T.C.
YÜZÜNCÜ YIL ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ
İSTATİSTİK ANABİLİM DALI

İKİ DURUMLU (BINARY) LOJİSTİK REGRESYON MODELİNE BAYESCİ BİR YAKLAŞIM: OECD ÖRNEĞİ

YÜKSEK LİSANS TEZİ

HAZIRLAYAN: Asuman YILMAZ
DANIŞMAN: Doç. Dr. H. Eray ÇELİK

VAN-2016

T.C.
YÜZÜNCÜ YIL ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ
İSTATİSTİK ANABİLİM DALI

İKİ DURUMLU (BINARY) LOJİSTİK REGRESYON MODELİNE BAYESCİ BİR YAKLAŞIM: OECD ÖRNEĞİ

YÜKSEK LİSANS TEZİ

HAZIRLAYAN: Asuman YILMAZ

VAN-2016

KABUL VE ONAY SAYFASI

İstatistik Anabilim Dalı'nda Doç. Dr. H. Eray ÇELİK danışmanlığında, Asuman Yılmaz tarafından sunulan "**İKİ DURUMLU (BINARY) LOJİSTİK REGRESYON MODELİNE BAYESCİ BİR YAKLAŞIM:OECD ÖRNEĞİ**" isimli bu çalışma Lisansüstü Eğitim ve Öğretim Yönetmeliği'nin ilgili hükümleri gereğince 26/05/2016 tarihinde aşağıdaki jüri tarafından oy birliği / oy çokluğu ile başarılı bulunmuş ve yüksek lisans tezi olarak kabul edilmiştir.

Başkan: Doç. Dr. H. Eray Çelik İmza:

Üye : Doç. Dr. Sinan Saraçlı İmza:

Üye : Yrd. Doç. Dr. Hatice TAŞKESEN İmza:

Fen Bilimleri Enstitüsü Yönetim Kurulu'nun ……../……../2016 tarih ve ……………………………. sayılı kararı ile onaylanmıştır.

Prof. Dr. Suat ŞENSOY
Enstitü Müdürü

TEZ BİLDİRİMİ

Tez içindeki bütün bilgilerin etik davranış ve akademik kurallar çerçevesinde elde edilerek sunulduğunu, ayrıca tez yazım kurallarına uygun olarak hazırlanan bu çalışmada bana ait olmayan her türlü ifade ve bilginin kaynağına eksiksiz atıf yapıldığını bildiririm.

Asuman YILMAZ

ÖZET

İKİ DURUMLU (BINARY) LOJİSTİK REGRESYON MODELİNE BAYESCİ BİR YAKLAŞIM: OECD ÖRNEĞİ

YILMAZ, Asuman
Yüksek Lisans Tezi, İstatistik Anabilim Dalı
Tez Danışmanı: Doç. Dr. H. Eray ÇELİK
Mayıs 2016, 105 sayfa

Bu çalışmanın amacı Bayesci yaklaşımın temel özelliklerini açıklamak lojistik regresyon analizini Bayesci ilkelere göre uygulamak, klasik ve Bayesci analiz tekniklerinin birbirlerine karşı zayıf ve güçlü yönlerinden, Bayesci istatistiğin en temel özelliği olan önsel ve sonsal dağılıma değinmektir. Aynı zamanda karmaşık istatistiksel analizlerin hesaplanmasını kolaylaştıran MCMC algoritmalarında Markov zincirleri Gibss örneklemesi gibi bazı simülasyon tekniklerine değinmektir.

Çalışmanın uygulam bölümünde OECD verilerinden bir lojistik regresyon modeli kurulmuştur. Bu model klasik ve bayesci regresyona göre analiz edilmiş, çıkan sonuçları karşılaştırılarak etkili bir analiz oluşturulmaya çalışılmıştır.

Anahtar kelimeler: Bayesci lojistik regresyon, Gibss örneklemesi, Lojistik regresyon, Markov zincirleri, MCMC, Önsel dağılım, Sonsal dağılım

ABSTRACT

A BAYESIAN APPROXIMATION TO BINARY LOGISTIC REGRESSION ANALYSIS: AN OECD SAMPLE

YILMAZ, Asuman
M. Sc., Statistics
Supervisor: Assoc. Prof. Dr. H. Eray ÇELİK
May 2016, 105 pages

The purpose of this study is to compare the Bayesian and the classical logit models. Weak and powerful aspects of this two statistical analysis tecniques are discussed. Moreover, prior and posterior distributions, which are the main features of the Bayesian statistics, and some simulation technics such as MCMC that simplifies the calculation of complex statistical analysis are mentioned.

In the application part of the study a logit equation was established based on OECD. The model was analyzed according to classical and Bayesian logit models, and the results were compared to make an effective analysis.

Key Words: Bayesian logit models, Gibss sampling, Logit models, Markov chain, MCMC, Prior distribution, Posterior distribution

ÖNSÖZ

Bu tez çalışmasında, her türlü ilgi ve yardımlarını esirgemeyen danışmanım Sayın Doç.Dr H.Eray Çelik'e ve tez jürimde bulunarak tezimin daha iyi bir duruma gelmesini sağlayan Sayın Doç. Dr. Sinan Saraçlı ve Sayın Yrd. Doç. Dr. Hatice TAŞKESEN'e teşekkürlerimi sunarım. Ayrıca tez yazma sürecinde yardımları için Sayın Yrd. Doç. Dr. Mahmut KARA' ya, Yüzüncü Yıl Üniversitesi İstatistik Bölümü öğretim üyelerine ve çalışma arkadaşlarıma teşekkür ederim. Eğitimimin her aşamasında maddi ve manevi desteklerini esirgemeyen aileme ve sevgili eşim Selçuk Yılmaz'a teşekkürü bir borç bilirim.

Hayatımda olduğu için her an şükrettiğim canım kızım Zeynep Özüm'e ...

Haziran, 2016

Asuman YILMAZ

İÇİNDEKİLER

Sayfa

Sayfa

ÇİZELGELER LİSTESİ

Çizelge **Sayfa**

ŞEKİLLER LİSTESİ

Şekil **Sayfa**

SİMGELER VE KISALTMALAR

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

Simgeler	**Açıklama**
ith	İthalat oranı
yaşmem	Yaşam memnuniyeti
sağlık	GSYİH'den sağlık harcamalarına ayrılan oran
Bin	Binom dağılımı
(β_j)	Bağımsız değişkenler
$S(\beta_j)$	Bağımsız değişkenlerin standart hatası
ST	Skor testi
G	Gözlenen değer
B	Beklenen değer
$p(\theta)$	Önsel dağılım

Kısaltmalar	**Açıklama**
GSYİH	Gayri Safi Yurtiçi Hasıla
MCMC	Markov zincirleri Monte Carlo
MC	Markov zincirleri
EÇO	En çok olabilirlik yöntemi
EKK	En küçük kareler yöntemi
OYF	Olasılık yoğunluk fonksiyonu

EKLER DİZİNİ

Ek **sayfa**

1. GİRİŞ

Temeli Bayes teoremine dayanan Bayesci istatistik son yıllarda tıp, hukuk, sosyoloji, çevre bilimi ve ekonometri gibi birbirinden farklı pek çok alanda klasik yaklaşıma bir alternatif olarak sıklıkla kullanılmaktadır (Congdon, 2003; Gelman ve ark., 2004).

Bilgisayar teknolojisi ve hesaplama tekniklerinin hızla gelişmesi üzerine büyük veri setleri ve çok karmaşık modeller için Bayesci çıkarımlar daha fazla kullanılmaya başlanmıştır (Gelman ve ark., 2004).

Bayesci istatistikte belirsiz parametreler, rassal değişken olarak düşünüldüğünden klasik istatistik teorisinden farklıdır (Ntzoufras, 2009). Bayesci istatistik, herhangi bir analiz yapılmadan önce araştırmacının mevcut bilgisini, deneyimini ve tecrübesini açıklamak için Bayesci yaklaşımın en önemli özelliklerinden biri olan önsel dağılımları kullanır (Ntzoufras, 2009; Ghosh ve ark., 2006). Daha önceden yapılan kuramsal çalışmalardan elde edilen sonsal dağılımın gelecekte herhangi bir çalışmada önsel dağılım olabilirliği Bayesci yaklaşımın subjektifliğinin en açık örneği olarak tanımlanmaktadır. Bayesci istatistikte hem önsel dağılım hem de sonsal dağılım birer sabit değil parametreye göre birer değişkendir (Judge, 1984; O’Hagan, 2003). Bayesci yaklaşımda parametreye ait önsel dağılım ile örnek bilgisi veya olabilirlik fonksiyonun birleştirilmesinden sonsal dağılım elde edilir. Parametreye ait tüm çıkarımlar bu sonsal dağılım kullanılarak oluşturulur (O’Hagan, 2003; Congdon, 2005; Zellner, 1976).

Bağımlı değişkenlerin kategorik olduğu durumlarda model parametrelerini tahmin etmek için klasik regresyon modelleri yerine genelleştirilmiş doğrusal modeller kullanılmaktadır (Nelder ve McCullagh, 1989; Agresti, 2002; Tektaş ve Günay, 2008). Genelleştirilmiş doğrusal modellerde; bağımsız değişkenlerle bağımlı değişken arasında ilişki kurmak için bağ fonksiyonları kullanılmaktadır. Bu bağ fonksiyonlarının seçimine göre model değişkenlik göstermektedir. Bağ fonksiyonu, lojit olarak tanımlandığında oluşan model, lojit model veya lojistik regresyon modeli olarak adlandırılmaktadır (Agresti ve Hotchock, 2005; Tatlıdil, 2002). Lojit modellerin parametre tahmininde genel olarak en çok olabilirlik yöntemi kullanılır. Fakat en çok olabilirlik yönteminin kullanıldığı durumlarda eğer asimptotluk varsayımı sağlanmaz ise yanlı sonuçlar elde edilmektedir. Bu durumda Bayesci lojit modellerde asimptotluk varsayımı aranmadığı için kullanılabilinir (Collet, 1991; Tektaş ve Günay, 2008).

Bu tez çalışmasının amacı Bayesci yaklaşımın temel özelliklerini açıklamak, iki durumlu lojistik regresyon analizini Bayesci esaslara göre uygulamak ve Bayesci iki durumlu lojistik regresyon analizi ile klasik lojistik regresyon analizinin birbirlerine benzer ve birbirlerinden farklı yönlerini göstermektir.

Bu tez çalışmasında Bölüm-2'de çalışmaya ait kaynak bildirişleri verilmiştir. Bölüm-3'te Bayesci çıkarımın temelleri ve bazı istatistiksel kavramlardan, Bayes Teoreminden, Bayesci çıkarımlarda belirsizlik ve olasılıktan, Bayesci çıkarımın temellerinden, Bayes yaklaşımının zorlukları ve üstünlüklerinden, Bayesci yaklaşımın en belirgin özelliği olan önsel dağılımdan, olabilirlik fonksiyonundan, önsel dağılım ile olabilirlik fonksiyonunu birleştirilmesinden elde edilen sonsal dağılımdan Bayesci istatistikte hesaplamaları kolaylaştıran Markov zincirlerinden, Monte Carlo iterasyonlarından ve Gibss örneklemesinden, Markov zincirlerinin yakınsaklık boyutlarından, Markov zincirine uygulanan yakınsaklık testlerinden, burn-in ve thinning kavramından bahsedilmiştir. Bölüm-4'de lojistik regresyon ve Bayesci iki durumlu lojistik regresyon analizine değinilmiştir. Bayesci lojistik regresyon analizinde bir veri genişletme metodu olan Albert ve Chib (1993), metodundan Bayesci lojistik regresyonu iki farklı yaklaşımla çözmeyi amaçlayan Helmes ve Hold metodu ile Groenewald ve Mokgatlhe yaklaşımlarından bahsedilmiştir. Bölüm -5'te tez çalışmasında kullanılacak olan materyal ve yöntem aktarılmıştır. Bölüm-6'da bazı OECD verileri derlenmiştir. Derlenen bu veriler Bayesci iki durumlu lojistik ve klasik iki durumlu lojistik regresyon yöntemleriyle incelenmiştir.

Tezin sonuç bölümünde ise analizlerden elde edilen sonuçlara bağlı olarak çıkarımlara yer verilmiştir. Tezde regresyon denklemlerinin analizinde R (3.0.3) programından yararlanılmıştır.

2. LİTERATÜR TARAMASI

Bayes Teoremi, 18. yüzyılın sonlarında Thomas Bayes tarafından ortaya atılmıştır (Hoff, 2009). Thomas Bayes, Binomial dağılım parametrelerinin sonuçları üzerine çalışmıştır (Stigler, 1983; Maryann, 1993). Laplace, 1774 yılında Bayes'ten bağımsız bir şekilde Binomial parametre tahminleri üzerine bazı yöntemler geliştirmiştir. Her iki bilim adamı da Binomial parametreler için Uniform önsel dağılımları kullanmışlardır (Marryann, 1993; Agresti ve Hitchock, 2005). Stigler (1986) ve Dale (1999) Bayes ve Laplace'ın çalışmalarını geliştirmişlerdir (Agresti ve Hitchock, 2005).

Bayesci istatistik üzerine diğer bir çalışma matematikçi ve istatistikçi olan Savage (1954) tarafından yapılmıştır. Savage bu çalışmada subjektiflik ve kişisel olasılık teorilerini ileri sürmüştür (Wakafield, 2012).

Bayes teoreminin 200 yıllık seyri boyunca Pierre, Fermat, Blaise ve Pascal gibi bilim adamlarının farklı çalışmalarından dolayı Bayesci istatistikle pek ilgilenmemiş, klasik olasılık popülerliğini sürdürmeye devam etmiştir. Daha sonra Fisher, Jeffreys, Bernardo, Ramsey, De Finetti, Lindley ve Stein gibi bilim adamları sayesinde klasik yaklaşıma bir alternatif olarak Bayesci yaklaşımla ilgilenilmeye başlanmıştır (Box ve Tiao, 1973).

Bayesci istatistik ile elde edilen bazı sonsal (posterior) dağılımların integrallerini hesaplamak hem çok zor hem de çok büyük zaman kaybına neden olmaktadır. Bu sebeplerden dolayı N. Metropolis nümerik problemlerin çözümü için Markov zincirlerini geliştirmiştir (Stigler, 1983). Daha sonra Metropolis ve diğerleri Markov Chain Monte Carlo (MCMC) algoritmalarını geliştirmiştir. MCMC algoritmaları, Bayesci yaklaşımın hareketlenmesine çok büyük katkı sağlamıştır (Stigler, 1983). Valentin Fredorovich (1971), Gibbs örneklemesini tanıtmıştır. Geman (1984), Gibbs örneklemesini daha da geliştirmiş ve yakınsaklık kriterlerini açıklamıştır (Hosmer ve Lemeshow, 2013). Gibss örneklemesi, MCMC gibi simülasyon tekniklerinin ilerlemesine ve modern Bayesci istatistikte karşılaşılan nümerik problemlerin daha kolay çözüme ulaşmasına yardımcı olmuştur (Stigler, 1983).

Fredorovich (1971), Bayesci yazılımların genelleştirilip ve geliştirilmesi üzerine önemli çalışmalar yapmıştır. Yazılımların gelişmesi ile zor ve karmaşık fonksiyonların çözümü kolaylaşmıştır. Hesaplama tekniklerinin artması ve çözümlere daha kolay ulaşılması Bayesci istatistiği popüler hale getirmiştir (Stigler, 1983; Congdon, 2003). Bayesci yaklaşımların durağanlığa ulaşmasında MC iterasyonlarının yakınsaklığı çok önemli bir konudur. İterasyonların durağanlığı ile ilgili net bir test bulunmamaktadır (Yüksel ve ark., 2013). Yakınsaklık tanıları hakkındaki tartışmalar için Cowles ve Carlin'in (1996) ve Brooks'un (1998) önemli çalışmaları vardır (Yüksel ve ark., 2013). O'Hagan (2003), Bernardo ve Smith (1994) Bayesci karar teorileri üzerine bir çalışma yapmıştır. O'Hagan (1994), farklı önsel seçimlerinin hassas sonuçları üzerine bir çalışmıştır (Wakefield, 2012).

Bayesci metotlarla, doğrusal regresyon ve zaman serilerinde kısmi korelasyonların hesaplanması üzerine Box ve Tiao (1973) ve Zelner'in (1961) önemli çalışmaları olmuştur (Congdon, 2003).

Good (1953), Uniform önsel dağılımları kullanarak çeşitli özellikteki hayvanların nüfus oranlarını kategorik verilerle tahmin etmiştir. Lindley (1964), kontenjans tablolarının özet ölçümlerini tahmin etmeye yoğunlaşmıştır. Daha sonra log odds oranı olarak bilinen log olabilirliklerin benzetim tabanlı sonsal dağılımlarını bulmuştur. Fisher, farklı boyutlar üstünde Uniform önsellerin farklı sonuçlara yol açtığını belirterek binomial parametreler için kullanımı değiştirmiştir. Fienberg (2005), Bayesci metotları yeniden ele almıştır (Agresti ve Hotchock, 2005).

Leonard ve Hsu (1994), Lindley ve Gook'un ilgilendiği kategorik veri analizleri için Bayesci yaklaşımı yeniden gözden geçirmiştirler (Congdon, 2005).

Bayesci iki durumlu verilerin analizinde Albert ve Chib (1993), gizil değişkenlerin modele dahil edildiği bir veri genişletme algoritması önermişlerdir (Congdon, 2005).

Zelner ve Rossi (1984), binary lojit regresyon modellerinde parametrelerin küçük sayıları için sonsal dağılımı hesaplamada nümerik integrasyon yöntemini kullanmışlardır. Büyük parametrelerde, çok değişkenli student t dağılımı ve Monte Carlo integrasyonuyla sonsal momentleri hesaplamışlardır (Collet, 1991).

Chen ve Dey (2000), ölçeklendirilmiş karma dağılımlar için bir lojistik regresyon yöntemi önermişlerdir. Diggle, Liang ve Zeger (1994) çok değişkenli kategorik sonuçlar için regresyon modelini kullanmışlardır (Collet, 1991).

Walters (1985), Uniform dağılımları kullanarak Bayesci binomial güven aralıklarını elde etmiştir (Agresti ve Hotchock, 2005).

Gewekw ve Keane (1997), karma (mixed) normal dağılımları ve sonsal odds oranlarının kullanıldığı karma oluşumlarının sayısına göre sonsal sonuçların nasıl çizileceğini göstermişlerdir (Lesage, 1999).

Nelder ve Wedderburn (1972), modellerin özel sınıfları için genelleştirilmiş doğrusal modelleri tanıtmışlardır (MCCullagh ve Nelder, 1989).

Bayesci istatistikte karmaşık çok boyutlu problemlerin çözümünde SAS (özellikle PROC GENMOG ve PROC MCMC kodlarını kullanır), BUGS (Bayesian inference using Gibbs Sampling) veya BUGS'ın yaygın bir versiyonu olan OpenBugs, R ve Matlab gibi yazılım programları önemli bir yere sahiptir (Hosmer ve Lemeshow, 2013).

Son yıllarda Bayesci lojit model hesaplamalarına Groenewald ve Mokgatlhe (2004), "*Bayesian Logistic Regression"* ve Holmes ve Held (2006), "*Bayesian Auxiliary Variable Models For Binary and Multinominal Regression Bayesian Analysis"* adlı çalışmalarıyla iki farklı yaklaşım önermişlerdir. Eldereny ve Rashwan (2012), "*The Comparison Betwen Results Of Application Bayesian And Maximum Likelihood Approaches On Logistic Regression Model For Prostate Cancer Data"*, Gelman ve ark., (2008) "*A Weakly Default Prior Distribution For Logistic And Other Regression Models"*, DuMouchel (2012), "*Multivariate Bayesian Logistic Regression For Analysis Of Clinical Study Safety Issues"* isimli çalışmalarıyla Bayesci lojistik regresyon analizini çok farklı alanlara uygulanabileceğini göstermişlerdir. Ayrıca Albert (2009), "*Bayesian Computation With R"* adlı kitabiyla ve bu kitapla tüm karmaşık ve zor işlemlerin R programında açık ve sade bir dille hesaplanmasını göstermiştir.

Türkiye'de Ekici (2009), "*İstatistikte Bayesyen ve klasik yaklaşımın farklılıkları"*, Tektaş ve Günay (2008), "*Bayesian Approach to Parameter Estimation in Binary Logit Models"*, Cengiz ve ark., (2013), "*Lojistik Regresyon Parametre Tahmininde Bayesci Bir Yaklaşım*," Yüksel ve ark., (2013), "*Doğrusal Regresyonda*

Markov Zincirleri Monte Carlo Yakınsama Kriterlerinin Karşılaştırılması", Olmuş ve Erbaş (2012), "*Bayes Ağlarda Kümeleme Analizini Kullanarak Meme Kanseri Tanısının Modellenmesi"*, Kaya (2012), "*Bayes Faktörü Bayesci Bilgi Ölçütü ve Sapma Bilgi Ölçütü Kullanımıyla Bayesci Model Seçiminin Bir Uygulaması"*, Avcı (2013), "*Marmara Üniversitesi Öğrencilerinin Kredi Kartına Sahip Olmalarını Etkileyen Faktörlerin Bayesci Lojistik Regresyon Yardımıyla İncelenmesi"* isimli çalışmaları yapmışlardır.

Ayrıca, Öztürk (2014), "*Genelleştirilmiş Doğrusal Karma Modellerde Bayesci Yaklaşımın Kullanımı"*, Bülbül (2015*), "E- Ticaret Sistemlerinde Reklam Ürünlerinin Bayes Teoremine Göre Yerleştirilmes"i*, Gayayr (2015), "*Bayesian Estimation For Multinominal Logistic Regression"*, Altındağ (2015) "*Bayesci Doğrusal Olmayan Yapısal Eşitlik Modeli* "adlı çalışmalar son yıllarda yazılan tezlerden bir kaçıdır.

3. TEMEL TANIM VE TEOREMLER

Tezin bu bölümünde Bayesci çıkarım ve bazı istatistiksel kavramlar, Bayesci çıkarımlarda belirsizlik ve olasılık, parametre, nokta tahmini, Bayes Teoremi, Bayesci yaklaşımın zorlukları ve üstünlükleri, önsel bilgi, Markov Zincirleri Monte Carlo İntegrasyonu, Gibbs örneklemesi gibi temel kavramlara kısaca değinilecektir.

3.1. Bayesci Çıkarım ve Bazı İstatistiksel Kavramlar

Bu kısımda Bayes teoreminin temelini oluşturan olasılık kavramına ve bazı olasılık dağılımlarına yer verilmiştir.

3.1.1. Olasılık

Olasılık teorisinin basitçe rassal bir olgu olarak tanımlanabilir. İstatistiksel sonuçlar olasılık teorisine bağlıdır. Olasılık teorisinin, klasik olasılık teorisi ve subjektif olasılık teorisi gibi birbirinden farklı tanımları mevcuttur (Savchuk ve Tsokos, 2011).

Klasik olasılık, deneyin olası sonuçları sonlu ve tüm olası sonuçlar eşit olasılığa sahip olmasıdır (Gut, 2012; Howson ve Urbach, 2006; Yılmaz ve Çelik, 2009). Subjektif olasılık ise, olayların meydana gelip gelmeyeceğinin belli bir güven derecesi ile tahmin edilmesidir. Subjektif olasılıkta klasik olasılıkta olduğu gibi 0 ve 1 arasında değerler almaktadır (Akın, 1996). Subjektif olasılık teorisinin temeli Bayesci istatistiğe dayanmaktadır. Subjektif olasılığı, Bernoulli, önem seviyesi veya inanç derecesi olasılığı olarak tanıtmıştır. Bu durumda subjektif olasılık teorisi önsel bilgi ile örtüşmektedir (Lynch, 2007; Howson ve Urbach, 2006; Casella ve Berger, 1990).

3.1.2. Bazı olasılık dağılımları

Aşağıdaki bölümde bu tez çalışmasının lojistik ve Bayesci lojistik regresyon analizlinde çok sık kullanılan Bernoulli, Binom, Uniform dağılımının karakteristik özellikleri kısaca ele alınmıştır.

