciler ise benzer bir alanda daha önce yapılmış çalışmalar sonucu elde edilen, çoğu zaman uzman görüşünden elde edilen bilgiye göre çıkarımı yapılan önsel dağılımları (elicited prior) kullanmayı önermişlerdir.

Bilgi Vermeyen Önsel Dağılımlar

Bilgi vermeyen önsel dağılımların, parametreyi açıklama gücü çok azdır yani veriden elde edilen bilgi dışında bilgiye kullanılmaz bu yüzden bu önsel dağılımlara bilgi içermeyen önsel dağılım denir. Elde edilen tahminlerle, klasik yaklaşımla elde edilen tahminler arasında çok fark olmayacağı söylenebilir. Bilgi içermeyen önsel dağılımlar verilerden elde edildiği için sübjektif olduğuna dair olan eleştirileri bertaraf etmiş olur. Bilgi vermeyen önsel dağılımlar; Uniform (düzgün) önsel dağılım, Jeffreys'in önsel dağılımı, referans önsel dağılım, belirsiz önsel dağılımlardır.

Düzgün Önsel Dağılım: Düzgün (Uniform) önsel dağılımda belirlenen bir aralıktaki parametreye aynı olasılık değerleri atanır böylece sübjektiflikten uzaklaşılması sağlanır.

Düzgün önsel olasılıkta eğer dönüşüm yapılıyorsa yeni bir ifade meydana getireceği için eşit olasılık özelliği kaybolabilir bu durumda bu önsel dağılımların dönüşümler altında sabit olmadığı söylenebilir. Bu dağılımın güçlü yanı ise, örneklem hacmi arttıkça düzgün dağılım olmasının etkilerinin azalmasıdır.

Jeffreys'in Önsel Dağılımı: Jeffreys'in önsel dağılım görüşü tam bilgisizlik durumuna inanır. Eğer olasılıklar eşit değilse bir olayın diğer olaya göre gerçekleşme olasılığı değişiyorsa bunun bir nedene dayandırılması gerektiğini söyler. Belirsizliği temsil etmesi içinde tek bir önsel dağılım olması gerekmez. Düzgün önsel dağılım gibi belirli bir aralık olsun ya da olmasın dönüşümler sonucu değişmez ve sabit kalır. Bu dağılım, parametrik olan herhangi bir model için bilgi içermeyen önsel dağılımın elde edilmesine imkân verir.

Bilgi Veren Önsel Dağılımlar

Eşlenik Aileleri: Önsel bilgi belirsiz değildir. Uygun matematiksel özelliklere sahip doğal eşlenik ailelerinden birinin üyesi durumundaki önsel dağılımlardır. Sonsal dağılımın kapalı bir formda matematiksel olarak elde edilmesi hedeflenir. Bu durumun sağlanması için de çoğu zaman önsel dağılımın eşlenik önsel olması istenir çünkü eşlenik önsel ile bulunan sonsal dağılım önsel dağılım ile aynı aileye sahiptir. Aynı aileden olması ise matematiksel hesaplamalarda büyük kolaylık sağlar. Eşlenik ailelerin her biri benzerlik fonksiyonuyla birleştirilebilen dağılımlardan oluşur.

Çizelge 5. 1 Eşlenik Aileleri (Karadağ, 2011)

Önsel	**Olabilirlik**	**Parametre(ler)**	**Sonsal**
Beta	Bernoulli/Binom	P	Beta
Gamma	Poisson/Üstel	Λ	Gamma
Normal	Normal D. (σ^2 biliniyor)	μ	Normal D.
Ters Gamma	Normal D. (μ biliniyor)	σ^2	Ters Gamma
Normal-Gamma	Normal D.	μ	Normal-Gamma
Pareto	Tek Biçimli	α,β	Pareto
Gamma	Gamma	α,β	Gamma

5.2 Sonsal Dağılım

$\theta = \left[\theta_1, \ldots, \theta_p\right]^T$ bilinmeyen parametre vektörünü ve $y = [y_1, \ldots, y_n]^T$ gözlemlenen verilerin vektörünü ifade eder. Bayes çıkarımı, Bayes teoremi tarafından verilen y'yi gözlemledikten sonra θ'nın sonsal olasılık dağılımına dayanır. Daha kolay anlaşılması için θ'nın her bir elemanının sürekli olduğu varsayımı altında:

$$p(\theta|y, I) = \frac{p(y|\theta, I)\pi(\theta|I)}{p(y|I)}, \tag{5.1}$$

şeklindedir. Burada iki önemli bileşen vardır: Olabilirlik işlevi $p(y|\theta, I)$ ve önsel dağılım $\pi(\theta|I)$. Önsel dağılım, y verileri gözlemlenmeden önce, θ için olasılık inançlarını temsil eder. Her ikisi de mevcut bilgi I'ya bağlıdır. Farklı bireylerin farklı bilgileri I olacaktır ve bu nedenle genel olarak önsel dağılımları (ve muhtemelen olasılık işlevleri) farklı olabilir. (5.1)'deki payda, $p(y|I)$, sağ tarafın parametre uzayı üzerinden bir ile bütünleşmesini sağlayan bir normalleştirme sabitidir. Çok önemli olmasına rağmen, işlemlerde kolaylık olması için, bu noktadan itibaren I'ya olan bağımlılık kaldırılmaktadır. Normalleştirme sabiti aşağıdaki gibidir:

$$p(y) = \int p(y|\theta)\pi(\theta)d\theta \tag{5.2}$$

bu değer θ'ya bağlı değildir. Orantı sabiti bilinen bir olasılık yoğunluk fonksiyonundan örnekleme yöntemleri bulunmaktadır. Bu yöntemler paydayı hesaplamadan sonsal dağılımı bulmamızı sağlar. Örneğin, eğer $\theta_1, \ldots, \theta_N$ bir örnekse $N^{-1}\sum_{i=1}^{N}\theta_i$ sonsal dağılımın bir tahminidir. Normalleştirme sabiti aynı zamanda modeli verilen gözlemlenen verilerin marjinal olasılığıdır, yani olabilirlik fonksiyonu ve önsel bilgidir. Sabit terim kaldırılırsa,

$$p(\theta|y) = \frac{p(y|\theta)\pi(\theta)}{p(y)}$$

$$p(\theta|y) \propto p(y|\theta) \times \pi(\theta)$$

şeklinde formülize edilebilir veya daha yaygın olarak,

Sonsal ∝ Olabilirlik× Önsel olarak gösterilir.

Bayesci çıkarım yaklaşımı için uygulamada dikkate alınması gereken bir dizi önemli konu vardır. En önemlisi önsel bilginin tanımlanmasıdır. İkinci olarak, önsel ve olabilirlik bileşenlerine karar verildikten sonra, (genellikle) çok değişkenli sonsal dağılımı özetlememiz gerekir ve bu özetleme, yüksek boyutlu olabilecek parametre uzayı üzerinde entegrasyon gerektirir. Bayes yaklaşımı çok doğal çıkarımsal özetler sağlar. Ancak, bu özetler integral hesaplamaları gerektirir ve çoğu model için bu integraller analitik olarak anlaşılmazdır. Bundan dolayı Bayesci yaklaşımda MCMC gibi teknikler kullanılır.

5.3 Bayesci İstatistikte Parametre Tahmini

Çok aşamalı modellerde ilgilenilen parametre sayısı kadar bilinmeyen parametre değeri vardır. Bayesci istatistikte çıkarım yapabilmek için ortak sonsal dağılımın bulunması gerekir bu da çok boyutlu integrallerin çözümünün yapılmasını gerektirmektedir. İntegral çözümleri zor olduğundan Bayesci yaklaşımda alternatif bir yöntem olan Markov zinciri ve Monte Carlo simülasyon (MCMC) yaklaşımı kullanılmaktadır. Pratikte kullanılan iki temel MCMC yöntemi vardır.

1. Metropolis ve Metropolis-Hastings Yöntemi
2. Gibbs Örnekleme

Koşullu önsel dağılımlar standart bir dağılıma sahip değilse Metropolis-Hastings yöntemi kullanılırken standart bir dağılımları varsa Gibbs örnekleme kullanılır.

