

**T.C.
MİMAR SİNAN GÜZEL SANATLAR ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ**

**BOYLAMSAL VE SAĞ KALIM VERİLERİNİN
BİRLEŞİK MODELLENMESİ ÜZERİNE
BİR ÇALIŞMA**

DOKTORA TEZİ

Zeynep ATLI

İstatistik Anabilim Dalı

İstatistik Programı

Tez Danışmanı:

Tez Eş Danışmanı:

HAZİRAN 2020

T.C.
MİMAR SİNAN GÜZEL SANATLAR ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

BOYLAMSAL VE SAĞ KALIM VERİLERİNİN
BİRLEŞİK MODELLENMESİ ÜZERİNE
BİR ÇALIŞMA

DOKTORA TEZİ

Zeynep ATLI
(20120897003)

İstatistik Anabilim Dalı

İstatistik Programı

Tez Danışmanı:

Tez Eş Danışmanı:

HAZİRAN 2020

Zeynep ATLI tarafından hazırlanan “BOYLAMSAL VE SAĞ KALIM VERİLERİNİN BİRLEŞİK MODELLENMESİ ÜZERİNE BİR ÇALIŞMA” adlı bu tezin Doktora tezi olarak uygun olduğunu onaylarım.

Tez Yöneticisi

Bu çalışma, jürimiz tarafından İstatistik Anabilim Dalında Doktora tezi olarak kabul edilmiştir.

Başkan :

Jüri Üyeleri :

.....

.....

Bu tez, Mimar Sinan Güzel Sanatlar Üniversitesi Fen Bilimleri Enstitüsü tez yazım kurallarına uygundur.

Bu tez çalışması TÜBİTAK BİDEB 2214-A Yurtdışı Doktora Sırası Araştırma Burs Programı tarafından 1059B141800367 numaralı proje ile desteklenmiştir.

ÖNSÖZ

Doktora süreci boyunca çalışmaya değerli fikirleriyle yön veren, her zaman yanımda olduğunu ve bana güvendiğini hissettiğim değerli tez danışmanım Prof. Dr. Gülay Başarır'a, bu konuda çalışmam için beni teşvik eden ve destekleyen, sorduğum her soruya sabırla cevap veren değerli eş danışmanım Prof. Dr. Mithat Gönen'e çok teşekkür ederim.

Tez izleme toplantılarında destekleyici konuşmalarından ve tezin şekillenmesindeki önemli katkılarından dolayı değerli hocalarım Prof. Dr. Aydın Erar'a ve Prof. Dr. Müjgan Tez'e çok teşekkür ederim.

Doktora eğitimim için geldiğim ve bu vesileyle yolumun iyi ki kesiştiği, birlikte çalışmaktan mutluluk duyduğum Mimar Sinan Güzel Sanatlar Üniversitesi İstatistik Bölümündeki hocalarıma ve Araştırma Görevlisi arkadaşlarıma bu süreçte yanımda oldukları için teşekkür ederim.

Son olarak bu süreci benimle birlikte yaşayan ve her zaman yanımda olduklarını hissettiren canım babama, anneme, ablama ve kardeşime çok teşekkür ederim. Siz olmasaydınız bu süreç benim için daha zor olurdu.

Haziran 2020

Zeynep ATLI

İÇİNDEKİLER

Sayfa

ÖNSÖZ	iv
İÇİNDEKİLER	v
KISALTMALAR.....	vii
ÇİZELGE LİSTESİ.....	viii
ŞEKİL LİSTESİ.....	ix
SEMBOL LİSTESİ.....	x
ÖZET	xi
SUMMARY	xii
1. GİRİŞ.....	1
2. BOYLAMSAL VERİ ANALİZİ.....	7
2.1 LME Modeli	8
2.2 LME Modelinde Parametre Tahmini	10
2.2.1 ML Yöntemi	10
2.2.2 REML Yöntemi	13
2.2.3 EM Algoritması	16
2.2.4 NR Algoritması	17
2.3 Rastgele Etkiler Parametrelerinin Tahmini.....	19
3. SAĞ KALIM (SURVIVAL) ANALİZİ	21
3.1 Sağ Kalım Analizinde Kullanılan Temel Fonksiyonlar.....	22
3.1.1 Sağ Kalım Fonksiyonu	22
3.1.2 Olasılık Yoğunluk Fonksiyonu	23
3.1.3 Tehlike Fonksiyonu	24
3.2 Sağ Kalım Fonksiyonunda Kullanılan Bazı Önemli Dağılımlar	26
3.2.1 Üstel Dağılım.....	26
3.2.2 Weibull Dağılımı	27
3.2.3 Log-Normal Dağılım	28
3.3 Cox Oransal Tehlike Modeli.....	28
3.3.1 Model Anlamlılığının Testi	31
3.4 Zamana Bağlı Değişkenler	32
3.5 Genişletilmiş (Extended) Cox Modeli.....	33
4. BOYLAMSAL VE SAĞ KALIM VERİLERİ İÇİN BİRLEŞİK MOD- ELLEME	35
4.1 Birleşik Model Yapısı.....	35
4.1.1 Sağ Kalım Alt Modeli	37
4.1.2 Boylamsal Alt Model	39

4.2 Birleşik Modelde Parametre Tahmin Yöntemleri.....	39
4.2.1 İki Aşamalı Yaklaşım	39
4.3 Birleşik Modelde ML Yöntemi	40
4.3.1 Birleşik Modelde EM Algoritması	41
4.3.2 Sayısal İntegrasyon Yöntemi.....	45
4.4 Birleşik Modelde Hipotez Testi.....	47
4.5 Birleşik Modelde Artık Analizi	48
5. BİRLEŞİK MODELDE PARAMETRE TAHMİNLERİ İÇİN BAYESÇİ YAKLAŞIM.....	51
5.1 Bayesçi Yaklaşım.....	52
5.2 Markov Zinciri Monte Carlo Yöntemi	55
5.2.1 Gibbs Örneklemesi	56
5.2.2 Metropolis Hasting (MH) Algoritması.....	57
5.3 Birleşik Modelde Olabilirlik Fonksiyonu ve Önsel Dağılımlar	58
5.4 Birleşik Modelde Sonsal Dağılımlar	60
5.4.1 Boylamsal Alt Model Parametreleri İçin Tam Koşullu Dağılımlar	60
5.4.2 Sağ Kalım Alt Modeli Parametreleri İçin Tam Koşullu Dağılımlar	63
5.5 Markov Zincirinde Yakınsamanın Değerlendirilmesi	63
6. UYGULAMA	67
6.1 Klasik Yaklaşım ile Parametre Tahminleri	68
6.2 Bayesçi Yaklaşım İle Parametre Tahminleri.....	70
6.2.1 Parametrelerin Önsel Seçimleri.....	70
6.2.2 Gibbs Örneklemesi ve MH Algoritması.....	71
6.2.3 Parametre Tahminleri	73
6.2.4 Yakınsama Tanıları	74
7. SİMÜLASYON	79
7.1 Birleşik Verinin Üretilmesi.....	79
7.2 Simülasyon Sonuçları.....	80
8. SONUÇ VE ÖNERİLER	87
KAYNAKLAR.....	91
EKLER	98
10. Matris İşlemleri	99
10.1 Matrisin Tersinin Türevi	99
10.2 Matris Determinantının Logaritmasının Türevi	99
11. Önsel Seçiminde Kullanılan Dağılımlara Ait Yoğunluk Fonksiyonları.....	100
11.1 Çok Değişkenli Normal Dağılım	100
11.2 Wishart ve Ters Wishart Dağılımı	100
11.3 Tek Düze Dağılım.....	100
ÖZGEÇMİŞ	101

KISALTMALAR

AIC	: Akaike Bilgi Kriteri (Akaike Information Criteria)
BIC	: Bayesçi Bilgi Kriteri (Bayesian Information Criteria)
BLUE	: Doğrusal En İyi Yansız Tahmin (Best linear unbiased estimation)
BLUP	: Doğrusal En İyi Yansız Kestirim (Best linear unbiased prediction)
EM	: Beklenti Maksimizasyonu (Expected Maximization)
LME	: Doğrusal Karma Etkiler (Linear mixed effects)
ML	: En Çok Olabilirlik (Maximum Likelihood)
MCMC	: Markov Zinciri Monte Carlo (Markov Chain Monte Carlo)
MH	: Metropolis Hasting
MSE	: Mean Square Error (Hata Kareler Ortalaması)
LR	: Olabilirlik Oran (Likelihood Ratio)
REML	: Kısıtlı En Çok Olabilirlik (Restricted Maximum Likelihood)
IW	: Ters Wishart (Inverse Wishart)

ÇİZELGE LİSTESİ

6.1	Tiroid veri setinde kullanılan değişkenler ve açıklamaları	69
6.2	Birleşik Model Analiz Sonuçları	69
6.3	Bayesçi Yaklaşım Parametre Tahminleri	73
6.4	Yakınsama Tanıları.....	75
7.1	$\sigma_1^2 = 0.1$ ve farklı örneklem genişlikleri (100,300,600) için ML ve Bayesçi yöntem için simülasyon sonuçlarının karşılaştırılması	81
7.2	$\sigma_1^2 = 0.96$ ve farklı örneklem genişlikleri (100,300,600) için ML ve Bayesçi yöntem için simülasyon sonuçlarının karşılaştırılması.....	82
7.3	$\sigma_1^2 = 10$ ve farklı örneklem genişlikleri (100,300,600) için ML ve Bayesçi yöntem için simülasyon sonuçlarının karşılaştırılması	83

ŞEKİL LİSTESİ

2.1	Rastgele seçilen 30 hastanın CD4 kan sayımlarının zaman boyunca değişimi	8
3.1	Sağ kalım eğrisi	23
3.2	λ parametresinin farklı değerleri için tehlike fonksiyonun grafiği	27
6.1	Tekrarlı Tiroglobulin ölçümlerinin zamana boyunca değişimi.....	69
6.2	Tüm hastalara ve cinsiyete göre Kaplan-Meire Eğrisi.....	70
6.3	β_0 , β_1 ve β_2 parametrelerine ait iz grafikleri ve sonsal dağılım histogram grafikleri.....	76
6.4	γ_1 ve γ_2 parametrelerine ait iz grafikleri ve sonsal dağılım histogram grafikleri.....	76
6.5	α ve σ_ϵ parametrelerine ait iz grafikleri ve sonsal dağılım histogram grafikleri.....	77
6.6	D11, D12 ve D22 varyans kovaryans parametrelerine ait iz grafikleri ve sonsal dağılım histogram grafikleri	77
6.7	Zincir I ve Zincir II için Otokorelasyon grafikleri.	78
7.1	$\sigma^2 = 0.1$ ve örneklem genişliği $n = (100, 300, 600)$ için Güven Aralıkları	84
7.2	$\sigma^2 = 0.96$ ve örneklem genişliği $n = (100, 300, 600)$ için Güven Aralıkları	84
7.3	$\sigma^2 = 10$ ve örneklem genişliği $n = (100, 300, 600)$ için Güven Aralıkları	85
7.4	$\sigma^2 = 0.1$ ve farklı örneklem genişlikleri $n = (100, 300, 600)$ MSE değerlerinin değişimi	86

SEMBOL LİSTESİ

β	: $(p + 1) \times 1$ boyutlu sabit etkiler parametre vektörü
b_i	: $n \times q_b$ boyutlu rastgele etkiler parametresi
X	: $n_i \times (p + 1)$ boyutlu sabit etkiler tasarım matrisi
Z	: $n \times q_b$ boyutlu rastgele etkiler tasarım matrisi
δ_i	: Olay gösterge fonksiyonu
$h_0(t)$: Temel tehlike fonksiyonu
$h_i(t)$: Tehlike fonksiyonu
$S(t)$: Sağ kalım fonksiyonu
w_i	: Zamandan bağımsız temel açıklayıcı değişken vektörü
γ	: Temel açıklayıcı değişkenlere ait parametre
y_i	: $n_i \times 1$ boyutlu açıklayıcı değişken vektörü
D	: $q_b \times q_b$ boyutlu rastgele etkiler varyans kovaryans matrisi
σ_ε^2	: $n_i \times 1$ boyutlu hata vektörü
$\ \cdot\ $: Öklid vektör normu
θ	: Tam parametre vektörü
$L(\cdot)$: Olabilirlik fonksiyonu
$\ell(\cdot)$: log olabilirlik fonksiyonu
$I(\cdot)$: Fisher bilgi matrisi
$H(\cdot)$: Hessian matrisi
α	: Birleşik modelde ilişki parametresi
T_i	: Gözlenen sağ kalım zamanları

BOYLAMSAL VE SAĞ KALIM VERİLERİNİN BİRLEŞİK MODELLENMESİ ÜZERİNE BİR ÇALIŞMA

ÖZET

Son yıllarda, özellikle klinik çalışmalardaki gelişmeler birleşik modellemeye olan ilgiyi artmıştır. Birleşik model boylamsal alt model ve sağ kalım alt modeli olmak üzere iki alt modelden oluşmaktadır. Bu modelde parametre tahminleri genellikle olabilirlik fonksiyonunun maksimizasyonuna dayanan en çok olabilirlik (ML) yöntemi ile elde edilir. Birleşik olabilirlik fonksiyonu karmaşık yapıya sahiptir ve ML yöntemi ile parametre tahminlerinin elde edilebilmesi için olabilirlik fonksiyonundaki yüksek boyutlu integralin alınması gerekir. Olabilirlik fonksiyonundaki integralin kapalı formda çözümü yoktur ve bu integrale genellikle Gauss Hermite yaklaşımı, uyarlanabilir Gauss Hermite yaklaşımı, Laplace yaklaşımı ve Monte Carlo yöntemleri gibi sayısal yöntemlerle yakınsama sağlanır.