3.1.2.1. Bernoulli dağılımı

Bir rassal deney yapıldığında, deneyin sonucu, başarılı- başarısız iyi kötü gibi sadece iki sonuç içeriyorsa bu deneye Bernoulli deneyi denir. Bernoulli dağılımının olasılık yoğunluk fonksiyonu aşağıdaki gibidir (Yılmaz ve Çelik, 2009).

$$p(x) = \begin{cases} p^x(1-p)^{1-x} & x = 0,1 \\ 0 & d.d \end{cases} \tag{3.1}$$

3.1.2.2. Binomial dağılımı

Binomial dağılım, istatistikte sık kullanılan kesikli bir dağılımdır. Bu dağılım p olasılığında n deneme sayısı için x başarı olasılığını sunar (Savchuk ve Tsokos, 2011; Yılmaz ve Çelik, 2009).

$$x \sim Bin(np)$$
$$P(x|n,p) = \binom{n}{x} p^x(1-p)^{n-x} \tag{3.2}$$

şeklinde gösterilir. Binomial dağılım çoğunlukla iki durumlu değişkenleri modellemek için kullanılmaktadır (Lynch, 2007). Binom dağılımı Bernoulli dağılımının genelleştirilmiş halidir (Yılmaz ve Çelik, 2009).

3.1.2.3. Uniform (Düzgün) dağılım

Uniform dağılım hataların yuvarlaştırılmasında ve başta Monte Carlo simülasyon tekniği olmak üzere simülasyon yöntemlerinde sıklıkla kullanılmaktadır (Yılmaz ve Çelik, 2009) .

X rasssal değişkeninin oyf’nu (3.3) eşitliğinde olduğu gibi tanımlanır.

$$f(x) = \begin{cases} \frac{1}{b-a} & a \leq x \leq b \\ 0 & \text{d.d} \end{cases} \tag{3.3}$$

Uniform dağılımın birikimli dağılım fonksiyonu;

$$F(x) = \begin{cases} 0, & x < a \\ \frac{x-a}{x-b}, & a \leq x < b \\ 1 \quad , & \text{x} \geq b \end{cases} \tag{3.4}$$

3.2. Bayesci Çıkarımlarda Belirsizlik ve Olasılık

Bayesci çıkarım istatistikte çok tartışılan konulardan biridir. Bu tartışmaların ana nedeni ise Bayesci çıkarımda belirsizliğin olasılık ile değerlendirilmesidir. Bayesci istatistiğin amacı olasılık ile belirsizliğin tatmin edici bir açıklamasını elde etmektir (Ekici, 2009).

Lidney'e göre; etrafımız belirsizliklerle sarılı ve bu belirsizlikler bizim hayatımızda çok büyük rol oynamaktadırlar. Bayesci paradigma, olasılık sayesinde araştırmacının takdir özelliği ile çevredeki belirsizlikleri anlamayı, manipüle etmeyi, kontrol etmeyi sağlar. Bayesci yaklaşıma göre her bilinmeyen nicelik olasılık olarak tanımlanabilir (Lindley, 1983; Ekici, 2009).

3.3. Parametre

Klasik yaklaşımda parametre sabit bir değer olmasına karşın Bayesci yaklaşımda parametre olasılık dağılımı ile ifade edilen bir rassal değişkendir. Bu durumda parametre tahmincisi olarak önsel dağılım belirlenir. Belirlenen bu önsel dağılım ile veriye ait olabilirlik fonksiyonun birleştirilmesinden sonsal dağılım elde edilir. Bayesci yaklaşımda parametreye ait tüm çıkarsamalar bu sonsal dağılımla yapılır (Casella ve Berger, 1990; Ekici, 2009).

3.4. Nokta Tahmini

Klasik yaklaşımda**,** nokta tahmininde $\hat{\theta}$ tahmincisi reel sayılarda herhangi bir noktadır, bunun için yapılan kestirim ise nokta tahmincisi olarak adlandırılır (Yılmaz ve Çelik, 2009).

Bayesci ve klasik yaklaşımda nokta tahminleri birbirinden farklıdır. Klasik yaklaşımda nokta tahmini, analizler sonucu elde edilen en iyi değer olarak tanımlanabilir. Bayesci yaklaşımda ise sonsal dağılımın ortalaması (posterior mean) nokta tahmincisi olarak incelenir (Ekici, 2009; Casella ve Berger, 1990).

Nokta tahmininde klasik ve Bayesci yaklaşımı birbirinden ayıran bazı temel özellikleden aşağıda bahsedilmiştir.

3.4.1. Yeterlilik

Klasik istatistiğe göre istatistiksel tahminlerin kalitesi, yeterlilik, etkinlik, yansızlık ve tutarlılık varsayımlarının ne kadarını karşıladığı ile incelenir (Casella ve Berger, 1990).

Klasik yaklaşımda yeterlilik, örneklemdeki bilginin tamamından yararlanmak olarak tanımlanabilir. Bu tanıma göre aritmetik ortalama yeterli bir tahminci iken; mod ve medyan yeterli tahminci değildir (Yılmaz ve Çelik, 2009;Howson ve Urbach, 2006).

Yeterlilik, özelliği hem klasik hem Bayesci yaklaşımda sağlanmaktadır. Literatürde ilgilenilen parametre ile ilgili tüm bilgiyi içeren istatistiğin yeterli olduğu kabul edilir (Howson ve Urbach, 2006). Bayesci istatistikte, sonsal dağılım parametre ile ilgili tüm bilgiye sahiptir. Dolayısı ile yeterli bir istatistiktir (Savchuk ve Tsokos, 2011; Ekici, 2009).

3.4.2. Tutarlılık

Tutarlılık, örneklem büyüklüğü sonsuza yaklaştığında tahmincilerin davranışlarını tanımlamaktadır. Tutarlılık tek bir tahminciden ziyade tahminci dizisine ait bir özelliktir (Casella ve Berger, 1990).

$$\lim_{n \to \infty} (P|\hat{\theta} - \theta| < \varepsilon) = 1 \,. \tag{3.5}$$

Eşitlik (3.5)'de tutarlılık, örneklem büyüklüğü sonsuza gittikçe tahminci değer ile gerçek değer arasındaki farkın epsilondan küçük olma olasılığının *1'* e yaklaşması anlamındadır (Casella ve Berger, 1990).

Tutarlılık, klasik yaklaşımın önemli varsayımlarındandır. Fakat Bayesci istatistik tutarlılık konusunda net değildir. Bir görüşe göre; temeli örneklem büyüklüğünün sonsuza yaklaşması fikri ile çeliştiğinden dolayı Bayesci yaklaşımda tutarlılık aranan bir özellik değildir (Howson ve Urbach, 2006). Diğer bir görüşe göre; parametrik Bayesci analizde daima tutarlı tahminciler elde edilir ve tutarlılık aranan bir özelliktir. (Ekici, 2009).

3.4.3. Yansızlık

Klasik yaklaşımda $\hat{\theta}$'nin beklenen değeri θ'ya eşit ise yani; $E(\hat{\theta})= \theta$ ise tahminci yansızdır. (Yılmaz ve Çelik, 2009; Çömlekçi, 1998). Yansızlık tahmin edicilerde aranan bir özellik olmasına rağmen bazen yanlı tahminciler tercih edilebilir (Akdi, 2014). Bayesci istatistikte hem, önsel dağılımın varlığından dolayı hemde, parametre tanımı klasik istatistikten farklı olduğundan dolayı tahminciler genel olarak yanlıdır (Savchuk ve Tsokos, 2011).

3.4.4. Etkinlik

Bir tahmincinin etkin olabilmesi için mümkün olan en küçük varyansa sahip olması şartı aranır. Genel olarak Hata kareler ortalaması ile tahmincinin etkinliği ölçülür (Yılmaz ve Çelik, 2009; Çömlekçi, 1998). Klasik yaklaşımcılara göre, Bayesci yaklaşımda sonsal dağılım önsel dağılımdan etkilendiğinden ve önsel dağılım subjektif olduğundan dolayı hata kareler ortalaması keyfi bir kıstasdır. Dolayısıyla Bayesci tahmincinin etkin olmadığı görüşü benimsenir (Box ve Tiao, 1976; Ekici, 2009). Fakat Bayesci yaklaşımcılara göre, sonsal dağılımın ortalaması sadece sonsal dağılımın ölçüsüdür ve ortalama her zaman için bilgi verici bir ölçek olarak kabul edilmemektedir.

Dolayısıyla Bayesci çıkarım değerlendirilirken keyfi olarak seçilen bir kıstasdan ziyade tüm dağılım göz önünde bulundurulmalıdır (Ekici, 2009).

3.5. Güven Aralığı

Klasik istatistikteki güven aralığı tanımı yerine Bayesci istatistikte güvenilir aralık, Bayesci aralık, en yüksek yoğunluk bölgesi veya en yüksek yoğunluk aralığı tanımlarını kullanılır (Box ve Tiao, 1976). Klasik istatistikte güven aralığı parametreden elde edilen tüm sonuçlara bağlı iken Bayesci istatistikte sonsal olasılık yoğunluk fonksiyonuna bağlıdır (Judge ve ark., 1985). Tanımlamada olduğu gibi yorumlamada da klasik yaklaşım ile Bayesci yaklaşım birbirinden farklıdır. Bu durum basit bir örnekle açıklanacak olursa; klasik yaklaşımda, *%95* güven düzeyi ile gerçek ortalamayı içerirken Bayesci yaklaşımda *%95* güven aralığında sonsal dağılımın ortalamasını içereceğine olan inanç derecesi *%95*''tir (Ekici, 2009).

3.6. Bayesci Çıkarımın Temelleri

Bu bölümde Bayesci çıkarımın temel taşı varsayılan Bayes teoremine, Bayesci yaklaşımın klasik yaklaşımdan farklarına, Bayesci yaklaşımın en önemli varsayımlarından olan önsel bilgiye, Bayesci yaklaşımda hesaplamaları kolaylaştıran Markov Zincirleri Monte Carlo (MCMC) yaklaşımına ve MCMC yönteminin koşullu dağılımlar için uygulanan özel bir şekli olan Gibbs örneklemesine değinilecektir.

3.6.1. Genel bilgi

Bayesci istatistik güçlü hesaplama tekniklerinin gelişmesi ile sağlık bilimleri, fen bilimleri ve sosyal bilimler gibi çok farklı alanlara uygulanmıştır (Bernardo, 1979; Maryyann, 1993).

Bayes teoremi 18.y.y'da ortaya atılmasına rağmen hesaplama zorlukları ve önsel bilginin subjektifliğinden dolayı 21.y.y'a kadar çok kullanılamamıştır. 21.y.y'da hesaplama zorluğunu ortadan kaldıracak hesaplama tekniklerinin çoğalması ve klasik yaklaşımın yetersiz kaldığı alanların artmasından dolayı Bayesci istatistiğe olan ilgi artmıştır (Ntzoufras, 2009).

Bayesci yaklaşımın iki temel amacı vardır. Birincisi, tümdengelim ve tümevarım arasındaki ilişkiyi açıklamaktır. Tümdengelim düşüncesi gelecekteki bir sonucu tahmin etmek için önceki bilginin sonuçlarını kullanır. Tümdengelim klasik istatistiğin gelişiminde önemli bir role sahiptir (O'Hagan, 2003; Savchuk ve Tsokos, 2011; Ekici, 2009). İstatistiksel olarak tümevarım popülasyondan çekilen bir örneklemin incelenmesi ile populasyonun genel özelliklerinin öğrenilmesi sürecidir. Bayesci istatistikte olasılık, tümevarım varsayımına dayanmaktadır. Başka bir ifade ile Bayesci istatistik yeterli deneme sayısı ile en yüksek olasılığa yani "*1*" olasılığına ulaşmayı amaçlar (Ekici, 2009; Hoff,2009).

İkincisi ise, subjektif olasılığı tanıtmaktır. Bayesci yaklaşım, bilinmeyen parametre değerlerini klasik yaklaşımda olduğu gibi sabit değerler olarak değil, rassal değerler olarak ele alır. Bu durum parametre üzerinde subjektif olasılık dağılımı oluşturur (Bernardo, 1979). Bayesci yaklaşımda bir parametre üstündeki subjektif olasılık dağılımı, bir kişinin parametre hakkındaki bilgisini özetler (O'Hagan, 2003). Bayesci istatistiğin temeli, sayılan önsel bilgi subjektiftir ve bilinmeyen parametreler hakkındaki tüm çıkarsamaların kaynağıdır. Önsel bilgi klasik yaklaşımla Bayesci yaklaşımın arasındaki en temel farktır. Önsel bilgi subjektif olduğundan dolayı klasik yaklaşımcılar hatalı sonuçlara yol açabileceğini düşünmekte ve Bayesci yaklaşımdan uzak durmaktadırlar (Ntzoufras, 2009). Bayesci yaklaşımda parametreye ait bütün çıkarımlar sonsal dağılımdan elde edildiğinden daha önce bahsedilmişti (Ntzoufras, 2009; Judge ve ark., 1984). Sonsal dağılımlar hesaplanırken olasılık dağılımlarının uzun ve karmaşık olması, hesaplamaların çok zaman alması gibi zorluklar Markov zincirleri ve Gibss Örneklemesi gibi simülasyon metotlarının ortaya çıkması ile son bulmuştur ki, bu durum Bayesci istatistiği çok popüler hale getirmiştir (Ntzoufras, 2009; O'Hagan, 2003; Cengiz ve ark., 2013).

Bayesci yaklaşımın temeli olarak kabul edilen Bayes teoremine aşağıda kısaca değinilecektir.

3.6.2. Bayes teoremi

İstatistikte koşullu olasılık geniş bir uygulama alanına sahiptir. Koşullu olasılığa dayalı olarak geliştirilen toplam olasılık ve Bayes teoremi özellikle karar problemlerinde çok önemlidir (Aytaç, 2012) . Bayes Teoremi içinoldukça önemli olan koşullu olasılık aşağıdaki gibi tanımlanır;

$\theta\ ve\ y,\ S$ örneklem uzayında iki olay ve $p(y) \neq 0$ ise y olayı verilmişken θ olayının koşullu olasılığı;

$$p(\theta \backslash y) = \frac{P(\theta \cap y)}{P(y)} \qquad (3.6)$$

şeklinde gösterilir. Eğer θ ve y bağımsız ise yani θ ve y gibi iki olaydan birinin gerçekleşmesi yada gerçekleşmemesi diğerinin gerçekleşme olasılığını etkilemiyorsa (3.6) denklemi aşağıdaki gibi yazılabilir (Schinazi, 2010; Miller, 2001; Yılmaz ve Çelik, 2009).

$$p(\theta \backslash y) = \frac{P(\theta \cap y)}{P(y)} = \frac{P(y)P(\theta)}{P(y)} = P(\theta) \qquad (3.7)$$

$\theta_{1,}\theta_{2,\ldots,}\theta_k$ olayları S örneklem uzayının bir parçasını oluşturuyorsa ve $i = 1, \ldots, k$ için $p(\theta_i) \neq 0$ ise S örneklem uzayında $p(y) \neq 0$ olan herhangi bir y olayı için $r = 1,2, \ldots, k$ iken şu şekilde yazılabilir.

$$p(\theta|y) = \frac{P(\theta,y)}{p(y)} = p(\theta|y) = \frac{P(\theta)P(y|\theta)}{p(y)} \qquad (3.8)$$

Eşitlik (3.8)'de $p(y) = \sum_{i=1}^{k} p(\theta) * p(y\,|\,\theta)$ θ'nın tüm olasılıkları toplamı (θ'nın kesikli veri olduğu durumda) veya $p(y) = \int(\theta) * p(y\,|\,\theta)\, d\theta$ (θ' nın sürekli değişken olduğu durumda) şeklinde yazılabilir (Gelman, 2004; Hoff, 2009). Bayesci yaklaşımın temel amacı sonsal dağılımı elde etmek olduğundan dolayı literatürde $p(y)$ toplam veya integral ifadesinin bire eşit olmasını sağlayan normalleştirme katsayısı kullanılır ve hesaplamada bu katsayı ihmal edilir (Hoff, 2009; Box ve Tiao, 1976).

Eşitlik (3.7)' de p bir olasılık yoğunluk fonksiyonu, θ bir parametre vektörü, y gözlem vektörü ve $p(y, \theta)$ birleşik olasılık yoğunluk fonksiyonudur. Bu durumda Bayes teoremi, $p(y, \theta)$ birleşik olasılık yoğunluk fonksiyonu y gözlem vektörü ve θ parametre vektörü için rassaldır (Gelman, 2004).

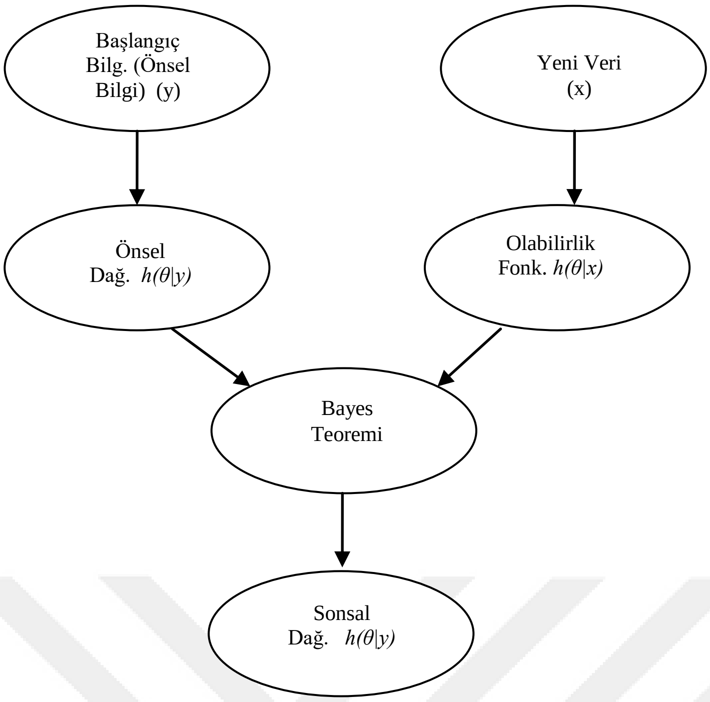
$p(y)$, $p(y) \neq 0$ olduğu takdirde ihmal edilebileceğinden dolayı (3.8) eşitliği aşağıdaki gibi yazılabilir.

$$p(\theta|y) \propto p(y|\theta)p(\theta) \tag{3.9}$$

İfade (3.9) da, genel olarak, $\propto$ ifadesi orantısal çarpım, $p(\theta)$ önsel dağılım, $p(y|\theta)$ örneklem bilgisi veya olabilirlik fonksiyonu ve $p(\theta|y)$ sonsal dağılım olarak adlandırılır. Bu durumda (3.9) ifadesi aşağıdaki gibi yazılabilir.

$$sonsal\ dağılım \propto (olabilirlik\ fonksiyonu)(önsel\ dağılım)$$

Eşitlik (3.9) da Bayes teoreminin özel bir şeklidir. Eşitlik (3.9)' dan da görüldüğü gibi $P(\theta|y)$ sonsal dağılımı $p(y|\theta)$ olabilirlik fonksiyonu ile $p(\theta)$ önsel bilginin birleşimidir. Bayesci yaklaşımda sonsal dağılım parametre hakkında yapılacak çıkarımların ve kararların temel taşıdır (Gelman, 2004; Box ve Tiao, 1973).



Şekil 3.1. Bayesci tahmin sürecinde Zelner'in olasılık değerlendirme şeması (Savchuk ve Stokos, 2011).

3.6.3. Bayesci yaklaşımın zorlukları ve üstünlükleri

Aşağıda Bayesci yaklaşımın bazı zorlukları ve üstünlükleri maddeler halinde kısaca sunulmuştur.

- Bayesci yaklaşım, klasik yaklaşımdan daha doğal ve kullanışlı çıkarımlar sunar (O'Hagan, 2003; Congdon, 2005).
- Bayesci metotlar, mevcut bilginin çoğunu kullanır ve klasik yaklaşımdan daha güçlü sonuçlar elde eder (O'Hagan, 2003; Congdon, 2005).
- Bayesci metotlar, MCMC ve Gibbs örneklemesi gibi simülasyon yöntemlerini kullanabildiğinden dolayı klasik metottan daha karmaşık problemleri rahatlıkla çözülebilir (O'Hagan, 2003; Gelman, 2004).
- Bayesci metotlar, önsel bilginin subjektifliğinden dolayı parametre tahmininde klasik metotlardan daha esnek varsayımlara sahiptir (O'Hagan, 2003; Lynch, 2007; Avcı, 2013).

- İstatistiksel çıkarımların genel amacı karar vermeyi kolaylaştırmaktır. Klasik yöntemler istatistiksel analizlerle ilgili dolaylı bilgiler verirken Bayesci modellemede var olan bilgiler doğru bir şekilde kullanıldığında en uygun sonuçları vermektedir. Dolayısıyla Bayesci metotlar karar verme problemleri için idealdir (O'Hagan, 2003; Ekici, 2009).

- Bayesci yaklaşım, örneklem büyüklüğü, hipotez testleri ve karmaşık modeller için model belirleme problemi gibi konularda klasik dağılımdan daha esnek varsayımlara sahiptir (Box ve Tiao, 1976; O'Hagan, 2003).

- Bayesci metotlarda, subjektif yaklaşımların kullanması ve parametrelerin rassal değişken olmasına karşın klasik metotlarda subjektiflik yoktur ve parametre sabit bir değişkendir. Bayesci ve klasik yaklaşım arasındaki en temel fark parametreden kaynaklanmaktadır (O'Hagan, 2003; Hoff, 2009).

- Bayesci metotlarda önsel bilgi vasıtası ile kullanılan ekstra bilgi güvenilir olmalıdır (O'Hagan, 2003; Savchuk ve Tsokos, 2011).

Çizelge 3.1. Bayesci ve klasik yaklaşımlar arasındaki farkların özeti (O'Hagan, 2003).

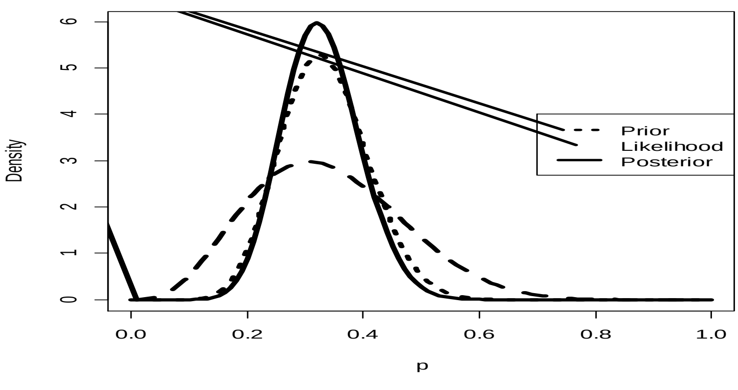
Klasik	Bayes
Olasılık üzerine	
Olasılık sınırlıdır.	Olasılık kişilerin inanç derecelerini ölçer.
Sadece tekrarlanabilir olaylar için geçerlidir.	Belirsizlik herhangi bir olay veya önerme için geçerlidir.
Parametre ile ilgili	
Parametreler rassal veya tekrarlanabilir değildir.	Parametreler belirsizdir.
Parametreler rassal değişken değildir. Fakat nicelikler sabittir.	Parametreler rassal değişkendir.
Çıkarımlarla ilgili	
Parametreler hakkında açıklamalar yapmaz.	Parametreler hakkında direk olasılık yorumu yapılır.
Örneklemle ilgili	
%5 anlamlılık seviyesinde hipotezi reddeder.	Hipotezin doğru olma olasılığı 0,05'tir

3.7. Önsel Bilgi

Önsel dağılım ilk olarak Richard Price (1763), tarafından tanıtılmış ve Bayesci çalışmalarda önsel dağılımın çeşitli tiplerinden bahsedilmiştir. Parametreler üzerindeki önsel dağılım inançlarla ilgili olduğundan aynı veri üzerine farklı kişilerin farklı önsel dağılımları olabilir (Marryann, 1993).

Bayesci çıkarımlar parametreye ait önsel yoğunluğu ve gözlemlenmiş veriyi kullandığından ve matematiksel gösteriminin eşitlik (3.9)'daki gibi olduğundan bahsedilmiştir.

Eşitlik (3.9)'da $P(\theta)$ önsel dağılımdır. Bayesci model çalışmalarında önsel dağılım veri işlenmeden veya görülmeden önce doğru olduğuna inanılan veya bilinen durum, önsel bilgi ise istatistiksel modelin bilinmeyen parametreleri için önsel dağılımın bir şekli olarak tanımlanabilir (O'Hagan, 2003). Bayesci yaklaşım, parametre üstündeki olasılık dağılımıyla popülasyon parametresi hakkındaki önsel bilgiyi açıklar (Savchuk ve Stokos, 2011; Gelman, 2002). $P(y|\theta)$ olabilirlik fonksiyonu ve $P(\theta|y)$ sonsal dağılımdır. Sonsal dağılım önsel bilgi ile olabilirlik fonksiyonun birleşimi olduğundan farklı önsel dağılımların kullanılmasından farklı sonsal dağılımların oluşacağı açıktır. Eğer veri sayısı yeterli ise sonsal dağılım ve önsel dağılım birbirine yakın olur (Marryann, 1993; O'Hagan, 2003).