5.3.1 Metropolis ve Metropolis-Hasting algoritması

Metropolis ve Metropolis-Hasting algoritmasında Markov zinciri ile ilişkili değişkenler oluşturulamamaktadır. Şimdiki değere rassal bir sayı eklenerek yeni bir değer önerilir ve bu değerin yeni değer olarak kabul edilip edilmeyeceğine karar verilir. Önerilen değerin olasılığı eski değerin olasılığından yüksekse yeni değerle devam edilir değilse eski değer kullanılmaya devam edilir. Bu değerler dizisi ile histogram oluşturulabilir ya da beklenen değer hesaplanabilir.

Metropolis Algoritması: Metropolis algoritması, belirtilen hedef dağılımına yakınsamak için bir kabul/red kuralına sahip rastgele bir yürüyüşün uyarlamasıdır. θ parametresinin olasılık yoğunluk fonksiyonu p(θ) olarak verilsin. Metropolis algoritması şu şekilde ilerler:

1. $p_0(\theta)$ başlangıç dağılımından $p(\theta_0|y) > 0$ olmak üzere θ_0 başlangıç noktasını alalım.
2. t= 1,2,… olmak üzere,
   a.) $J_t(\theta^*|\theta^{t-1})$ olmak üzere t zamanında sıçrama(jumping) ya da teklif (proposal) dağılımından bir θ^* örneği alınır. Sıçrama dağılımı simetrik olmalıdır ve tüm θ_a, θ_b ve t için $J_t(\theta_a|\theta_b) = J_t(\theta_b|\theta_a)$ koşulunu sağlamalıdır.
   b.) Yoğunluk oranları hesaplanır:

$$r = \frac{p(\theta^*|y)}{p(\theta^{t-1}|y)} \tag{5.3}$$

   c.) $\theta^t = \begin{cases} \theta^* & \min(r,1)\text{olasılığıyla} \\ \theta^{t-1}, & diğer\ durumlarda \end{cases}$

Eğer aday nokta yoğunluk fonksiyonunu arttırırsa (r>0 ise) kabul edilir yani $\theta^t = \theta^*$ olarak alınır ve tekrar bir aday noktası çekilir. Aday nokta yoğunluk fonksiyonunu azaltırsa (r<0 ise) r olasılığı ile kabul edilir, aksi halde reddedilir. $\theta^t = \theta^{t-1}$ olduğunda yani, sıçrama kabul edilmez ise zincir hareket etmez ama algoritmada bir yineleme olarak sayılır ve yeni bir aday nokta seçilir. Bu şekilde bir Markov zinciri oluşturulur ve yeteri kadar yineleme yapılarak zincir durağan dağılıma yaklaşır ve elde edilen örneklemler p(θ) dağılımından örneklemlere karşılık gelir.

Metropolis-Hasting Algoritması: Metropolis-Hasting algoritması, yukarıda sunulan temel Metropolis algoritmasını iki şekilde genelleştirir. İlk olarak, J_t sıçrama kurallarının artık simetrik olması gerekmez; yani, $J_t(\theta_a|\theta_b) = J_t(\theta_b|\theta_a)$ şartı yoktur. İkincisi, sıçrama kuralındaki asimetriyi düzeltmek için, (5.3)'deki r oranı, bir oran oranı ile değiştirilir:

$$r = \frac{p(\theta^*|y)/J_t(\theta^*|\theta^{t-1})}{p(\theta^{t-1}|y)/J_t(\theta^{t-1}|\theta^*)}$$

r oranı her zaman tanımlanır, çünkü θ^{t-1}'den θ^*'a bir sıçrama ancak $p(\theta^{t-1}|y)$ ve $J_t(\theta^*|\theta^{t-1})$ sıfır olmadığında gerçekleşebilir.

Sıçrama Kuralı ile Simülasyonların Verimliliği Arasındaki İlişki

İdeal Metropolis-Hastings sıçrama kuralı, hedef dağılımdan alınan θ^* ile örneklemedir; yani, tüm θ için $J(\theta^*|\theta) \equiv p(\theta^*|y)$'dir ve θ^t yinelemeleri $p(\theta|y)$'den bağımsız çekilişler dizisidir. Ancak genel olarak, yinelemeli simülasyon doğrudan örneklemenin mümkün olmadığı problemlere uygulanır.

İyi bir sıçrama dağılımı aşağıdaki özelliklere sahiptir:

- Herhangi bir θ için $J(\theta^*|\theta)$'den örnek almak kolaydır.
- r oranını hesaplamak kolaydır.
- Her sıçrama parametre alanında makul bir mesafe kat eder (aksi takdirde rastgele yürüyüş çok yavaş hareket eder).
- Sıçrayışlar çok sık reddedilmez (aksi takdirde rastgele yürüyüş hareketsiz durmak için çok fazla zaman harcar).

5.3.2 Gibbs Örnekleme

Gibbs örneklemesi, birçok çok boyutlu problemde faydalı bulunan bir özel Markov zincir algoritmasıdır. θ'nin alt vektörleri cinsinden tanımlanan *alternatif koşullu örnekleme* olarak da adlandırılır. θ parametre vektörünün d bileşene veya $\theta = \theta_1, \cdots, \theta_d$ şeklinde alt vektöre bölündüğü varsayılsın. Gibbs örnekleyicisinin her yinelemesi, her bir alt kümeyi diğerlerinin değerine koşullu olarak bağlı olarak θ alt vektörleri arasında döngü yapar. t yinelemede d adım vardır. Her t yinelemesinde, θ'nın d alt vektörlerinin bir sıralaması seçilir ve sırayla, her θ_j^t, θ'nın diğer tüm bileşenleri verilen koşullu dağılımdan örneklenir:

$$p(\theta_j|\theta_{-j}^{t-1}, y),$$

burada θ_{-j}^{t-1}, θ'nın θ_j hariç tüm bileşenlerini mevcut değerlerinde temsil eder:

$$\theta_{-j}^{t-1} = (\theta_1^t, \ldots, \theta_{j-1}^t, \theta_{j+1}^{t-1}, , \ldots, \theta_d^{t-1})$$

Böylece, her θ_j alt vektörü, θ'nin diğer bileşenlerinin en son değerine bağlı olarak güncellenir, ki bunlar, halihazırda güncellenmiş bileşenler için yineleme t değerleri ve diğerleri için yineleme t − 1 değerleridir. Standart istatistiksel modelleri içeren birçok problem için, parametrelerin koşullu sonsal dağılımlarının çoğundan veya tamamından doğrudan örnekleme yapmak mümkündür. Bir dizi koşullu olasılık dağılımı kullanılarak modeller oluşturulabilir. Genellikle bu koşullu dağılımlar bilinen bir forma sahiptirler böylece rastgele sayılar simülasyonla kolay bir şekilde üretilebilirler. Gibbs örneklemesi her zaman yeni bir değere gider ve önemli bir farkı önerilen bir dağılıma gerek duymamasıdır ancak parametre uzayı anlaşılması zor veya parametreler yüksek ilişkili ise Gibbs örneklemesi başarılı olmayabilir.

5.4 Bayesci Semiparametrik Regresyon Tahmini

Markov zinciri Monte Carlo'ya dayanan Bayesci semiparametrik regresyon modeli, cezalandırılmış spline'lar ve karma modeller arasında uygun bir bağlantı kullanan karma bir model olarak tanımlanır. Karma doğrusal modelde amaç hem sabit etkilere hem de rastgele etkilere ilişkin parametre kestirimlerinin elde edilmesidir. Bu bölümde Gibbs örneklemesi kullanarak cezalandırılmış spline'larla Bayes tahmini anlatılmıştır. Doğrusal karma bir modelin genel formu aşağıdaki şekilde tanımlanır:

$$y_i = \sum_{j=0}^{p} \beta_j x_{ji} + g(x_{p+1,i}) + \varepsilon_i, \quad i = 1,2, \ldots, n \tag{5.4}$$