Gauss Hermite kuralı ve uyarlamalı Gauss Hermite kuralı ile integrale yakınsama sağlamak kolaylık sağlasa da gözlem sayısı arttıkça bu yöntemlerle integrale yakınsama sağlamak işlem yükünü arttırdığından zorlaşmaktadır. Bu çalışmanın amacı birleşik modelde parametre tahminlerinde karşılaşılan hesaplama güçlüklerine çözüm aramaktır. Bu amaç doğrultusunda parametre tahminlerinde Bayesçi yöntemin kullanılması önerilmiştir. Bayesçi yöntemlerle parametre tahminlerinin elde edilebilmesi için parametrelerin sonsal dağılımından örneklemelerin üretilmesini sağlayan Markov Zinciri Monte Carlo (MCMC) yöntemlerinden Gibbs örnekleme ve Metropolis-Hasting algoritmasından yararlanılmıştır. Uygulama bölümünde Tiroid veri seti kullanılmış ve birleşik modelde parametre tahminleri Bayesçi yöntemle elde edilmiştir. Ayrıca Bayesçi yöntemin ve klasik yöntemin tahmin performanslarının karşılaştırılması amacıyla simülasyon çalışması yapılmıştır.

Anahtar kelimeler: Boylamsal veri analizi, sağ kalım analizi, boylamsal ve sağ kalım verilerinin birleşik modellenmesi, Bayesçi yaklaşım, MCMC yöntemleri, Gibbs örnekleme, Metropolis-Hasting algoritması.

A STUDY ON JOINT MODELLING OF LONGITUDINAL AND TIME TO EVENT DATA

SUMMARY

Joint modelling of longitudinal and time to event data have received more attention in conjunction with development in clinical studies. Joint model comprises of two submodel, longitudinal submodel and survival submodel. Parameter estimation obtain Maximum likelihood(ML) approach which based on maximization of likelihood function in joint modelling. Likelihood function has complex form and in order to obtain parameter estimation, high dimension integral needs to evaluated. In joint model, the integral has no closed form solution and convergence is usually achieved by numerical methods such as Gauss Hermite approach, adaptive Gauss Hermite approach and Monte Carlo methods.

However, it is easy to approximate integral with Gauss Hermite and Laplace approximation when the integral dimension is low. Computation complexity increases when dimension of integral is high. The aim of this study seek for solution to calculation complexity in parameter estimation. In accordance with this purpose, we propose Bayesian approach which helps with computational complexity. To obtain the model parameter estimation, we use Markov Chain Monte Carlo algorithm, which is combined with Gibbs sampling and Metropolis-Hasting algorithm. In application part thyroid data set is used and parameter estimation is obtained by Bayesian methods. In order to compare Bayesian and classical method, simulation study was designed.

Keywords: Longitudinal data analysis, survival analysis, joint modelling of longitudinal and time to event data, Bayesian approach, MCMC methods, Gibbs sampling, Metropolis-Hasting algorithm

1 GİRİŞ

Klinik çalışmalarda genellikle bireylere ait, eşit ya da eşit olmayan zaman aralıklarında alınan tekrarlı ölçüm verileriyle birlikte, bireylerde ilgilenilen olayın görülmesi ve sağ kalım zamanı gibi bilgileri içeren sağ kalım verileri de kaydedilir. Bireylerden alınan tekrarlı ölçüm verilerindeki zaman boyunca meydana gelen değişimler hastalık süreci ile ilgili bilgi verirken aynı zaman da bireyin sağ kalım süreci ile ilişkili olabilir. Örneğin, HIV (Human Immunodeficiency Virüs) ile ilgili çalışmalarda HIV virüsü taşıyan hastalar, AIDS tanısı konulana kadar ya da ölüm gerçekleşene kadar takip edilir. Bu hastalara ait demografik bilgiler, uygulanan tedavi yöntemleri gibi zaman boyunca sabit temel değişkenlerle birlikte bireylerin bağışıklık sistemi durumları ile ilgili bilgi veren CD4 (Cluster of Different 4) hücre sayımları ya da viral yük (viral RNA) gibi zamana bağlı ölçümleri de kaydedilir. Benzer şekilde kanser tedavilerinde ilgilenilen kanser türü ile ilgili ilişkili olan boylamsal ölçümlerle birlikte bireye ait ölüm ya da tümörün tekrarlanması gibi sağ kalım bilgileri de kaydedilir (Tsiatis ve Davidian, 2004). Bu tip çalışmalarda odaklanılan klinik soru uygulanan tedavi yönteminin bireylerin sağ kalım sürelerine olan etkisi üzerine olabileceği gibi boylamsal süreçten elde edilen çıktılarda zaman boyunca meydana gelen değişimlerini incelemek ya da boylamsal ölçümlerde meydana gelen değişim ile bireyin sağ kalım olayı yaşaması arasındaki ilişkiyi incelemek üzerine olabilir.

Klasik yaklaşımda zamana bağlı boylamsal ölçümlerde meydana gelen değişim ile bireyin sağ kalım olayı yaşama riski arasındaki ilişki Cox model ile incelenebilir. Cox modelde boylamsal ölçümlerin sürekli ve ölçüm hatası ile ölçülen değişken olduğu bilgisi göz ardı edilir ve gözlenen boylamsal ölçümler modele açıklayıcı değişken olarak eklenir. Bu durum sağ kalım modelinden elde edilen regresyon parametrelerinin tahminlerinin yanlı olarak elde edilmesine neden olur (Prentice, 1982). Bu nedenle boylamsal süreçten gelen tekrarlı ölçümlerin sürekli olarak tahmin edilmesine ve iki sürecin eş zamanlı olarak analiz edilmesine olanak sağlayan birleşik modele ihtiyaç

duyulur. Birleşik model, boylamsal süreç ve sağ kalım süreci arasındaki ilişkiyi açıklamada kullanılan güçlü bir yöntemdir. Birleşik modellere genellikle boylamsal süreçten gelen zamana bağlı değişkenler ölçüm hatası ile ölçüldüğünde boylamsal süreçte bilgilendiri sansürleme göz önünde bulundurulduğunda (örneğin bireyin hastalığın seyrinin kötü gitmesi nedeniyle çalışmadan ayrılması) ve boylamsal süreç ve sağ kalım süreci arasındaki ilişki incelenmek istendiğinde ihtiyaç duyulur (Diggle ve diğ., 2002; McCrink ve diğ., 2013).

Boylamsal ve sağ kalım verilerinin birleşik olarak modellenmesi için geliştirilen yöntemler ilk olarak CD4 hücre sayımları ile sağ kalım verileri arasındaki ilişkinin ortaya çıkarılması üzerine yapılan AIDS/HIV ile ilgili klinik denemeler sırasında yapılmıştır (De Gruttola ve Tu, 1994; Tsiatis ve diğ., 1995; Faucett ve Thomas, 1996; Self ve Pawitan, 1992; Lavalley ve Degruttola, 1996). Özellikle klinik çalışmalardaki gelişmeler ve birleşik model ile elde edilen parametre tahminlerinin alternatif yöntemlere göre daha etkili ve yansız olması son yıllarda birleşik modele olan ilgiyi arttırmıştır.

Birleşik modelleme ile ilgili yapılan derleme makaleler için Tsiatis ve Davidian (2004), Ibrahim ve diğ. (2010), McCrink ve diğ. (2013); Yu ve diğ. (2004) makaleleri incelenebilir. Hickey ve diğ. (2016), birleşik modelde son yıllarda yapılan çalışmalarla ilgili geniş bir literatür taramasına yer vermiştir. Rizopoulos (2012a); Elashoff ve diğ. (2016) tarafından boylamsal ve sağ kalım verisinin birleşik modellenmesi üzerine yazılmış kitaplar bulunmaktadır.

Self ve Pawitan (1992) birleşik modellemede parametre tahminleri için iki aşamalı yöntemi önermişlerdir. İki aşamalı yaklaşım, olabilirlik yaklaşımına göre daha kolay olsa da parametre tahminlerindeki yanlılığı tamamen ortadan kaldırmaz. De Gruttola ve Tu (1994) sağ kalım zamanının parametrik olarak modellendiği birleşik modeli önermişlerdir. Sağ kalım zamanının normal dağılıma sahip olduğunu varsayarak birleşik modelde parametre tahminlerini EM (Expected-Maximization) algoritması ile eş zamanlı olarak kolaylıkla elde etmişlerdir. Ancak bu yöntem sadece sağ kalım zamanının normal dağılıma sahip olduğu varsayıldığında geçerlidir. Tsiatis ve diğ. (1995) parametre tahminlerinde iki aşamalı yöntemi önermişlerdir fakat Self ve Pawitan (1992)'dan farklı olarak boylamsal alt modelde parametre tahminleri için EM algoritmasından yararlanmışlardır.

Wulfsohn ve Tsiatis (1997) parametre tahminlerinde tam olabilirlik yaklaşımını kullanarak boylamsal ve sağ kalım sürecine ait parametre tahminlerini eş zamanlı olarak elde etmişlerdir. Tam olabilirlik yaklaşımı ile iki aşamalı tahmin yönteminden kaynaklanan parametre tahminlerindeki yanlılığı azaltmayı amaçlamışlardır. Faucett ve Thomas (1996) Markov Zinciri Monte Carlo (MCMC) yöntemi Gibbs örneklemesi ile model parametrelerini tahmin etmişlerdir.

Tsiatis ve Davidian (2001) ve Song ve diğ. (2002) rastgele etki parametrelerinin dağılımının yanlış belirlenme problemi üzerine yoğunlaşmışlardır. Rastgele etkiler üzerinde dağılım varsayımının yapılmadığı yarı parametrik model kullanmışlardır. Tsiatis ve Davidian (2001) parametre tahminlerinde koşullu skor (conditional score) yaklaşımını önermişlerdir. Song ve diğ. (2002) parametre tahminlerinde EM algoritmasından yararlanmışlardır.

R. Brown ve G. Ibrahim (2003) parametre tahminlerindeki dayanıklılığı arttırmak için boylamsal alt model parametrelerinin dağılımlarının tanımsız bırakıldığı yeni yarı parametrik model önermişlerdir. Böylelikle modelde esneklik sağlamayı ve sağlam kestiriciler elde etmeyi amaçlamışlardır. Parametre tahminlerinde Bayeşçi yöntemlerden yararlanmışlardır.

Wu ve diğ. (2010) parametre tahminlerinde Laplace yaklaşımı ve Monte Carlo Beklenti Maksimizasyonu (MCEM) yöntemlerinden yararlanmışlardır. Çalışmalarında iki yöntemin performansını gerçek veri seti uygulaması ve benzetim çalışmalarıyla karşılaştırmışlardır.

Birleşik modelde son yıllarda yapılan çalışmalar boylamsal çıktılarının çok değişkenli olması ve yarışan riskler durumu üzerinedir. Brown ve diğ. (2005) çok değişkenli boylamsal yanıtlar ve sağ kalım çıktıları için birleşik modellemede, boylamsal çıktıların doğrusal bir modelle modellenemediği durumlar için boylamsal alt modelde parametrik olmayan model kullanılması önermişlerdir. Boylamsal alt model için kübik B-spline modeli ve sağ kalım alt modeli için oransal Cox modeli kullanmışlardır. Modelde parametre tahminleri için MCMC algoritmasını kullanmışlardır.

Pantazis ve Touloumi (2007) birleşik modelde rastgele etki parametrelerinin dağılımlarının yanlış belirlenmesinin parametre tahminleri üzerinde etkisi üzerine yoğunlaşmışlardır. Çalışmalarında parametrelerin dağılımlarının yanlış belirlenmesine

karşı doğrusal modele göre daha dayanıklı olan bir model önermişlerdir. Önerdikleri modelde sabit etki parametre tahminlerinin dağılımının yanlış belirlendiği durumlarda diğer parametrelere göre daha az etkilendiklerini benzetim çalışmalarıyla göstermişlerdir.

Rizopoulos ve diğ. (2008) ve Huang ve diğ. (2009) çalışmalarında tekrarlı ölçümlerin sayısı arttıkça rastgele etki parametrelerinin dağılımının yanlış belirlenmesinin model parametre tahminleri ve parametre standart hataları üzerinde daha az etkiye sahip olduğunu göstermişlerdir.