Şekil 3.2. Önsel dağılım ile olabilirlik fonksiyonunun birleştirilmesinden elde edilen sonsal dağılımın R program çıktısı.

Bayesci yaklaşımın en büyük varsayımı önsel bilgi olduğundan, önsel dağılımın doğru seçimi Bayesci tahmin için oldukça önemlidir. Bayesci yaklaşımda önsel dağılımın seçiminde aşağıdaki durumlar dikkate alınır;

a)Eşlenik prensipler

b)Bilginin eksikliği

c)Bilginin tamlığı kriteri (Savchuk ve Stokos, 2011).

Önsel bilgi Bayesci yaklaşımın hem güçlü hem zayıf yanıdır (Statisticat, LLC, 2014; O'Hagan, 2003). Verinin doğru ve güvenilir bir şekilde toplanması ile önsel pozisyonlardaki farklılıklar çözülebilir ve ortak görüşe varılabilir. Eğer önsel bilgi belirsiz ve yetersiz ise sonsal dağılım büyük oranda olabilirlik fonksiyonuna bağlı olacaktır.

3.7.1. Bilgi veren önsel dağılım

Önsel bilgi, önceki verileri, çalışmaları, deneyimleri kapsayabilir ki bu bilgi veren önseldir (Zellner, 1996; O'Hagan, 2003).

Bilgi veren önsel dağılımlarda sonsal dağılımı ağırlıklı olarak önsel dağılım etkiler. Bilgi veren önsel dağılımın matematiksel yapısı sonsal dağılımların daha etkili ve yansız olması bakımından sade olmalıdır (Statisticat, LLC, 2014).

Önsel dağılımın; θ önseli $\theta \epsilon R$ ise normal veya t dağılımı, $\theta \epsilon (0, \infty)$ ise Gamma dağılımı ve $\theta \epsilon (0,1)$ ise Beta dağılımı olarak seçilmesi genel olarak dağılıma olan güveni artırır. (http://isites.harvard.edu/fs/docs/icb.topic196222.files/unit2.pdf).

3.7.2. Bilgi vermeyen önsel dağılım

Kass ve Wasserman (1996), bilgi vermeyen önselleri iki farklı şekilde yorumlamaktadır. Bu yorumlardan ilki bilgi vermeyen önseller ve bilgi eksikliğini temsil eder. İkincisi bilgi vermeyen önsel yetersiz bilgi kullandığında hatalı sonuç verir (Syversen, 1998).

Box ve Tiao (1973) ve Bernardo ve Smith (1994), bilgi vermeyen önsel dağılımları, veri ile ilgili bilgi verilmemesi olarak değil az bilgi verilmesi olarak tanımlarlar ve bilgi vermeyen önsel dağılımın sonsal dağılım üstünde minimum etkiye sahip olduğunu belirtilmektedir. Bayesci yaklaşımcılar genellikle bu tanımı kabul eder ve genel olarak bilgi vermeyen önsel dağılımın uniform dağıldığı kabul edilir (Syversen, 1998; Judge ve ark., 1984).

Başlıca bilgi vermeyen önsel dağılımlar, yaygın/dağınık, bilgi zayıf, etkisiz , Jeffrey's düzgün ve Referans önsel dağılımlardır (Box ve Tiao, 1973).

3.7.3. Eşlenik önsel dağılımlar

Bayesci yaklaşımda önsel dağılım ve sonsal dağılım aynı ailedense dağılımın ailesi için eşlenik denilmektedir. Eşlenik dağılımlar çok karmaşık modelleri düzenlemede kullanılabilinir (Gelman ve ark., 2004; Ntzoufros, 2009). Eşlenik önsel dağılımlar ile ilgili basit bir örnek aşağıda verilmiştir.

$$\begin{aligned} &y \sim Bin(n,p) \\ &\theta \sim Beta(\alpha,\beta) \\ &\theta \,|\, y \sim Beta(y+\alpha, n-y+\beta) \end{aligned} \tag{3.10}$$

θ için beta önseli eşlenik önseldir.

Eşlenik önsel dağılımların karmaşık olmayan modeller için hem uygulaması kolay hemde sonuçları güvenilirdir (Gelman ve ark., 2004).

Çizelge 3.2. Bazı eşlenik önsel dağılımlardan sonsal dağılımların elde edilmesi (Ntzoufras, 2009).

Olabilirlik Fonk.	**Önsel dağılım**	**Sonsal dağılım**
Bernoulli	Beta	Beta
Negatif binom	Beta	Beta
Poisson	Gamma	Gamma
Üstel	Gamma	Gamma
Normal	Gamma	Gamma
Normal	Normal	Normal

3.8. Olabilirlik Fonksiyonu

Olabilirlik fonksiyonu Fisher tarafından tanıtılmıştır. İstatistiksel yaklaşımlarda çok önemli bir yere sahiptir (Bayarri ve Degrott, 1992).

Olabilirlik, verinin bileşik olasılık fonksiyonudur. $y = (y_1, \dots, y_n)$ verilerinden bağımsız olarak elde edildiği kabul edilsin. Bu takdirde olabilirlik fonksiyonu;

$$l(\theta|Y) = p(y_1, \dots, y_n \,\big|\, \theta) = \prod_{i=1}^{n} p(y_i \,\big|\, \theta) \tag{3.11}$$

Bayesci modeli tanımlamak için önsel model ve olabilirlik birlikte kullanılır. Olabilirlik $p(y \,\big|\, \theta)$ örneklemiyle tanımlanan mevcut bilgiyi kapsar.

$$p(Y|\theta) = \prod_{i=1}^{n} p(y_i|\theta) \tag{3.12}$$

Y verisi $p(y|\theta)$ olabilirlik fonksiyonuyla $p(\theta \,\big|\, y)$ sonsal dağılımını etkiler (Gelman ve ark., 2004).

3.9. Sonsal Dağılım

Bayes teoremi önsel bilgi ile veriyi birleştirerek sonsal dağılımı oluşturur. Sonsal olasılık dağılımı θ parametresi hakkındaki tüm bilgiyi kapsar.

Bayesci yaklaşımın bir avantajı simülasyonla uygulanabildiğinden, dağılım geniş nümerik değerlerle özetlenilebilmektedir. Yaygın olarak bu özetleme işlemi dağılımın medyan, mod ve ortalaması kullanılarak yapılır. Örneğin ortalama parametrenin beklenen sonsalıdır. Mod, verinin en çok tekrar eden olasılık değeridir. Mod özellikle çok karmaşık modellerin çözümünde önemlidir (Gelman ve ark., 2004).

Sonsal dağılımla verinin sonsal aralıkları ve kantilleri de hesaplanabilir (Gelman ve ark., 2004). Büyük örneklemlerde sonsal dağılım öncelikle normallik yaklaşımına bağlıdır. n örnek büyüklüğü arttıkça sonsal dağılım normal dağılıma yaklaşır (Box ve Tiao, 1973).

3.10. Markov Zincirleri Monte Carlo İntegrasyonu (MCMC)

MCMC yaklaşımı, ilk olarak ikinci dünya savaşı yıllarında fizikçiler tarafından kullanılmıştır. Daha sonra 18.yy'da Buffon's bir örnekte parametre tahminini elde etmek için MCMC yöntemini kullanmıştır. Fakat işlem sayısı çok ve karmaşık olduğundan dolayı son yıllara kadar çok tercih edilmemiştir. Son yıllarda güçlü bilgisayarların gelişmesi ile MCMC yöntemi çok farklı alanlara uygulanmıştır. Bu uygulamalardan birkaç tanesi şöyle sıralanabilir; Large, Carlin ve Gelfand (1992) tarafından HIV virüsünün modellenmesi, Buck, Litton ve Strephesns (1993) arkeolojik kazı tahminlerinde, Anderws, Berger ve Smith (1993) otomobiller için yakıt ekonomisi gibi alanlarda MCMC yöntemlerini kullanmaya başlamışlardır (Cowles ve Carlin, 1996; Gelman, 2004).

MCMC, stokastik süreçlerin simülasyonu için genel bir metottur. Bu yüzden istatistiğin her alanında uygulanır. Karmaşık veya çok boyutlu integrallerde sonsal dağılımın hesaplanması zor olduğundan, özellikle Bayesci istatistikte çok sık kullanılır (Ntzoufras, 2009; O'Hagan, 2003). Sonsal dağılımlardan örneklem oluşturan MCMC teknikleri Bayesci modelleri değerlendirmek için etkili bir yol sağlar ((Ntzoufras, 2009; Geyer, 1992; Richey, 2011). MCMC yöntemi basitçe benzetim ile örneklem çekme veya simülasyon metodu olarak tanımlanabilir (Geyer, 1992; Hasting, 1970; Gelman, 2002). MCMC bir sonraki örnek değerini elde etmek için mevcut örnek değerini kullanarak rassal bir örneklem oluşturur ve bu oluşum Markov zinciri olarak tanımlanır. Markov zincirinin temel düşüncesi yeterince uzun ve yeterince büyük bir Markov süreci yaratmaktır. Markov zincirinde simülasyonun amacı durağan (stationary) bir Markov süreci oluşturmaktır (Gelman, 2004; Walsh, 2004).

MCMC yöntemleri Markov zincirlerini kullanan Monte Carlo integrasyonları ile çözüme gitmeyi amaçlar. Sık kullanılan MCMC yöntemleri, karmaşık istatistiksel metotların analizi için popüler olan Gibbs örneklemesi ve Metropolis Hasting algoritmasıdır (Gelman, 2004). Bu çalışmada sadece Gibbs örneklemesinden bahsedilecektir.

3.10.1. Markov zincirleri

$\mathrm{S} = \{1,2,\dots,\mathrm{N}\}$ sonlu bir durum uzayı ile verilen, Markov zinciri $X_t \varepsilon\, S, t = 1,2,\dots\}$ rassal değişkenlerin dizisi ile tanımlanan stokastik süreçtir. $\mathrm{X}_t\,, t$ zamanda bir rassal değişkenin değeri ve muhtemel X değerlerinin bir örnek uzayıdır (Richey, 2011; Walsh, 2004).

Eğer örnek uzayının farklı değerleri arasındaki geçiş olasılıkları sadece rassal değişkenin geçerli veya o andaki durumuna bağlıysa bu rassal değişken bir Markov süreci belirtir. Aşağıdaki eşitliğe Markov özelliği denir (Walsh, 2004).

$$P\big(X_{t+1} = s_j \big| X_0 = s_k, \dots, X_t = s_i\big) = P(X_{t+1} = s_j \,\big|\, X_t = s_i) \tag{3.13}$$

Örneklem uzayı kesikli ya da sürekli parametreye sahip olabilir. Örneklem uzayı kesikli olan Markov sürecine Markov zinciri denir. Bir Markov rassal değişkeni ile geleceği tahmin etmek için geçmişteki bilgilerden ihtiyaç duyulan sadece rassal değişkenin şu andaki durumudur. Yani diğer bir deyişle bir Markov sürecinin geleceği hakkında bilmek istediğimiz herşey, onun geçmiş değerlerinden bağımsız olarak şu anki değerinde özetlenmiştir. Önceki durumların bilgisi geçiş olasılığını değiştirmeyecektir (Walsh, 2004; Önalan, 2010). Markov süreci ile oluşturulan Markov zinciri $(X_{0,} X_{1,} \dots, X_n)$ rassal değişken dizini ifade eder.

Bir Markov zincirini tanımlamak için sadece başlangıç olasılıkları dağılımı ve geçiş olasılıklarını bilmek yeterlidir. Belirli bir zincirin durum uzayındaki süreci s_i durumundan s_j durumuna tek bir adımda geçme olasılığı olan geçiş olasılığı $P(i,j) = P(i \to j)$ ile tanımlanır (Walsh, 2004; Gilks, 1996; Önalan, 2010).

$$P(i,j) = P(i \to j) = P(X_{t+1} = s_j \,\big|\, X_t = s_i) \tag{3.14}$$

$P(i \to j)$ notasyonu i durumundan j durumuna geçiş anlamına gelmektedir.

$$p_j(t) = P\big(X_t = s_j\big) \tag{3.15}$$

İfade (3.15)'da p t adımda durum uzay olasılıklarının satır vektörü olasılık ise t zamanda $j.$ durumundaki zinciri anlatır. $p(0)$ özel başlangıç vektörüyle zincir başlar. Genel olarak başlangıç vektörünün 1 elemanı hariç tüm değerleri sıfırdır (Walsh, 2004). MC simülasyonu, yaklaşık dağılımdan θ'nın değerini çizmeye çalışan genel bir metottur ve sonsal dağılıma en iyi şekilde yaklaşmayı amaçlar. Örneklem dağılımı ile sonraki değere bağlı olarak çizilen sıralı bir metottur. Bu yüzden bir Markov zincirinden çizilir. Markov zinciri bir olasılık teorisi için tanımlandığında $\theta^1, \theta^{2,}, \ldots$ rassal değişkenler dizisinde θ^t değeri sadece θ^{t-1} değerine bağlıdır (Gelman, 2004; Walsh, 2004). MCMC algoritmaları verilen veri seti için aday modellerin sınıflarının muntazam genişlemesine izin verir (Cowles ve Carlin, 1996).

3.10.2. Gibbs örneklemesi

Gibbs örneklemesini ilk olarak Hasting (1970), istatistikte artan nümerik problemlerin çözümü için daha sonra Geman (1980), tam koşullu dağılımlarla birleşik dağılımların belirlenmesinde kullanmıştır. Gibbs örneklemesi genel olarak birleşik dağılımın analitik yöntemlerle direk belirlenmesinin zor olduğu durumlarda veya değişkenlerin çok büyük sayılar içerdiği karmaşık stokastik modellerin çözümlenmesinde uygulanır (Gelfand ve Smith, 1990; Hasting, 1970).

Gibss örneklemesinin ana amacı hedef dağılıma yakınsayan bir Markov zinciri oluşturmaktır. Böylece çok değişkenli dağılımlarda koşullu dağılımlarla örneklem elde etmek karmaşık bileşik dağılımlarla örneklem elde etmekten daha kolay olur. Bu yüzden incelenen rassal değişkenlerin birleşik olasılık dağılımında n boyutlu bir vektör oluşturmak yerine n tane tam koşullu dağılımdan ardışık olarak n tane rassal değişken oluşturmak daha kolaydır (Walsh, 2004; Önalan, 2010).

Gibbs örneklemesi, geçiş dağılımını tam koşullu dağılım ile şekillendiren bir MCMC örneğidir. Dağılımın $p(\theta)$ birleşik dağılım olduğu kabul edildiği takdirde; $\theta = (\theta_1, \ldots, \theta_n)'$ şeklinde gösterilir. Bu durumda θ_i bileşenlerinin her biri sabit bir vektör veya matristir ve tam koşullu dağılım aşağıda gösterildiği gibidir (Gelfand ve Smith, 1990; Gamerman, 2006).

$p_i(\theta_i) = p(\theta_i|\theta_{-i}), = p(\theta_i \mid \theta_1, \dots, \theta_{i-1}, \theta_{i+1}, \dots, \theta_d) \quad i = 1, \dots, d$

Tam koşullu dağılımların elde edilebileceği ve örneklem çekilebildiği durumda θ_i için tam koşullu dağılım aşağıdaki gibidir.

$$p_i(\theta_i) = p_i(\theta_i \mid \theta_{-i}) = \frac{p(\theta_i, \theta_{-i})}{\int p(\theta_i, \theta_{-i}) d\theta_i} \tag{3.16}$$

Eşitlik (3.16)'de $p(\theta)$ birleşik dağılımından örneklem çekmek amaçlanmaktadır. Bu durumda; Gibbs örneklemesi ile örnekleme elde etmek için tam koşullu dağılımdan ardışık olarak örneklem çekilir. Gibbs örnekleme algoritmasının adımları aşağıdaki gibidir.

1. Adım: $j = 1$ iterasyon sayısı ile başlanır ve $\theta^{(0)} = (\theta_1^0, \dots, \theta_n^0)'$ başlangıç değerleri oluşturulur.

2. Adım: $\theta^{(j-1)}$ kullanılarak parametre değerlerinin aşağıda belirtilen biçimde ardışık olarak üretilmesi sonucunda yeni $\theta^j = (\theta_1^{(j)}, \dots, \theta_n^{(j)})'$ vektörü elde edilir.

$\theta_1^{(j)} \sim p(\theta_1 / \theta_2^{j-1}, \dots \theta_n^{j-1})$

$\theta_2^{(j)} \sim p(\theta_2 / \theta_1^{(j)}, \theta_3^{j-1}, \dots, \theta_n^{(j-1)})$

.

.

.

$\theta_k^{(j)} \sim p(\theta_k / \theta_1^{(j)}, \theta_3^{j-1}, \dots, \theta_{k-1}^{(j-1)})$

.

3. Adım: Yakınsama sağlanıncaya kadar zincir *j*'nin değerini 1 artırır ve 2. adıma geri döner. Yakınsama sağlandığında, döngü durdurulur. İterasyon sayısı arttıkça zincir denge koşuluna yaklaşır ve yakınsama yaklaşık olarak gerçekleşir (Gelfand ve Smith, 1990; Gamerman, 2006).

3.10.3. MCMC yöntemlerinde yakınsamanın belirlenmesi

Karmaşık istatistiksel problemlerin çözümünde MCMC yönteminin kullanılması bazı problemleri de beraberinde getirmiştir. MCMC yöntemlerinin kullanıldığı Bayesci yaklaşımlarda parametre ile ilgili tüm çıkarsamalar simülasyon ile elde edilen sonsal dağılım üstünden yapılır (Burke, 2012; Walsh, 2004; Yüksel ve ark., 2013). Simülasyon ile oluşturulan örneklemde iki konu oldukça önemlidir. Bunlardan ilki Markov Zincirinin durağanlığa ya da istenen sonsal dağılıma ulaşması ikincisi ise Markov zincirini durağanlığa ulaştıracak olan iterasyon sayısının belirlenmesidir. Bayesci istatistikle yapılan analizlerin geçerli olması için sonsal dağılımın mutlaka yakınsama kriterini sağlaması gerekir (Cowles ve Carlin, 1996; Avcı, 2013). Yakınsaklıkla ilgili tek bir yaklaşım bulunmamaktadır. Çok çeşitli yöntemlerle sonsal dağılımın yakınsaklığı incelenebilir. Yakınsaklık testlerinin birbirine üstünlüğü de yoktur aynı zamanda kesin yakınsaklığı gösteren bir testte yoktur (Cowles ve Carlin, 1996; Yüksel ve ark., 2013).

Bu bölümde otokorelasyonla yakınsaklığın belirlenmesi, iz (trace) grafikleri, olasılık yoğunluk grafikleri gibi bazı yakınsaklığı belirleme yöntemlerinden kısaca bahsedilecektir. Birkaç yakınsaklık testi ise tablo ile özetlenecektir.

3.10.4. MCMC algoritmasında burn-in kavramı

Burn–in ya da Warm-up problemi basitçe zincir dengeye ulaşmadan iterasyonnun ne kadarının atılması gerektiğini sorgular (Geyer, 1992).

Burn-in kavramı, Markov zincirinin ilk kısmının, yani başlangıç değerinin otokorelasyonu azaltmak için örneklemden atılmasıdır. Burn–in periyodunun uzunluğu her zincir için farklıdır. Eğer zincirde otokorelasyon önemsiz ise durağan bir dağılıma ulaşmak için çok fazla iterasyon değerinin atılmasına gerek yoktur (Geyer, 1992; Burke, 2012). Burn-in kavramı, otokorelasyon hesaplaması ile de ilgilidir. Eğer otokorelasyon ihmal edilebilir seviyede ise dengeli ya da durağan bir zincire ulaşmak için çok sayıda iterasyonun atılmasına gerek yoktur. Genel olarak yeterince uzun bir zincirde başlangıç değerinin *%1* veya *%2*'sini örneklemden atmak yeterlidir (Geyer, 1992).

3.10.5. MCMC algoritmasında seyreltme kavramı

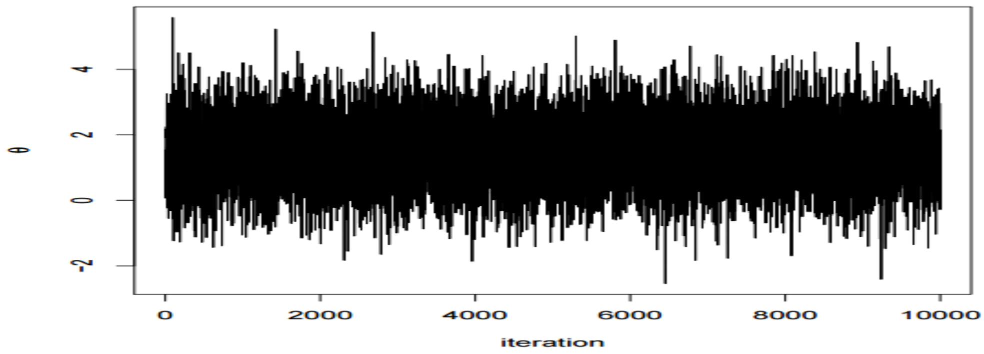
İterasyon değerleri arasındaki otokorelasyon yüksek ise örneklemdeki otokorelasyonu azaltmak için seyreltme kavramı kullanılır. Seyreltme otokorelasyondan kurtulmak için simülasyonla elde edilen örneklemden sistematik olarak m. terimin kullanılıp diğer iterasyon değerlerinin atılmasıdır. Bir Markov zincirine seyreltme terimi uygulandığında zincir hızlanmaz fakat otokorelasyon azaldığı için tahminler daha güvenilir hale gelir (Hosmer ve Lemeshow, 2013; Burke, 2012).

3.10.6. MCMC yakınsaklık boyutları

MCMC metotları kullanıldığında sonsal sonuçlar için zincirin durağan bir dağılıma yakınsaması ve sonsal sonuçlar hakkında ne kadar bilginin zincirde mevcut olduğu önemlidir (Burke, 2012).

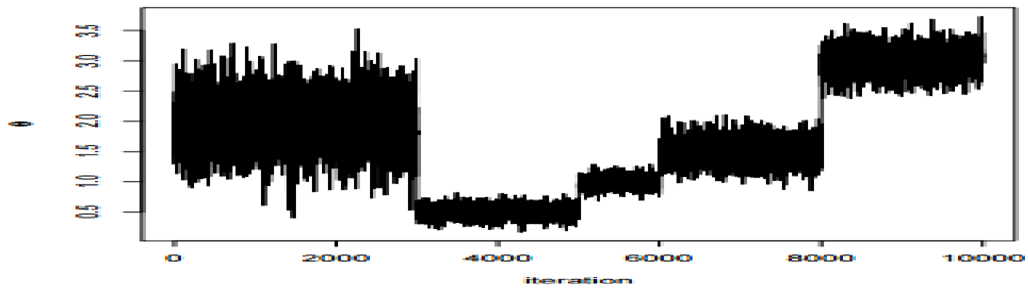
MCMC algoritmaları, verilen veri setleri için modellerin simülasyon ile sınıfının genişlemesine izin verdiğinden dolayı avantajlı bir yöntemdir fakat simülasyon ile veri genişletme işlemi otokorelasyona neden olabilir. Bu durumda MCMC algoritmalarının güvenirliğine ve yakınsaklığına karar vermek oldukça zordur. Örneklemin yakınsak (durağan) bir dağılıma ulaşıp ulaşmadığına nasıl karar verileceği çok önemlidir (Cowles and Carlin, 1996).

MCMC algoritmalarının yakınsaklığını incelemede çeşitli metotlar vardır. MCMC iz grafikleri bu metotlardan biridir (Cowles and Carlin, 1996). İz grafikleri zincirdeki iterasyon değerlerinin çıktılarıdır. İz grafikleri incelenerek zincirde yakınsama sağlanıp sağlanmadığına, burn-ini artırmaya gerek olup olmadığına karar verilmektedir (Hosmer ve Lemeshow, 2013). İz grafiklerini daha iyi kavramak için Şekil 3.3. ve Şekil 3.4. grafikleri incelenmelidir.