Bu denklemde $\sum_{j=0}^{p} \beta_j x_{ji}$ ifadesi modelin p boyutlu ortak değişken fonksiyonunun doğrusal kabul edilen parametrik kısmıdır. $g(x_{p+1,i})$, parametrik olmayan kısmı ve $\varepsilon_1, \varepsilon_2, \ldots \varepsilon_n$ gözlemlenemeyen hata terimlerini temsil eder. Bu hata terimlerinin sıfır ortalama ve $\sigma_\varepsilon^2 I_n$ kovaryansıyla bağımsız ve özdeş normal dağılımlı olduğu kabul edilirken, σ_ε^2 ise bilinmeyen varyansı gösterir. (5.4)'deki model için q dereceli cezalandırılmış spline'ı kullanarak aşağıdaki modele ulaşılır:

$$y_i = \sum_{j=0}^{p} \beta_j x_{ji} + \sum_{j=1}^{q} \beta_{p+j} x_{p+1,i}^{j} + \sum_{k=1}^{K} u_k \left(x_{p+1,i} - k_k\right)_+^q + \varepsilon_i, \qquad i = 1,2,\dots,n \qquad (5.5)$$

$k_1, \dots, k_k \quad a < k_1 < \cdots < k_k < b$ olmak üzere iç düğümlerdir. Modelin matris formundaki gösterimi aşağıdaki gibidir.

$$Y = X\beta + Zu + \varepsilon \qquad (5.6)$$

$X\beta$ spline'ın saf polinom bileşenidir. Zu, $\sigma_u^2 Q$ kovaryansıyla cezalandırılmış kesikli fonksiyon bileşenidir. X, sabit etkilere ilişkin tasarım matrisi, β, sabit etkilere ilişkin parametre vektörü, Z, rastgele etkilere ilişkin tasarım matrisi, u rastgele etkilere ilişkin parametre vektörü, ε rastgele hatalar vektörüdür. Olabilirlik fonksiyonu $f(y \mid \beta, u, \sigma_u^2, \sigma_\varepsilon^2)$ şeklinde belirtilir. $(\beta, u, \sigma_u^2, \sigma_\varepsilon^2)$ parametre vektörü olmak üzere, karma bir model u'nun önseli $N(0, \sigma_u^2 I)$ 'dır. Bu modeli bayesci modele genişletebilmek için $\beta, \sigma_u^2, \sigma_\varepsilon^2$ önsel dağılımlarının bulunması gerekir. β hakkında çok az şey bilindiği için önsel olarak uygun olmayan tek düze dağılımı seçmek mantıklıdır ya da daha uygun bir önsel seçmek istenirse β için σ_β^2 ile $N(0, \sigma_\beta^2 I)$ önseli kullanılabilir burada normal dağılım β aralığında tek düzedir. Sonuç olarak $\pi_0(\beta) \equiv 1$ kullanılır. σ_ε^2'nun önseli içinse A_ε ve B_ε parametreleriyle ters gamma dağılımı kullanılır ve $IG(A_\varepsilon, B_e)$ şeklinde belirtilir (Rupert vd., 2003). $IG(A_\varepsilon, B_e)$'nın olasılık yoğunluk fonksiyonu:

$$\pi_0(\sigma_\varepsilon^2) = \frac{\beta_\varepsilon{}^{A_\varepsilon}}{\Gamma(A_\varepsilon)} (\sigma_\varepsilon^2)^{-(A_\varepsilon+1)} exp\left(-\frac{B_\varepsilon}{\sigma_\varepsilon^2}\right) \qquad (5.7)$$

Ayrıca σ_u^2 için de aynı şekilde ters gamma dağılımı, önsel olarak kullanılır yani, $\sigma_u^2 \sim IG(A_u, B_u)$ şeklindedir.

- Eğer $A_\varepsilon > 1$ ise rastgele değişkenlerin ortalaması sonlu ve $\frac{B_\varepsilon}{(A_\varepsilon - 1)}$'e eşittir.
- Eğer $A_\varepsilon > 2$ ise varyansı sonludur ve $\frac{B_\varepsilon^2}{(A_\varepsilon-1)^2(A_\varepsilon-2)}$'ye eşittir.

Burada, $A_\varepsilon, B_\varepsilon, A_u, B_u$ önselleri belirten ve istatistikçi tarafından seçilmesi gereken hiperparametrelerdir. Eşlenik bir önsel kullanıldığında sonsal dağılım aynı aileden olurken verilerden eklenen bilgileri yansıtan farklı hiperparametrelere sahip olacaktır. Genel bir önsel dağılım için bu hesaplama açısından çok karmaşıktır ve sonsal zor bir forma sahip olabilir. Eğer eşlenik önseller sonsal hiperparametrelerle ilişkilendirilirse sonsalın hesaplanması daha kolay olmaktadır. Bu parametrelerin önseller için uygun olmasının şartı kesinlikle pozitif olmalarıdır. A_ε ve B_ε sıfır olsaydı $[\sigma_\varepsilon^2]$ 'nun uygun olmayan önseli $\frac{1}{\sigma_\varepsilon^2}$ ile orantılı olup bu da uygun olmayan tekdüze önsele sahip $log(\sigma_\varepsilon)$'ye eşit olmaktadır. Bu nedenle A_ε ve B_ε sıfıra yakın değerler seçilirse (örneğin, bu değer 0.1 alınsın) bilgilendirici olmayan ama uygun olan bir önsel seçilmiş olur. Aynı mantık A_u ve B_u için de geçerlidir. Oluşturulan model, rastgele değişkenlerin her seviyedeki dağılımların önceki seviyelerdeki rastgele değişkenler tarafından belirlendiği bir hiyerarşide düzenlendiği hiyerarşik bir Bayes modelidir. Böylece her bir seviyedeki dağılımlar bir önceki seviyelerdeki rastgele değişkenler tarafından belirlenir. Hiyerarşinin en altında bilinen hiperparametreler bulunur. Bir sonraki seviyede, dağılımları hiperparametreler tarafından belirlenen sabit etki parametreleri ve varyans bileşenleri bulunur. Bunun üzerindeki seviyede, dağılımları varyans bileşenleri tarafından belirlenen rasgele etkiler u ve ε vardır. En üst seviye, y verilerini içerir ve hiyerarşik model tanımlanmış olur.$\propto$ ile gösterilen sabit orantısallık hariç sonsal dağılım aşağıdaki denkleme eşittir (Rupert vd. 2003).

$$[\beta, u, \sigma_u^2, \sigma_\varepsilon^2 \mid y] \propto [y \mid \beta, u, \sigma_\varepsilon^2][u \mid \sigma_u^2][\sigma_u^2][\beta][\sigma_\varepsilon^2] \tag{5.8}$$

<u>Sonsal Dağılım</u>: (5.8)'deki denklemde (β, u)'ya bağlı olan kısım yalnız bırakılırsa verilen $(\sigma_u^2, \sigma_\varepsilon^2)$'nun **(**$\beta, u$)'nun koşullu sonsalı olduğu görülür ve aşağıda verilen denklemle orantılı olduğu söylenir.

$$exp\left\{\frac{1}{2\sigma_\varepsilon^2}(\|y - X\beta - Zu\|)^2 + \frac{\sigma_\varepsilon^2}{\sigma_u^2}\|u\|^2\right\} \tag{5.9}$$

Eşitlikteki parantez içindeki terim, (β, u)'nun negatif olmayan karesel bir fonksiyonudur ve (5.9)'daki bu denklem çok değişkenli normal olasılık yoğunluk fonksiyonuyla orantılıdır. Karesel tamamlama metoduyla formül şu şekilde yazılabilir.

$$[\beta, u \mid \sigma_\varepsilon^2, \sigma_u^2, y] \sim N\left\{\left(C^T C + \frac{\sigma_\varepsilon^2}{\sigma_u^2} D\right)^{-1} C^T y, \sigma_\varepsilon^2 \left(C^T C + \frac{\sigma_\varepsilon^2}{\sigma_u^2} D\right)^{-1}\right\} \qquad (5.10)$$

Formülde C= [X, Z] ifadesine eşittir ve X, Z rastgele değişkenlerinin olasılık yoğunluğunu ifade eder. ([] ifadesi Markov zincirlerinde olasılık yoğunluğunu temsil eder). D köşegenleri K tane düğüm noktasını gösteren diagonal bir matristir geri kalan p+1 elemanı sıfırdır p+1 ise polinom katsayılarını gösterir. MCMC zincirinin bir parçası olarak, (5.10) eşitliğinde verilen kovaryans matrisi ve ortalama ile, çok değişkenli normal dağılıma göre $(\sigma_\varepsilon^2, \sigma_u^2)$ şimdiki değerlerinden β ve u üretilir. σ_ε^2 için tam koşullu dağılımı aşağıdaki denklemle orantılıdır.

$$(\sigma_\varepsilon^2)^{-(n/2+A_\varepsilon+1)} exp\left\{-\frac{1}{\sigma_\varepsilon^2}\left(\frac{1}{2}\|y - X\beta - Zu\|^2 + \beta_\varepsilon\right)\right\} \qquad (5.11)$$

(5.7) ve (5.11) eşitliklerinin ortak sonucu olarak σ_ε^2 için

$$[\sigma_\varepsilon^2 \mid y, \beta, u, \sigma_u^2] \sim IG\left(A_\varepsilon + \frac{1}{2}n, \beta_\varepsilon + \frac{1}{2}\|y - X\beta - Zu\|^2\right) \qquad (5.12)$$

Ve aynı şekilde σ_u^2 için

$$[\sigma_u^2 \mid y, \beta, u, \sigma_\varepsilon^2] \sim IG\left(A_u + \frac{1}{2}n, \beta_u + \frac{1}{2}\|u\|^2\right) \qquad (5.13)$$

Sonsaldan örnekleme için aşağıdaki üç adım N kez tekrarlanır.