Rizopoulos ve Ghosh (2011) çok değişkenli birleşik modelde bireylere ait boylamsal değişimleri açıklamak için boylamsal alt modelde spline tabanlı yaklaşım kullanarak bireylerin boylamsal ölçümlerinde meydana gelen değişimlerini esnek bir modelle açıklamayı amaçlamışlardır ve parametre tahminleri için Bayesçi yaklaşım kullanmışlardır. Boylamsal alt model için doğrusal bir model yerine kübik spline modelini kullanmışlardır. Li ve diğ. (2012) iki değişkenli boylamsal çıktılarının ve yarışan riskler için boylamsal alt model için yarıparametrik model önermişlerdir. Dayanıklı kestiriciler elde edebilmek için rastgele etkilerin dağılımını tanımsız bırakmışlardır Brown (2009). Bir hastalığın teşhisi ya da durumunu belirlemek için kullanılan ve bireylerden alınan kan, idrar gibi biyolojik göstergelerdeki değişimi ifade eden biyobelirteçlerdeki değişim ile bireylerin sağ kalım olayı yaşama riski arasındaki ilişkiyi ortaya çıkartmak için birleşik modelde kullanılan standart ilişki yapısından farklı ilişki yapıları kullanmışlardır. Boylamsal alt modelde kübik B-spline model kullanmışlardır ve parametre tahminlerinde MCMC algoritmalarından yararlanmışlardır.

Rizopoulos (2012a) boylamsal ve sağ kalım modellerinin birleşik modellenmesinde parametre tahminlerinde rastgele etkileri boyutu arttıkça karşılaşılan hesaplama zorluklarını azaltmak için yapay-uyarlanabilir (pseudo-adaptive) Gauss Hermite yaklaşımını önermişlerdir.

Elashoff ve diğ. (2008), Williamson ve diğ. (2008) ve Huang ve diğ. (2011a) birleşik modellemede çoklu başarısızlık problemlerine odaklanmışlardır.

Tang ve Tang (2015) çok değişkenli birleşik modellemede ölçüm hatalarının çarpık normal dağılımdan geldiği ve rastgele etkiler üzerinde dağılım varsayımının

yapılmadığı yarı parametrik model önermişlerdir. Parametre tahminlerinde Gibbs örnekleme ve Metropolis-Hasting (MH) algoritmasını birlikte kullanmışlardır.

Türkiye’de birleşik modelleme ile ilgili az sayıda çalışma bulunmaktadır. Ülgen ve Asar (2017) boylamsal ve sağ kalım verilerinin birleşik modellenmesi üzerine derleme bir makale yayınlamışlardır. Konar ve diğ. (2018) yoğun bakım ünitesinde yatan hastalardan tekrarlı olarak alınan C-Reaktif Protein ölçümleri ile bireylerin ölüm olayı yaşaması arasındaki ilişkinin incelenmesi ve tekrarlı ölçümlerdeki değişim üzerinde etkili olan değişkenlerin ortaya çıkarılması için birleşik modellemeden yararlanmışlardır.

Bu çalışmanın amacı birleşik modellemede, parametre tahminlerinde karşılaşılan hesaplama güçlüklerinin üstesinden gelebilmek için Bayesçi yöntem ile parametre tahminlerinin elde edilmesidir. Bu amaç doğrultusunda çalışma, sekiz bölümden oluşmaktadır. İkinci bölümde boylamsal veri analizi anlatılmıştır. Boylamsal veri analizinde sıklıkla kullanılan doğrusal karma etkiler modelinden (LME) ve parametre tahminlerinden detaylı olarak bahsedilmiştir.

Üçüncü bölümde sağ kalım analizi detaylı olarak anlatılmıştır. Sağ kalım analizinde kullanılan temel fonksiyonlardan bahsedilmiş ve sağ kalım zamanının belirlenmesinde kullanılan dağılımlar ve sağ kalım analizinde kullanılan modeller incelenmiştir.

Dördüncü bölümde boylamsal ve sağ kalım verilerinin birleşik olarak modellenmesinden bahsedilmiş ve birleşik modelin teorik alt yapısı anlatılmıştır.

Beşinci bölümde birleşik modelde Bayesçi yöntemlerle parametre tahminlerinden bahsedilmiştir. Önsel dağılım kavramından bahsedilmiş ve birleşik modelde parametreler için belirlenen önsel dağılımlara yer verilmiştir. Modelde yer alan her parametre için tam koşullu sonsal dağılımlar elde edilmiştir. Bayesçi yöntemde sonsal dağılımların güncellenmesinde kullanılan MCMC yöntemlerinden Gibbs örnekleme ve MH algoritması anlatılmıştır. Bayesçi yöntemde parametrelerin yakınsama kontrolleri için kullanılan yakınsama tanılarından bahsedilmiştir.

Altıncı bölüm uygulama bölümüdür. Bu bölümde Memorial Sloan Kettering Kanser Araştırma Merkezinden alınan 341 tiroid kanseri hastasına ait veri seti kullanılmıştır. Veri seti En Çok Olabilirlik (ML) yöntemi ve Bayesçi yaklaşım ile analiz edilmiş ve parametre tahminleri elde edilmiştir. MCMC benzetiminden elde edilen sonsal

örneklerin yakınsama kontrolleri iz grafikleri ve Geweke tanısı ve Gelman ve Rubin yakınsama tanıları ile yapılmıştır.

Yedinci bölümde önerilen yöntemin performansının değerlendirilmesi amacıyla simülasyon çalışması yapılmıştır ve sonuçlar özetlenmiştir. Tez çalışmasının son bölümünde Bayesçi yaklaşım ile elde edilen sonuçlar yorumlanmış ve ileriki çalışmalar için neler yapılabileceği konusu tartışılmıştır.

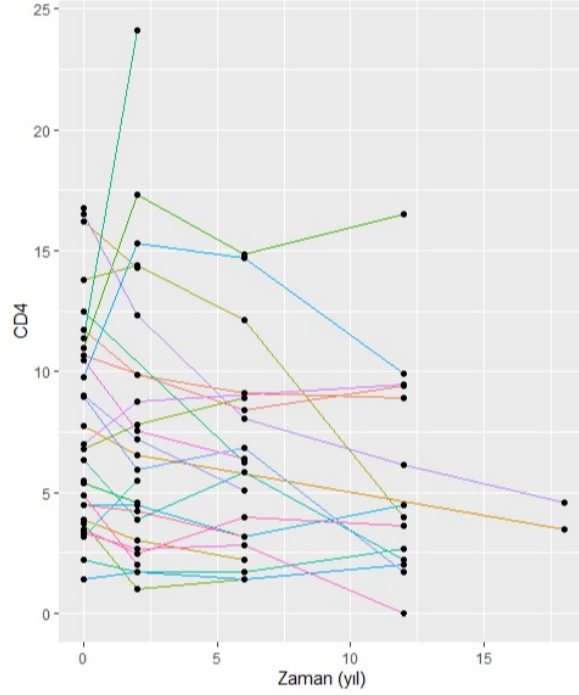


2 BOYLAMSAL VERİ ANALİZİ

Boylamsal veri, aynı bireylerden farklı izlenme zamanlarında alınan tekrarlı ölçüm verileridir. Boylamsal veri analizinde, tekrarlı ölçümlerin alındığı deney birimlerinden birim ya da birey olarak bahsedilebilir. Bu çalışmada klinik çalışmalar üzerinden örnekler verileceğinden birimlerden birey olarak bahsedilecektir. Boylamsal veriyi diğer veri türlerinden ayıran en önemli özellik, çalışma boyunca bireylerden tekrarlı ölçümlerin alınması sonucunda, bireylere ait yanıt değişkenindeki değişimin görülmesine olanak sağlamasıdır. Boylamsal veri analizinde amaç, yanıt değişkeninde zaman boyunca meydana gelen değişimleri ve bu değişimlere etki eden etkenleri incelemektir (Fitzmaurice ve diğ., 2012).

Boylamsal veri, aynı birimlerden tekrarlı olarak alındığı için kümelenmiş yapıdadır. Kümeler, aynı birimden farklı zamanlarda alınan ölçümlerden oluşur. Aynı kümede yer alan ölçümler arasında ilişki bulunurken farklı birimlerden alınan ölçümler arasında heterojen bir yapı mevcuttur (Fitzmaurice ve diğ., 2012). Örneğin, AIDS ile ilgili çalışmalarda, CD4 kan ölçümleri çalışma boyunca tekrarlı olarak alınır ve alınan CD4 ölçümleri bireylerde hastalık sürecinin takibinde kullanılır. Şekil 2.1'de rastgele olarak seçilen 30 hastanın CD4 kan ölçümlerinin zaman boyunca değişimlerini göstermektedir. Bireylerden alınan tekrarlı kan ölçümleri arasında değişkenlik mevcuttur aynı bireylerden alınan kan ölçümleri arasında ise ilişki bulunmaktadır. AIDS veri seti R programı JM paketinde açık olarak bulunmaktadır.

Boylamsal veri analizi ile bu heterojenliğin derecesi belirlenebilir ve tekrarlı ölçümlerde azalmanın meydana geldiği ortalama süresinin tahmini yapılabilir. Ölçümlerin birbirinden bağımsız olduğu ve ölçümler arasında homojen varyans varsayımının yapıldığı klasik regresyon modelleri, boylamsal veri analizi için yetersiz kalmaktadır. Bu nedenle, boylamsal veri analizinde, aynı bireylerden alınan ölçümler arasındaki ilişkinin ve bireyler arası farklılıklardan kaynaklanan heterojenliğin göz önünde bulundurulduğu modellere ihtiyaç duyulmaktadır.



Şekil 2.1: Rastgele seçilen 30 hastanın CD4 kan sayımlarının zaman boyunca değişimi

Boylamsal veri analizinde Laird ve Ware (1982) tarafından önerilen LME modeli genellikle tercih edilen modellerden biridir. LME modelinle, açıklayıcı değişkenler arasındaki ilişki, doğrusal regresyon modelinde olduğu gibi sabit etkiler olarak da bilinen regresyon parametreleri ile açıklanır. Aynı birimlerden alınan tekrarlı ölçümler arasındaki bağımlılık ise birey-özel rastgele etkiler parametreleriyle açıklanır (Diggle ve diğ., 2002; Fitzmaurice ve diğ., 2012).

2.1 LME Modeli

LME modeli her birimin, kendine ait rastgele eğim ve rastgele kesim parametresine sahip olduğu fikrine dayanır. y_{ij} , i . birimin ($i = 1, \dots, n$) t_{ij} zamanında alınan j . yanıt değişkeninin değerini, ($j = 1, \dots, n_i$) gösterebilir. Birimlerden, izlenme zamanı boyunca tekrarlı olarak alınan yanıt değişkenine ait değişim, doğrusal bir modelle eşitlik 2.1'deki gibi yazılabilir.

$$y_{ij} = (\beta_0 + b_{i0}) + (\beta_1 + b_{i1})t_{ij} + \varepsilon_{ij} \quad (2.1)$$

Modelde, β_0 ve β_1 ortalama değişimi gösteren sabit etki parametreleridir. b_{i0} ve b_{i1} sırasıyla i . bireye ait rastgele kesim ve rastgele eğim parametreleridir ve ortalama

yanıt deęişkeninden sapmalarını göstermektedir. Modelde bireyler arası deęişim sadece ölçüm deęerlerinin başlangıç noktasındaki farklılıklarından deęil aynı zamanda bireylerin zaman içersindeki ölçüm deęerlerindeki farklılıklardan kaynaklanmaktadır.

Eşitlik 2.1’de verilen modelin matris yapısı aşağıdaki gibi yazılabilir.

$$\begin{pmatrix} y_{i1} \\ y_{i2} \\ \vdots \\ y_{in_i} \end{pmatrix} = \begin{pmatrix} 1 & t_{i1} \\ 1 & t_{i2} \\ \vdots & \vdots \\ 1_{ni} & t_{in_i} \end{pmatrix} (\beta_0 \quad \beta_1)^\top + \begin{pmatrix} 1 & t_{i1} \\ 1 & t_{i2} \\ \vdots & \vdots \\ 1_{ni} & t_{in_i} \end{pmatrix} (b_{0i} \quad b_{1i})^\top + \begin{pmatrix} \varepsilon_{i1} \\ \varepsilon_{i2} \\ \vdots \\ \varepsilon_{in_i} \end{pmatrix} \quad (2.2)$$

Model, birimlere ait rastgele kesim ve rastgele eğim parametrelerini gösteren b_{0i} ve b_{1i} parametrelerini içermektedir. Bu durumda rastgele kesim ve eğim parametrelerine ait $\sigma_{b_0}^2$ ve $\sigma_{b_1}^2$ varyanslarını içeren D_i varyans kovaryans matrisi,

$$\begin{pmatrix} b_{0i} \\ b_{1i} \end{pmatrix} \sim N(0, D_i) \quad , D_i = \begin{pmatrix} \sigma_{b_0}^2 & \sigma_{b_0b_1} \\ \sigma_{b_0b_1} & \sigma_{b_1}^2 \end{pmatrix}$$

şeklindedir.