Şekil 3.3. İz grafiği görüntüsü (Good Mixing), (Lam, 2011).

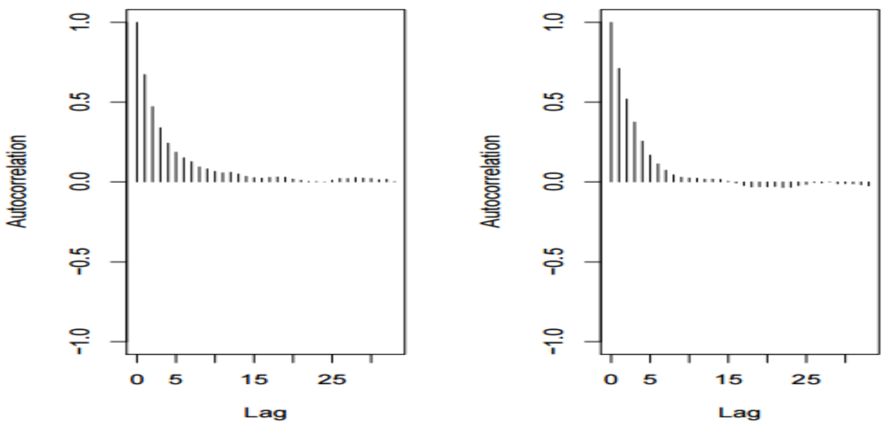
Şekil 3.3.'deki görüntüde, salınım fazla olduğundan yakınsamanın gerçekleştiği yani zincirin çok hızlı bir şekilde sonsal dağılıma ulaştığı anlaşılmaktadır. (Ekici, 2005; Hosmer ve Lemeshow, 2013).



Şekil 3.4. İz grafiği görüntüsü (Bad Mixing), (Lam, 2011).

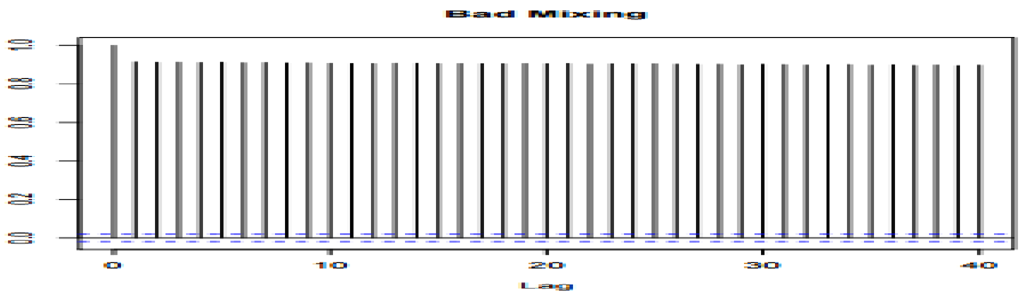
Şekil 3.4.'deki salınım şekil 3.3'deki salınım ile karşılaştırıldığında çok kötü olduğu görülmektedir. Dolayısı ile yakınsama gerçekleştirilememiştir (Sahlin, 2011).

Markov zincirinde simülasyonla elde edilen değerler birbirinden bağımsız değildir. Markov zincirindeki bu değerler arasındaki bağımlılık veya ilişki otokorelasyonla ölçülür (Hosmer ve Lemeshow, 2013; Ekici, 2005). MCMC örnekleminde yakınsamayı hızlandırmak ve otokorelasyonu azaltmak önemlidir (Cowles ve Carlin, 1996).



Şekil 3.5. Otokorelasyonsuz görüntüler (Lam, 2011).

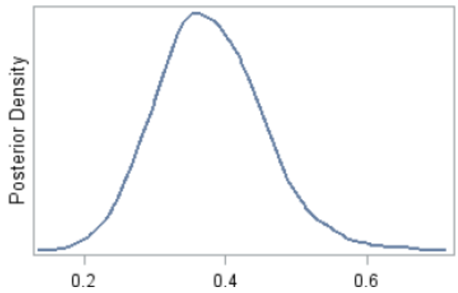
Markov zincirinde simülasyonla elde edilen değerler arasında otokorelasyon yoksa yani yakınsama sağlanmışsa Şekil 3.5.'deki çıktıdan da görüldüğü gibi zincir 0 değerine yaklaşır (Cowles ve Carlin, 1996).



Şekil 3.6. Otokorelasyonlu görüntü (Lam, 2011).

Yakınsama sağlanamamışsa, yani zincir yüksek oranda otokorelasyona sahip ise otokorelasyon görüntüsü Şekil 3.6. da görüldüğü gibi olmaktadır (Lam, 2011).

Markov zincirlerinde yakınsamanın sağlanıp sağlamadığına karar vermek için kullanılan bir diğer yöntem olasılık yoğunluk grafikleridir. Şekil 3.7. yakınsamanın sağlandığı olasılık yoğunluk grafiğine aittir.



Şekil 3.7. Yakınsamanın sağlandığı olasılık yoğunluk grafiği .

Yukarıdaki yöntemlerin yanı sıra Markov zincirlerinde yakınsaklığı inceleyen başka testler de bulunmaktadır. Fakat zincirin yakınsadığını belirleyen kesin bir test bulunmamaktadır. Var olan tüm testler ve yorumları ise Çizelge 3.3. de verilmiştir (Yüksel ve ark., 2013).

Çizelge 3.3. Markov yakınsaklık testleri (Yüksel ve ark., 2013).

İsim	Tanım	Testin Yorumlanması
Geweke	Markov zincirinin önceki ve sonraki kısımlarının ortalamalarını karşılaştırarak ortalama tahminlerinin yakınsamaya sahip olup olmadığını test eder.	Z test istatistiğine dayalı iki yönlü testtir. Büyük mutlak değerli Z değeri yakınsamanın sağlanmadığını göstermektedir.
Heidelberger-Welch (durağanlık testi)	Markov zincirinin bir durağan zincir olup olmadığını test eder. Başarısızlık daha uzun bir Markov zincirine ihtiyaç olduğunu gösterir.	Cramer-von Mises istatistiğine dayalı tek yönlü testtir. Küçük p değerleri yakınsamanın sağlanmadığını göstermektedir.
Heidelberger-Welch (yarı-genişlik testi)	Örneklem büyüklüğünün ortalama tahmini için gerekli hassasiyeti karşılamada yeterli olup olmadığını gösterir. Başarısızlık daha uzun bir Markov zincirine ihtiyaç olduğunu gösterir.	Eğer yarı-genişlik istatistiği, önceden belirlenmiş doğruluk ölçüsünden daha büyükse yakınsamanın sağlanmadığını göstermektedir.
Raftery-Lewis	Yüzdeliklerin istenen doğruluğuna ulaşmak için ihtiyaç duyulan örnek sayısını rapor ederek tahmini yüzdeliklerin doğruluğunu değerlendirir. Başarısızlık daha uzun bir Markov zincirine ihtiyaç olduğunu gösterir.	Eğer gerekli olan toplam örneklem, Markov zinciri örneğinden daha az ise bu yakınsamanın sağlanmadığını gösterir.
Otokorelasyon	Markov zinciri örnekleri arasındaki bağımlılığı ölçer.	Uzun gecikmeler arasındaki yüksek korelasyon, zayıf karışımı dolayısıyla yakınsamanın sağlanmadığını gösterir.
Markov Zinciri hata yaklaşımı	Markov zinciri (MC) standart hatası hesaplanır.	MC hata standart sapmanın %5'inden küçükse yakınsama vardır.
Etkili Örneklem Büyüklüğü	Otokorelasyonla ilgilidir. Markov zincirinin karışımını ölçer.	Etkili bir örnek büyüklüğü ile simüle edilen örnek büyüklüğü arasındaki büyük fark, zayıf karışımı dolayısıyla yakınsamanın sağlanmadığını gösterir.

4. GENELLEŞTİRİLMİŞ DOĞRUSAL MODELLER

Bağımlı değişkenlerin kategorik olduğu, normallik varsayımını sağlamadığı durumlarda model parametrelerini tahmin etmek için doğrusal modeller yerine genelleştirilmiş doğrusal modeller (GDM) kullanılmaktadır. GDM, Nelder ve Wedderburn (1972), tarafından geliştirilmiştir. Özellikle son yıllarda tıpta, mühendislikte, fen ve sosyal bilimler gibi pek çok uygulamalı alanda kullanılmaktadır (MCCullagh ve Nelder, 1989; Tektaş ve Günay, 2008; Şenel ve ark., 2009).

GDM, sistematik bileşen, rassal bileşen ve bağ fonksiyonu olmak üzere üç bileşenden oluşmaktadır (Barak, 2005; Agresti, 2002). Bu kavramlar kısaca aşağıda aktarılmıştır.

Rassal bileşen; olasılık dağılımının bağımlı değişkenin temelini oluşturduğunu kabul eder. Bağımlı değişkenin normal, poisson ve binomial gibi üstel bir aileden geldiğini varsaymaktadır (Agresti, 2002; Hair ve ark., 2006).

Sistematik bileşen; doğrusal tahminci olarak kabul edilir. $x_1, \dots, x_p$ bağımsız değişkenleri ve β parametresi için η doğrusal tahmincisi (4.1) eşitliğinde gösterildiği gibidir (Agresti, 2002; Hair ve ark., 2006; Lee ve ark., 2006).

$$\eta = \sum_{j=1}^{n} x_j \beta_j \qquad j = 1,2, \dots, n \tag{4.1}$$

Bağımlı değişkenlerin doğrusal kombinasyonu doğrusal tahminci olarak adlandırılmaktadır (Agresti, 2002).

Bağ fonksiyonu; farklı model şekillerinin uyum sağladığı rassal bileşen ve sistematik bileşen arasındaki teorik bağlantıdır (Hair ve., ark 2006; Maryyann, 1993). $i = 1, \dots, n$ için $\mu_i = E(y_i)$ olarak tanımlandığında bağ fonksiyonu $\eta_i \; ile \; \mu_i$ arasındaki bağıntıyı $\eta_i = g(\mu_i)$ eşitliği ile sağlar. Başka bir ifade ile $i.$ gözlemin ortalaması ve doğrusal tahmincisi arasında $g(.)$ bağ (link) fonksiyonu ile bir ilişki bulunmaktadır (Agresti, 2002 ; Hair ve ark., 2006; Lee ve ark., 2006).

$$\eta_i = g(\mu_i) = x\beta \tag{4.2}$$

Eşitlik (4.2)'de gösterilen $g(.)$ bağ fonksiyonu monoton ve diferansiyellenebilir bir fonksiyondur.

4.1. Bazı Bağ Fonksiyonları

GDM ailesinde kullanılan pek çok bağ fonksiyonu vardır. Bu bölümde bağ fonksiyonlarının bazılarının genel özelliklerine kısaca değinilecektir.

4.1.1. Lojit bağ fonksiyonu

Bağımlı değişken Bernoulli dağılımı gösterdiğinde; kullanılan bağ fonksiyonudur. $g(\mu) = \log{(p/1-p)}$ olarak tanımlanır. GDM'de kullanılan lojit bağ, lojit model olarak adlandırılmaktadır (Agresti, 2002).

4.1.2. Doğrusal bağ fonksiyonu

Bağımlı değişken normal dağılım gösterdiğinde kullanılan bağ fonksiyonudur. Doğrusal modellerde kullanılan bu bağ fonksiyonu $\eta = g(\mu)$ şeklinde tanımlanmaktadır (Ntzoufras, 2009).

4.1.3. Probit bağ fonksiyonu

Lojit bağ fonksiyonuna benzeyen probit bağda, bağımlı değişken iki durumludur ve $g(\mu) = \phi^{-1}(\mu)$ şeklinde tanımlanır. ϕ^{-1} standart normal birikimli dağılımın tersi olarak tanımlanır. GDM 'de kullanılan probit bağ, probit model olarak adlandırılmaktadır (Ntzoufras, 2009).

4.2. Lojistik Regresyon Analizi

Lojistik regresyon ile ilgili ilk çalışmalar 1944, 1953 ve 1955 yıllarında Berkson tarafından yapılmıştır (Çokluk, 2010; Avcı, 2013).

Regresyon yöntemleri bir bağımlı değişken ile bir veya daha fazla bağımsız değişken arasındaki ilişkiyi incelemektedir. Bu yöntemler arasında en çok bilinen basit veya çoklu regresyon analizidir. Regresyon tekniği bağımlı değişkenin sayısal bir değişken olması ve normal dağılım göstermesi, hataların normal ve birbirinden bağımsız dağılması gibi varsayımlar sağlandığı zaman kullanılmaktadır (Alpar, 2011; Alvin, 2008; Özdamar, 2013; İnal ve ark., 2006).

Bağımlı değişken, kategorik niteliksel veri şeklinde olduğu zaman lojistik regresyon analizi kullanılmaktadır. Lojistik regresyon, lojit bağ fonksiyonu ve binomial rassal bileşenlerle GDM ailesinin bir üyesidir. Lojistik regresyon modeli aynı zamanda lojit model olarak da adlandırılmaktadır (Agresti, 2002; Tatlıdil, 2002; İnal ve ark., 2006).

Eğer bağımlı değişken iki değer alıyorsa iki durumlu, ikiden çok değer alıyorsa çok durumlu lojistik regresyon analizi kullanılmaktadır. Bağımlı değişkendeki kategorilerin sınıflayıcı veya sıralayıcı olmasına göre de kullanılan faklı regresyon teknikleri bulunmaktadır. Uygulamada genel olarak iki durumlu lojistik regresyon analizi kullanılmaktadır. Bu çalışmada da iki durumlu lojistik regresyon analizi kullanılacaktır (Alpar, 2011; Hosmer ve Lemeshow, 2013; Burmaoğlu ve ark., 2009).

Lojistik regresyon ilk olarak biyomedikal çalışmalarında kullanılmasına rağmen, değişkenlerin tipi ve dağılımı ile ilgili varsayımların az olması, sonuçların kolaylıkla yorumlanabilmesinden dolayı son 20 yıldır sosyal bilimler, pazarlama gibi pek çok alanda kullanılmaya başlanmıştır (Alpar, 2011; Agresti, 2002; Çokluk, 2010).

4.2.1. İki durumlu lojistik regresyon analizi

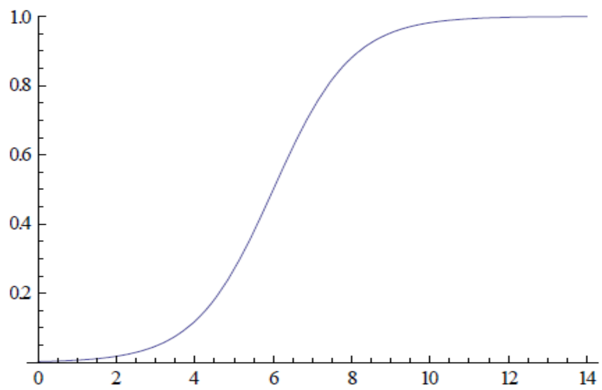
Klasik regresyon analizinde bağımlı değişkenin iki durumlu olması yani 0-1 arası olasılık değeri alması doğrusal olasılık modeli olarak adlandırılmaktadır (İnal ve ark., 2006; Tatlıdil, 2002). Doğrusal olasılık modeli regresyon varsayımlarının çoğunu sağlamadığından dolayı model analizlerinde doğru sonuçlar vermez. Bu durumda modeli anlamlı hale getirmek için çeşitli dönüşümler kullanılır, bu dönüşümlerden en sık kullanılan lojit dönüşümlerdir. Lojit dönüşümler başarı olasılığı p ve başarısızlık olasılığı $1 - p'$ nin doğal logaritmasıdır. Lojit dönüşümlerin kullanılması ile oluşan modeller ise lojit modellerdir. Lojit modeller doğrusal olmayan fakat uygun dönüşümlerle doğrusallaştırılabilen bir regresyon modelidir (Tatlıdil, 2002; Hosmer ve Lemeshow, 2013; Oğuzlar, 2005).

Basit bir lojistik regresyon modelinin;

$$y = x' \beta + e \tag{4.3}$$

Eşitlik (4.3)'de gibi olduğu kabul edilsin. Bu durumda $x_i' = [1, x_{i1}, \dots, x_{ik}]\ ve\ \beta'$=$[\beta_0, \dots, \beta_k]$ şeklindedir. Ayrıca y bağımlı değişkeni 0 veya 1 değerini alır. Lojistik regresyon modellerinde veya lojit modellerde bir olayın ortaya çıkma olasılığı $P(y_i = 1) = p$ ve bir olayın ortaya çıkmama olasılığı $P(y_i = 0) = 1 - p$ olarak ifade edilir (Tatlıdil, 2002; Barak, 2005; Hosmer ve Lemeshow).

İki durumlu lojit modelde bağımlı değişken grafiği S veya ters S şeklinde olan bir fonksiyondur. Bu fonksiyon lojistik fonksiyon olarak adlandırılır ve Şekil 4.1. deki gibi gösterilir (Hair ve ark., 2006).



Şekil 4.1. Bağımsız ve bağımlı değişken arasındaki lojistik ilişkinin şekli (Hair ve ark., 2006).

İstenen bir olayın gerçekleşme olasılığını hesaplamada kullanılan lojistik regresyon fonksiyonu aşağıdaki gibidir;

$$p = \frac{\exp\,(x'\,\beta)}{1+exp(x'\,\beta)} \tag{4.4}$$

veya

$$\frac{p}{1-p} = \exp\,(x'\,\beta) \tag{4.5}$$

Eşitlik (4.4) lojit model olarak adlandırılır (İnan ve ark, 2006). Eşitlik (4.5)'deki $\frac{p}{1-p}$ ifadesi, odds oranı olarak adlandırılır. Odds oranı, bir olayın gerçekleşme olasılığının gerçekleşmeme olasılığına oranınıdır. Bu oran *0* ile +∞ arasındaki tüm değerleri alabilir. Birden büyük bir odds oranı, bağımsız değişkendeki bir birimlik artışın, ilgilenilen olgunun ortaya çıkma şansını arttırdığını, birden küçük bir odds oranının ise bu şansı azalttığını açıklamaktadır. Bir olayın odds oranının doğal logaritmasının alınması lojit dönüşüm olarak adlandırılır. Lojit dönüşüm $-\infty$ ile $+\infty$ arasındaki tüm değerleri alabilir. Odds oranı ile lojit dönüşüm arasında doğrusal bir ilişki vardır ve model parametreleri lojit dönüşümle yorumlanır. Lojit dönüşüm aşağıdaki gibi gösterilir (Agresti, 2002; Alpar, 2011; Burmaoğlu ve ark., 2009; İnan ve ark., 2006).

$$ln(\frac{p}{1-p}) = x'\,\beta \tag{4.6}$$

Eşitlik (4.6) doğrusal tahminci için lojit bağ fonksiyonudur (Agresti, 2002; Hair ve ark., 2006).

Lojit bağ fonksiyonunda hata terimi binom dağılımlı olup, analiz bu teorik temele dayanmaktadır (Tatlıdil, 2002).

Çizelge 4.1. İki durumlu bağımlı değişken için olasılık odds ve lojit oranları (Hair ve ark., 2006).

Olasılık	Odds	Log Odds (lojit)
.10	.111	-2.97
.00	00	Hesaplanamaz
1.00	Hesaplanamaz	Hesaplanamaz

$y_i \epsilon (0,1)$, $p(y_i = 1/x_i)$ =p_i olan iki durumlu lojistik modele ilişkin varsayımlar kısaca şöyledir.

- $y_1, \dots, y_n$ değerleri istatistiksel olarak bağımsızdır.
- Açıklayıcı değişkenler (x_k) birbirinden bağımsızdır,

Ayrıca, modelin sonuç değişkeninin sınırlarını genişletmek için uygulanan $ln(\frac{p}{1-p})$ lojit dönüşümün bazı özellikleri de şöyle sıralanabilir;

- p arttıkça $lojit(p)$'de artar,
- p, 0-1 arasında iken $lojit(p)$ tüm gerçel değerleri alır (Tatlıdil, 2002).

4.2.2. Lojit modelde katsayıların tahmin edilmesi

Klasik regresyon analizinde model katsayılarının tahmin edilmesinde en sık kullanılan yöntem en küçük kareler (EKK) yöntemidir. EKK yönteminde hata kareler toplamının minumum olması amaçlanır. Genel olarak EKK yöntemi, bağımlı değişken ile bağımsız değişken(ler) arasındaki ilişki doğrusal olduğunda kullanılmaktadır. Fakat lojistik regresyon analizinde bağımlı değişken ile bağımsız değişken(ler) arasındaki ilişki doğrusal değildir. Dolayısı ile model parametrelerini EKK yöntemi ile tahmin etmek zordur. Bu durumda EKK yöntemi yerine genel olarak en çok olabilirlik (EÇO) yönteminin kullanılmaktadır. Bugün kullanılan istatistiksel yazılım programlarının bir çoğunda lojistik regresyon modelinin katsayıları EÇO yöntemi ile tahmin edilmektedir (Wakafield, 2012; Hair ve ark., 2006; Alpar,2011). EÇO yöntemi, EKK yönteminin aksine hata kareler toplamını minimize etmek yerine bir olayın gerçekleşeceği olabilirliğini (likelihood) maksimize etmeyi amaçlar. Diğer bir deyişle bir olayın olma ihtimalini en çok yapmaya çalışır (Alpar, 2011; Çokluk, 2010).

EÇO yöntemi, kullanıldığında değişken sayısının en az 10 katı kadar gözlem sayısı olması önerilmektedir. Buna karşılık, veride bağımsız değişkenler arasında çoklu bağlantı olması durumlarında da büyük örneklerin kullanılması önerilir (Alpar, 2011; Çokluk, 2010).

İki durumlu lojistik regresyon modelinde $y = 0, 1$ olmak üzere iki değer alabileceğinden ve $y'nin$ 1 değerini alma olasılığının p, $y'nin$ 0 değerini alma olasılığının $1 - p$ olduğu daha önceki bölümde aktarılmıştır.

İki durumlu lojit model için olabilirlik fonksiyonu aşağıda gösterildiği gibidir;

$$p_i = p(y_i/x_i) = P_i^{y_i}(1 - p_i)^{1-y_i}, \qquad \text{i=1,2,…,n için} \tag{4.7}$$

Eşitlik (4.7) aşağıdaki gibi yazılabilir.

$$p_i = p(y_i/x_i) = \left[\frac{\exp(x'\beta)}{1+exp(x'\beta)}\right]^{yi} \left[\frac{1}{1+exp(x'\beta)}\right]^{1-yi} \tag{4.8}$$

Eşitlik (4.8)'in *n* gözlem için olabilirlik fonksiyonu aşağıdaki gibidir.

$$l(y/x) = p(y/x) = \prod_{i=1}^{n} \left[\frac{\sum \exp(x'\beta)}{\sum 1+exp(x'\beta)}\right]^{yi} \left[\frac{1}{\sum 1+exp(x'\beta)}\right]^{1-yi} \tag{4.9}$$

Eşitlik (4.9)'un logaritması alınarak aşağıdaki eşitlik elde edilir.

$$\log l(y/x) = \sum_{i=1}^{n} y_i \, lnp_i + (1 - y_i)\ln(1 - p_i) \tag{4.10}$$

$l(y/x)$ en büyük yapan β değeri $\log l(y/x)$ ifadesinide en büyük yapar. Dolayısı ile $y = (y_{1,} \dots y_n)$ örnekleminin gözlenme olabilirliğini en büyük yapan parametre tahminlerini elde etmek için (4.10) eşitliğinin birinci türevinin alınıp sıfıra eşitlenmesi yeterlidir.

$$\frac{\partial logL}{\partial \beta} = \sum_{i=1}^{n}[Y_i\text{-}\frac{\sum \exp(x'\ \beta)}{\sum 1+exp(x'\ \beta)}]x_{ij} \qquad i = 1, \dots, N,\ j = 1, \dots K \qquad (4.11)$$

şeklinde olup *K* tanedir.

β' lar doğrusal olmadığından dolayı (4.10) ifadesinin direk çözümü bulunamaz. Bu yüzden Newton-Raphson yöntemi gibi yöntemlerle nümerik çözümleri elde edilir (Uçar, 2004).

4.2.3. Katsayıların önemliliğinin test edilmesi

Lojit modellerde parametrelere ait katsayılar test edildikten sonra katsayıların önemliliği veya anlamlılığı test edilmelidir. Katsayı tahminleri ile modeldeki bağımsız değişkenlerle bağımlı değişken arasındaki ilişkinin önemli olup olmadığı test edilir. Yani bir değişken, modelde yokken ve modelde varken elde edilen tahminciler karşılaştırılır (Alpar, 2012; Çokluk, 2010; Oğuzlar, 2005).