1) Çok değişkenli normal dağılımdan β ve u örneklemi alınır.

$$N\left\{\left(C^T C + \frac{\sigma_\varepsilon^2}{\sigma_u^2} D\right)^{-1} C^T y, \sigma_\varepsilon^2 \left(C^T C + \frac{\sigma_\varepsilon^2}{\sigma_u^2} D\right)^{-1}\right\}$$

2) $IG\left(A_u + \frac{1}{2}K, \beta_u + \frac{1}{2}\|u\|^2\right)$ dağılımından σ_u^2 örneklemi türetilsin

3) $IG\left(A_\varepsilon + \frac{1}{2}n, \beta_\varepsilon + \frac{1}{2}\|y - X\beta - Zu\|^2\right)$ dağılımından σ_ε^2 örneklemi türetilsin

4) 1.adıma geri dönülür ve adımlar tekrarlanır.

(5.10)'daki formülde $\sigma_\varepsilon^2/\sigma_u^2$ ifadesi düzleştirme parametresine karşılık gelmektedir ve λ ile gösterilir. λ, hata ve değişkenlik arasındaki değişim oranını göstermektedir.

Düzleştirme parametresinin düşük değerleri regresyon eğrisinin daha kıvrımlı olmasına neden olurken yüksek değerleri daha az kıvrımlı bir hale gelmesine neden olur. $\lambda \to \infty$ iken regresyon eğrisi düz bir çizgi halini alırken, λ= 0 olduğunda ise cezalandırılmamış regresyon eğrisi tahminine dönüşür. Markov zincirinde cezalandırılmış spline tahmini için başlangıç noktaları $\left(\beta^{(0)}, u^{(0)}\right)$ olarak alınabilir ya da cezalandırılmış spline karma modelinden $\left(\beta^0, u^0, \left(\sigma_\varepsilon^{(0)}\right)^2, \left(\sigma_u^{(0)}\right)^2\right)$ tahmin edilebilir. Ancak zincir hızlı bir şekilde durağan dağılıma yaklaştığı için ayrıca Markov zincirinde yanma periyodu olarak tanımlanan başlangıçtaki yinelemelerin atılması durumuna dayanarak başlangıç değerleri atılabilir bu yüzden başlangıç değerlerinin kesin seçimi önemli değildir. Burada dikkat edilmesi gereken β, u, y verildiğinde σ_ε^2 ve σ_u^2 bağımsız olmasıdır. Bu nedenle algoritmanın 2. ve 3. adımlarının net etkisi gereken β, u, y biliniyorken onun koşullu dağılımından σ_ε^2 ve σ_u^2 örneklemesidir. Ayrıca σ_ε^2 ve σ_u^2 bağımsız olması sebebiyle 2. ve 3. adımlarının sırasının değiştirilmesinin algoritma üzerinde etkisi bulunmamaktadır.

Bayes tahmini parametre vektörünün herhangi bir alt vektörünün kare hatası kaybı altında parametrenin sonsal dağılımının ortalamasıdır. Parametreler hakkındaki belirsizlik sonsalın kovaryans matrisi ile ölçülebilir. Bir parametre kümesinin sonsal ortalamasını ve kovaryansını tahmin etmenin en kolay yolu Markov zincirinde N kez tekrarla bu parametrelerin örnek ortalaması ve kovaryansını almaktır. Daha verimli sonuçlar elde etmek için Rao-Blackwell teknikleri kullanılmaktadır.

5.5 Lineer Karma Modeller

Bu bölümde birden fazla varyans bileşenine sahip doğrusal karma modellerin genişletilmesi anlatılmıştır. Birden fazla varyans bileşenine sahip Gauss doğrusal karma modeli:

$$y = X\beta + Zu + \varepsilon, \qquad Cov\begin{bmatrix} u \\ \varepsilon \end{bmatrix} = \begin{bmatrix} G & 0 \\ 0 & \sigma_\varepsilon^2 I \end{bmatrix}, \tag{5.14}$$

Bu eşitlikte

$$G \equiv blockdiag\sigma_{u\ell}^2 I, \quad 1 \le \ell \le L \quad ve \quad u = [u_1^T, \dots, u_L^T]^T \tag{5.15}$$

$Cov(u_\ell) = \sigma_{u\ell}^2 I$ olacak şekilde u'nun bir bölümüdür. $q_\ell, 1 \le \ell \le L$, olmak üzere u_ℓ'deki tam sayıları göstersin ve $\sigma_{u\ell}^2$'nin önsel dağılımı için $IG(A_{u\ell}, B_{u\ell})$ olarak ele alınsın. Yukarıda verilen denklem (5.14) için uygun bir Gibbs örnekleme şeması aşağıdaki gibidir.

1) Çok değişkenli normal dağılımından (β, u) türetilir.

$$N\{(C^T C + \sigma_\varepsilon^2 B)^{-1} C^T y, \sigma_\varepsilon^2 (C^T C + \sigma_\varepsilon^2 B)^{-1}\},$$
$$C \equiv [X\ Z] \text{ ve } B \equiv blockdiag(0, G^{-1})' dir.$$

2) $1 \le \ell \le L$, $\sigma_{u\ell}^2$, $IG\left(A_{u\ell} + \frac{1}{2} q_\ell, \beta_{u\ell} + \frac{1}{2}\|u_\ell\|^2\right)$'den türetilir.
3) $IG\left(A_\varepsilon + \frac{1}{2} n, \beta_\varepsilon + \frac{1}{2}\|y - X\beta - Zu\|^2\right)$'den σ_ε^2 türetilir.
4) 1. adıma geri dönülür ve adımlar tekrarlanır.

5.6 Genelleştirilmiş lineer karma modeller

Ortak değişkenlere verilen yanıt değişkeninin dağılımının Gauss olmadığını, bunun yerine başka bir üstel ailede bir dağılım örneğin iki kategorili Binom dağılımı olduğu varsayılsın. (β, u) biliniyorken y'nin olasılık yoğunluk fonksiyonu,

$$[y \mid \beta, u] = exp\{y^T(X\beta + Zu) - 1^T b(X\beta + Zu) + 1^T c(y)\} \tag{5.16}$$

$u \sim N(0, G)$ olduğu varsayılmak üzere, (5.15)'deki $\sigma_{u\ell}^2$ $1 \le \ell \le L$ için örnekleme şeması, Gauss dağılım için aynı şekildedir.

$$\left[\sigma_{u\ell}^2 \mid \beta, u, y, \sigma_{\ell'}^2, 1 \le \ell' \le L, \ell' \ne \ell\right] \sim IG\left(A_{u\ell} + \frac{1}{2} q_\ell, \beta_{u\ell} + \frac{1}{2}\|u_\ell\|^2\right)$$

Ayrıca,

$$[\beta, u \mid y, \sigma_{u1}^2, \dots, \sigma_{uL}^2] \propto exp\left\{y^T(X\beta + Zu) - 1^T b(X\beta + Zu) - \frac{1}{2} u^T G^{-1} u\right\} \tag{5.17}$$