Rastgele etkilerin ve sabit etkilerin birlikte yer aldığı Laird ve Ware (1982) tarafından önerilen LME modelinin matris gösterimi ile;

$$\begin{aligned} y &= X\beta + Zb + \varepsilon \\ b &\sim N(0, D) \\ \varepsilon &\sim N(0, \sigma^2 I_n) \end{aligned} \quad (2.3)$$

eşitlik 2.3’teki gibi yazılabilir. Burada, y , $(n \times 1)$ boyutlu yanıt deęişkeni vektörüdür. X ve Z sırasıyla $(n \times (p + 1))$ boyutlu sabit etkiler ve $(n \times q)$ boyutlu rastgele etkiler tasarım matrisini, β $((p + 1) \times 1)$ boyutlu sabit etkiler vektörü, b $(q \times 1)$ boyutlu rastgele etkiler vektörünü göstermektedir. b rastgele etkiler parametresinin genellikle 0 ortalama ve D , $(q \times q)$ boyutlu varyans kovaryans matrisi ile normal dağıldığı varsayılır. Doğrusal regresyonda olduğu gibi, β_j , $j = 1, \dots, p$ sabit etkiler parametresi, dięer tüm deęişkenler sabitken, x_j açıklayıcı deęişkeninde bir birim

arttığında y yanıt değişkenindeki ortalama değişimi gösterir. b_i , rastgele etkiler parametresi ise, i . birimin β sabit etkiler parametresinden sapmasını gösterir. ε_i , hata teriminin 0 ortalama ve $\sigma^2 I_{ni}$ varyans ile normal dağıldığı varsayılır. b_i rastgele etkiler parametresi ve ε_i hata teriminin bağımsız olduğu kabul edilir, $Cov(b_i, \varepsilon_i) = 0$ (Fitzmaurice ve diğ., 2012). y yanıt değişkeninin, b rastgele etkiler parametrelerine göre koşullu ortalaması,

$$E(y | b) = X\beta + Zb \quad (2.4)$$

şeklindedir. y yanıt değişkeninin ortalaması,

$$\begin{aligned} E(y) &= E(X\beta + Zb + \varepsilon) \\ &= X\beta \end{aligned} \quad (2.5)$$

varyansı,

$$\begin{aligned} Var(y) &= Var(X\beta + Zb + \varepsilon) \\ &= ZVar(b)Z^T + Var(\varepsilon) \\ &= ZDZ^T + \sigma^2 I_{ni} \\ &= ZDZ^T + R \end{aligned} \quad (2.6)$$

olmak üzere, y yanıt değişkeni, $X\beta$ ortalama ve $V = (ZDZ_i^T + R)$ varyans kovaryans matrisi ile normal dağılıma sahiptir, $y \sim N(X\beta, V)$.

2.2 LME Modelinde Parametre Tahmini

LME modelinde, doğrusal regresyon modelinden farklı olarak, β sabit etki parametresinin, σ_i^2 varyans bilşenlerinin ve b_i rastgele etkiler parametrelerinin tahmin edilmesi gerekir. Parametre tahmin edicileri için genellikle olabilirlik fonksiyonunun maksimizasyonuna dayanan, ML yöntemi ve Kısıtlı en çok olabilirlik (REML)yöntemi kullanılır. Bu bölümde, bu iki tahmin yöntemi üzerinde durulacaktır.

2.2.1 ML Yöntemi

Eşitlik 2.3 'te verilen LME modelinde,

$$y = X\beta + Zb + \varepsilon \quad (2.7)$$

yanıt deęiřkeni y 'nin $X\beta$ ortalama ve V varyans kovaryans matrisi ile normal daęıldıęı varsayımı altında $y \sim N(X\beta, V)$, yanıt deęiřkeni y 'nin olasılık yoęunluk fonksiyonu,

$$p(y | \beta, V) = (2\pi)^{-\frac{n}{2}} |V|^{-\frac{1}{2}} \times \exp \left(-\frac{1}{2} (y - X\beta)^\top V^{-1} (y - X\beta) \right) \quad (2.8)$$

olmak üzere,

olabilirlik fonksiyonu,

$$L(\theta | y) = (2\pi)^{-\frac{n}{2}} |V|^{-\frac{1}{2}} \times \exp \left(-\frac{1}{2} (y - X\beta)^\top V^{-1} (y - X\beta) \right) \quad (2.9)$$

řeklinindedir. ML ynteminde olabilirlik fonksiyonunun β ve σ_i^2 varyans bileřenlerine gře maksimize edilmesiyle parametre tahmin edicileri elde edilebilir (Diggle ve dię., 2002; Fitzmaurice ve dię., 2012).

Eřitlik 2.9'daki ifadenin logaritması alındıęında log-olabilirlik fonksiyonu,

$$\ell(\theta | y) = \frac{1}{2} \log(|V|) - \frac{1}{2} (y - X\beta)^\top V^{-1} (y - X\beta) \quad (2.10)$$

řeklinde yazılabilir (Fitzmaurice ve dię., 2012; Diggle ve dię., 2002).

V varyans kovaryans matrisinin bilindięi varsayımı altında eřitlik 2.10'da verilen log olabilirlik fonksiyonunun β , sabit etkiler parametresine gře birinci dereceden turevi alınıp 0'a eřitlendięinde;

$$\frac{\partial \ell}{\partial \beta} \Big|_{\beta=\hat{\beta}} = X^\top V^{-1} y - X^\top V^{-1} X \hat{\beta} = 0 \quad (2.11)$$

β sabit etkiler parametresinin ML tahmin edicisi

$$\hat{\beta} = (X^\top V^{-1} X)^{-1} X^\top V^{-1} y, \quad (2.12)$$

řeklinde elde edilir. V varyans kovaryans matrisi bilinmedięi durumda β parametresinin ML tahmin edicisi elde edilmeden gnce σ_i^2 varyans bileřenlerinin ML tahmin edicilerinin bulunması gerekir. Log olabilirlik fonksiyonu, σ_i^2 , $i = 1, \dots, r$ varyans bileřenlerine gře birinci dereceden turevi alınarak ML tahmin edicileri elde edilebilir. Log olabilirlik fonksiyonunun varyans bileřenlerine gře turevine geęmeden gnce iřlemlerde kolaylık aısından varyans kovaryans matrisi V ,

$$V = \sum_{i=1}^q Z_i Z_i^T \sigma_i^2 + \sigma_\varepsilon^2 I_{n_i} \quad (2.13)$$

şeklinde genişletilebilir. Rastgele hata terimi, rastgele etkiler terimine dahil edilerek $b_0 = \varepsilon$ $q_0 = n$ ve $Z_0 = I_n$ şeklinde tanımlandığında, eşitlik 2.13'te yer alan ifade aşağıdaki gibi yazılabilir (Searle ve diğ., 2009).

$$V = \sum_{i=0}^q Z_i Z_i^T \sigma_i^2 \quad (2.14)$$

V varyans kovaryans matrisinin σ_i^2 varyans bileşenine göre birinci dereceden türevi,

$$\frac{\partial \ell}{\partial \sigma_i^2} \Big|_{\sigma_i^2 = \hat{\sigma}_i^2} = -\frac{1}{2} \text{tr}(V^{-1} Z_i Z_i^T) + \frac{1}{2} (y - X\beta)^T V^{-1} Z_i Z_i^T V^{-1} (y - X\beta) \quad (2.15)$$

şeklinde yazılabilir. Birinci dereceden türevler sıfıra eşitlendiğinde,

$$\frac{1}{2} \text{tr}(\hat{V}^{-1} Z_i Z_i^T) = \frac{1}{2} (y - X\hat{\beta})^T \hat{V}^{-1} Z_i Z_i^T \hat{V}^{-1} (y - X\hat{\beta}) \quad (2.16)$$

eşitliği elde edilir. Eşitlik 2.16'nın sağ tarafında yer alan $\hat{V}^{-1}(y - X\hat{\beta})$ ifadesi daha sade bir şekilde, P , $(n \times n)$ boyutlu simetrik kare matrisi,

$$P = V^{-1} - V^{-1} X (X^T V^{-1} X)^{-1} X^T V^{-1} \quad (2.17)$$

olmak üzere,

$$\hat{V}^{-1}(y - X\hat{\beta}) = \hat{V}^{-1} - \hat{V}^{-1} X (X^T \hat{V}^{-1} X)^{-1} X^T \hat{V}^{-1} = \hat{P} y \quad (2.18)$$

olarak ifade edildiğinde varyans bileşenlerinin ML tahmin edicisi aşağıda verilen eşitlik 2.19'daki gibi elde edilir. Burada varyans kovaryans matrisi V 'nin tahmin edicisi \hat{V} kullanıldığından, P matrisi yerine eşitlikte \hat{P} kullanılmıştır (McCulloch ve Neuhaus, 2005).

$$\text{tr}(\hat{V}^{-1} Z_i Z_i^T) = y^T \hat{P} Z_i Z_i^T \hat{P} y \quad (2.19)$$

σ_i^2 varyans bileşenleri için denklemler elde edildikten sonra log olabilirlik fonksiyonunun β parametresine göre birinci dereceden türevi aşağıdaki gibi alınıp 0'a eşitlendiğinde,

$$\frac{\partial \ell}{\partial \beta} \Big|_{\beta=\hat{\beta}} = -2X^T V^{-1} y + 2X^T V^{-1} X \beta = 0 \quad (2.20)$$

β parametresinin ML tahmin edicisi

$$\hat{\beta} = \left(X^T \hat{V}_i^{-1} X \right)^{-1} X^T \hat{V}_i^{-1} y \quad (2.21)$$

şeklinde elde edilir. Eşitlik 2.21'de β sabit etkiler parametresinin ML tahmin edicisi için elde edilen denklemde, varyans kovaryans matrisinin tahmini kullanıldığından, ML tahmin edicisi en iyi doğrusal yansız tahmin edicisi (BLUE) özelliği taşımaz. $\hat{\beta}$ ve \hat{V} tahminleri için elde edilen denklemlerde yer alan varyans kovaryans matrisinin tersi V^{-1} , varyans bileşenlerine göre doğrusal olmadığından bu denklemin çözümü için Newton-Raphson (NR) yöntemi, Beklenti maksimizasyonu (EM), Fisher skorlama algoritması gibi sayısal yöntemlerden yararlanılır (Faucett ve Thomas, 1996; West ve diğ., 2014).

2.2.2 REML Yöntemi

LME modelinde, ML yöntemi, varyans bileşenlerinin tahmininde serbestlik derecesini göz önünde bulundurmaz. Bu nedenle sabit etkiler parametrelerinin tahmininden kaynaklanan serbestlik derecesi kaybının da hesaba katıldığı REML yönteminden yararlanılır. Örneğin, $y_i = x_i \beta + \varepsilon_i$ regresyon modelinde, hata terimi varyansının ML tahmin edicisi;

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^n (y_i - x_i^T \hat{\beta})^2}{n}$$

olarak ifade edilir (Searle ve diğ., 2009; Faraway, 2016). Hata terimlerinin ML tahmin edicisi, β parametrelerinin tahmininden kaynaklanan serbestlik derecesini hesaba katmadığından, elde edilen varyans tahmin edicisi, σ^2 'nin yansız bir tahmin edicisi değildir. σ^2 varyans parametresinin yansız tahmin edicisi REML yönteminden yararlanılarak,

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^n (y_i - x_i^T \hat{\beta})^2}{n - p} \quad (2.22)$$

şeklinde elde edilir. Burada p , x açıklayıcı değişken vektörünün boyutunu belirtmektedir (Searle ve diğ., 2009; Rizopoulos, 2012b).

LME modelinde, varyans bileşenleri tahmin edicilerinin elde edilmesinde karşılaşılan yanlılık sorununu ortadan kaldırmak için Harville (1974) tarafından geliştirilen REML yöntemini Laird ve Ware (1982) LME modelinde varyans bileşenlerinin tahmini için kullanmışlardır (Searle ve diğ., 2009).

REML yönteminde varyans bileşenlerinin tahmininde yanlılık sorununu ortadan kaldırmak için β parametrelerinden bağımsız y vektörü yerine y 'nin elemanlarının doğrusal bir dönüşümü olan $A'y$ dönüşümü kullanılır. Sütunları X matrisine dik, tam ranklı ($n \times (n - p)$) boyutlu ve $A = (I - X(X^T X)^{-1} X^T)$ özelliklerini taşıyan herhangi bir A matrisi seçilir.

Eşitlik 2.3'te verilen LME modeli, sütunları X matrisine dik olan A matrisi ile çarpıldığında

$$A^T y = A^T X \beta + A^T Z b \quad (2.23)$$

β parametrelerinden bağımsız y bağımlı değişken vektörünün doğrusal dönüşümünden oluşan $A^T y$ vektörü edilmiş olur. REML yöntemi, β sabit etkiler parametrelerinden bağımsız olarak elde edilen olabilirlik fonksiyonunun maksimizasyonuna dayanan bir yöntemdir.