Katsayıların önemliliği, olabilirlik oran testi, Wald testi ve Skor test yöntemlerinden biri ile incelenir (Özdamar, 2013; Wakefield, 2012).

4.2.3.1. Olabilirlik oranı testi

Olabilirlik oran testinde önce model oluşturularak bağımsız değişken(lerin) önemliliği G istatistiği ile incelenir (Alpar, 2011; Wakefield, 2012; Çokluk, 2010).

$$LR = G = -2\ln\left(\frac{L(değişken\ modelde\ olmadığuda)}{L(değişken\ modelde\ olduğunda)}\right) \qquad (4.12)$$

Eşitlik (4.12) de bağımlı değişkenin modelde olduğu ve bağımlı değişkenin modelde olmadığı durumun karşılaştırıldığı açıktır (Çokluk, 2010).

LR ki kare dağılır ve serbestlik derecesi her iki modeldeki serbestlik derecesi arasındaki farka eşittir. Elde edilen test değerinin küçük olması değişkenlerin model için bir anlam ifade etmediği anlamına gelmektedir (Alpar, 2012; Wakefield, 2012; Çokluk, 2010).

4.2.3.2. Wald testi

Wald testide bağımsız değişkelerin önemlilik oranını (4.13) ifadesi ile inceler.

$$W = {\beta_j}^2 / S(\beta_j)^2 \sim X^2 \tag{4.13}$$

Eşitlik (4.13)'deki $S(\beta_j)$ bağımsız değişkenlerin standart hatasıdır. Wald testide ki-kare dağılımına uymaktadır (Alpar, 2011; Wakefield, 2012; Albayrak, 2006).

Wald istatistiği modelde bulunan tüm bağımsız değişkenler için tek tek anlamlılık testi yapmaktadır (Albayrak, 2006).

4.2.3.3. Skor testi

Skor testinde hesaplama işlemleri daha kolay yapılır ve genel olarak matris işlemleri ile hesaplanır. Skor testinin avantajı Wald ve olabilirlik testinden daha kolay hesaplanmasıdır (Çokluk, 2010; Alpar, 2011).

$$ST = \frac{\sum_{i=1}^{n} x_i (y_i - \bar{y})}{\sqrt{y_i (1-\bar{y}) \sum_{i=1}^{n} (x_i - \bar{x})}} \tag{4.14}$$

Skor testi standart normal dağılıma uymaktadır (Alpar, 2012; Wakefield, 2011).

4.2.4. Uyum iyiliği ve sapma ölçütleri

Lojistik regresyonda, elde edilen modelin sonuç değişkenini tanımlamakta ne kadar etkili olduğu modelin uyum iyiliği ile incelemektedir. Modelin veriye uyumu Pearson ki-kare testi, sapma istatistikleri ve Hosmer- Lemeshow testi gibi yöntemlerle değerlendirilir (Alpar, 2011).

Hosmer-Lemeshow testi ile kestirilen olasılıklar küçükten büyüğe doğru sıralanır ve sıralanan olasılıklar k alt gruba (k genel olarak 10 alınır) bölünerek işlem yapılır. Hosmer- Lemeshow testinin avantajı yorumlanmasının kolay olmasıdır.

Dezavantajı ise küçük örneklemler için önemli sapmaları belirleyememesidir. (Çokluk, 2010). Hosmer ve Lemeshow testi (4.15) eşitliği ile hesaplanır.

$$\hat{C}=\sum\frac{(G-B)^2}{B} \tag{4.15}$$

Eşitlik (4.15)'teki G: gözlenen ve değer ve B: beklenen değerdir ve $\hat{C}$ ki- kare dağılımına uymaktadır (Alpar, 2011).

Model uyum iyiliğinin değerlendirilmesinde kullanılan Pearson ki- kare testi ve sapma istatistikleri modelde beklenen değer ile gözlemlenen değer arasında anlamlı bir ilişki olup olmadığını incelemektedir. Pearson ki- kare ve sapma istatistikleri (4.16) ve (4.17) eşitliği ile hesaplanır (Alpar, 2011; Tatlıdil, 2002).

$$\chi^2 = \sum_{i=1}^{j} r_i^2 \tag{4.16}$$

$$D = \sum_{i=1}^{j} d_j^2 \tag{4.17}$$

Eşitlik (4.16)' da verilen $r_i = \frac{y_{i-\widehat{p_i}}}{\sqrt{\widehat{p_i}(1-\widehat{p_i})}}$ ve eşitlik (4.17)'de verilen $d_j = sign(y_i - \widehat{p_i})\sqrt{-2[(-y_i \ln(\widehat{p_i})) - (1-y_i)\ln(1-\widehat{p_i})}$ şeklinde tanımlanır.

4.3. Lojistik Regresyon Analizine Bayesci Bir Yaklaşım

Bu bölümde varsayımları daha esnek olan Bayesci lojistik regresyon analizi kısaca aktarılacaktır.

4.3.1. Genel bilgi

Bağımlı değişken kategorik olduğu zaman parametre tahmini için genel olarak EÇO yöntemi kullanıldığından daha önce bahsedilmişti. EÇO yöntemi asimptotik varsayımlara bağlı olarak parametre tahmini yapmaktadır. Örneklem sayısı yeterli olmadığında asimptotik varsayıma dayalı sonuçlar çok güvenilir olmamaktadır. Bu durumda Bayesci yöntemler tercih edilmektedir (Tektaş ve Günay, 2008).

$y_1, \dots, y_n$ n adet bağımsız ikili rassal değişken olarak kabul edilsin. *(y=1)*, p_i başarı olasılığı ile Bernoulli dağılımı gösterir. p_i başarı olasılığı ile bağımsız değişken(ler) arasındaki ilişki bağ fonksiyonu ile kurulur. İkili regresyon modeli bağımlı değişkenlerin bir fonksiyonu olarak bağımsız değişkenin olasılığını açıklamakta kullanılır. Genel olarak aşağıdaki gibi gösterilmektedir (Albert ve Chib, 1993; Souza ve Migon, 2004; Tektaş ve Günay, 2006).

$$yi \,|\, pi = Ber(p_i),$$

$$p_i = \mathrm{P}(y_i = 1) = F(x_i'\beta) \quad i = 1, \dots, n \tag{4.18}$$

$\beta, (\beta_1, \dots, \beta_k)'$ $k * 1$ bilinmeyen parametre vektörü; x_i' , $(x_{i1} \dots, x_{ik})$ bilinen bağımsız değişkenlerin bir vektörü ve F(.), x_i' doğrusal yapısı p_i olasılığı ile bilinen birikimli kümülatif bağ fonksiyonudur. $0 < F(.) < 1$ şeklinde tanımlanan sürekli, monoton artan bir bağ fonksiyonudur. Eğer F(.) bağ fonksiyonu lojistik birikimli dağılım fonksiyonu ise lojit model elde edilir (Albert ve Chib, 1993; Souza ve Migon, 2004).

$y = (y_1, \dots, y_n)'$, n birimlik örneklem için olabilirlik fonksiyonu eşitlik (4.19)'daki gibidir.

$$l(\beta|y, x) = \prod_{i=1}^{n} F(x_i'\beta)^{y_i}\left(1 - F(x_i'\beta)\right)^{1-y_i} \tag{4.19}$$

Bayesci yaklaşımda β parametresi hakkında önsel bilgiyi anlatan *p(β)* önsel dağılımı ile olabilirlik fonksiyonun birleştirilmesinden elde edilen sonsal dağılım eşitlik (4.20) ve (4.21)'deki gibidir.

$$p(\beta|y, x) \propto p(\beta)\; l(\beta|y, x) \tag{4.20}$$

$$p(\beta|y, x) \propto p(\beta) \prod_{i=1}^{n} F(x_i'\beta)^{y_i}\left(1 - F(x_i'\beta)\right)^{1-y_i} \tag{4.21}$$

Eşitlik (4.21) karmaşık bir fonksiyondur ve direk çözümü zordur. Bu yüzden Laplace metodu veya nümerik integrasyon yolu ile yaklaşımlar elde edilebilirken son yıllarda simülasyona bağlı metotlar hızla artmıştır (Souza ve Migon, 2004). Bu simülasyon metotlarından en sık kullanılanlar Zelner ve Rossi (1984), metodu ve Gibbs örneklemesidir. Zelner ve Rossi (1984), parametrelerin küçük sayıları için sonsal dağılımı hesaplamada nümerik integrasyon yöntemini kullanmışlardır. Büyük parametreler için de çok değişkenli student t dağılımı ile Monte Carlo integrasyonunu kullanarak sonsal dağılımları hesaplamışlardır (Groenewald ve Mokgatlhe, 2004; Albert ve Chib, 1993).

Daha sonraki yıllarda Albert ve Chib (1993), β'nın sonsal dağılımını hesaplamak için bir yöntem tanıtmışlardır. Bu yöntem, basitçe gizil değişkeninin normal dağıldığı durumlarda modele dahil edilmesi veri genişletme algoritması olarak tanımlanabilir. Bu yöntem ayrıntılı olarak Albert ve Chib metodunda aktarılacaktır (Collet, 1991; Souza ve Migon, 2004; Albert ve Chib, 1993). Veri genişletme algoritması ve Gibbs örneklemesi β'nın sonsal dağılımını hesaplamakta birlikte kullanılır. Gibbs örneklemesinin tek biçimli dağılım üstünde hesaplaması kolay olduğundan lojistik regresyon analizinde sık kullanılır (Groenewald ve Mokgatlhe, 2004).

4.3.2. Albert ve Chib metodu

Doğrusal regresyon modellerinin Bayesci analizinde model parametrelerini tahmin etmek için eşlenik önsel dağılımlar kullanıldığından dolayı model çıkarımı yapmak kolaydır (Albert ve Chib, 1993; Congdon, 2005). GDM'in Bayesci analizinde eşlenik önsel dağılımlar kullanılmadığından dolayı model çıkarsaması ve simülasyon çalışması yapmak zordur. Bu yüzden Albert ve Chib (1993), önce probit modellerde daha sonra lojit modellerde β'nın sonsal dağılımını hesaplamak için simülasyona bağlı bir yaklaşım önermişlerdir. Bu yaklaşımda u gizil değişken setinin modele eklenmesiyle lojit ve probit modellerin veri seti genişletilmektedir (Albert ve Chib, 1993; Dektaş ve Günay, 2008). y üzerinde tanımlı iki düzeyli probit regresyon modelini gizil veri içindeki doğrusal regresyon modeli ile birleştiren bu yöntem, gizil veriyi modellemek için normal dağılımların karışımını kullanarak daha ayrıntılı çözümlemeler yapılmasını sağlamaktadır (Albert ve Chib, 1993).

Albert ve Chib (1993), tahmin problemlerinin önemli bir parçasını oluşturan parametrelerin serbestlik dereceleri ile Bayesci çerçevede bir *t* bağ fonksiyonu önermişlerdir. $F(.)$ bağ fonksiyonu *t* dağılımlar ailesinin bir üyesidir (Lesage, 1999; Albert ve Chib, 1993). Bu bağ fonksiyon ailesinin seçimine göre tahminlerin duyarlılığı değişir. İki durumlu modellerde genel olarak lojit bağ fonksiyonu kullanılmaktadır. Albert ve Chib (1993), serbestlik derecesininde tahmin aşamasına dahil olduğu bir *t* dağılımı kullanmayı uygun görmüşler ve lojistik dağılımın serbestlik derecesi *7* olan *t* dağılım ailesinden, probit modelin, normal birikimli dağılım fonksiyonunun serbestlik derecesi 100 olan *t* dağılım ailesinden geldiğini ileri sürmüşlerdir (Congdon, 2005; Lesage, 1999; Albert ve Chib, 1993; Tektaş ve Günay, 2008).

4.3.3. İki durumlu lojit modele Bayesci bir yaklaşım

Bayesci lojit model, küçük örneklemler için daha güvenilir sonuçlar verdiği ve parametre tahminlerini daha kolay yorumladığından son yıllarda sıklıkla kullanılmaya başlanmıştır (Santos ve ark., 2009). Lojit modelde bağ fonksiyonu birikimli lojistik dağılımdır. Lojit modelde parametrelerin yorumu $\frac{p}{1-p}$ odds oranı ve odds kavramına bağlıdır. Binomial verilerde odds, $p(y=1)=p$ başarı olasılığı ve $p(y=0)=1-p$ başarısızlık olasılığının karşılaştırılması olarak tanımlanabileceği daha önceki bölümlerde aktarılmıştı (Ntzoufras, 2009).

Albert ve Chib (1993), tarafından probit ve lojit modeller için geliştirilen yöntem daha sonraki yıllarda lojit modeller için Groenewald ve Mokgatlhe ile Holmes ve Held tarafından iki farklı şekilde uygulanmıştır. Bu bölümde önce Groenewald ve Mokgatlhe metodu daha sonra Holmes ve Held metodu incelenecektir.

Lojit bağ fonksiyonu kullanılarak ikili lojistik birikimli dağılım fonksiyonu aşağıda gibi gösterilebilir (Tsai, 2004; Souza ve Migon, 2004).

$$log\left[\frac{p}{1-p}\right] = log\left[\frac{p(y=1)}{p(y=0)}\right] = F(x_i'\beta) \tag{4.22}$$

$$\frac{p}{1-p} = expF(x_i'\beta) \tag{4.23}$$

$$p = p(y_i) = \frac{expF(x_i'\beta)}{1+\exp\ F(x_i'\beta)},\ \ 0 < p < 1 \tag{4.24}$$

$F_z(.)$, lojistik dağılıma sahip Z rassal değişkenin birikimli dağılım fonksiyonudur. Z rassal değişkenin olasılık yoğunluk fonksiyonu eşitlik (4.25)'de olduğu gibi yazılır.

$$f_z(z) = \frac{\exp(z)}{(1+\exp(z))^2} \quad -\infty < z < +\infty \tag{4.25}$$

$p(y_{i=}1) = p_i$ başarı olasılığı eşitlik (4.26)'da olduğu gibi yazılır;

$$p_i = \int_{-\infty}^{x'_i\beta} \frac{\exp\,(z)}{(1+\exp(z))^2}\ \mathrm{dz} \tag{4.26}$$

$$p_i = P\left(U < \frac{exp(x_i'\beta)}{1+exp(x_i'\beta)}\right) \tag{4.27}$$

U, (0,1) arasında tek değişkenli düzgün bir dağılımdır. $u = (u_{1,}, \ldots, u_n)$ bağımsız gizil değişkenlerin tanıtılmasıyla $y = (y_1, \ldots, y_n)$ verisi verildiğinde, β ve u'nun bileşik sonsal olasılık yoğunluk fonksiyonu (4.28)'de olduğu gibi yazılır (Gronewald ve Mokgatlhe, 2004).

$$p(\beta, u|y) \ \propto\ p(\beta) * \prod_{i=1}^{n} \left\{ I\left(u_i \leq \frac{exp(x_i'\beta)}{1+exp(x_i'\beta)}\right) I(y_i = 1) +\ I\left(u_i > \frac{exp(x_i'\beta)}{1+exp(x_i'\beta)}\right) I(y_i = \right.$$

$$\left. 0 / 0 \leq ui \leq 1 \right. \tag{4.28}$$

$$\propto\ p(\beta)\ \prod_{i=1}^{n} \left\{ I\left((x'_i\beta) \geq \log\left(\frac{u_i}{1-u_i}\right)\right) I(y_i = 1) + I\left(x'_i\beta < \log\left(\frac{u_i}{1-u_i}\right)\right) I(y_{i=}0) \right\}$$

$$I(0 \leq u_i \leq 1) \tag{4.29}$$

Eşitlik (4.29) açıktır ki, β ve y verildiğinde u_i düzgün dağılır. Bu durumda $u_i{}'$nin koşullu dağılımı;

$$u_i \mid \beta, y \backsim \begin{Bmatrix} Uniform\left(0, \frac{\exp(x'_i\beta)}{1+exp(x'_i\beta)}\right)(y_i = 1)\ ise \\ Uniform\left(\frac{\exp(x'_i\beta)}{1+\exp\,(x'_i\beta)}, 1\right)(y_i = 0)\ ise \end{Bmatrix},\ i = 1,2,\dots,n \tag{4.30}$$

Eşitlik (4.30)'da ; $y_i = 1$ durumu için aşağıdaki eşitsizlik yazılabilir.

$$\sum_{j=0}^{p} x'_i\beta \geq log \frac{u_i}{1-u_i} \tag{4.31}$$

Bu durumda $y_i = 0$ ve $x_{ik} > 0$ ve $y_i = 1$ ve $x_{ik} < 0$ varsayımını sağlayan tüm i'ler için;

$$\beta_p \geq \frac{1}{x_{ik}}(log \frac{u_i}{1-u_i} - \sum_{j\neq k}^{n} x'_i\beta\ ) \tag{4.32}$$

$y_i = 0$ ve $x_{ik} > 0$, $y_i = 1$ ve $x_{ik} < 0$ ve $x_{ik} \neq 0$ varsayımını sağlayan tüm i'ler için;

$$A_k = \{i\colon \big((y_i = 1) \cap (x_{ik} > 0)\big) \cup (y_i = 0) \cap (x_{ik} < 0))\} \tag{4.33}$$

$$B_k = \{i\colon \big((y_i = 0) \cap (x_{ik} > 0)\big) \cup (y_i = 1) \cap (x_{ik} < 0))\} \tag{4.34}$$

β için $\pi(\beta) \propto 1$ biçiminde düz (diffuse) bir önsel dağılım olduğu kabul edilirse β_k'nın koşullu dağılımı düzgün dağılımdır (Gronewald ve Mokgatlhe, 2004).

.

$$\beta_k \mid \beta_{(-k)}, u, y \sim Uniform\big(a_{k,}b_k\big),\ \ k = 0,1,\dots,p \tag{4.35}$$

Eşitlik (4.35)'de belirtilen $a_k \; ve \; b_k$ (4.36) ve (4.37) eşitliğinde olduğu gibi tanımlanır;

$$a_k = \max_{i \in A_k} \left[\frac{1}{x_{ik}} \left(\log \frac{u_i}{1-u_i} - \sum_{j \neq k}^{n} \beta x_{ij} \right) \right] \tag{4.36}$$

$$b_k = \min_{i \in A_k} \left[\frac{1}{x_{ik}} \left(\log \frac{u_i}{1-u_i} - \sum_{j \neq k}^{n} \beta x_{ij} \right) \right] \tag{4.37}$$

Bu durumda Gibbs örneklemesinden seçilen örneklemler tek biçimli dağıldığı için yorumu daha kolaydır. Eğer $\beta' lar$ (4.35) eşitliği kullanılarak sırayla elde edilirken Z' nin n tane değeri eşitlik (4.31)'den çekilir. Bu durumda herhangi bir (β_k) önsel dağılımı ile bağımsız ise koşullu dağılım a_k ve b_k arasında budanmış $p(\beta_k)$ olur (Gronewald ve Mokgatlhe, 2004).

Bayesci lojit modeller için bir başka yaklaşım Holmes ve Held (2006), tarafından önerilmiştir. Holmes ve Held'e göre $F(x_i'\beta)$ lojistik dağılımında eğer $e_i \sim p(e_i)$ olarak ele alınırsa lojistik regresyon modeli elde edilmektedir. Bu durumda, β'nın güncellenmesi için eşlenik önsel koşulu sağlanmaz ve modele $\lambda_i, i = 1, \dots, n$ değişken seti eklenir. Bu durumda model aşağıdaki gibi yazılabilir (Holmes ve Held, 2006) ;

$$y_i = \begin{Bmatrix} 1 & \quad u_i > 0 \\ 0 & \quad u_i < 0 \end{Bmatrix} \tag{4.38}$$

$$u_i = x'_i \beta + e_i \tag{4.39}$$

$$e_i \sim N(\, 0, \lambda_i) \tag{4.40}$$

$$\lambda_i = (2\varphi_i)^2 \tag{4.41}$$

$$\varphi_i \sim KS \tag{4.42}$$

$$\beta \sim \pi(\beta) \tag{4.43}$$

Bu durumda φ_i, *i=1,2,…,n* için Kolmogrov-Smirnov dağılımı bağımsız rassal değişkenlerdir. Bu durumda e_i marjinal lojistik dağılımla ölçeklendirilmiş karma normal dağılımdır (Holmes ve Held, 2006).

β için önsel dağılım $P(\beta) = N(b, v)$ olarak tanımlanırsa, $z\ ve\ \lambda$ bilindiğinde β'nın tam koşullu dağılımı normaldir ve Eşitlik (4.44)'teki gibi gösterilir;

$$\beta \,\big|\, u, \lambda \sim N(B, V)$$

$$B = V(v^{-1}b + x'Wu) \tag{4.44}$$

$$V = (v^{-1} + x'Wx)^{-1}$$

$$W = k\ddot{o}\c{s}(\lambda_1^{-1}, \dots, \lambda_n^{-1})$$

gizil değişkenleri tam koşullu dağılımlardır ve λ_i'nin her bir varyans değeri için budanmış normal dağılım gösterir.

$$u_i \,\big|\, \beta, y_i, \lambda_i \sim \begin{Bmatrix} N(x_i'\beta, \lambda_i) I(u_i > 0), y_i = 1 \\ N(x_i'\beta, \lambda_i) I(u_i \le 0), y_i = 0 \end{Bmatrix} \tag{4.45}$$

Bu işlemler belirlendikten sonra Bayesci lojistik regresyondan sırayla $(\beta|u, \lambda), (u|\beta, \lambda)\ ve\ (\lambda \,\big|\, u, \beta)$ dağılımından örneklem çekmek daha kolaydır (Holmes ve Held, 2006).

5. MATERYAL VE YÖNTEM

Tezde Ekonomik Kalkınma ve İşbirliği Teşkilatı (OECD)'ye bağlı 34 ülkeye ait bazı veriler kullanılmıştır. Veriler 2013 yılı veya verisi bulunan en son yıla aittir. Veriler OECD'nin resmi web sitesi olan OECD.org'dan alınmıştır.

Tezde kullanılan lojistik regresyon denklemi (5.1) eşitliğinde gösterilmiştir. Hem Bayesci lojistik regresyon analizinde hem de klasik lojistik regresyon analizinde aynı veri seti kullanılmıştır. Her iki analizden elde edilen sonuçlar karşılaştırılmıştır. Lojistik regresyon analizinde parametreleri tahmin etmek için en çok olabilirlik yöntemi kullanılmıştır. Bayesci lojit modelde parametre tahminleri için Gibss örneklemesi yaklaşımı ile ve sonsal dağılımın yakınsaklığı incelenmiştir. Parametre ile ilgili tüm çıkarsamalar sonsal dağılımla yapıldığından sonsal dağılımın yakınsama kriterlerini sağlaması oldukça önemli olduğundan yakınsama kriterleri sonuçları verilmiştir (Yüksel ve ark., 2013; Cowles ve Carlin, 1996).

$$y = \alpha + \beta_1 nüfus + \beta_2 ith + \beta_3 yaşmem + \beta_4 sağlık + \beta_5 doğum + \beta_6 ihr + \beta_7 kadın\_mv + \beta_8 kadın\_iş + e \quad (5.1)$$

Eşitlik (5.1) var olan değişkenlerin ayrıntılı açıklaması 6.2.1. bölümünde verilmiştir.

Hem klasik hem Bayesci lojistik regresyon analizinde verileri analiz etmek için R (3.0.3) yazılım programı kullanılmıştır.

Çizelge 5.1. Kullanılan değişkenler ve kodları

Kodu	**Değişken**	**Yıl**
Nüfus	Nüfus	2013
İhr	İhracat oranı	2013
İth	İthalat oranı	2013
Kad_Mv	Kadın milletvekili oranı	2013
Kad_İş	Çalışan kadın oranı	2013
Doğum	15-49 yaş arası her bir kadının sahip olduğu ortalama çocuk sayısı	2013
Yaşmem	Yaşam memnuniyeti	2013
Sağlık	GSYİH'dan sağlık harcamalarına ayrılan oran	2013

6. UYGULAMA

Bu bölümde OECD verilerine lojistik ve Bayesci lojistik regresyon analizi uygulanmış ve elde edilen sonuçlar karşılaştırılmıştır.