Koşullu yoğunluk herhangi bir standart aileye ait olmadığı için buradan örnekleme yapmak kolay değildir. Bu gibi durumlarda, örnekleme şemasını tamamlamak için uyarlamalı reddetme örneklemesi (Gilks ve Wild 1992, Rupert vd. 2003) veya Metropolis–Hastings algoritması (Metropolis vd.1953, Hastings 1970, Rupert vd. 2003) gibi daha kapsamlı algoritmalar gerekir. Burada Metropolis-Hasting algoritmasına yer verilmiştir. Metropolis-Hastings algoritması, hedef dağılımın yoğunluğu yalnızca orantılılık sabitine kadar bilindiğinde, belirli bir durağan dağılıma sahip bir Markov zinciri oluşturmak için kullanılabilir. Örneğin burada, $[\beta, u \mid y, \sigma_{u1}^2, \ldots, \sigma_{uL}^2]$. Metropolis-Hastings algoritması, kişinin uygun bir dağılımdan örneğin, normal dağılımdan örnek almasına ve ardından hedef dağılımdaki olasılığına bağlı olarak bu yeni gözlemi "kabul etmesine" izin vermektedir. Genellikle Metropolis-Hastings algoritması direk kullanılmaz ve bunun yerine Metropolis-Hastings adımları daha karmaşık bir MCMC'ye yerleştirilir.

Zincirdeki (β, u)'nun mevcut değeri β_c, u_c olsun. (β, u)'nun deneme değerleri olan yeni β_t, u_t oluşturulur, ki bu değerler β_c, u_c ortalamasıyla normal dağılımlıdır. r oranı (5.17)'deki denkleme göre hesaplanan (β_t, u_t)'nin aynı şekilde bulunan (β_c, u_c)'ye oranıdır ve $\min(1, r)$ olasılığı ile (β_t, u_t) kabul edilir, yani, (β_c, u_c) yerine (β_t, u_t) değeri dikkate alınır. Aksi takdirde (β, u)'nun mevcut değeri korunur. Deneme değerlerinin (β_t, u_t) kovaryansı için, (β, u)'nun cezalandırılmış yarı-olasılık tahmininin yaklaşık koşullu kovaryans matrisinin bir çarpanı kullanılabilir. τ skaler çarpan olmak üzere τ yeterince küçük olursa deneme değerleri oldukça yüksek olasılıkla kabul edilir. $\tau \rightarrow 0$'a yaklaşıyorsa (β_t, u_t)'de (β_c, u_c)'ye yaklaşır ve kabul olasılığı bütüne (unity) yakınsar. τ küçükse, deneme değerleri (β_t, u_t), (β_c, u_c)'ye yakın olduğundan, neredeyse tüm deneme değerleri kabul edilse bile zincir yavaş hareket edecektir. Denemeler sonucunda τ değerinin $0.2 \leq \tau \leq 0.4$ aralığında oldukça iyi sonuçlar verdiği bulunmuştur. Son olarak Bayesci genelleştirilmiş doğrusal karma model için geliştirilen algoritma adımları aşağıda gibidir (Rupert vd. 2003).

1) Skaler çarpan ayar parametresi τ'nun değeri belirlenir. $0.2 \leq \tau \leq 0.4$ aralığı önerilmektedir.
2) Cezalı yarı-olasılık yöntemi aracılığıyla $\beta_c, u_c, \sigma_{u1}^2, \ldots, \sigma_{uL}^2$'nin ilk değerleri tahmin edilir.

$$\sum (C^T WC + B)^{-1} C^T WC (C^T WC + B)^{-1}, \;\; C = [X, Z] \;\; ve \; W = dia\{b''(X\beta_c + Zu_c)\}$$

3) $\begin{bmatrix} \beta_t \\ u_t \end{bmatrix} \sim N\left(\begin{bmatrix} \beta_c \\ u_c \end{bmatrix}, \tau\sum\right)$ olmak üzere r oranı aşağıdaki şekilde hesaplanır:

$$r = \frac{exp\left\{y^T(X\beta_t + Zu_t) - 1^T b(X\beta_t + Zu_t) - \frac{1}{2}u_t^T G^{-1} u_t\right\}}{exp\left\{y^T(X\beta_c + Zu_c) - 1^T b(X\beta_c + Zu_c) - \frac{1}{2}u_c^T G^{-1} u_c\right\}}$$

Sonrasında 0 ile 1 arasında rastgele tekdüze değişken u örneklenir.

a) Eğer u ≤ r ise (β_c, u_c) yerine (β_t, u_t) kullanılır.

b) Eğer u >r ise (β_c, u_c) değiştirilmeden bırakılır.

4) $1 \le \ell \le L$, $IG\left(A_{u\ell} + \frac{1}{2}q_\ell, \beta_{u\ell} + \frac{1}{2}\|u_\ell\|^2\right)$ 'den $\sigma_{u\ell}^2$ üretilir

5) 3.adıma dönülür ve adımlar yinelenir.

5.7 Rao-Blackwell Yöntemi ile Tahmin

(β, u)'nun sonsal tahminini elde etmenin alternatif bir yöntemi, aşağıda verilen formülde koşullu ortalamanın aritmetik ortalamasının hesaplanmasıyla Markov zincirinden tahmin edilmesidir.

$$[\beta, u \mid \sigma_\varepsilon^2, \sigma_u^2, y] \sim N\left\{\left(C^T C + \frac{\sigma_\varepsilon^2}{\sigma_u^2} D\right)^{-1} C^T y, \sigma_\varepsilon^2 \left(C^T C + \frac{\sigma_\varepsilon^2}{\sigma_u^2} D\right)^{-1}\right\} \qquad (5.18)$$

(β, u)'nun koşullu ortalamalarının aritmetik ortalamasının alınmasıyla daha küçük Monte Carlo varyansı elde edilir. N, MCMC algoritmasındaki denemelerin sayısını göstermek üzere, (β, u)'nun sonsal kovaryans matrisinin etkili bir tahmini aşağıdaki adımlarla elde edilir:

1) Öncelikle (5.10)'daki (β, u)'nun N koşullu ortalamasının kovaryansı yani $Cov[E\{(\beta, u) \mid \sigma_u^2, \sigma_\varepsilon^2, y\}]$ tahmin edilir:

2) N tane kovaryans matrisin ortalaması yani $E[\{(\beta, u) \mid \sigma_u^2, \sigma_\varepsilon^2, y\}]$ bulunur.

3) Koşullu ve koşulsuz ortalamalar ve kovaryans matrisleri arasında aşağıdaki eşitlikler geçerlidir:

$$E(y) = E\{E(y|x)\};$$
$$Cov(y) = E\{Cov(y|x)\} + Cov\{E(y|x)\}$$

bu iki eşitliğin toplamının alınmasıyla, $Cov\{(\beta, u)|\, y\}$'nin etkili bir tahmini elde edilmiş olur. $IG(a, b)$ dağılımının ortalaması, $a > 1$ için $b/(a-1)$'dir. 2.adımda σ_u^2'nun koşullu ortalaması, $\left(B_u + \frac{1}{2}\|u\|^2\right)/\left(A_u + \frac{1}{2}K - 1\right)$ şeklindedir. u'nun N kez simüle edilmiş değerlerinin ortalamasının alınması ile σ_u^2 tahmin edilir. Eğer σ_u^2'nun sonsal olasılık yoğunluğunun tahmini isteniyorsa, IG olasılık yoğunluğunun u'nun N deneme değerleri üzerinden ortalaması alınabilir. Aynı yol izlenerek σ_ε^2'nun tahmini ve sonsal olasılık yoğunluğu da bulunabilmektedir.