Eşitlik 2.23'de verilen y vektörünün doğrusal dönüşümlerinden oluşan $A'y$ vektörü, 0 ortalama ve $A^T V A$ varyans kovaryans matrisi ile normal dağılıma sahip olduğu varsayılır, $A^T y \sim N(0, A^T V A)$. Yanıt değişkeni vektörü y 'nin doğrusal bir dönüşümü olan $A^T y$ vektörüne ait olabilirlik fonksiyonu,

$$L_R(A^T V A | A^T y) = \frac{1}{(2\pi)^{\frac{n-p}{2}} |A^T V A|^{\frac{1}{2}}} \exp\left(-\frac{1}{2} (A^T y)^T (A^T V A)^{-1} A^T y\right) \quad (2.24)$$

şeklinde yazılabilir. Eşitlik 2.24'te verilen olabilirlik fonksiyonunun logaritması alındığında log olabilirlik fonksiyonu aşağıdaki gibi elde edilir.

$$\begin{aligned} \ell_R(A^TVA | A^Ty) &= -\frac{n-r}{2} \ln(2\pi) - \frac{1}{2} \ln |A^TVA| \\ &\quad - \frac{1}{2} (A^Ty)^T (A^TVA)^{-1} A^Ty \end{aligned} \quad (2.25)$$

Bu ifadeyi sadeleştirmek için Eşitlik 2.17’de verilen P simetrik kare matrisi A^T matrisi için

$$P = A(A^TV^{-1}A)^{-1}A^T \quad (2.26)$$

şeklinde yazıldığında Eşitlik 2.25’te verilen log olabilirlik fonksiyonunun daha sade formu,

$$\ell_R(A^TVA | A^Ty) = -\frac{n-r}{2} \ln(2\pi) - \frac{1}{2} \ln |A^TVA| - \frac{1}{2} y^T P y \quad (2.27)$$

şeklinde elde edilir (McCulloch ve Neuhaus, 2005; Searle ve diğ., 2009).

REML yönteminde, Eşitlik 2.25’te verilen log olabilirlik fonksiyonunun varyans bileşenleri parametrelerine REML tahmin edicilerini elde edebilmek için öncelikle P ’ye göre birinci dereceden türevlerinin alınması gerekmektedir.

$$\begin{aligned} \frac{\partial P}{\partial \sigma_i^2} &= -A^T (A^TVA)^{-1} \frac{\partial A^TVA^T}{\partial \sigma_i^2} (A^TVA)^{-1} A^T \\ &= -A^T (A^TVA)^{-1} A \frac{\partial V}{\partial \sigma_i^2} A^T (A^TVA)^{-1} A^T \\ &= -P \frac{\partial V}{\partial \sigma_i^2} P \\ &= -P Z_i Z_i^T P \end{aligned} \quad (2.28)$$

$$\frac{\partial \ln |A^TVA|}{\partial \sigma_i^2} = -\frac{1}{2} \text{tr} \left((A^TVA)^{-1} A^T Z_i Z_i^T A \right) \quad (2.29)$$

elde edilir.

Tüm bu işlemler yapıldıktan sonra varyans bileşenlerinin log olabilirlik fonksiyonuna göre birinci dereceden türevi,

$$\begin{aligned} \frac{\partial \ell_R}{\partial \sigma_i^2} &= -\frac{1}{2} \text{tr} \left((A^TVA)^{-1} A^T Z_i Z_i^T A \right) + \frac{1}{2} P Z_i Z_i^T P \\ &= -\frac{1}{2} \text{tr} (P Z_i Z_i^T) + \frac{1}{2} y^T P Z_i Z_i^T P y \end{aligned} \quad (2.30)$$

şeklinde elde edilir. Eşitlik 2.30'deki birinci türevler alınarak elde edilen denklem 0'a eşitlendiğinde varyans bileşenleri σ_i^2 'nin REML tahmin edicisi

$$\text{tr}(PZ_iZ_i^T) = y^T PZ_iZ_i^T P y \quad (2.31)$$

olarak bulunur. Eşitlik 2.19'da verilen ML tahmin edicileri ile karşılaştırıldığında, REML ve ML tahmin ediciler için eşitliğin sağ tarafı eşitken, eşitliğin sol tarafında REML tahmin edicileri için V^{-1} 'in yerini P aldığı görülmektedir (Searle ve diğ., 2009).

2.2.3 EM Algoritması

EM algoritması, parametrelerin ML ve REML tahminlerinin elde edilmesinde kullanılan iteratif bir yöntemdir. Dempster ve diğ. (1977) veri de kayıp gözlem olması durumunda ML tahminlerinin elde edilebilmesi için, diğer iteratif yöntemlerin özel bir durumu olan, EM algoritmasını önermişlerdir.

LME modelinde, b_i , $i = 1, 2, \dots, q$ rastgele etkiler vektörü gözlenebilseydi, rastgele etkilerin normal dağılıma sahip olduğu varsayımı altında $b_i \sim N(0, D)$, ML tahmin edicisi, $\text{var}(\tilde{b}) = \frac{b_i^T b_i}{q_i}$ eşitliği ile elde edilebilirdi (McCulloch ve Neuhaus, 2005). Gerçek hayatta, rastgele etkilerin gerçek değerleri gözlenemediğinden, gözlenemeyen bu değerlerin yerine koşullu beklenen değerleri kullanılır (Searle ve diğ., 2009). Böylelikle $\text{var}(\tilde{b})$ eşitliğinde yer alan $b_i^T b_i$ yerine koşullu beklenen değerler kullanılarak varyans tahmini elde edilir.

LME modelinde, EM algoritması ile gözlenemeyen b_i ve $b_i^T b_i$ yerine, (y_i, b_i) tam veri vektöründen elde edilen koşullu beklenen değerler kullanılır. Bu aşama EM algoritmasının ilk aşaması olan beklenti aşamasını oluşturur. Maksimizasyon aşamasında ise, beklenti aşamasında hesaplanan koşullu beklenen değerler, modelde kayıp veri olarak ifade edilen b_i ve $b_i^T b_i$ yerine koyularak olabilirlik fonksiyonu maksimize eden ML ya da REML tahminleri elde edilir. EM algoritmasının gerçekleştirilebilmesi için birleşik olabilirlik fonksiyonunun ve $E(b_i^T b_i | y)$ koşullu beklenen değerlerin detaylı çıkarımları için Searle ve diğ. (2009)'a bakılabilir.

LME modelinde EM algoritmasının adımları şu şekilde gerçekleşir.

Adım 0: $m = 0$ için $\beta^{(0)}$ ve $\sigma_i^{2(0)}$ başlangıç değerleri belirlenir.

Adım 1 (E adımı): y tamamlanmamış veri vektörü bilindiğinde $b_i^\top b_i$ 'nin koşullu beklenen değer aşağıdaki gibi hesaplanır.

$$E(b_i^\top b_i | y) = \sigma_i^{4(m)} (y - X\beta^{(m)})^\top V^{(m)-1} Z_i Z_i V^{(m)-1} (y - X\beta^{(m)}) + \text{tr}(\sigma_i^{2(m)} I_{q_i} - \sigma_i^2 Z_i^\top (V^{(m)})^{-1} Z_i) \quad (2.32)$$

β parametresi için,

$$X\beta^{(m+1)} = X\beta^{(m)} X(X^\top X)^{-1} X^\top (V^{(m)})^{-1} (y - X\beta^{(m)}) \quad (2.33)$$

şeklinde hesaplanır.

Adım 2 (M adımı): İkinci adım olan M adımında tam veriye dayalı olabilirlik fonksiyonu maksimum yapan $\sigma_i^{2(m+1)}$ ve $X\beta^{(m+1)}$ parametre değerleri hesaplanır.

Adım 3: Yakınsama sağlanırsa, $\hat{\sigma}_i^2 = \sigma_i^{2(m+1)}$ ve $X\hat{\beta} = X\hat{\beta}^{(m+1)}$ olarak kabul edilir. Yakınsama sağlanmadığında birinci adıma dönülür (Searle ve diğ., 2009).

REML yönteminde σ_i^2 varyans bileşenlerinin tahmini için $A^\top X = 0$ olacak şekilde β parametrelerinden bağımsız y 'nin doğrusal bir dönüşümü olan $A^\top y$ dönüşümünden yararlanır. ML tahmin edicileri için elde edilen eşitliklerde y yerine, $A^\top y$, Z yerine $A^\top Z$, V varyans kovaryans matrisi yerine, $A^\top V A$ yazılarak EM algoritmasının adımları gerçekleştirilir.

2.2.4 NR Algoritması

NR algoritması, doğrusal olmayan fonksiyonların maksimizasyonunda kullanılan kök bulma yöntemidir (McCulloch ve Neuhaus, 2005). $f(\theta)$ fonksiyonunu maksimize eden,

$$\frac{\partial f(\theta)}{\partial \theta} = 0 \quad (2.34)$$

türevinin kökünü bulmaya çalıştığımızı varsayalım. Kök bulmak için $\frac{\partial f(\theta)}{\partial \theta}$ ifadesi θ_0 etrafında aşağıdaki gibi genişletilebilir,

$$\frac{\partial f(\theta)}{\partial \theta} = f'(\theta) = f'(\theta_0) + \frac{\partial^2 f(\theta)}{\partial \theta \partial \theta'} (\theta - \theta_0) \quad (2.35)$$

Bu eşitlik sıfıra eşitlenirse,

$$f'(\theta_0) + \frac{\partial^2 f(\theta)}{\partial \theta \partial \theta'} (\theta - \theta_0) = 0 \quad (2.36)$$

buradan,

$$\theta = \theta_0 - \left[\frac{\partial^2 f(\theta)}{\partial \theta \partial \theta'} \right]^{-1} f'(\theta_0) \quad (2.37)$$

kökü elde edilir (McCulloch ve Neuhaus, 2005).

Eşitlik 2.37'den yararlanarak θ parametresinin tahmin değerleri iteratif olarak elde edilebilir.

$$\theta^{(m+1)} = \theta^{(m)} - \left[\frac{\partial^2 f(\theta)}{\partial \theta \partial \theta'} \right]^{-1} \Big|_{\theta=\theta^{(m)}} f'(\theta^{(m)}) \quad (2.38)$$

LME modelinde, parametre tahminleri için Newton Raphson yöntemini uygulanmak istendiğinde, Eşitlik 2.38'da verilen ifadede $f(\theta)$ yerine Eşitlik 2.10 'da verilen log olabilirlik fonksiyonu yerleştirilerek, log olabilirlik fonksiyonunun maksimizasyonu yapılabilir. Log olabilirlik fonksiyonunun σ^2 parametresine göre birinci dereceden türevi,

$$\frac{\partial \ell}{\partial \sigma_i^2} = -\frac{1}{2} \text{tr}(\hat{V}^{-1} Z_i Z_i') + \frac{1}{2} (y - X\hat{\beta})^\top \hat{V}^{-1} Z_i Z_i^\top \hat{V}^{-1} (y - X\hat{\beta}) = 0 \quad (2.39)$$

şeklindedir. β parametresine göre ikinci türevi,

$$\frac{\partial^2 \ell}{\partial \beta \partial \beta^\top} = \ell_{\sigma^2} = -X'V^{-1}X \quad (2.40)$$

olup, σ_i^2 varyans bileşenlerine göre ikinci türevi,

$$\begin{aligned} \frac{\partial \ell}{\partial \sigma^2 \partial \sigma^2^\top} = \ell_{\sigma^2 \sigma^2} = & -\frac{1}{2} \text{tr}(V^{-1} Z_j Z_j^\top V^{-1} Z_i Z_i^\top) - \frac{1}{2} (y - X\beta)^\top \\ & \times V^{-1} Z_j Z_j' V^{-1} Z_i Z_i^\top V^{-1} Z_j Z_j^\top V^{-1} (y - X\beta)^\top \end{aligned} \quad (2.41)$$

şeklinde elde edilir. Eşitlik 2.38'ten yararlanarak $\sigma^2 = \sigma^{2(m)}$ olmak üzere $(m + 1)$. adım için aşağıdaki gibi düzenlenebilir.

$$\sigma^{2(m+1)} = \sigma^{2(m)} (\ell_{\sigma^2 \sigma^2}) \ell_{\sigma^2} \quad (2.42)$$

Newton Raphson algoritmasında, ML ve REML tahmin edicileri elde edilirken, varyans bileşenleri için negatif değerlerin elde edilmesi problemi ile karşılaşılabilir. Bu durumu önlemek amacıyla σ^2 yerine σ tahmin edilip, elde edilen tahmin değerinin karesi alınarak σ^2 varyans bileşenleri için tahminler elde edilebilir. Varyans bileşenlerinin negatif elde edilmesinin önüne geçebilmek amacıyla bir diğer öneri ise $0 < \alpha \leq 1$ olmak üzere Newton Raphson denklemini,

$$\sigma^{2(m+1)} = \sigma^{2(m)} - \alpha (\ell_{\sigma^{2(m)} \sigma^{2(m)}})^{-1} \ell_{\sigma^2} \quad (2.43)$$

şeklinde düzenlemektir (McCulloch ve Neuhaus, 2005).