6.1. Model Tanımlama

İki durumlu lojistik regresyon analizinde parametre tahminlerini elde etmek için R yazılım programı kullanılmıştır. R programında klasik lojistik regresyon için glm fonksiyonu Bayesci iki durumlu lojistik regresyon için MCMCpack kütüphanesinde bulunan MCMClogit fonksiyonlarından ve Bayesci tahmin yönteminde elde edilecek parametrelerin yakınsama durumunu incelemek için R yazılım programında yer alan coda fonksiyonundan faydalanılmıştır.

Bu bölümde önce çözümlenecek veri kümesindeki değişkenlerden ve OECD'nin kısa tarihçesi ile amaçlarından bahsedilecek daha sonra uygulama yapılacaktır.

6.1.1. Ekonomik kalkınma ve işbirliği teşkilatı (OECD)

OECD kurulmadan önce, 1948 yılında Marshall planı çerçevesinde Avrupa Ekonomi ve İşbirliği Teşkilatı (OECC) kurulmuştur. OECC'nin kurulma amacı ikinci dünya savaşı sonrasında savaş yıkıntıları içindeki Avrupa ülkelerinin ekonomilerinin desteklenmesi ve onarılmasıdır. OECC'nin görevini tanımlaması üzerine 14 aralık 1960 yılında imzalanan Paris sözleşmesi ile OECC'nin yerine ve daha geniş bir görev tanımıyla OECD kurulmuştur. OECD'nin 20'si kurucu üye olmak üzere toplam 34 üyesi bulunmaktadır. Türkiye OECD'nin kurucu üyelerindendir (Altaş ve Turgan, 2008; OECD Çevresel Performans incelemeleri. Türkiye, 2008; http://www.mfa.gov.tr/iktisadi-isbirligi_ve-gelisme-teskilati-_oecd_.tr.mfa)

OECD dünya ekonomisinin yönetimine alt yapı oluşturan, küreselleşmenin yol açtığı sosyoekonomik ve çevresel problemlere karşı birlikte mücadele edilen öncü kuruluşlardan biridir. OECD için ekonomik konularda uzmanlaşmış hükümetler arası bir istişare kuruluşu olduğu söylenebilir. OECD ekonomik büyüme, mali istikrar, ticaret ve yatırım teknolojileri, yenilik girişimcilik ve kalkınma alanındaki işbirliği yoluyla refahın sağlanması ve yoksullukla mücadele konusunda hükümetlere yardımcı olmayı amaçlanmaktadır (http://www.mfa.gov.tr/iktisadi-isbirligi_ve-gelisme-teskilati-_oecd_.tr.mfa; http://www.oecd.org/about/).

OECD aynı zamanda örgütün toplantıları, üye ülkelerin üzerinde uzlaştığı konuları, duyuruları, sosyal, ekonomik, çevresel konularda yaptığı araştırmaları, elde ettiği istatistikleri, hazırladığı rapor veya belgelerin dağıtımını / yayılımını da yapmaktadır (OECD Çevresel Performans incelemeleri. Türkiye, 2008).

6.1.2. Değişkenlerin tanımlanması

Bu çalışmada 34 OECD ülkesinin AB üyesi olup olmaması ile OECD ülkelerinin nufusları, kadın milletveki oranları, yaşam memnuniyeti oranları, GSYH'dan sağlık harcamaları için ayrılan bütçe oranları, ithalat oranları, 15- 49 yaş arası her kadına düşen çocuk sayısı oranları, ihracat oranları ile model kurulmuştur. Veri setinde yer alan değişkenlerin verileri 2013 yılına ait olup, 2013 yılına ilişkin verisi olmayan değişkenler için en son ölçüm yapıldığı yıl verileri alınmıştır.

AB üyeliği *(y)*, her bir OECD ülkesinin AB üyesi olup olmamasını belirtir. İki durumludur. Her bir OECD ülkesi AB üyesi ise $y=1$, AB üyesi değil ise $y=0$ şeklinde tanımlanmıştır.

Nüfus (x_1); OECD ülkelerinin 2013 yılı toplam nüfusları,

Kad_Mv (x_2); ele alınan dönem içerisindeki her bir OECD ülkesindeki kadın parlamenter oranı,

Kad_iş (x_3); her bir OECD ülkesinde 15-64 yaş aralığındaki kadınların iş gücüne katılım oranları,

İth (x_4); ele alınan dönemdeki her bir OECD ülkesinin ithalat oranı,

İhr (x_5); ele alınan dönemde ki her bir OECD ülkesinin ihracat oranı,

Yaşmem (x_6); yaşam memnuniyeti veya yaşam doyumu, kişinin bulunduğu şartlardan hoşnut olması yani, bireyin mutluluğu olarak tanımlanabilir. Daha açık bir ifade ile yaşam memnuniyeti bireyin kendi belirlediği kıstaslara göre yaşam kalitesi hakkındaki genel değerlendirmesidir (Tümlü ve Recepoğlu, 2013; Akın ve Yalnız, 2015). Modelde ele alınan x_6 değişkeni her bir OECD ülkesindeki insanlardan yaşamından memnun olanların oranı,

Sağlık (x_7); 2013 yılında her bir OECD ülkesinin gayri safi yurt içi hasıladan (GSYİH) sağlık harcamaları için ayrılan pay,

Doğum (x_8); 2013 yılı her bir OECD ülkesindeki 15-49 yaş arasındaki kadın başına düşen ortalama çocuk sayıdır.

Klasik lojistik regresyon için model;

$$log\left(\frac{p}{1-p}\right) = \propto + \beta_1 nufus + \beta_2 ith + \beta_3 yaşmem + \beta_4 sağlık + \beta_5 doğum + \beta_6 ihr + \beta_7 KAD_{MV} + \beta_8 kad_{iş} + e \quad (6.1)$$

Bu çalışmada kullanılan OECD verilerinin parametre tahminlerini elde etmek için klasik lojistik regresyon analizinde EÇO yöntemi ve Bayesci lojit modellerde Gibss örnekleme yöntemi, önsel dağılım olarak ise bilgi vermeyen önsel dağılımlardan olan düzgün dağılım kullanılmış ve sonuçlar aşağıda verilmiştir. Uygulanan hipotez testlerinde yanılma düzeyleri (α) *0.05* olarak alınmıştır.

Çizelge 6.1. EÇO yöntemi ile elde edilen parametre tahminleri

Parametre	Tahmin	Standart hata	z	p
β_0	10.04930	7.5753	1.327	0.088
Nüfus	0.04387	0.02705	1.622	0.1401
Yaşmem	-5.69552	3.24655	-1.754	0.0314
Saglık	1.55750	1.05062	1.482	0.405
İth	0.08177	0.08394	0.974	0.4132
İhr	0.20735	0.16268	1.275	0.29
Kad_mv	0.38593	0.23629	1.633	0.066
Kad_iş	-0.06910	0.15646	-0.442	0.6877
Doğum	-0.80534	2.87410	-0.280	0.644

Yukarıdaki Çizelge 6.1. de parametre tahminleri, p değerleri, z değerleri ve standart hata değerleri verilmiştir. Çizelgede ki her bir değişken için hipotez testi aşağıda gösterilmiştir. Çizelge 6.1. 'den de görüldüğü gibi neredeyse tüm parametrelerin p anlamlılık düzeyleri 0.05'ten büyüktür. Dolayısıyla model klasik lojistik regresyon analizi ile incelendiğinde modeldeki yaşam memnuniyeti hariç hiçbir değişken anlamlı değildir.

Hipotez testleri;

H_0: Nüfus değişkenin modele katkısı anlamlıdır,

$H_{1:}$ Nüfus değişkeninin modele katkısı anlamsızdır, şeklinde kurulur.

Yaşmem, İth, İhr, Kad_mv, Kad_iş, doğum parametreleri içinde aynı şekilde hipotez kurulur.

Çizelge 6.2. Lojistik regresyon analizinde EÇO yöntemi ile elde edilen sonuçlar

$MAC\ R^2$	p	Olabilirlik
0.66	0.0003	29.51

H_0: Model uyumlu değildir.

H_1: Model uyumludur.

Çizelge 6.2.'deki model çıktılarında yer alan p değeri incelendiğinde $0,05$'ten küçük olduğu görülmektedir. Böylece model $\%\ 95$ olasılıkla uygun olduğu söylenebilir. Dolayısıyla H_0 hipotezi reddedilir. Ayrıca olabilirlik oran testinin model için anlamlı olup olmadığına karar vermek için ki-kare tablo değeri de incelenebilir.

Klasik regresyon analizinde R^2, bağımsız değişkenlerin bağımlı değişkendeki değişimin yüzde kaçını açıkladığını gösterir ve $[0,1]$ arasında değerler alır. R^2 değerinin 1'e yaklaşması modelin açıklama gücünün yüksek olması anlamına gelir ve R^2 belirlilik katsayısını hesaplamada EKK yöntemi kullanılır. Lojistik regresyon analizinde parametre tahminlerinde EKK yöntemi kullanılmadığından dolayı klasik regresyon analizinde olduğu gibi direk R^2 değeri hesaplanamaz (Kalaycı, 2010; Albayrak, 2006; Alpar, 2012). Bu yüzden lojistik regresyon analizinde belirlilik katsayısı temsili bir katsayıdır, genel olarak küçük çıkmaktadır ve modelin uyum iyiliğini belirlemede pek kullanılmamaktadır (Alpar, 2011; Albayrak, 2006; Hosmer ve Lemeshow, 2013). Lojistik regresyon için çok çeşitli R^2 değerleri olmasına karşın uygulamada genel olarak MAC Fadden R^2 değeri kullanılmaktadır. Bu çalışmada da MAC Fadden R^2 değeri kullanılmıştır. MAC Fadden R^2 eşitlik (6.2)'de olduğu gibi hesaplanır (Alpar, 2011; Menard, 2002).

$$R^2{}_{McF} = 1 - \frac{\ln \hat{L}(M_\beta)}{\ln (\hat{L}(M_\alpha)} \tag{6.2}$$

M_α: Sadece sabitin bulunduğu model

M_β : Bağımsız değişkenlerin bulunduğu model

Eşitlik (6.2) modelindeki R^2 değerinin *0.20* ile *0.40* arasında çıkması model uyumu için yeterli olduğu düşünülmektedir (Alpar, 2011; Menard, 2002). Genel olarak model uyumunu değerlendirmektense farklı modellerin performansını değerlendirmek için kullanılmaktadır.

Çizelge 6.3 Değişkenler arası korelasyon matrisi

	Yaşmem	Nüfus	Kad_mv	Saglık	İhr	İth	Ab	Doğ	Kad_iş
Yaşmem	1.000	-0.320	0.620	0.430	0.030	-0.093	-0.320	0.540	0.670
Nüfus	-0.320	1.000	-.0260	0.110	-0.593	-0.420	-0.110	-0.070	-0.430
Kad_mv	0.620	-0.260	1.000	0.490	-0.046	0.080	0.100	0.174	0.620
Saglık	0.430	0.110	0.491	1.00	-0.385	-0.230	0.110	0.074	0.600
İhr	0.032	-0.590	-0.046	-0.380	1.000	0.705	0.260	-0.217	0.063
İth	-0.093	-0.420	0.083	-0.230	0.705	1.000	0.440	-0.203	0.036
Ab	-0.320	-0.110	0.107	0.110	0.260	0.440	1.000	-0.420	-0.074
Doğ	0.540	-0.070	0.174	0.074	-0.217	-0.230	-0.420	1.000	0.163
Kad_iş	0.640	-0.430	0.620	0.609	0.630	0.036	-0.070	0.163	1.000

Çizelge 6.3. de değişkenler arasındaki ilişkiyi anlamak amacıyla korelasyon hesaplanmıştır. Bu ilişki nüfus, parlamentodaki kadın milletvekili sayısı GSYİH’dan sağlık harcamalarına ayrılan bütçe oranı, ihracat oranı, ithalat oranı, 15-49 yaş arası her

kadına düşen ortalama çocuk sayısı ve kadınların iş gücüne katılım oranı olan bağımsız değişkenlerle, her bir OECD ülkesinin AB üyesi olması veya olmaması bağımlı değişkeni ile aralarında bir ilişki olup olmadığını anlamaya yardımcı olmaktadır.

Çizelge 6.4. Lojistik regresyon analizinde değişkenler arası güven aralıkları

Güven aralıkları	%2.5	%97.5
Sabit	-0.1298063936	26.311945985
Nüfus	-0.0006682814	0.005369104
Yaşmem	-0.6454193527	-0.049603356
Kad_mv	-0.0005035578	0.047768913
Sağlık	-0.0682819373	0.172320909
İhr	-0.0046511593	0.0157777568
İth	-0.006552836	0.0152250879
Doğ	-0.6330541244	0.389227823
Kad_iş	-0.01788650302	0.027224204

Çizelge 6.4. 'e göre sadece yaşam memnuniyeti değişkeni *% 95* güven aralığında *0* değerini içermediğinden dolayı anlamlıdır (Avcı, 2013).

R programında Bayesci lojit modeller için elde edilen sonuçlar aşağıda verilmiştir. Bayesci tahminleri oluşturabilmek için ilk olarak 10000 iterasyon sayısı ile işlem yapılmış ve yakınsamanın sağlanması ilk 200 zincir örneklemden atılmış seyreltme sayısı ise 1 olarak alınmıştır. Fakat analiz yapıldığında yakınsama sağlanamamış ve yakınsamanın sağlanması için iterasyon sayısı artırılmıştır. Yakınsamanın sağlanmadığına dair otokorelasyon grafikleri, iz grafikleri ve sonsal olasılık yoğunluk grafikleri Ek-1'de sunulmuştur. R progmanında iterasyon sayısı ile ilgili bir kısıtlama olmadığından dolayı iterasyon sayısı 100000'e zincirden atılan örneklem sayısı 1000'e ve seyreltme sayısı 5'e çıkartılarak aynı işlemler tekrarlanmıştır. Yapılan iterasyon sonucunda belirlenen örnekleme sayısı ile yakınsamanın sağlandığı istatistiksel olarak kanıtlanmıştır.

Çizelge 6.5 Bayesci yöntem ile elde edilen istatistikler

Parametre	$\hat{\beta}$	Exp($\hat{\beta}$)	SS	%2.5	%97.5	Örnek	Başlama
Sabit	11.90223	147.72	8.47170	-3.11615	29.9593	1000	99001
Nüfus	0.05229	1.053	0.02201	0.01377	0.0998	1000	99001
İhr	0.16783	1.182	0.09245	-0.01837	0.3316	1000	99001
İth	0.15183	1.163	0.09890	-0.00556	0.3775	1000	99001
Kadın_mv	0.41872	1.520	0.15591	0.117633	0.78240	1000	99001
Doğum	0.420212	1,522	3.54912	-0.89632	5.20511	1000	99001
Yaşmem	-6.84683	0.001	2.60164	-12.9435	-2.45507	1000	99001
Kadın_iş	0.05399	1.055	0.16803	-0.25042	0.41285	1000	99001
Sağlık	1.42337	4.151	0,033442	0.09428	2.68094	1000	99001

Çizelge 6.5.'de Bayesci lojit model parametrelerinin tahmin değerleri standart sapması (SS), MC hata değerleri Markov zincirinde elde edilen örneklemin başlangıç kısmı çıkarılmalıdır. Bu uygulamada iterasyon sayısı 100000 alınmıştır ve ilk 1000 terim yakınsamayı sağlamak için modelden çıkarılmış, 99001 örneklem daha çekilerek parametre tahmini yapılmıştır.

Çizelge 6.5'deki tahminci değerleri 6.1 eşitliğinde yerine yazıldığı takdirde;

$$log\left(\frac{p}{1-p}\right) = 11.902223 + 0.05229nufus + 0.16783ihr + 0.15182ith + 0.41872kadın\_mv + 0.420212doğum - 6.84683yaşmem + 0.05399kadın_{iş} + 1.42337sağlık \quad (6.2)$$

Eşitlik (6.2) elde edilir. Çizelge 6.5.'de Bayesci lojit modelde her bir parametre için *% 2.5* ve *% 97.5* güven aralığı değerleri verilmiştir. Çizelge 6.5'e göre nüfus, kadın_mv (parlemontadaki kadın parlementer oranı), yaşmem (yaşam memnıniyeti), sağlık (GSYiH'dan sağlık harcamaları için atrılan oran) parametreleri *% 2.5-% 97.5* güven aralığında *0* değerini içermemektedir. Yani % 2.5- % 97.5 güven aralığında bu parametreler anlamlıdır (Avcı, 2013). Dolayısıyla *% 5* önem düzeyinde nüfus, yaşmem, sağlık, kadın_mv parametreleri OECD ülkesinin AB üyesi olması üzerinde etkilidir Tekrar çizelge 6.5 incelendiğinde OECD ülkelerinin AB üyesi olmasında nufus değişkeni 1.05 kat yaşmem değişkeni 0.001 kat meclisteki kadın milletvekili oranı 1.52

kat ve GSYİH'dan sağlık harcamaları için ayrılan oran 4.52 kat etkili olduğu söylenebilir.

Çizelge 6.6. Bayesci sonsal özetler için SS ve MChata değerleri

Parametre	SS	MC Hata	SS*%5
Sabit	8.47170	0.24620	0.42355
Nüfus	0.022001	0.000675	0.00110
İhr	0.09245	0.00220	0.00462
İth	0.09890	0.002575	0.00494
Kadın_mv	0.15591	0.004214	0.00779
Doğum	3.54912	0.0924600	0.17745
Yaşmem	2.60164	0.0654260	0.13008
Kadın_iş	0.16803	0.005172 9	0.00840
Sağlık	0.66853	0.015882	0.033442

Bayesci istatistikte kullanılan MCMC yönteminin doğru sonuç vermesi için yakınsamanın gerçekleşmesi gerekir. Tekrar Çizelge 6.5. incelendiğinde her bir parametre için MC hata değeri, SS değerinin *%5*'inden küçük olduğu için Thumb kuralına göre yakınsama sağlanmaktadır (Hosmer ve Lemeshow, 2013). Yakınsamanın sağlandığına dair diğer kanıtlar ise parametrelere ait iz grafikleri, sonsal yoğunluk grafikleri ve otokorelasyon grafikleridir. Daha sonraki bölümde verilecek olan bu grafikler incelendiğinde tüm parametreler için yakınsamanın sağlandığı görülecektir.

Çizelge 6.7 Klasik ve Bayesci yaklaşımda bazı değerlerin karşılaştırılması

Ölçekler	Klasik	Bayes
Genel Yüzde	60.6	60.6
Doğru sınıflandırma oranı	89.7	94.2
McFadden R^2	0.66	0.53
AIC	38.440	32.740
BIC	51.909	46.209

.

Çizelge 6.7’ten anlaşımlaktadır ki 34 OECD ülkesinden AB üyesi olanların oranı 60.6’dır. tekrar çizelge 6.7 incelendiğinde görülmektedir ki Bayesci lojit model ile oluşturulan doğru sınıflandırma oranı daha yüksek ve McFadden R^2 katsayısına göre kurulan her iki modelde uyumludur. Aynı zamanda çizelge 6.7’de klasik ve Bayesci yöntemi karşılaştırmak için AIC ve BIC değerleri verilmiştir. Her iki değerde Bayesci yöntem ile kurulan modelde daha küçük olduğu için Bayesci yöntemin kullanılmasının daha uygun olduğu söylebilir.

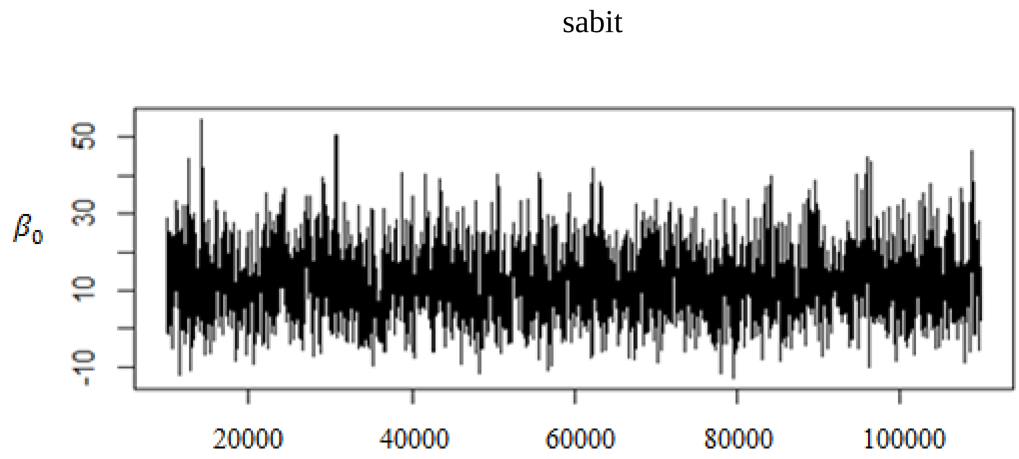
Çizelge 6.8. Bayesci ve klasik model için güven aralıkları karşılaştırması

Parametre	%2.5-%97.5 (BI)	%2.5-%97.5 (FI)
Sabit	(-3.116165- 30.5146)	(-0.1298063938-26.319459859)
Nüfus	(0.013775- 0.0998)	(-0.0006682814-0.005369104)
İhr	(-0.018378- 0.33163)	(-0.0046511593-0.015777568)
İth	(-0.005556-0.37753)	(-00061552836-0.015250879)
Kadın_mv	(0.117633-0.72809)	(-0.00050335578- 0.047768913)
Doğum	(-0.896325-5.20511)	(-0.6330541244-0.389227823)
Yaşmem	(-12.943501- -2.4550)	(0.6454193527- -0.049603356)
Kadın_iş	(-0.250428- 0.41285)	(-0.0178650302- 0.027224204)
Sağlık	(0.094288-2.68094)	(-0.0682819373-0.172320909)

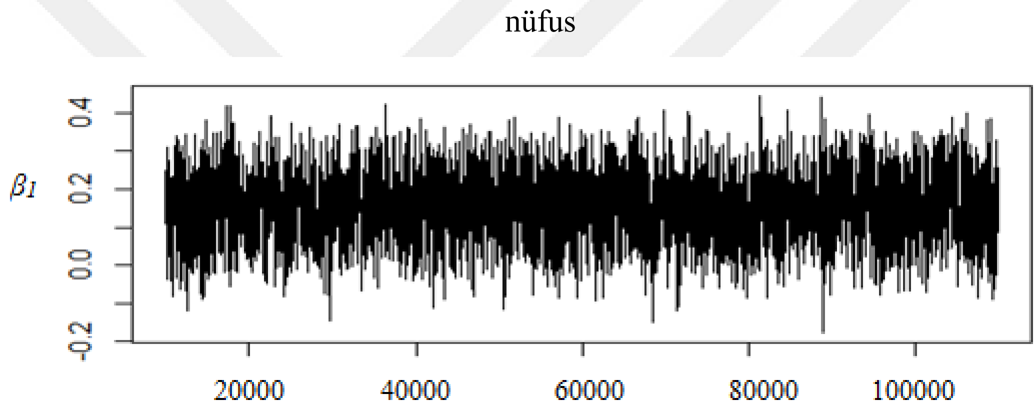
Çizelge 6.8. ‘de tüm parametrelerin *% 95* güven aralığında klasik ve Bayesci güven aralıkları karşılaştırılmıştır. Çizelge 6.8. incelendiğinde Bayesci ve klasik yaklaşımla elde edilen güven aralıkları birbirinden farklıdır. Bu farklılığın temel sebebi örneklem sayısının az olmasıdır (Santos ve ark., 2009).

6.2. Modelde Uygulanan Yakınsaklık Testleri

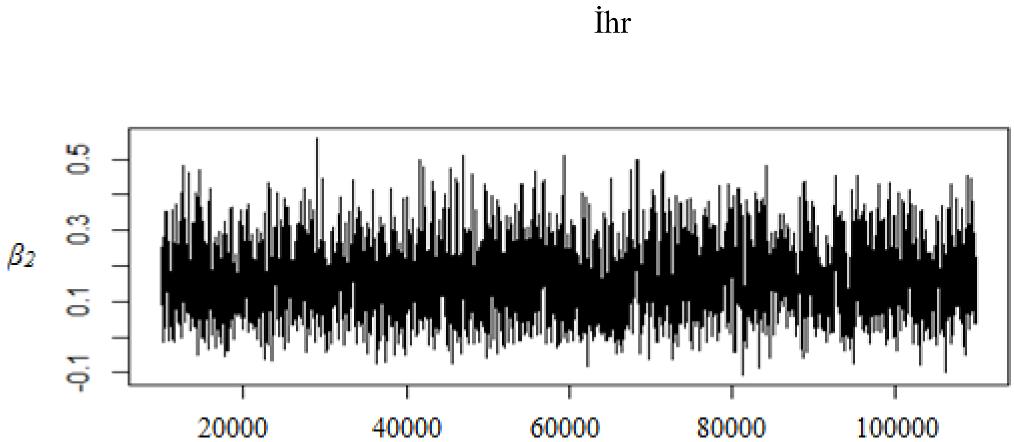
Tezin bu bölümünde Bayesci yöntemde Gibbs örnekleme algoritması elde edilen yakınsaklık sonuçlarını araştırmak için iz grafikleri, Geweke yakınsaklık testi, parametrelere ilişkin sonsal olasılık yoğunluk grafikleri, parametre zincirlerine bağlı olarak otokorealsyon katsayıları ve otokorelasyon grafikleri incelenecektir.