5.8 Bayes Faktörleri

Bayes faktörlerini kullanarak elde edilen karma model Bayesci cezalı spline semiparametrik regresyon modeli ile katsayıları bilinen Bayesci cezalı spline semiparametrik regresyon modeli arasında seçim yapılması gerektiğinde hipotez testi ile bu iki model karşılaştırılarak seçim yapılabilir.

$$H_0: Y = X\beta^0 + Zu^0 + \varepsilon$$
$$H_1: Y = X\beta + Zu + \varepsilon$$

Burada β^0 ve u^0 bilinen değerlerdir. H_0'ın H_1'e göre Bayes faktörü β_{01} aşağıdaki şekilde hesaplanmaktadır:

$$\beta_{01}(Y) = \frac{m(Y \mid H_0)}{m(Y \mid H_1)}$$

$m(Y \mid H_i), i = 0{,}1$ H_i modeli altında Y'nin marinal olasılık yoğunluk fonksiyonudur. $m(Y \mid H_0)$, Marjinal olasılık yoğunluk fonksiyonu:

$$m(Y \mid H_0) = \int f(Y \mid \beta^0, u^0, \sigma_\varepsilon^2)\pi_0(\sigma_\varepsilon^2)\, d\sigma_\varepsilon^2$$

$$= (2\pi)^{-n/2} \frac{\beta_\varepsilon^{\alpha_\varepsilon}}{\Gamma(\alpha_\varepsilon)} \int (\sigma_\varepsilon^2)^{-n/2} \, exp\left(\frac{\beta_\varepsilon}{\sigma_\varepsilon^2}\right) (\sigma_\varepsilon^2)^{-(\alpha_\varepsilon+1)} exp\left(-\frac{(Y - X\beta^0 - Zu^0)^2}{2\sigma_\varepsilon^2}\right) d\sigma_\varepsilon^2$$

$$= (2\pi)^{-n/2} \frac{\beta_\varepsilon^{\alpha_\varepsilon}}{\Gamma(\alpha_\varepsilon)} \int (\sigma_\varepsilon^2)^{-\left(\frac{n}{2}+\alpha_\varepsilon+1\right)} \, exp\left(\frac{-\beta_\varepsilon + \frac{1}{2}\left(Y - X\beta^0 - Zu^0\right)^2}{2\sigma_\varepsilon^2}\right) d\sigma_\varepsilon^2$$

$$= (2\pi)^{-n/2} \frac{\beta_\varepsilon^{\alpha_\varepsilon}}{\Gamma(\alpha_\varepsilon)} \int (\sigma_\varepsilon^2)^{-\left(\frac{n}{2}+\alpha_\varepsilon+1\right)} \left(\beta_\varepsilon + \frac{1}{2}\left(Y - X\beta^0 - Zu^0\right)^2\right)^{\left(\frac{n}{2}+\sigma_\varepsilon+1\right)}$$

$$\left(\beta_\varepsilon + \frac{1}{2}(Y - X\beta^0 - Zu^0)^2\right)^{\left(\frac{n}{2}+\sigma_\varepsilon+1\right)} \, exp\left(\frac{-\beta_\varepsilon + \frac{1}{2}(Y - X\beta^0 - Zu^0)^2}{2\sigma_\varepsilon^2}\right) d\sigma_\varepsilon^2$$

$$= (2\pi)^{-n/2} \frac{\beta_\varepsilon^{\alpha_\varepsilon}}{\Gamma(\alpha_\varepsilon)} \int \left(\frac{\beta_\varepsilon + \frac{1}{2}(Y - X\beta^0 - Zu^0)^2}{(\sigma_\varepsilon^2)^{\left(\frac{n}{2}+\alpha_\varepsilon+1\right)}}\right)^{\left(\frac{n}{2}+\alpha_\varepsilon+1\right)}$$

$$exp\left(\frac{-\beta_\varepsilon + \frac{1}{2}(Y - X\beta^0 - Zu^0)^2}{2\sigma_\varepsilon^2}\right) \left(\beta_\varepsilon + \frac{1}{2}(Y - X\beta^0 - Zu^0)^2\right)^{-\left(\frac{n}{2}+\alpha_\varepsilon+1\right)} d\sigma_\varepsilon^2$$

$$= (2\pi)^{-n/2} \frac{\beta_\varepsilon^{\alpha_\varepsilon}}{\Gamma(\alpha_\varepsilon)} \int \left(\frac{\beta_\varepsilon + \frac{1}{2}(Y - X\beta^0 - Zu^0)^2}{\sigma_\varepsilon^2}\right)^{\left(\frac{n}{2}+\alpha_\varepsilon+2\right)-1}$$

$$exp\left(\frac{-\beta_\varepsilon + \frac{1}{2}(Y - X\beta^0 - Zu^0)^2}{2\sigma_\varepsilon^2}\right) \left(\beta_\varepsilon + \frac{1}{2}(Y - X\beta^0 - Zu^0)^2\right)^{-\left(\frac{n}{2}+\alpha_\varepsilon+1\right)} d\sigma_\varepsilon^2$$

$$m(Y \mid H_0) = (2\pi)^{-n/2} \frac{\beta_\varepsilon^{\alpha_\varepsilon}}{\Gamma(\alpha_\varepsilon)} \, \Gamma\left(\frac{n}{2} + \alpha_\varepsilon + 1\right) \left(\beta_\varepsilon + \frac{1}{2}(Y - X\beta^0 - Zu^0)^2\right)^{-\left(\frac{n}{2}+\alpha_\varepsilon+1\right)}$$

$m(Y \mid H_1)$, marjinal olasılık yoğunluk fonksiyonu:

$$m(Y \mid H_1, \sigma_\varepsilon^2, \delta, \gamma) = (2\pi\sigma_\varepsilon^2)^{-n/2} \left(\prod_{i=1}^{p+q+1} (1 + \delta d_i)\right)^{-1/2} \left(\prod_{i=p+q+2}^{n} (1 + \gamma d_i)\right)^{-1/2}$$

$$exp\left\{-\frac{1}{2\sigma_\varepsilon^2}\left(\sum_{i=1}^{p+q+1} \frac{s_i^2}{1 + \delta d_i} + \sum_{i=p+q+2}^{n} \frac{s_i^2}{1 + \delta d_i}\right)\right\}$$

$$m(Y \mid H_1) = \int m(Y \mid H_1, \sigma_\varepsilon^2, \delta, \gamma) \pi_0(\sigma_\varepsilon^2, \delta, \gamma) d\sigma_\varepsilon^2 d\delta d\gamma$$

$$m(Y \mid H_1) = \frac{\beta_\varepsilon^{\alpha_\varepsilon}}{\Gamma(\alpha_\varepsilon)}(2\pi)^{-n/2}\,\Gamma\left(\frac{n}{2}+\alpha_\varepsilon\right)$$

$$\int\left(\prod_{i=1}^{p+q+1}(1+\delta d_i)\right)^{-1/2}\left(\prod_{i=p+q+2}^{n}(1+\gamma d_i)\right)^{-1/2}\pi_0(\delta,\gamma)$$

$$\left(\beta_\varepsilon+\frac{1}{2}\left(\sum_{i=1}^{p+q+1}\frac{s_i^2}{1+\delta d_i}+\sum_{i=p+q+2}^{n}\frac{s_i^2}{1+\gamma d_i}\right)\right)d\delta d\gamma$$

5.9 Uygulama

Simülasyon çalışması R programında yer alan "mtcars" datası için yapılmıştır. 32 farklı araba markasına ait veri setinde galon başına mil ile bir dizi açıklayıcı değişken arasındaki ilişki için modeller belirlenmek isteniyor. Bağımlı değişken, galon başına mil mpg olmak üzere açıklayıcı değişkenler disp: motorun üretebileceği toplam güç miktarını ölçen araç, hp: brüt beygir gücü qsec: çeyrek mil çalışma süresi, wt: aracın ağırlığı arasındaki ilişkiyi anlamak için sonsal olasılığı en yüksek olan modeller belirlenmiştir. Örneklem hacmi n=32, yineleme sayıları sırasıyla 50.000, 40.000. 30.000, 20.000, 10.000 olarak alınmış ve sonsal olasılıklara göre modeller karşılaştırılmıştır. Zincirde doğru bir yakınsama elde etmek için yineleme sayılarının yarısı yanma periyodu olarak belirlenmiştir. Seyreltme oranı 2 olarak alınmış yani elde edilen sonsal örneklemlerden 2 tanesinden 1'i alınarak sonsal olasılıklar elde edilmiştir.