2.3 Rastgele Etkiler Parametrelerinin Tahmini

Henderson ve diğ. (1959), LME modelinde sabit etkiler parametrelerinin en iyi doğrusal yansız tahmin edicisi (BLUE) ve rastgele etkiler parametrelerinin en iyi doğrusal yansız tahmin edicisinin (BLUP) birlikte elde edildiği karma model eşitlikleri olarak da bilinen Henderson eşitliklerini geliştirmişlerdir. Rastgele etkiler parametresi b ve boylamsal yanıt değişkeni y 'nin normal dağıldığı varsayımı altında b ve y 'nin birleşik olabilirlik fonksiyonu

$$L(y, b) = (2\pi)^{-\frac{n}{2}} |R|^{-\frac{1}{2}} \exp \left((y - X\beta - Zb)^\top R^{-1} (y - X\beta - Zb) \right) \times (2\pi)^{-\frac{q}{2}} |D|^{-\frac{1}{2}} \exp \left(-\frac{1}{2} b^\top D^{-1} b \right) \quad (2.44)$$

şeklindedir (Henderson ve diğ., 1959). Eşitlik 2.44'te verilen birleşik yoğunluk fonksiyonunun β sabit etkiler parametresine ve b rastgele etkiler parametresine göre maksimizasyonu yapılarak β ve b parametrelerinin tahmin edicileri elde edilebilir (Henderson ve diğ., 1959).

Eşitlik 2.44'de yer alan fonksiyonun logaritması alınıp,

$$\begin{aligned}
\ell(y, b) &= -\frac{1}{2} \log |R| - \frac{1}{2} \log |D| - \frac{1}{2} \left((y - X\beta - Zb)^\top R^{-1} (y - X\beta - Zb) + b^\top D^{-1} b \right) \\
&= -\frac{1}{2} \log |R| - \frac{1}{2} \log |D| - \frac{1}{2} \left(y^\top r^{-1} y - 2R^{-1} X\beta - 2y^\top R^{-1} Zb \right. \\
&\quad \left. + \beta^\top X^\top R^{-1} X\beta + 2\beta^\top X^\top R^{-1} Zb + b^\top Z^\top R^{-1} Zb + b^\top D^{-1} b \right)
\end{aligned} \tag{2.45}$$

sabit etkiler parametresi β ve b , rastgele etkiler parametresine göre birinci dereceden türevi alındığında,

$$X^\top R^{-1} y = X^\top R^{-1} X \hat{\beta} + X^\top R^{-1} Z \tilde{b} \tag{2.46}$$

ve

$$Z^\top R^{-1} y = Z^\top R^{-1} X \hat{\beta} + (Z^\top R^{-1} Z + D^{-1})^{-1} \tilde{b} \tag{2.47}$$

eşitlikleri elde edilir. Elde edilen bu eşitliklere Henderson eşitlikleri denir (Henderson ve diğ., 1959; Searle ve diğ., 2009).

Henderson eşitliklerinden yararlanarak $\hat{\beta}$ 'nin BLUE tahmin edicisi

$$\hat{\beta} = BLUE(\beta) = (X^\top V^{-1} X)^{-1} X^\top V^{-1} y \tag{2.48}$$

ve b rastgele etkiler parametresinin BLUP tahmin edicisi,

$$\tilde{b} = BLUP(b) = DZ^\top V^{-1} (y - X\hat{\beta}) \tag{2.49}$$

şeklinde elde edilir (Henderson ve diğ., 1959; McCulloch ve Neuhaus, 2005; Searle ve diğ., 2009). Henderson denklemlerinden elde edilen eşitliklerde, parametre tahminleri hesaplanırken V^{-1} varyans kovaryans matrisinin tersi yerine R ve D matrislerin tersleri ile işlem yapılır. R ve D matrisleri genellikle köşegen matrislerdir ve V^{-1} matrisinin tersinin alınmasına göre bu matrislerin tersinin alınması çok daha kolaydır. Henderson eşitliklerinden elde edilen parametre tahmin edicileri varyans bileşenlerinin ML ve REML tahminlerinin hesaplanmasında kullanılan sayısal yöntemlerde de kullanılabilirler (Searle ve diğ., 2009).

3 SAĞ KALIM (SURVIVAL) ANALİZİ

Sağ kalım analizi, ilgilenilen olay gerçekleşene kadar geçen sürenin incelenmesinde kullanılan bir yöntemdir. İlgilenilen olay, tedaviye cevap verme, ölüm, bir hastalığın tekrarlaması gibi olaylar olabilir. Sağ kalım analizinde, çalışma başlangıcından ilgilenilen olay gerçekleşene kadar geçen zaman genellikle, sağ kalım zamanı ya da olay zamanı olarak bilinir. Olayın gerçekleşmesi ise, genellikle ilgilenilen olay ölüm, hastalık tekrarlaması gibi negatif deneyimlerden oluştuğundan başarısızlık olarak bilinir. Sağ kalım analizinde amaç sağ kalım verisi üzerinden, iki farklı tedavi yöntemi uygulanan bireylerin sağ kalım eğrilerini karşılaştırmak ve sağ kalım zamanı ile bu süreye etki eden faktörleri açıklamaktır (Kleinbaum, 1998).

Sağ kalım analizini diğer yöntemlerden ayıran en önemli özellik sağ kalım zamanının kesin olarak bilinmediği sansürlü veri tipinin bulunmasıdır. Bireyin çalışmadan kendi isteği ile ayrıldığı (withdraw), bireyin herhangi bir nedenle takipten çıktığı (lost to follow-up, örneğin birey çalışma sırasında başka bir şehire taşınmış ve hasta ile iletişim kesilmiş olabilir.) ya da ilgilenilen olayın, çalışma için belirlenen son zaman noktasına kadar bireyde gözlenmediği durumlarda birey sansürlü olarak kaydedilir. Sansürlü veri, olayın meydana geliş zamanına göre, sağdan sansürlü veri, soldan sansürlü veri, aralıklı sansürlü veri olmak üzere genellikle üç gruba ayrılır.

Sağdan sansürlü veri, çalışmada bulunan bireylerin, çalışma için belirlenen son zaman noktasına kadar, herhangi bir olay yaşamaması, bireyin kendi isteğiyle çalışmadan ayrılması, çalışma sırasında, ilgilenilen olay ortaya çıkmadan bireyin ölüm ya da ortaya çıkan başka bir hastalık nedeniyle çalışmadan ayrılması gibi durumlarda ortaya çıkar. Çalışmalarda genellikle sağdan sansürlü veri ile çalışılır. Sağdan sansürlü veri kendi içinde I.tür sansürleme ve II tür sansürleme olmak üzere ikiye ayrılır. I. tür sansürleme, çalışmada bulunan bireylerin, önceden belirlenen sabit bir durdurma noktasından önce, bireyin sağ kalım olayı yaşamaması durumunda kaydedilen veri türüdür. II. tür sansürleme de ise, araştırmacı tarafından önceden belirlenen başarısızlık

sayısına ulaşıldığında çalışmanın sonlandırılır. Çalışmanın sonlandırıldığı zamana kadar ilgilenilen olayı yaşamayan bireyler sansürlü olarak kaydedilir (Lawless, 2011). Soldan sansürlü veri, bireyin gerçek sağ kalım zamanı, gözlenen sağ kalım zamanından küçük olduğu durumlarda ortaya çıkar. Örneğin, bireylerde HIV virüsü pozitif çıkana kadar gözlemlendiği bir çalışmada, yapılan ilk testte HIV pozitif çıkan bireyler soldan sansürlü veridir. Çünkü bu bireylerde virüsün ne zaman vücuda yerleştiği tam olarak bilinmemektedir (Kleinbaum, 1998).

Aralıklı sansürlü veri ise, ilgilenilen olay iki zaman noktası arasında gerçekleştiği durumlarda ortaya çıkar. Örneğin, HIV virüsü taşıyan hastalara, AIDS teşhisinin konulabilmesi için düzenli olarak belirli kan testleri uygulanır. Hastaya yapılan kan testinde hastanın AIDS hastalığına yakalandığı belirlenirse ve hastalığın ortaya çıkma zamanının ilk kan testi ile son kan testi arasındaki bir zamanda olduğu biliniyorsa, aralıklı sansürlü veri olarak kaydedilir (Kleinbaum, 1998; Rizopoulos, 2012b).

3.1 Sağ Kalım Analizinde Kullanılan Temel Fonksiyonlar

Sağ kalım zamanı, ölüm, hastalığın tekrarlaması gibi ilgilenilen olay gerçekleşene kadar geçen süreyi ifade eden rastgele değişkendir ve tüm rastgele değişkenler gibi dağılıma sahiptir. Sağ kalım zamanının dağılımı genellikle, sağ kalım fonksiyonu, olasılık yoğunluk fonksiyonu ve tehlike fonksiyonu ile ifade edilir. Bu fonksiyonlardan herhangi biri elde edildiğinde diğer fonksiyonlar da elde edilebilir (Lee ve Wang, 2003).

3.1.1 Sağ Kalım Fonksiyonu

Herhangi bir bireye ait sağ kalım zamanı rastgele değişkeni T^* ile gösterilsin. Bir bireyin t zamanından daha uzun yaşama olasılığını ifade eden Sağ kalım fonksiyonu $S(t)$,

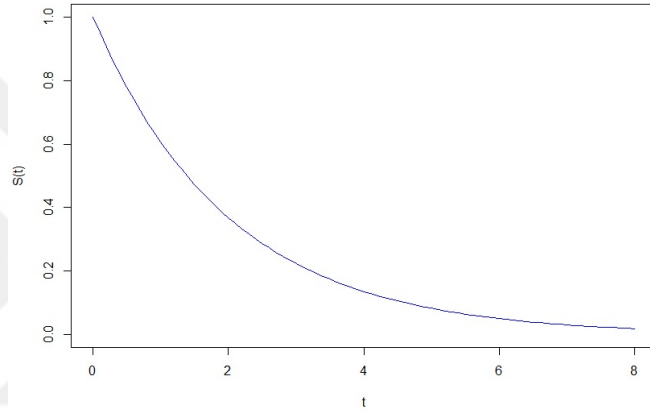
$$S(t) = P(T^* > t) \quad (3.1)$$

şeklinde gösterilir (Lee ve Wang, 2003; Lawless, 2011).

Sağ kalım fonksiyonu $S(t)$, T^* rastgele değişkeninin birikimli yoğunluk fonksiyonu ile ifade edilebilir.

$$S(t) = 1 - P(T^* < t) = 1 - F(t) \quad (3.2)$$

Sağ kalım fonksiyonu $S(t)$ azalan bir fonksiyondur. Çalışma başlangıcı $t=0$ zamanında tüm bireyler hayatta olduğu için sağ kalım fonksiyonu $S(t = 0) = 1$ 'dir. Zaman ilerledikçe hayatta kalan birey sayısı azalacağından $S(t)$ sağ kalım olasılığı azalır (Lee ve Wang, 2003; Kleinbaum, 1998).



Şekil 3.1: Sağ kalım eğrisi

$S(t)$, sağ kalım fonksiyonunun grafiği sağ kalım eğrisi olarak bilinir. Dik bir sağ kalım eğrisi, düşük bir sağ kalım oranını gösterirken, daha düz bir sağ kalım eğrisi ise yüksek sağ kalım oranını ya da uzun sağ kalım zamanını gösterir. Sağ kalım fonksiyonu yardımıyla aynı zamanda medyan değerleri bulunabilir ve iki ya da daha fazla grup karşılaştırılması yapılabilir. Az sayıda birey ile çalışırken, daha uzun ya da daha kısa yaşam süreleri, sağ kalım zamanlarının ortalamasının daha yüksek ya da daha düşük olmasına neden olur. Bu nedenle sağ kalım fonksiyonunun dağılımı tanımlamak için genellikle medyan değeri kullanılır (Lee ve Wang, 2003).

3.1.2 Olasılık Yoğunluk Fonksiyonu

Sağ kalım zamanı T^* 'nin olasılık yoğunluk fonksiyonu, $[t, t + \Delta t)$ gibi kısa bir zaman aralığında bireylerin olay yaşama olasılığını ifade eder.

$$f(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T^* < t + \Delta t)}{\Delta t} \quad (3.3)$$

Sağ kalım zamanı olasılık yoğunluk fonksiyonu;

- $f(t)$ negatif olmayan bir fonksiyondur, $f(t) \geq 0$.

- $\int_{\mathbb{R}} f(t) dt = 1$

özelliklerini sağlar. (Lee ve Wang, 2003)

3.1.3 Tehlike Fonksiyonu

Tehlike fonksiyonu, bireyin belirli bir t zamanına kadar hayatta olduğu biliniyorken, $[t, t + \Delta t)$ gibi kısa bir zaman aralığında, olay yaşama riskini gösteren fonksiyondur. Tehlike fonksiyonu genellikle $h(t)$ ile gösterilir ve sağ kalım zamanı T 'nin koşullu başarısızlık oranını verir.

$$h(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T^* < t + \Delta t \mid T^* \geq t)}{\Delta t} \quad (3.4)$$

Tehlike fonksiyonu birikimli yoğunluk fonksiyonunu ve olasılık yoğunluk fonksiyonu yardımıyla aşağıdaki gibi ifade edilebilir.

$$h(t) = \frac{f(t)}{1 - F(t)} \quad (3.5)$$

Tehlike fonksiyonu aynı zamanda anlık başarısızlık oranı olarak da bilinir. Tehlike fonksiyonu, zaman boyunca artabilir, azabilir ya da sabit kalabilir. Tehlike oranı sıfıra eşit ya da sıfırdan büyük olabilir fakat negatif olmayan bir fonksiyondur (Kleinbaum, 1998).