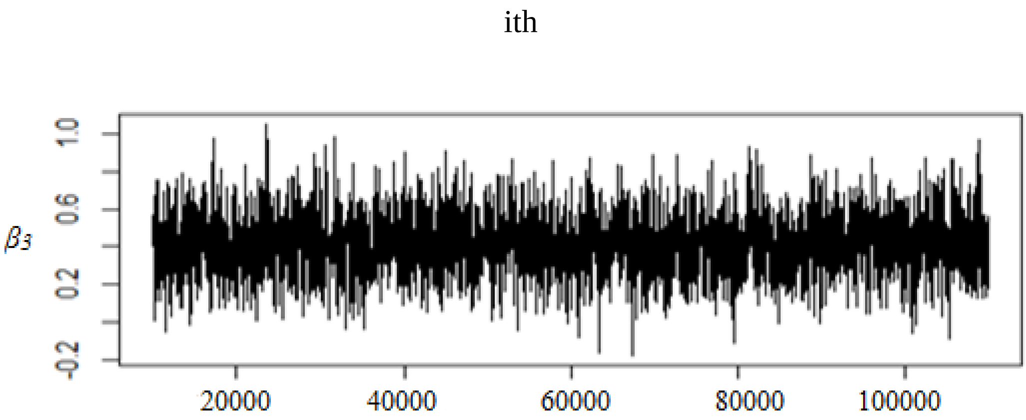
Şekil 6.1. Sabit parametrenin iz grafiği (β_0).



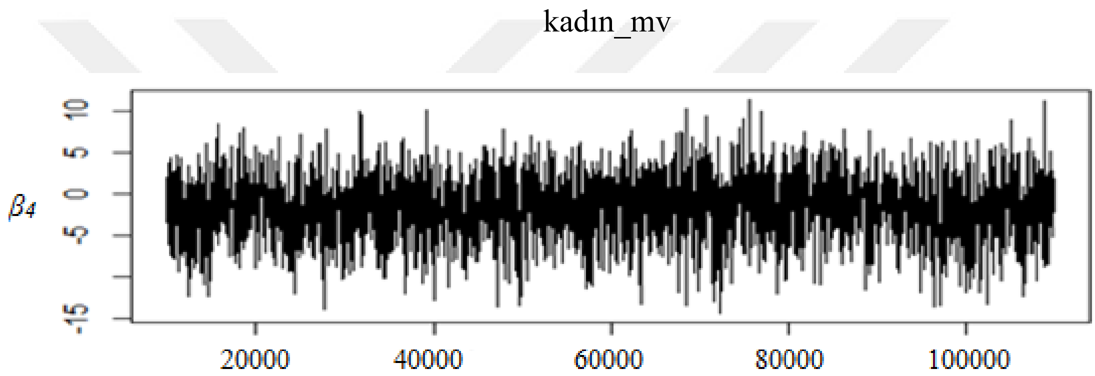
Şekil 6.2. β_1 parametresine ait iz grafiği.



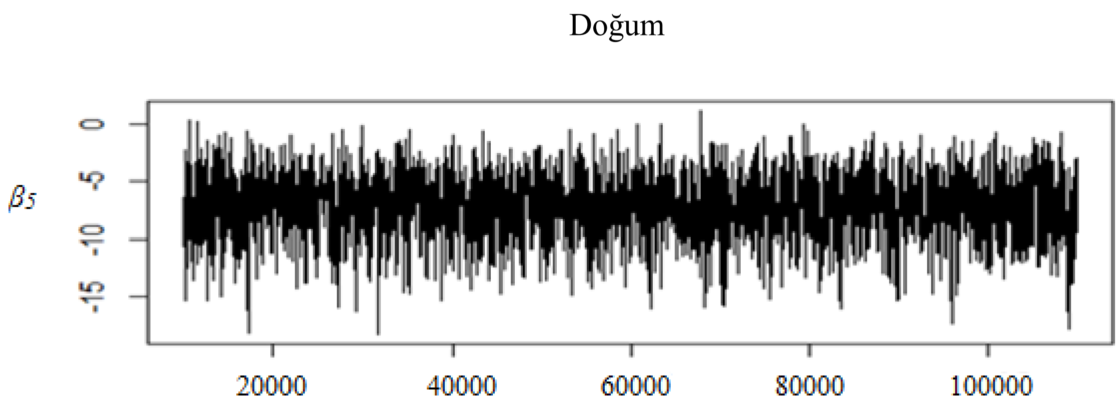
Şekil 6.3. β_2 parametresine ait iz grafiği.



Şekil 6.4. β_3 parametresine ait iz grafiği.

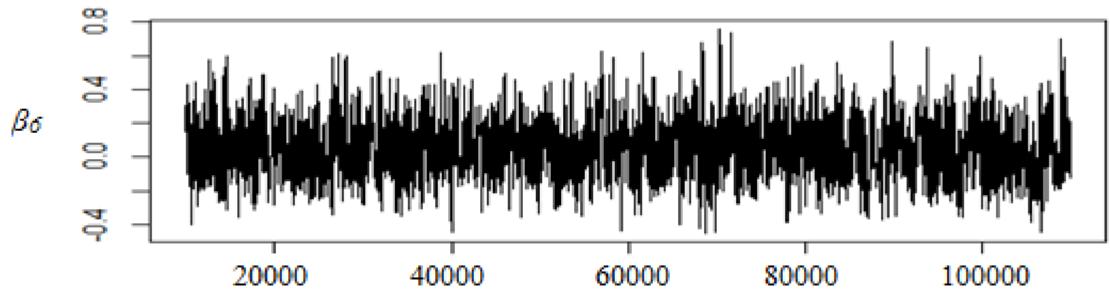


Şekil 6.5. β_4 parametresine ait iz grafiği.



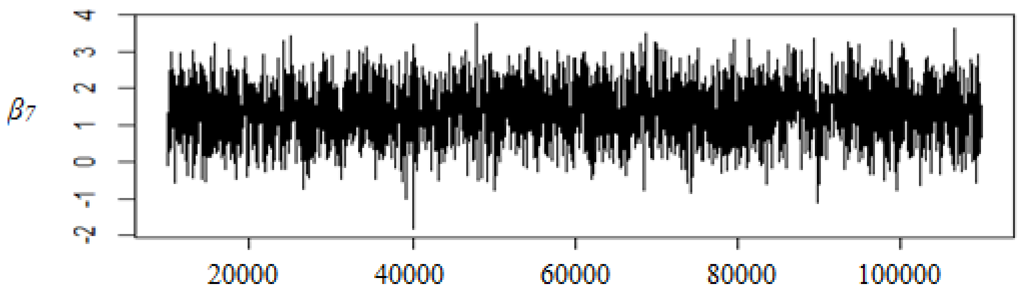
Şekil 6.6. β_5 parametresine ait iz grafiği.

Yaşmem



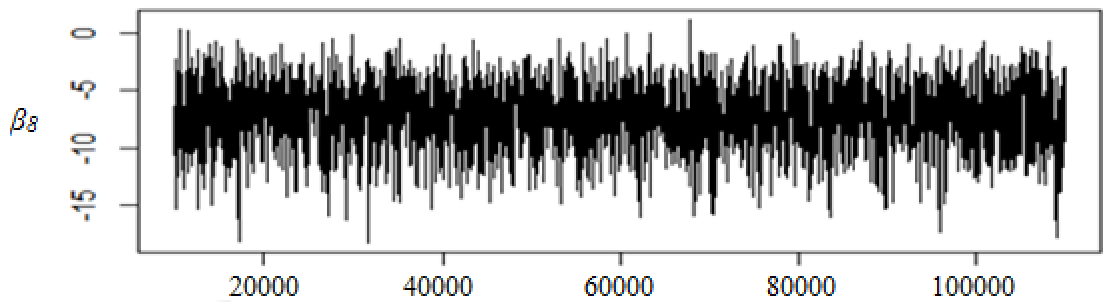
Şekil 6.7. β_6 parametresine ait iz grafiği.

kadın_iş



Şekil 6.8. β_7 parametresine ait iz grafiği.

sağlık



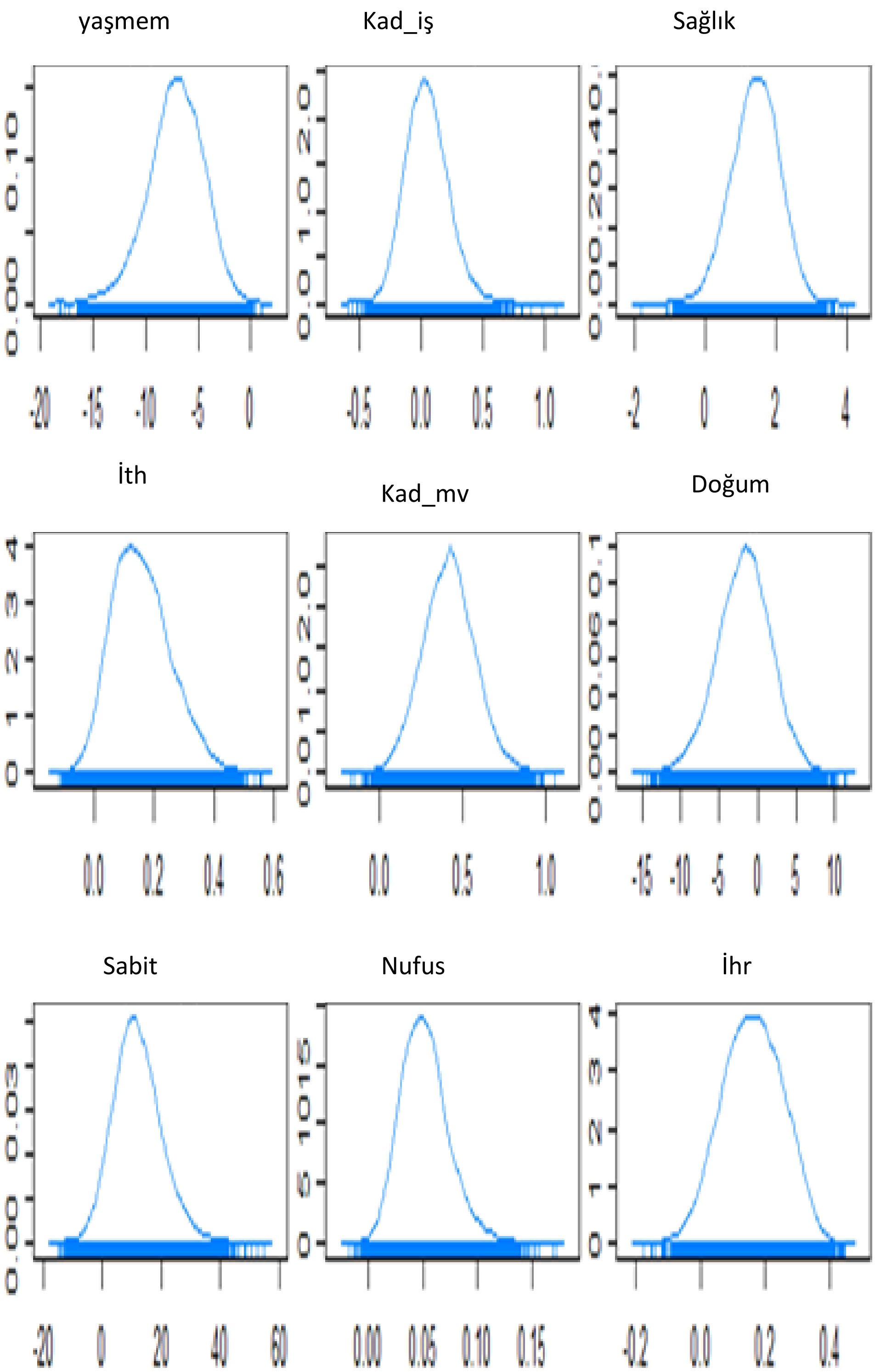
Şekil 6.9. β_8 parametresine ait iz grafiği.

Yukarıdaki Şekil 6.1. ve Şekil 6.9. arasındaki tüm parametreler şekil 3.4'teki (Bknz sayfa:28) good mix'e benzediğinden dolayı yakınsama sağlanmıştır. Şekil 6.1 ile şekil 6.9 arasındaki tüm grafikler Markov zincirinin sonsal dağılımı ne kadar yavaş veya hızlı aradığını gösterir (Burke, 2012; Ekici, 2005; Sahlin, 2011). Zincirlerde çok büyük saçılımlar gözlenmemektedir. Bu durum her bir zincirin başlangıç değerinin etkisinden kurtulduğu ve her bir zincirin yakınsama sağladığını göstermektedir.

Çizelge 6.9 Geweke yakınsaklık testi

Değişken	Z-testi
Sabit	0.69381
Nufus	-1.34032
İhr	1.0187
İth	-1.7552
Yaşmem	0.31675
kad_mv	0.36204
sağlık	0.01868
doğum	-0.07091
kad_iş	-0.99034

Geweke yakınsaklık testi ile ilgili açıklama çizelge 3.3'te (Bknz sayfa:31) aktarılmıştır. Bu modelde hipotez testleri %95 güven aralığında kurulduğundan dolayı *Z* tablo değeri 1.96'dır. Çizelge 6.9'daki tüm değişkenlerin *Z* test istatistiği değeri ise 1.96 değerinden küçük olduğundan dolayı yakınsama sağlanmıştır (Santos ve ark., 2009).



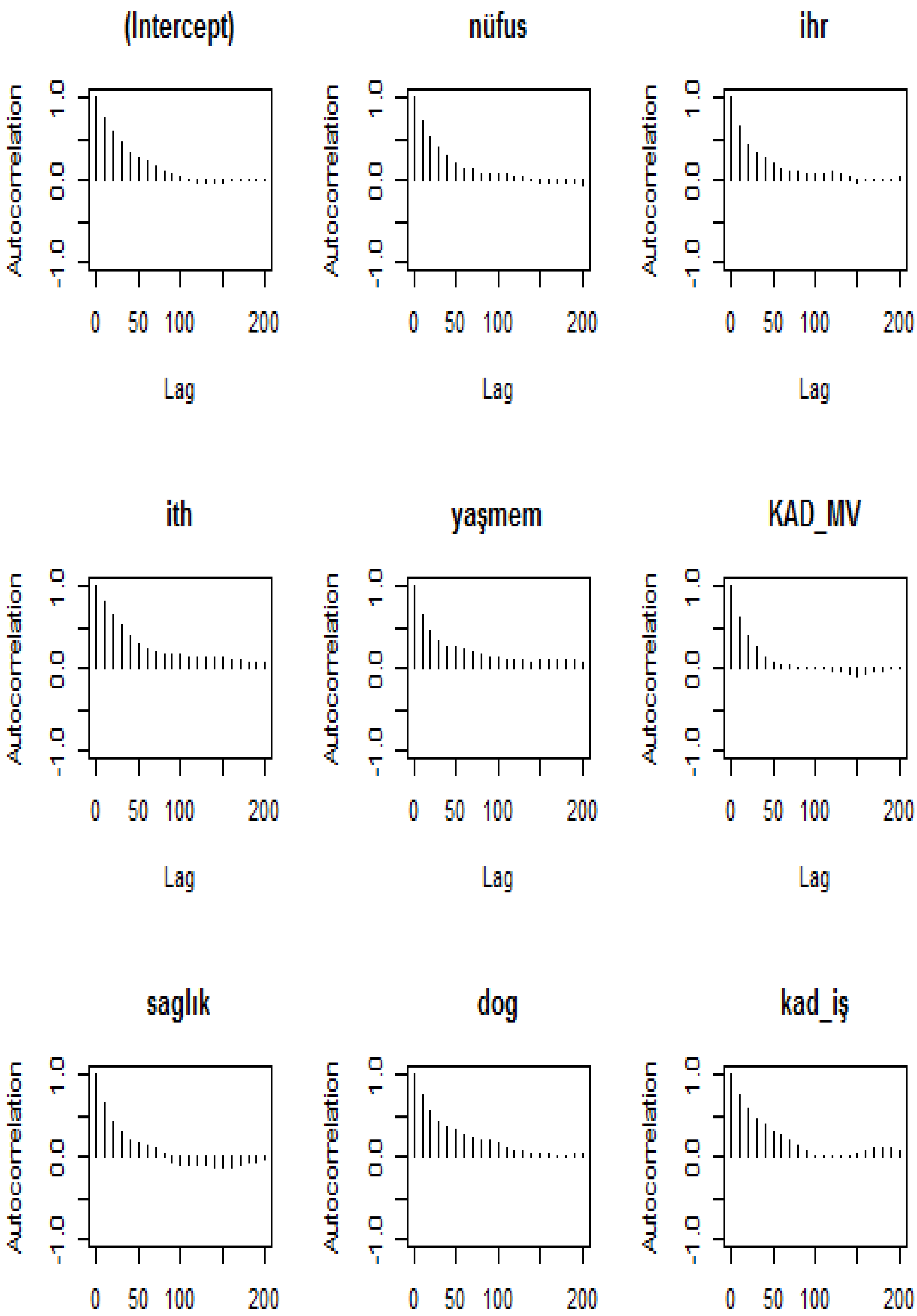
Şekil 6.10. Parametrelere ilişkin sonsal olasılık yoğunluk grafikleri.

Bayesci yaklaşımda sonsal olasılık yoğunluk grafiği çan şeklinde olduğunda veya normal dağılıma benzediğinde sonsal dağılıma ulaşıldığı kabul edilmektedir (Sahlin, 2011; Ekici, 2005; Burke, 2012). Şekil 6.10'deki tüm değişkenlerin sonsal olasılık yoğunluk grafikleri şekil (3.7)'ye benzediğinden dolayı tüm değişkenlerde yakınsama sağlanmış ve sonsal dağılıma ulaşılmıştır. (Bknz: sayfa: 30)

Çizelge 6.10. Parametre zincirlerine bağlı olarak otokorelasyon katsayıları.

Parametre	Gecikme 5	Gecikme 25	Gecikme50	Gecikme 100
Sabit	0.86586	0.511320	0.27623	0.00792
Nufus	0.80662	0.36080	0.14406	0.01928
İhr	0.79421	0.36243	0.27671	0.07451
İth	0.89179	0.57603	0.32634	0.18475
Yaşmem	0.80270	0.36078	0.23428	0.18475
kadın_mv	0.77748	0.30195	0.07429	-0.00855
sağlık	0.77668	0.27528	0.11000	-0.10776
doğum	0.85132	0.44375	0.26843	0.05701
kadın_iş	0.85022	0.48021	0.27234	0.014596

Çizelge 6.10.'da ki modelde her bir parametre için zincir otokorelasyon katsayıları hesaplanmıştır. Yani modeldeki her bir zincir için örneklemler arası bağımlılık ölçülerek her bir simulasyon değerinin bir önceki değere bağlı olduğu zincir değerler oluşturulmuştur. Modeldeki zincir otokorelasyonu yeterince uzun çalıştığı takdirde sonsal dağılımda istenen halini alacak yani yakınsama sağlanacaktır. Bu modelde 100. gecikmede tüm parametrelerin sıfır değerine yaklaşmaktadır.



Şekil 6.11. Parametrelere iliskin otokorelasyon grafikleri

Otokorelasyon grafikleri Markov zincirinde ki her bir örnek değeri arasındaki bağımlılığı ölçmektedir. Düşük korelasyon yakınsamanın sağlandığı anlamına gelmektedir (Sahlin, 2011; Ekici, 2005; Burke, 2012). Şekil 6.11.'de her bir değişken için otokorelasyon grafiği Şekil 3.5'e (Bknz sayfa:29) benzemektedir. Yani belli bir değerden sonra sıfır değerine yaklaşmaktadır ki bu yakınsamanın sağlandığı anlamına gelmektedir.

7. SONUÇ

Bu tez çalışmasında, bağımlı değişken kategorik olduğundan dolayı lojistik regresyon analizi kullanılmıştır. Lojistik regresyonda model parametreleri EÇO yöntemi ile tahmin edilmiştir. Fakat EÇO yöntemi örneklem sayısı yetersiz olduğunda yansızlık, etkinlik ve normallik gibi varsayımları sağlamaz. Dolayısı ile eğer örneklem sayısı yeterli değil ise EÇO güvenilir sonuçlar vermez. Bu durumda klasik lojistik regresyon modeline bir alternatif olarak örneklem büyüklüğü ile ilgili herhangi bir sınırlama getirmeyen, Bayesci lojistik regresyon kullanılabilir (Cengiz ve ark., 2012; Hosmer ve Lemeshow, 2013).

Tezin uygulama bölümünde bağımlı değişkenin iki durumlu olduğu OECD örneği hem lojistik regresyon analiz ile hem de Bayesci lojistik regresyon analizi ile incelenmiştir. Elde edilen sonuçlardan görülmüştür ki OECD örneği lojistik regresyon ile incelendiğinde sadece yaşam memnuniyeti %95 güven aralığında sıfır değerini içermemektedir. Dolayısı ile her bir OECD ülkesinin AB üyesi olup olmaması üzerinde sadece yaşam memnuniyeti parametresi etkilidir. Bayesci lojit modelde ise her bir OECD ülkesin, toplam nüfusu , mecliste bulunan kadın parlamenter oranı, GSYİH 'dan sağlık harcamaları için ayırdıkları bütçe oranları, yaşam memnuniyet oranları *% 95* güven aralığında 0 değerini içermemektedir. Dolayısıyla Bayesci lojit modele göre bu parametreler bir OECD ülkesinin AB üyesi olup olmaması üzerinde etkili olduğu söylenebilir. Ayrıca lojistik regresyon analizinde elde edilen güven aralıkları da Bayesci lojistik regresyonda elde edilen güven aralıkları arasında fark vardır. Bayesci lojit model ve klasik lojit modelin sonuçlarının bu kadar farklı çıkmasındaki temel sebep örneklem sayısının küçük olmasıdır. Eğer örneklem sayısı yeterince büyük olursa ve önsel dağılım belirsiz önsel dağılım seçilirse lojistik regresyon analizinin ve Bayesci lojistik regresyon analizinin birbirine yakın sonuçlar vermesi beklenir (Santos ve ark., 2009). Örneklem sayısı küçük olduğunda Bayesci lojit ve lojit modeller benzer sonuçlar verse bile parametre tahmininde kullanılan EÇO yönteminde modeldeki her bir parametrenin en az 10 katı kadar gözlem sayısı olmasından eğer gözlem sayısı yeterli değil ise EÇO yöntemi asimptotik varsayımları sağlamayacağından dolayı elde edilen sonuçların çok güvenilir olmadığından ve bu durumda Bayesci lojit modellerin kullanılmasının daha doğru olacağından tezin daha önceki bölümlerinde bahsedilmişti. Bu durum uygulamada da açıkça görülmektedir.

Bayesci yaklaşımda Gibbs örneklemesinin doğru sonuç vermesi için sonsal dağılıma yakınsamanın gerçekleşmesi gerekmektedir. Yakınsamanın sağlanıp sağlanmadığını netleştirmek için örneklemdeki her bir parametrenin yakınsama durumları incelenmesi gerekir (Avcı, 2013; Cengiz,2013). Bu amaçla modeldeki her bir parametre için sonsal olasılık yoğunluk grafikleri, iz grafikleri ve otokorelasyon grafikleri incelenmiş ve görülmüştür ki modelde her bir parametre zinciri belli bir noktadan sonra durağanlığa ulaştığı için yakınsama sağlanmıştır. Bu durumda OECD örneğini Bayesci lojit model ile incelemede bir sakınca yoktur.

Sonuç olarak uygulamadan elde edilen bilgiler ışığında; açık bir şekilde görülmektedir ki yakınsama sağlandığı zaman özellikle küçük örneklemlerde Bayesci lojistik regresyon analizi klasik lojistik regresyon analizinden daha kullanışlı bir yöntemdir.