Çizelge 5. 2 50.000 Yineleme İçin Sonuçlar

	disp	hp	wt	qsec	Sonsal Olasılık (%)	Kümülatif (%)
Model-1	0	1	1	0	42.91	42.91
Model-2	0	0	1	1	42.41	85.32
Model-3	0	0	1	0	5.68	91.00
Model-4	0	1	1	1	3.14	94.14
Model-5	1	1	1	0	2.00	96.14
Model-6	1	0	1	1	1.55	97.69
Model-7	1	0	1	0	1.32	99.01

Tabloda sonsal olasılığı %1 ve üzeri sonuç veren 7 model gösterilmiştir. Birinci model brüt beygir gücü ile ağırlık değişkenlerini içermektedir. Modeller içerisinde %42.91 ile 1. model en yüksek sonsal olasılığa sahip modeldir. İkinci model ağırlık ve çeyrek mil çalışma süresini içeren modeldir ve %42.41 yüzdesine sahiptir. Üçüncü model en net bilgiyi veren modeldir çünkü sadece ağırlık değişkenini içermektedir fakat %5.68 ile düşük bir yüzdeye sahiptir. Yineleme sayısı arttıkça önerilen model sayısı da artmaktadır ama özellikle üçüncü modelden sonra sonsal olasılık yüzdelerinde ciddi bir düşüş yaşanmaktadır.

Çizelge 5. 3 40.000 Yineleme İçin Sonuçlar

	disp	hp	wt	qsec	Sonsal Olasılık (%)	Kümülatif (%)
Model-1	0	1	1	0	42.60	42.60
Model-2	0	0	1	1	42.48	85.08
Model-3	0	0	1	0	6.12	91.20
Model-4	0	1	1	1	3.11	94.31
Model-5	1	1	1	0	1.81	96.12
Model-6	1	0	1	1	1.54	97.66
Model-7	1	0	1	0	1.29	98.95

40.000 yineleme ise ilk iki model için sonsal olasılık yüzdeleri birbirine çok yakındır. Sonsal modellerin %85.09'unu ilk iki model içermektedir. Ağırlığı tek başına içeren üçüncü modelin ise sonsal olasılığı %6.12'ye çıkmıştır.

Çizelge 5. 4 30.000 Yineleme İçin Sonuçlar

	disp	hp	wt	qsec	Sonsal Olasılık (%)	Kümülatif (%)
Model-1	0	1	1	0	42.16	42.16
Model-2	0	0	1	1	41.93	84.09
Model-3	0	0	1	0	9.12	93.21
Model-4	0	1	1	1	2.23	95.44
Model-5	1	1	1	0	1.25	96.69
Model-6	1	0	1	1	1.09	97.79
Model-7	1	0	1	0	1.01	98.80
Model-8	1	0	1	0	1.00	99.80

30.000 yineleme ise ilk iki model için sonsal olasılıklar yine birbirine çok yakındır. Kümülatif olasılıkları %84.09'a düşmüştür. Ağırlığı tek başına içeren üçüncü modelin ise sonsal olasılığı %9.12'ye çıkmıştır.

Çizelge 5. 5 20.000 Yineleme İçin Sonuçlar

	disp	hp	wt	qsec	Sonsal Olasılık (%)	Kümülatif (%)
Model-1	0	1	1	0	45.06	45.06
Model-2	0	0	1	1	40.10	85.16
Model-3	0	0	1	0	5.22	90.38
Model-4	0	1	1	1	3.70	94.08
Model-5	1	1	1	0	2.12	96.20
Model-6	1	0	1	1	1.56	97.76
Model-7	1	0	1	0	1.22	98.98

20.000 yineleme için hp ve wt açıklayıcı değişkenlerini içeren modelin tüm modeller içindeki sonsal olasılığı %45.06'ya yükselmiştir. Ağırlık wt ve qsec'i içeren model, %40.10'a düşmüştür. Bu iki model sonsal kitlenin %85.16'sını oluşturmaktadır. Yalnızca ağırlığı içeren modelin sonsal olasılığı ise tekrar %5.22'e düşmüştür.

Çizelge 5. 6 10.000 Yineleme İçin Sonuçlar

	disp	hp	wt	qsec	Sonsal Olasılık (%)	Kümülatif (%)
Model-1	0	1	1	0	45.08	45.08
Model-2	0	0	1	1	39.76	84.84
Model-3	0	0	1	0	5.24	90.08
Model-4	0	1	1	1	2.72	92.80
Model-5	1	1	1	0	2.48	95.28
Model-6	1	0	1	1	2.16	97.44
Model-7	1	0	1	0	1.64	99.08

10.000 yineleme sonuçlarına bakıldığında 20.000 yineleme ile birbirine çok yakın sonuçlar verdiği söylenebilir.

Sonuç olarak simülasyon çalışmasıyla farklı açıklayıcı değişkenlerden oluşan modeller belirlenmiştir. En yüksek olasılığa sahip iki model brüt beygir gücü ve ağırlığı içeren birinci modelle ağırlık ve çeyrek mil çalışma süresini içeren ikinci modeldir. Yalnızca ağırlığı içeren model ise üçüncü sıradadır bu modelin sonsal olasılığının yinelemenin 40.000 ve 30.000 olduğu simülasyonlarda yükseldiği görülmüştür.

6. SONUÇ

Çalışmada regresyon analiz yöntemlerinden nonparametrik regresyon ve semiparametrik regresyon anlatıldıktan sonra Bayesci yaklaşımla semiparametrik regresyon açıklanmıştır. Nonparametrik ve semiparametrik regresyona neden ihtiyaç duyulduğundan, düzleştirme ve tahmin yöntemlerinden bahsedilmiştir.

İstatistikte klasik yaklaşıma alternatif bir yöntem olarak sunulan Bayesci yaklaşımda önsel ve sonsal dağılımların nasıl belirlenmesi gerektiği, tahmin yöntemi olarak Markov zinciri ve Monte Carlo simülasyon yöntemlerinden Metropolis ve Metropolis-Hasting algoritması ve Gibbs örneklemenin adımları anlatılmıştır. Bayesci semiparametrik regresyon için önseller ve sonsallar hiyerarşik bir model kullanarak belirlenmiştir. Daha etkili bir yöntem olarak Rao-Blackwell yönteminin adımları gösterilmiştir.

Uygulamada bir simülasyon çalışması yapılmış, farklı yineleme değerleri için regresyon modelleri elde edilmiştir. Sonsal olasılık değeri üzerinden farklı açıklayıcı değişkenlerle kurulabilecek modeller gösterilmiştir. İlk üç modelden sonra yüzdelerde ciddi bir düşüş olduğu görülmüştür. İlk model brüt beygir gücü ve ağırlık değişkenlerini, ikinci model ise ağırlık ve çeyrek mil çalışma süresini içermektedir. Üçüncü model ise en umut verici model olabilir çünkü tek değişken ağırlığı içermektedir ama düşük bir yüzdeye sahiptir. 40.000 ve 30.000 yineleme ile elde edilen sonuçlarda üçüncü modelin yüzdesinin arttığı görülmüştür. Her bir yinelemede ise ilk iki model en yüksek yüzdelere sahiptir.

Sonuç olarak regresyon analizinin varsayımlarının sağlanmadığı durumlarda kullanılan alternatif yöntemler anlatılmıştır. Farklı bir bakış açısıyla istatistiği ele alan Bayesci istatistik tartışılmış, Bayesci yaklaşımla semiparametrik regresyon analizi ele alınmış ve bir simülasyon çalışmasının sonuçları tartışılmıştır.

KAYNAKLAR

Anonymous. Web sitesi: https://bookdown.org/egarpor/PM-UC3M/, Erişim tarihi: 22.05.2022

Anonymous. Web sitesi: https://www.bauer.uh.edu/rsusmel/phd/ec1-27.pdf, Erişim Tarihi: 15.11.2022.

Anonymous. 2017. Web sitesi: http://faculty.washington.edu/yenchic/17Sp_302/R12.pdf, Erişim Tarihi: 01.12.2022

Anonymous. 2018. Web sitesi: https://artowen.su.domains/courses/200/lec06.pdf, Erişim tarihi:15.11.2022

Anonymous. 2018. Web sitesi: https://stat.ethz.ch/R-manual/R-patched/library/stats/html/ksmooth.html, Erişim Tarihi: 15.11.2022

Anonymous. 2018. Web sitesi: https://www.researchgate.net/publication/324859941BNSP_an_R_Package_for_Fitting_Bayesian_Semiparametric_Regression_Models_and_Variable_Selection, Erişim tarihi: 01.11.2021

Anonymous. 2021. Web sitesi: http://users.stat.umn.edu/~helwig/notes/smooth-notes.html#local-regression, Erişim tarihi: 15.02.2022

Anonymous. 2022. Web sitesi: https://cran.r-project.org/web/packages/bayesm/bayesm. pdf, Erişim Tarihi: 10.12.2022

Anonymous. 2022. Web sitesi: https://statswithr.github.io/book/, Erişim Tarihi: 16.03.2022

Anonymous. Web sitesi: https://search.rproject.org/CRAN/refmans/BNSP/html/print.mvrm.html, Erişim tarihi: 10.10.2021.