Tehlike fonksiyonunun olasılık yoğunluk fonksiyonu, sağ kalım fonksiyonu ve birikimli tehlike fonksiyonu arasındaki ilişki aşağıdaki gibi gösterilebilir.

$$h(t) = \frac{f(t)}{S(t)} \quad (3.6)$$

Bu ilişki aynı zamanda koşullu olasılık yardımıyla da yazılabilir. B bilindiğinde A olayının meydana gelme olasılığı, $P(A | B) = \frac{P(A \cap B)}{P(B)}$ ile elde edilebiliyordu. Bu durumda, $h(t)$ yoğunluk fonksiyonunun pay kısmında yer alan $P(t \leq T < t + \Delta t | T \geq t)$ olasılığı,

$$\begin{aligned}
 P(t \leq T^* < t + \Delta t | T^* \geq t) &= \frac{P(t \leq T^* < t + \Delta t \cap T^* \geq t)}{P(T^* \geq t)} \\
 &= \frac{P(t \leq T^* < t + \Delta t)}{P(T^* \geq T^*)} \\
 &= \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T^* < t + \Delta t) \Delta t}{P(T^* \geq t) \Delta t} \\
 &= \lim_{\Delta t \rightarrow 0} \frac{F(t + \Delta t) - F(t) \Delta t}{P(T^* \geq t) \Delta t}
 \end{aligned} \tag{3.7}$$

şeklinde yazılabilir. Daha önce belirtildiği gibi $f(t)$, $[t, t + \Delta t)$ gibi kısa bir zaman aralığı için bireylerin olay yaşama olasılığını verir. Δt gibi küçük bir zaman aralığı için olasılık yoğunluk fonksiyonu,

$$f(t) \Delta t \simeq P(t \leq T^* < t + \Delta t) = F(t + \Delta t) - F(t) \tag{3.8}$$

olarak ifade edilebilir. Bu durumda Eşitlik 3.7'de yer alan ifade tekrar düzenlendiğinde tehlike fonksiyonu aşağıdaki gibi yazılabilir (Collett, 2015; Lee ve Wang, 2003).

$$h(t) = \frac{f(t) \Delta t}{S(t) \Delta t} = \frac{f(t)}{S(t)} \tag{3.9}$$

Birikimli olasılık yoğunluk fonksiyonunun türevi yoğunluk fonksiyonuna eşittir.

$$f(t) = \frac{d}{dt} [1 - S(t)] = -S'(t) \tag{3.10}$$

Eşitlik 3.9'da yerine koyulularak tehlike fonksiyonu eşitlikteki gibi yazılabilir,

$$h(t) = -\frac{S'(t)}{S(t)} = -\frac{d}{dt} \log s(t) \tag{3.11}$$

$h(t)$ tehlike fonksiyonunun integrali alındığında,

$$-\int_0^t h(x) dx = -\log S(t) \tag{3.12}$$

elde edilir. Bu eşitlik birikimli tehlike fonksiyonunu verir.

$$H(t) = -\log(S(t)) \quad (3.13)$$

ya da sağ kalım fonksiyonu, birikimli tehlike fonksiyonu kullanılarak,

$$S(t) = \exp\left(-H(t)\right) = \exp\left[-\int_0^t h(x)dx\right] \quad (3.14)$$

olarak yazılabilir (Collett, 2015; Lee ve Wang, 2003).

3.2 Sağ Kalım Fonksiyonunda Kullanılan Bazı Önemli Dağılımlar

Sağ kalım zamanının modellenmesinde, uygun olasılık dağılımının seçimi önemlidir. Sağ kalım zamanının dağılımının belirlenmesinde genellikle Üstel dağılım, Weibull dağılımı, Lognormal dağılım kullanılır.

3.2.1 Üstel Dağılım

Üstel dağılım, sabit tehlike fonksiyonu ile tanımlanır. Yüksek λ değeri, yüksek risk ve kısa sağ kalım süresini, düşük λ değerleri ise düşük risk ve uzun sağ kalım süresini ifade eder. $\lambda > 0$ olmak üzere, sağ kalım süresinin dağılımı, üstel dağılım olarak belirlendiğinde, sağ kalım süresi olasılık yoğunluk fonksiyonu,

$$f(t) = \begin{cases} \lambda e^{-\lambda t} & t \geq 0, \lambda > 0 \\ 0 & \end{cases} \quad (3.15)$$

sağ kalım fonksiyonu,

$$S(t) = e^{-\lambda t} \quad t \geq 0 \quad (3.16)$$

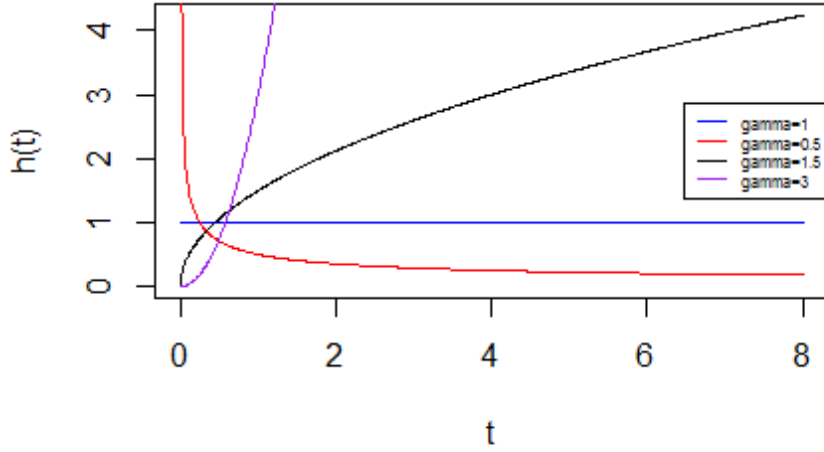
ve tehlike fonksiyonu,

$$h(t) = \lambda \quad t \geq 0 \quad (3.17)$$

şeklinde tanımlanır. Sağ kalım süresinin dağılımı üstel dağılım olarak seçildiğinde, tehlike fonksiyonu zaman boyunca sabittir (Lee ve Wang, 2003).

3.2.2 Weibull Dağılımı

Weibull dağılımı, üstel dağılımın genelleştirilmiş halidir. Üstel dağılımdan farklı olarak tehlike fonksiyonu, sabit bir tehlike oranını ifade etmez. Weibull dağılımı, şekil parametresi λ ve ölçek parametresi ρ olmak üzere iki parametreye sahiptir. Şekil parametresi $\lambda = 1$ olarak belirlendiğinde sabit tehlike oranını, $\lambda > 1$ artan tehlike oranını, $\lambda < 1$ azalan tehlike oranını belirtir.



Şekil 3.2: λ parametresinin farklı değerleri için tehlike fonksiyonun grafiği

Olasılık yoğunluk fonksiyonu,

$$f(t) = \lambda \rho (\lambda t)^{\rho-1} e^{-(\lambda t)^\rho} \quad t \geq 0, \quad \lambda, \rho > 0 \quad (3.18)$$

sağ kalım fonksiyonu,

$$S(t) = e^{-(\lambda t)^\rho} \quad (3.19)$$

ve tehlike fonksiyonu,

$$h(t) = \lambda \rho (\lambda t)^{\rho-1} \quad (3.20)$$

şeklinde gösterilir (Lee ve Wang, 2003; Lawless, 2011).

3.2.3 Log-Normal Dağılım

Log-normal dağılım en basit şekilde bir değişkenin logaritması alındıktan sonra normal dağılım göstermesi olarak tanımlanabilir. Logaritması alınan sağ kalım süresi μ ortalama ve σ^2 varyans ile normal dağılıma sahip ise sağ kalım süresinin olasılık yoğunluk fonksiyonu

$$f(t) = \frac{1}{(2\pi)^{-\frac{1}{2}}\sigma} \exp \left[-\frac{1}{2} \left(\frac{\log t - \mu}{\sigma} \right)^2 \right] \quad t > 0 \quad (3.21)$$

şeklinde yazılabilir. $\phi(\cdot)$ normal dağılımın birikimli yoğunluk fonksiyonunu göstermek üzere birikimli yoğunluk fonksiyonu,

$$F(t) = \phi \left(\frac{\log t - \mu}{\sigma} \right) \quad (3.22)$$

şeklinde yazılabilir.

Sağ kalım fonksiyonu ise,

$$S(t) = 1 - \phi \left(\frac{\log t - \mu}{\sigma} \right) \quad (3.23)$$

şeklinde dir. Tehlike fonksiyonu daha önce de belirtilen fonksiyonlar arası ilişkiden yararlanarak $h(t) = \frac{p(t)}{S(t)}$ eşitliği yardımıyla elde edilebilir (Lee ve Wang, 2003).

$$h(t) = \frac{\frac{1}{(2\pi)^{-\frac{1}{2}}\sigma_t} \exp \left[-\frac{1}{2} \left(\frac{\log t - \mu}{\sigma_t} \right)^2 \right]}{1 - \phi \left(\frac{\log t - \mu}{\sigma} \right)} \quad (3.24)$$

3.3 Cox Oransal Tehlike Modeli

Sağ kalım analizinde, bireylerin sağ kalım süresine etki eden faktörleri ortaya çıkarmak için genellikle Cox (1972) tarafından geliştirilen Cox oransal tehlike modeli kullanılır. Cox modeli oluşturulurken sağ kalım fonksiyonu ve tehlike fonksiyonunu tahmin

edilmeden önce sansür bilgisi de modele dahil edilmediler. T_i , gözlenen sağ kalım zamanı olmak üzere, T_i^* , gerçek olay zamanı ile C_i , sansürlenme zamanının minimum değeri $\min(T_i^*, C_i)$, gözlenen sağ kalım zamanını verir. Bu sürece ek olarak olay gösterge fonksiyonu, $\delta_i = I_i(T_i^* \leq C_i)$ şeklinde gösterilir. Birey çalışma sırasında ilgilenilen olayı yaşadıysa 1, yaşamadıysa 0 değerini alır. $I(\cdot)$ gösterge fonksiyonudur. Cox oransal tehlike modeli genellikle;

$$h_i(t | w_i) = \lim_{\Delta t \rightarrow 0} P(t \leq T^* < t + \Delta t | T^* \geq t, w_i) \Delta t = h_0(t) \exp(\gamma^T w_i) \quad (3.25)$$

şeklinde gösterilir (Kalbfleisch ve Prentice, 2011).

Modelde, $h_i(t)$, i . bireyin t zamanındaki tehlike fonksiyonunu gösterir. $h_0(t)$, tüm değişkenler sıfır değerini aldığında $w_i = 0$, bireylerin olay yaşama riskini gösteren temel tehlike fonksiyonudur. $w_i = (w_{i1}, \dots, w_{ip})$, bireylerin sağ kalım olayı yaşamalarına etkisi olduğu düşünülen açıklayıcı değişken vektörüdür. γ , w_i açıklayıcı değişkenlerine ait regresyon parametresini gösterir. Cox modelde, $\exp(\gamma)$, t zamanında w_i açıklayıcı değişkenindeki bir birimlik artışa veya azalışa karşılık gelen tehlike oranındaki değişimi gösterir. Cox oransal tehlike modelinde, açıklayıcı değişken w_i , zamandan bağımsız değişkendir. w_i ve w_k değişken değerlerine sahip iki bireyin tehlike oranları $\frac{h_i(t)}{h_k(t)}$ oranı sabittir. Bu nedenle Oransal tehlike olarak isimlendirilir. Modelin sağ tarafında yer alan $\exp(\gamma^T w_i)$ ifadesi göreceli (relative) riski ifade ettiği için aynı zamanda bu model göreceli risk modeli (relative risk model) olarak da bilinmektedir (Kalbfleisch ve Prentice, 2011; Rizopoulos, 2012b).

$$\frac{h_i(t | w_i)}{h_k(t | w_k)} = \exp(\gamma^T (w_i - w_k)) \quad (3.26)$$

Oransal Cox modelinde sağ kalım fonksiyonu,

$$S(t | x) = \exp\left(-\int_0^t h_0(u) \exp(\gamma^T w_i) du\right) \quad (3.27)$$

şeklinde yazılabilir.

Olasılık yoğunluk fonksiyonu,

$$h(t | x) = h_0(t) \exp(\gamma^\top w_i) \exp\left(-\int_0^t h_0(u) \exp(\gamma^\top w_i) du\right) \quad (3.28)$$

şeklinde yazılabilir.