KAYNAKLAR

Agresti, A., 2002. ***Categorical Data Analysis***. Published by John Wiley & Sons, 721s. USA.

Agresti, A., Hitchock, D., 2005. Bayesian inference for categorical data analysis. ***Statistical Methods Applications,*** **14:** 297-330.

Akın, A., Yalnız, A., 2015. Yaşam memnuniyeti ölçeği (YMÖ) Türkçe formu: geçerlilik ve güvenirlik çalışması. ***Elektronik Sosyal Bilimler Dergisi***, **1**:54

Akın, F., 1996. ***Olasılık***. Ekin Kitabevi, ISBN:9757338923, 252s.

Albayrak, A. S., 2006. ***Çok Değişkenli İstatistik Teknikleri***. Asil yayın dağıtım, 497s. Ankara.

Albert, H. J., Chib, S., 1993. Bayesian analysis of binary polychotomous response data. ***American Statistical Association***, **88**(422): 669-679.

Alpar, R., 2011. ***Uygulamalı Çok Değişkenli İstatistiksel Yöntemler***. Detay yayıncılık, 853s. Ankara.

Altaş, D., Giray, T. S., 2008. Avrupa birliği ve OECD'ye üyelikte etkili olan ekonomik ve demografik özelliklerin incelenmesi. ***Marmara Üniversitesi İ.İ.B.F,*****1:** 2149-1844.

Alvin, Alvin C. R., Schaalje B. G., 2007. ***Linear Models in Statistics***. Published by John Wiley & Sons, 679s. Canada.

Avcı, E., 2013. Marmara üniversitesi öğrencilerinin kredi kartı sahibi olmalarını etkileyen faktörlerin Bayesci lojistik regresyon yardımı ile incelenmesi**.** ***İstatistik Araştırma Dergisi***, **10**(2):42-55.

Barak, A., 2005. ***Sıralı ve Multinominal Logit Modeller Üzerine Bir Uygulama***. (yüksek lisans tezi, basılmamış), Hacettepe Üniversitesi, Fen Bilimleri Enstitüsü Ankara.

Bayarri, M. J., Degrott, M., H., 1992**.** Difficulties and ambiguties in the defination of a llikehood function**.** ***Statistical Methods and Applications,*** **1**(1):1-15.

.

Bernardo, J. M., Degroft, M. H., Lindley, D. V., Smith, A. F. M., 1979. ***Bayesian Statistics***. Bayesian Methodology in Biostatistics**,** 324s. Spain.

Box, P. E. G., Tiao, G. C., 1973. ***Bayesian Inference in Statistical Analysis***. Addison-Wesley Publishing Company, 588s. California.

Burke, O., 2012. ***MCMC Output Analysis***. Department of Statistics, South Parks.

Burmaoğlu, S., Oktay, E. Özen., 2009. Birleşmiş Milletler kalkınma programı beşeri kalkınma endeksi verilerini kullanarak diskriminant ve lojistik regresyon analizinin sınıflandırma performanslarının karşılaştırılması. ***Kara Harp Okulu Savunma Bilimleri Dergisi,*** **8**(2):23-49.

Cowles, M. K., Carlin, B.P.M., 1996. Markov chain monte carlo convergence diagnostics: a comparative review**.** ***Journal of the American Statistical Association***, **91**(434): 883-904.

Casella, G., Berger, L. R., 1990. ***Statistical İnference***. Duxbury Press, 667s.

Cengiz, M., Şenol, T., Murat, N., Terzi, Y., 2013. Lojistik regresyon parametre tahmininde Bayesci bir yaklaşım. ***Akü Febid,*** **12** : 15-22.

Collet, D., 1991. ***Modelling Binary Data***. Chapman & Hall Publication, 381s.

Congdon, P., 2003. ***Applied Bayesian Modelling***. Wiley Series in Probability Statistics, England, 465s. England.

Congdon, P., 2005. ***Bayesian Models for Categorical Data***. Wiley Series in Probability Statistics, 447s. England.

Çokluk, Ö., 2010. Lojistik regresyon analizi kavram ve uygulama. ***Kuram ve Uygulamada Eğitim Bilimleri***, **10**(3): 1357-1407.

Çömlekçi, N., 1998. ***Temel İstatistik***. Nisan kitab evi, 497s.Eskişehir.

Ekici, O., 2005. **Bayesyen Regresyon.** (yüksek lisans tezi, basılmamış), İstanbul Üniversitesi, Sosyal Bilimler Ensitütüsü, İstanbul.

Ekici, O., 2009. İstatistikte bayesyen ve klasik yaklaşımın farklılıkları. ***Balıkesir Üniversitesi Sosyal Bilimler Dergisi,*** **12**: 89-101.

Gamerman, D., 2006. ***Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference***. Chapman and Hall Publication**,** 343s. Canada.

Gelfand, A. E., Smith, F. M., 1990. Sampling-based approaches to calculating marginal densities. ***Journal of the American Statistical Association,*** **85**:398-409.

Gelman, A., 2004. ***Bayesian Data Analysis*.** Chapman and Hall Publication/CLC, 695s. Canada.

Gelman, A., 2002. Prior distribution**. *Encyclopedia of Environmetrics***, **3**: 1634-1637.

Geyer, J., 1992. Practical markov chain monte carlo. ***Statistical Science***, **7**: 473-511.

Ghosh, K. G., Delampady, M., Samanta, T., 2006***. An İntroduction to Bayesian Analysis***, Springer, 356s. USA.

Gilks, W. R., 1996. ***Markov Chain Monte Carlo in Practice*.** Chapman & Hall Publication , 481s.

Groenewald, C., N., Mokgatlhe, L., 2004. Bayesian logistic regression, ***Computational Statistics and Data Analysis***, **48**: 857-868.

Gut, A., 2012***, Probability : A Gruduate Course.*** Springer Texts in Statistics, 619s. USA.

Hair, F. T., William, C. B., Babin, B. T., Anderson, E. R., 2006. ***Overview of Multivariate Methods***. Published by John Wiley & Sons , 721s.

Hastings, W. K., 1970. Monte carlo sampling methods using markov chains and their applications***. Biometrika***, **57**: 97-101.

Hoff, P. N., 2009. ***A First Course in Bayesian Statistical Methods***. Springer , 265s. London.

Holmes, C. G., Held, L., 2006. Bayesian auxiliary variable models for binary and multinominal regression Bayesian analysis. ***Bayesian Analysis***, **1**(1):145-148.

Hosmer, D. W., Lemeshow, S., 2013. ***Applied Logistic Regression***. Published by John Wiley & Sons, 518s. USA.

Howson, C., Urbach, P., 2006. ***Reasoning the Bayesian Approach***. Publication Data, 340s. USA.

İnal, M., Topuz, M., Oktay, U., 2006. Doğrusal olasılık ve lojit modellerle parametre tahmini. ***Sosyoekonomi***, **1**: 49-85.

Jude, G. G., 1984. ***The Theory and Pratice of Econometris*.** Library oj Congress Cataloging in Publication Data**,** 996s. Canada.

Kalaycı, Ş., 2010. *SPSS uygulamalı çok değişkenli istatistik teknikleri*. ***Asil Yayın****, 426s.* Ankara.

Lam, P., 2011. Erişim Adresi: http://www.people.fas.harvard.edu/~plam/teaching/methods Erişim Tarihi: 09.10.2015.

http://www.glicko.net/research/glickman-vandyk.pdf, Erişim Tarihi: 09.12.2015.

http://www.oecd.org/env/country-reviews/42198785.pdf, Erişim Tarihi:09.09.2015.

http://isites.harvard.edu/fs/docs/icb.topic196222.files, Erişim Tarihi, 09.11.2015

Lee ,Y., Nelder, J. A., Pawition, Y., 2006. ***Generalized Linear Models with Random Effects***. Chapman and Hall Publication, 411s. NewYork.

Lessage, J., 1999. ***Applied Data Econometrics Using Matlab***. Published by John Wiley & Sons, 348s. USA.

Lindley, V., D., 1983. Theory and pratice of Bayesian statistics. ***Journal of the royal statistical society***, **32**: 1-11.

Lynch, S. M., 2007. ***Introduction to Applied Bayesian Statistics and Estimation for Social Scientistics***. Springer, 376s. USA.

Maryann, G., 1993**.** ***A Bayesian Approach to Estimating the Regression Coefficients of A Multinominal Logit***. The University of Texas H.S.C. at Houston Sch. of Public Health, 132s**.** Texas**.**

Menard, S., 2002. ***Applied Logistic Regression Analysis***. Sage Publications, Second Edition, 120s**.** London.

Miller, I. M., Miller, M., John, E., Freund, 2001. ***Matematiksel İstatistik***. 6. Baskıdan çeviren Ümit Şenesen, Literatür yayınları. 644s. İstanbul.

Mccullgh, P., Nelder, A., 1989. ***Generalized Linear Models***. Chapman & Hall Publication , Second Edition, 526s. London.

Ntzoufras, I., 2009. Bayesian ***Modeling Using WinBUGS***. Wiley & Sons, Inc Publication, 491s. USA.

OECD Çevresel performans incelemeleri, http://www.oecd.org/env/country-reviews/42198785.pdf, Erişim Tarihi, 09.09.2015.

O'Hagan, A., Luce, R. B., 2003. ***A Primer on Bayesian Statistics in Healty and Outcomes Research***. Medtap International, 72s.

Oğuzlar, A., 2005. Lojistik regresyon analizi yoluyla suçlu profilinin belirlenmesi. ***İktisadi ve İdari Bilimler Dergisi***, **19**: 21-36.

Özdamar, K., 2013. ***Paket Programlar ile İstatistiksel Veri Analizi***. Kaan Kitapevi, 346s. Eskişehir.

Önalan, Ö., 2010. ***Stokastik Süreçler***. Avcıol Basım Yayın, 371s. İstanbul.

Richey, M., 2011. The eciltution of markov chain monte carlo methods. ***The American Mathematical Montly***, **5** (117): 383-413.

Sahlin, K., 2011**.** ***Estimating Convergence of Markov Chain Monte Carlo Simulations, Mathematical Statistics.*** Stockholm University, Master thesis, 65s.

Santos, M. A., Moala, F. A., Tachibana, V. M., 2009. Approximate Bayesian methods for logistic regression model. ***Sao Paulo***, **27**(2): 288-300.

Savchuk, V. P., Tsokos, C. T., 2011. ***Bayesian Theory and Methods Applications***. Atlantis press, 354s. USA.

Schinazi, R. B., 2010. ***Probability with Statistical Applications***. Springer, 349s. London.

Souza, D. P., Migon, H. S., 2004. Bayesian binary regression model: an aplication to in- hospital death after amı prediction**.** ***Versao Impressa,*** **24**(2): 253-267.

Statisticat, LLC, 2014. ***htpp://Bayesian inference.com***. Erişim Tarihi: 09.09.2014

Stigler, S., 1983. Who discovered Bayes's theorem. ***The American Statistican***, **37**: 290-296.

Syversen, A. R., 1998. Noninformative Bayesian priors interpretation and problems with construction and applications**.** ***Lectture6.***

Şenel, T., Cengiz, M. A., Savaş, N., Terzi, Y., 2009. Çoklu doğrusal regresyon seçiminde genelleştirilmiş toplamsal modellerin kullanılması. ***EÜFBED- Fen Bilimleri Enstitüsü***, **2**(2): 217-228.

Tatlıdil, H., 2002. ***Uygulamalı Çok Değişkenli İstatistiksel Analiz***. Ziraat Matbaacılık, 423s. Ankara.

Tektaş, D., 2006. ***İki Düzeyli Lojit ve Probit Modellerde Parametre Tahminleri***. (yüksek lisans tezi, basılmamış), Hacettepe Üniversitesi, Fen Bilimleri Enstitüsü, Ankara.

Tektaş, D., Günay, S. A., 2008. Bayesian approach to parameter estimation in binary logit models. ***Hacettepe Journal of Mathematics and Statistics***, **37**(2):167-176.

Tsai, C. H., 2004. Bayesian inference in binomial logistic regression: a case study of the 2002 taipei mayoral election**.** Election Study Center , ***National Changchi University*****, 94**(3):103-123.

Uçar, Ö., 2004**. *Nitel verilerin analizinde lojit ve probit modeller***. (yüksek lisans tezi, basılmamış), Hacettepe Üniversitesi, Fen bilimleri Ensitüsü, Ankara.

Ülker, Tümlü, G., Recepoğlu, E., 2013. Üniversite akademik personelinin psikolojik dayanıklılık ve yaşam doyumu arasındaki ilişki. ***Yüksek Öğretim ve Bilim Dergisi***, **3**(3): 205-213.

Wakefield, W., 2012. ***Bayesian and Frequentist Regression Methods*****.** Springer, 696s. USA.

Walsh, B., 2004. Markov Chain Monte Carlo and Gibbs Sampling**. *Lecture notes for EEB*, 581**:1-24.

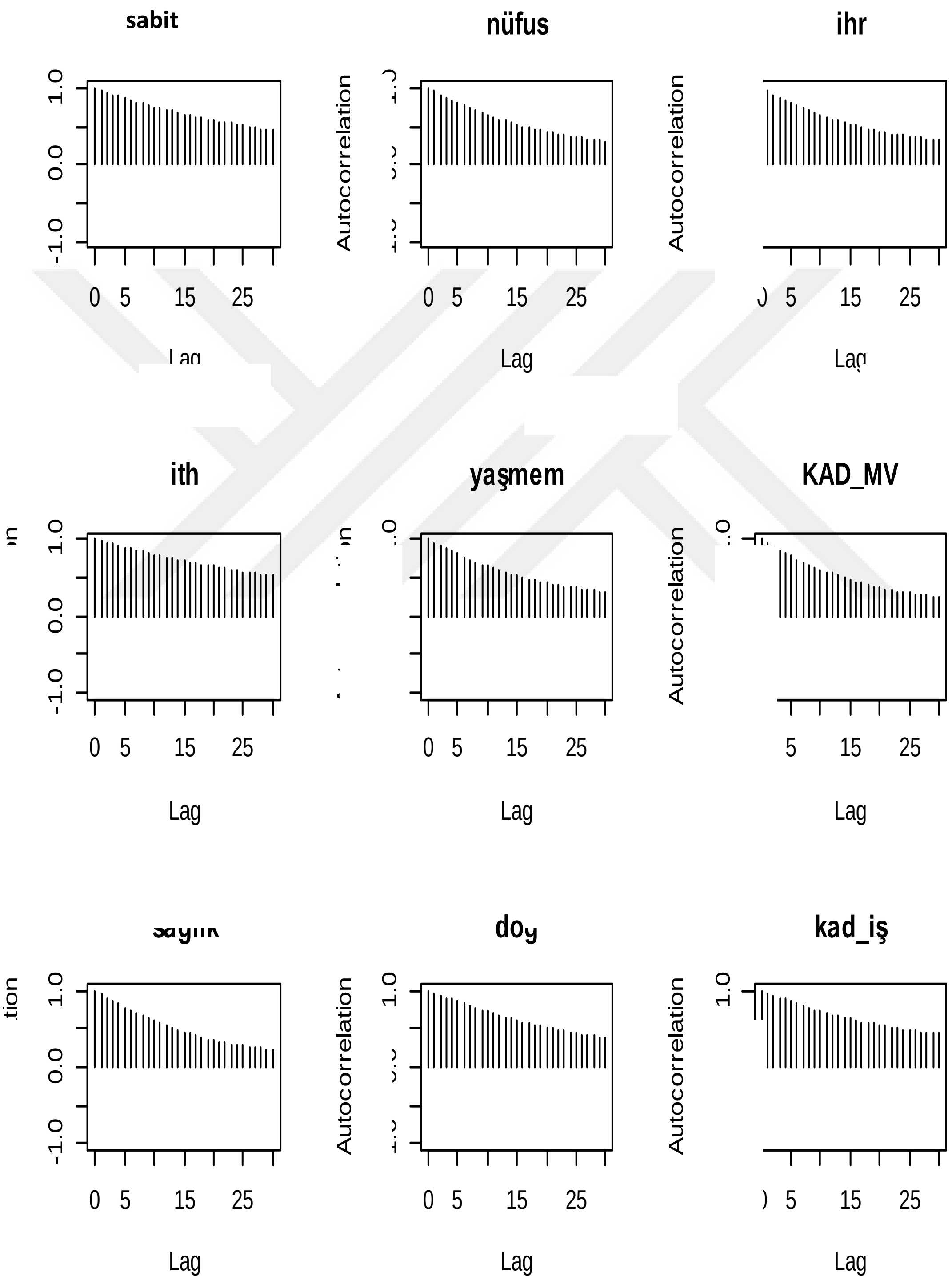
Yılmaz, V., Çelik, E., 2009**. *Matematiksel İstatistik Ders Notları II***, (yayınlanmamış).

Yüksel, T., Şenel, T., Şenel, E., Terzi,Y., 2013**.** Doğrusal regresyonda markov zincirleri monte carlo yakınsama kriterlerinin karşılaştırılması**. *AAOJ, Scientific Science***, **1**:27-33.

Zellner, A., 1976. ***An Introduction to Bayesian Inference in Econometric***. A Wiley Interscience Publication, 431s. USA.

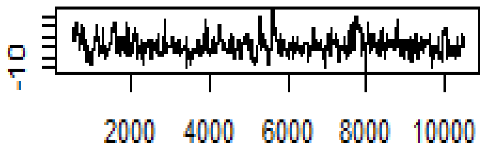
EKLER

EK1. Parametrelere ilişkin yakınsamanın sağlanmadığı otokorelasyon grafikleri

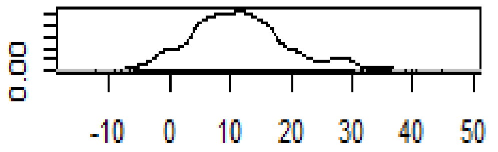


EK2. Parametrelere ilişkin yakınsamanın sağlanmadığı sonsal olasılık yoğunluk ve iz grafikleri

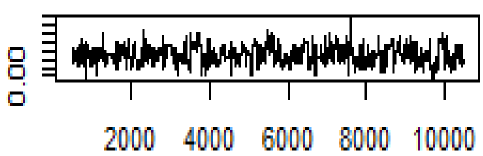
sabit



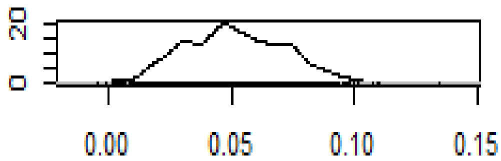
sabit



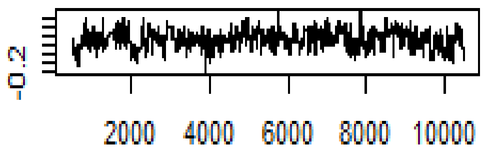
nufus



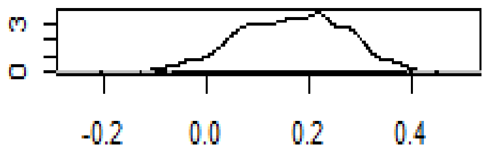
nufus



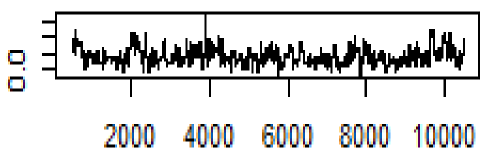
ihr



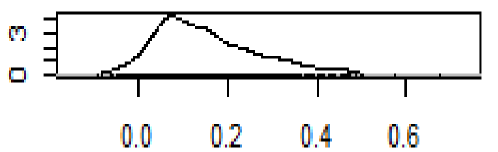
ihr



ith

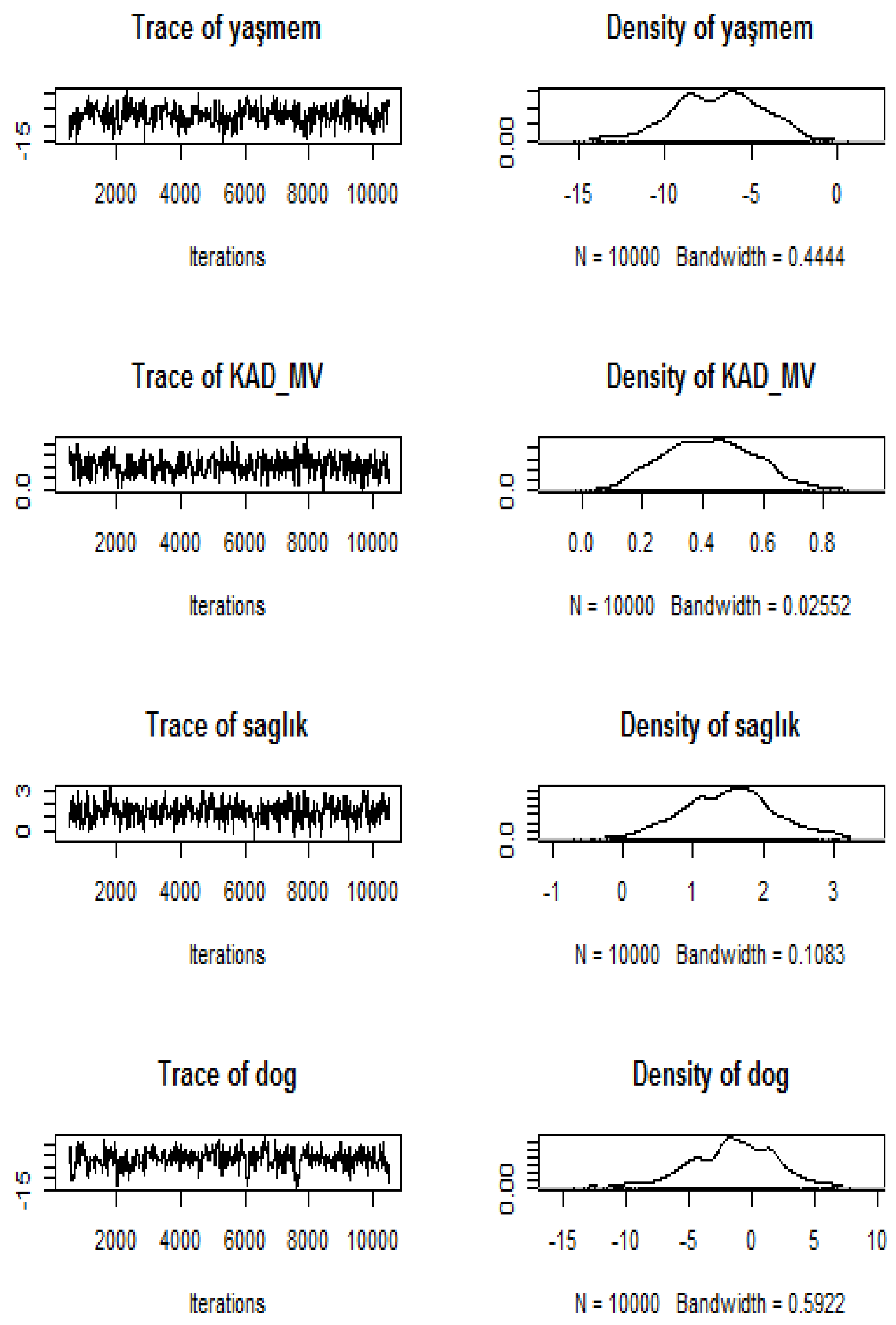


ith



yaşmem

yaşmem


Trace of yaşmem
Iterations
Density of yaşmem
N = 10000 Bandwidth = 0.4444
Trace of KAD_MV
Iterations
Density of KAD_MV
N = 10000 Bandwidth = 0.02552
Trace of saglık
Iterations
Density of saglık
N = 10000 Bandwidth = 0.1083
Trace of dog
Iterations
Density of dog
N = 10000 Bandwidth = 0.5922

ÖZ GEÇMİŞ

Nisan 1988'de Iğdır'da doğdu. İlk ve orta öğrenimini Iğdır, Aralık ilçesinde tamamladı. 2006-2010 yılları arasında Elazığ Fırat Üniversitesi İstatistik Bölümünde okudu. Aynı üniversitenin Matematik Bölümünde çift ana dal yaptı. 2010 yılında her iki bölümden de mezun oldu. 2012 yılı Ekim ayından itibaren Yüzüncü Yıl Üniversitesi Fen Fakültesi İstatistik Bölümünde araştırma görevlisi olarak görev yapmaktadır. Evli ve bir çocuk annesidir.