Anonymous. Web sitesi: https://www.wsl.ch/fileadmin/user_upload/WSL/Services Produkte/Lehre_an_Hochschulen/Smoothing_and_NonparametricRegression/Nonpara-Reg.pdf, Erişim Tarihi: 15.11.2022

Akman N.B. 2018. Parametrik Olmayan Regresyon Analizinde Bootstrap Yöntemi. Yüksek Lisans Tezi, Sivas Cumhuriyet Üniversitesi, Sosyal Bilimler Enstitüsü, Ekonometri Anabilim Dalı, Sivas.

Aydın D. 2005. Semiparametrik Regresyon Modellemede Spline Düzeltme Yaklaşımı ile Tahmin ve Çıkarsamalar. Doktora Tezi, Anadolu Üniversitesi, Fen Bilimleri Enstitüsü, İstatistik Anabilim Dalı, Eskişehir.

Bertoli W., Oliveira R. P., Achcar J. A. 2022. A New Semiparametric Regression Frameork For Analyzing Non-Linear Data. Analytics 1, s.15-26.

Çevik M. 2009. Doğrusal Olmayan Bayesçi Regresyon ve Yüksek Frekanslı Ses Sistemlerinde Bir Uygulama. Yüksek Lisans Tezi. Yıldız Teknik Üniversitesi, Fen Bilimleri Enstitüsü, İstatistik Anabilim Dalı, İstanbul.

D. Chan, R. Kohn, D. Nott, and C. Kirby. 2006. Locally adaptive semiparametric estimation of the mean and variance functions in regression models. Journal of Computational and Graphical Statistics, 15(4), 915–936. Demir S., Toktamış Ö. 2010. On the Nadaraya-Watson kernel regression estimators. Hacettepe Journal of Mathematics and Statistics, 39 (3), s.429 – 437.

Ekici O. 2005. Bayesyen Regresyon ve WinBUGS ile Bir Uygulama. Yüksek Lisans Tezi, İstanbul Üniversitesi, Sosyal Bilimler Enstitüsü, Ekonometri Anabilim Dalı, İstanbul.

Ekici O. 2009. İstatistikte Bayesyen ve Klasik Yaklaşımın Kavramsal Farklılıkları. Balıkesir Üniversitesi Sosyal Bilimler Enstitüsü Dergisi, 12(21) s.89-101

Fahrmeir L., Lang S. 2000. Bayesian Semiparametric Regression Analysis of Multicategorical Time-Space Data. Sonderforschunhbereich 386, 202.

Gasım N.2013. Bayesyen Model ile Doğrusal Regresyon Modellerinin Karşılaştırılması Üzerine Bir Uygulama. Yüksek Lisans Tezi. Dokuz Eylül Üniversitesi, Sosyal Bilimleri Enstitüsü, Ekonometri Anabilim Dalı, İzmir.

Gelfand A.E., Kottas A. 2003. Bayesian semiparametric Regression for Median Residual Life. 30, 651-665.

Genç A., Karadavut Ufuk ve Palta Çetin. 2010. Lineer Olmayan Bayesci Regresyon ve Tarım Alanında Uygulama, Tubav Bilim Dergisi, C.3.S.3, s.250-258.

Goztepe K. (2018). R kullanımı ve Uygulama Alanları.

Güner A. 2014. Bayesci Yaklaşımda Eşlenik Aileleri Önseli ile Jeffreys Önselinin Karşılaştırılması. Yüksek Lisans Tezi. Eskişehir Osmangazi Üniversitesi, Fen Bilimleri Enstitüsü, İstatistik Anabilim Dalı, Eskişehir.

Hardle W., Müller M., Sperlich S., Werwatz A. 2004. Nonparametric and Semiparametric Models, New York.

Hepsağ A. 2007. Parametrik Olmayan Regresyon Tahmin Yöntemleri: İMKB'de İşlem Gören Hisse Senetlerine Ait Piyasa Riskinin Tahmini Üzerine Bir Uygulama. Yüksek Lisans Tezi, İstanbul Üniversitesi, Sosyal Bilimler Enstitüsü, Ekonometri Anabilim Dalı, İstanbul.

Hinne M., Gronau Q.F., Bergh D., Wagenmakers E. 2020. A Conceptual Introduction to Bayesian Model Averaging. Advances in Methods and Practices in Psychological Science, 3(2), 200-215.

Kneib T., Lang S., Brezger A. (2004). Bayesian semiparametric regression based on mixed model methodology: A tutorial, Department of Statistics, University of Munich.

Lenk P.J., 1999. Bayesian inference for semiparametric regression using a Fourier representation, Bölüm(4), s.863-879.

Machkouri M., Reding L. Fan X. 2020. On the Nadaraya–Watson kernel regression estimator for irregularly spaced spatial data. Journal of Statistical Planning and Inference (205), s.92–114

Mohasisen A., Abdulhussain A. 2013. A note on Bayes Semiparametric Regression. Mathematical Theory and Modeling 3(12).

Mouel A., Mohaisen A. 2017. Study on Bayes Semiparametric Regression. Journal of Advances in Applied Mathematics 2(4), 197-207.

Ong V.M.H, Mensah D.K., Nott D.J. 2017. A Variational Bayes approach to a semiparametric regression using Gaussian process priors. Electronic ournal of Statistics (11) 4258-4296.

Öngen Bilir K.B. 2016. Bayesyen Markov Zinciri Monte Carlo Simülasyonu. Doktora Tezi. Uludağ Üniversitesi, Sosyal Bilimleri Enstitüsü, Ekonometri Anabilim Dalı, İstatistik Anabilim Dalı, Bursa.

Öztürk O. 2011. Bayesian semiparametric models for nonignorable missing data mechanisms in logistic regression, Middle East Technical University, Master of Science in Statistics Department, Ankara.

Pelenis J. 2012. Bayesian Semiparametric Regression, Vienna.

Ruppert, D., Wand, M.P. ve Carroll, R.J. 2003. Semiparametric Regression, New York.

Smith M., Mathur S., Kohn R. 2000. Bayesian Semiparametric Regression: An Exposition and Application to Print Advertising Data. Journal of Business Research 49(3), 229-244.

Tabakan G. 2009. Yarı Parametrik Regresyonda Tahmin Metodları. Doktora Tezi, Çukurova Üniversitesi, Fen Bilimleri Enstitüsü, İstatistik Anabilim Dalı, Adana.

Tezcan N. 2009. Tahmine Parametrik Olmayana Regresyon Yöntemiyle Yaklaşım. Doktora Tezi, İstanbul Üniversitesi, Sosyal Bilimler Enstitüsü, Sayısal Yöntemler Anabilim Dalı, İstanbul.

Thorson J.T., Branch T.A. 2016. Faster estimation of Bayesian models in ecology using Hamiltonian Monte Carlo. Methods in ecology and evolution. 8(3) s.339-348.

Turanlı M., Bağdatlı S. 2011. Semiparametrik Regresyon. Öneri, 9(35), 207-213.

Türkan S. 2012. Yarı Parametrik Regresyon Modelinde Etkili Gözlem Analizi. Doktora Tezi, Hacettepe Üniversitesi, Fen Bilimleri Enstitüsü, İstatistik Anabilim Dalı, Ankara.

Türkan S., Toktamış Ö. 2013. Cezalandırılmış Eğrisel Çizgi Regresyonunda Karışık Doğrusal Model Yaklaşımı. Nevşehir Bilim ve Teknoloji Dergisi, 34-40.

Wakefield, J., 2013. Bayesian and Frequentist Regression Methods, New York.

Wang C., Lin X., 2022. Bayesian Semiparametric Regression Analysis of Multivariate Panel Count Data. Stats, 5, 477–493.

Wasserman, L. (2006). All of Nonparametric Statistics. Springer.