Cox oransal tehlike modelinde parametre tahminleri için kısmi olabilirlik yönteminden yararlanır. k tane farklı başarısızlık zamanının olduğu bir çalışmada $t_{(1)} < t_{(2)} < \dots < t_{(k)}$ sıralı başarısızlık zamanları, $x_{(1)}, x_{(2)}, \dots, x_{(k)}$ başarısızlık zamanlarına karşılık gelen açıklayıcı değişkenler olsun. $R(t_i)$ risk kümesi, $t_{(i)}$ zamanına kadar çalışmada olan risk altındaki bireylerden oluşur. Belirli bir $t_{(i)}$ başarısızlık zamanında $R(t_i)$ risk kümesine koşullu olarak bireylerin başarısızlık olasılıkları

$$\frac{\exp(\sum_{j=1}^p \gamma_j^\top w_i)}{\sum_{t \in R(t_i)} \exp(\sum_{j=1}^p \gamma_j^\top w_i)} \quad (3.29)$$

şeklinde gösterilir. Bu analizde temel tehlike fonksiyonu nüans parametresi olarak kabul edilir (Cox, 1972).

Kısmi olabilirlik fonksiyonu aşağıdaki gibi yazılabilir.

$$L(\gamma) = \prod_{i=1}^k \frac{\exp(\sum_{j=1}^p \gamma_j^\top w_i)}{\sum_{t \in R(t_i)} \exp(\sum_{j=1}^p \gamma_j^\top w_i)} \quad (3.30)$$

Kısmi olabilirlik fonksiyonunun payında yer alan ifade yalnızca olay yaşayan bireylerin bilgilerini içerirken paydadaki ifade ise henüz olay yaşamamış risk altındaki bireylerin bilgisini içermektedir. Eşitlik 3.30'de yer alan kısmi olabilirlik fonksiyonunun logaritması alındığında

$$\begin{aligned} \ell(\gamma) &= \sum_{i=1}^k \sum_{j=1}^p \gamma_j^\top w_i - \sum_{i=1}^k \log \left[\sum_{t \in R(t_i)} \exp(\sum_{j=1}^p \gamma_j^\top w_i) \right] \\ &= \sum_{i=1}^k \left(\gamma^\top w_i - \log \left(\sum_{t \in R(t_i)} \exp(\sum_{j=1}^p \gamma_j^\top w_i) \right) \right) \end{aligned} \quad (3.31)$$

kısmi log olabilirlik fonksiyonu elde edilir (Cox, 1972; Lee ve Wang, 2003).

Eşitlik 3.31'de yer alan kısmi log olabilirlik fonksiyonunun γ parametresine göre birinci dereceden türevlerini veren $U(\gamma)$ skor fonksiyonu sıfıra eşitlendiğinde γ parametresinin kısmi en çok olabilirlik tahmin edicisi elde edilir.

$$U(\gamma) = \frac{\partial \ell(\gamma)}{\partial \gamma} = \sum_{i=1}^k w_i - \sum_{i=1}^k \frac{\sum_{j \in R_{(i)}} \exp(\gamma^T w_j) w_j}{\sum_{j \in R_{(i)}} \exp(\gamma^T w_j)} = 0 \quad (3.32)$$

Log olabilirlik fonksiyonunun ikinci dereceden türevlerinin negatif değerleri $I(\gamma)$ gözlenen bilgi matrisini verir (Klein ve Moeschberger, 2006).

$$I(\hat{\gamma}) = - \sum_{i=1}^n \frac{\partial^2 \ell_i(\gamma)}{\partial \gamma \partial \gamma} \Big|_{\gamma=\hat{\gamma}} \quad (3.33)$$

Olabilirlik fonksiyonunu maksimize eden γ parametresinin tahmin edicisi elde etmek için Newton Raphson gibi iteratif yöntemlerden yararlanır. Newton Raphson yönteminde $s = 0, 1, 2, \dots$, iterasyon sayısını göstermek üzere γ_0 için başlangıç değeri seçilir. Başlangıç değerinin $\gamma_0 = 0$ olarak seçilmesi parametreye yakınsama sağlamak için genellikle yeterlidir. $s + 1$ iterasyon için γ_{s+1} parametresinin tahmini $I(\gamma)$ gözlenen bilgi matrisi ve $U(\gamma)$ yardımıyla aşağıdaki gibi elde edilir.

$$\gamma_{s+1} = \hat{\gamma}_s + I^{-1}(\hat{\gamma}_s)U(\hat{\gamma}_s) \quad (3.34)$$

İterasyona log olabilirlik fonksiyonundaki değişim azalıncaya kadar devam edilir. γ parametresine yakınsama sağlandığında $I^{-1}(\hat{\gamma})$ bilgi matrisinin tersinin tersi ile $\hat{\gamma}$ parametresine ait varyans kovaryans matrisi elde edilebilir. Bu matrisin köşegen elemanlarının kökleri γ parametresinin standart hatalarını verir (Collett, 2015).

3.3.1 Model Anlamlılığının Testi

Kurulan modelde, $H_0 : \gamma = \gamma_0$ yokluk hipotezinin testi için Olabilirlik oran (LR) testi, Wald testi ve Skor testinden yararlanır.

Model katsayıların testi için kullanılan LR testinin istatistiği ise aşağıdaki gibi ifade edilir.

$$LR = -2 \left\{ \ell(\hat{\gamma}_0) - \ell(\hat{\gamma}) \right\} \quad (3.35)$$

LR testine ait test istatistiği p serbestlik dereceli χ_p^2 dağılımına sahiptir.

Wald test istatistiği,

$$W = (\hat{\gamma} - \gamma_0)^T I(\hat{\gamma})(\hat{\gamma} - \gamma_0) \quad (3.36)$$

şeklindedir. Yokluk hipotezi altında, test istatistiğinin asimptotik dağılımı p serbestlik dereceli χ_p^2 dağılımı gösterir. p test edilen parametre sayısını ifade etmektedir. γ_j parametresinin anlamlılığının testi için Wald test istatistiği aşağıdaki gibidir.

$$Z = \frac{(\gamma_j - \hat{\gamma}_j)}{sh(\hat{\gamma}_j)} \quad (3.37)$$

Skor testi test istatistiği $S(\gamma)$ skor fonksiyonu ve $I(\gamma)$ gözlenen bilgi matrisinden yararlanarak aşağıdaki gibi elde edilir.

$$U = S^T(\hat{\gamma}_0) \left\{ I(\hat{\gamma}_0) \right\}^{-1} S(\hat{\gamma}_0) \quad (3.38)$$

3.4 Zamana Bağlı Değişkenler

Oransal Cox modelinde, bireylerin sağ kalım olayı yaşama riski üzerinde etkisi olabileceği düşünülen değişkenlerin cinsiyet, tedavi yöntemi, temel yaş gibi zaman boyunca sabit değişkenler olduğu varsayılır. Bazı durumlarda risk üzerinde etkili olan değişkenler çevresel faktörler ya da izlenme zamanı boyunca bireylerden alınan klinik ölçümler gibi zamana bağlı değişkenler olabilir. $x_i(t)$, i . bireyin t . zamandaki değişken vektörünü gösterebilir. $X_i(t) = \{x_i(s); 0 \leq s < t\}$, ise i . bireyin t zamanına kadar olan değişkenlerini ifade eder. Bu durumda tehlike fonksiyonu t zamanına kadar olan değişkenlere koşullu olarak,

$$h(t; X(t)) = P\{T^* \in [t, t + \Delta t) \mid X(t), T^* \geq t\} \quad (3.39)$$

şeklinde yazılır (Kalbfleisch ve Prentice, 2011). Zamana bağlı değişkenler içsel değişkenler (internal ya da endogenous) ve dışsal değişkenler (external ya da exogenous) olmak üzere ikiye ayrılabilir. Zamana bağlı herhangi bir değişken,

$$P\{T^* \in [u, u + \Delta u) \mid X(u), T^* \geq u\} = P\{T^* \in [u, u + \Delta u) \mid X(t), T^* \geq u\} \quad (3.40)$$

koşulunu sağlıyorsa bu değişkene dışsal değişken denir. t zaman noktasındaki herhangi bir dışsal değişken değeri, u zamanında meydana gelen olaydan etkilenmemektedir, $t > s$. Örneğin, astım hastalarının astım atağı geçirme sıklığı ile hava kirliliği arasında ilişki bulunabilir. Hava kirliliği değişkeni dışsal değişkendir çünkü bireyin herhangi bir s zamanında astım atağı geçirmesinin, t zamanındaki hava kirliliği seviyesi değişimine etkisi bulunmamaktadır (Rizopoulos, 2012b; Kalbfleisch ve Prentice, 2011).

İçsel değişkenler ise, çalışma sırasında bireylerden alınan, kan hücre sayımları, kolesterol düzeyleri gibi biyobelirteç ya da klinik parametreleri içeren ve bireyin sağ kalım zamanı ile ilgili bilgi içeren zamana bağlı değişkenlerdir. İçsel değişkenlerin en önemli özelliği bireyin çalışmada yer aldığı yani sağ kalım olayı yaşamadığı ya da sansürlenmediği sürece gözlenebilen değişken olmasıdır. Kısaca bireyin s zamanında, ilgilenilen olayı yaşaması, $t \geq s$ zamanında bireye ait değişken değerinin bulunmaması durumuna karşılık gelir.

3.5 Genişletilmiş (Extended) Cox Modeli

Eşitlik 3.25'te verilen Cox oransal tehlike modeli, modelde zamana bağlı değişkenler yer aldığına,

$$h_i(t | w_i, y_i(t)) = h_0(t) \exp \left(\gamma^T w_i + \alpha y_i(t) \right) \quad (3.41)$$

şeklinde genişletilebilir.

Modelde, w_i Cox oransal tehlike modelindeki gibi cinsiyet, tedavi yöntemi gibi zaman boyunca sabit olan değişkenleri ifade eder. $y_i(t)$, zamana bağlı değişkenleri içeren açıklayıcı değişken vektörüdür. γ ve α parametreleri sırasıyla zamandan bağımsız ve zamana bağımlı açıklayıcı değişkenlere karşılık gelen regresyon parametrelerini ifade eder. Cox oransal tehlike modelinden farklı olarak, $y_i(t)$ açıklayıcı değişkeni zamana bağlı olduğundan, bireyler için tehlike oranı da zamana bağlıdır (Collett, 2015). Genişletilmiş Cox modelinde i ve k bireyi için tehlike oranı;

$$\frac{h_i(t | w_i, y_i(t))}{h_s(t | w_k, y_k(t))} = \exp \left(\gamma^T (w_i - w_k) \right) + \exp \left(\alpha (X_w(t) - Y_k(t)) \right) \quad (3.42)$$

şeklinde ifade edilir. γ ve α parametrelerinin tahmin edicileri, Cox oransal tehlike modelinde olduğu gibi kısmi en çok olabilirlik yöntemi ile elde edilir. Genişletilmiş Cox modeli teorik olarak, modelde dışsal değişkenlerin yer aldığı durumlar için geçerlidir yani biyobelirteç gibi bireye ait ölçümlerin yer aldığı içsel değişkenler için geçerli değildir. Bunun nedeni, genişletilmiş Cox modelde, hastadan alınan biyobelirteçlerin ziyaret süreleri arasında sabit olduğu varsayımının yapılmasıdır. Halbuki kan sayımı ölçümleri, kolesterol düzeyi gibi biyobelirteçler zaman boyunca değişkenlik gösteren değişkenlerdir. Ziyaret süreleri arasında bu ölçümlerin sabit kaldığı varsayımı, parametre tahminleri ve standart hataların yanlış tahmin edilmesine neden olabilir (Rizopoulos, 2012b).



4 BOYLAMSAL VE SAĞ KALIM VERİLERİ İÇİN BİRLEŞİK MODELLEME

Bu bölümde, boylamsal ölçümler ve sağ kalım çıktıları arasındaki ilişkinin belirlenmesinde kullanılan birleşik modelden ve model parametrelerinin tahmininde kullanılan yöntemlerden bahsedilecektir.

4.1 Birleşik Model Yapısı

Boylamsal çalışmalarda, zamana bağlı tekrarlı ölçümler toplanırken aynı bireylere ait ölüm, ilgilenilen hastalığın tekrarlanması gibi sağ kalım sürecine ait bilgiler de kaydedilir. Sağ kalım süreci ile ilişkili olduğu düşünülen tekrarlı ölçümlerin sayısı bireyden bireye farklılık göstermektedir ve bu gözlemler genellikle ölçüm hatası içermektedir. Modelde zamana bağlı değişkenler yer aldığına, sağ kalım sürecinin modellenmesinde genellikle Bölüm 3'te bahsedilen genişletilmiş Cox modeli kullanılır. Sağ kalım sürecinde genişletilmiş Cox modelin kullanılmasının dezavantajlarından biri, ölçüm hatası içeren boylamsal değişkenlerin modele dahil edilmesidir. Ölçüm hatası içeren boylamsal değişkenler yanlış parametre tahminlerinin elde edilmesine neden olabilir (Prentice, 1982; Wulfsohn ve Tsiatis, 1997). Diğer dezavantajı ise, genişletilmiş Cox modelinde, bireylerden alınan tekrarlı ölçümlerin birbirini takip eden iki zaman aralığında sabit kaldığı varsayımdır. Bu varsayım tedavide uygulanan doz miktarı gibi dışsal değişkenler için geçerli olsa da bireyden alınan biyobelirteç gibi sürekli değerler için geçerli bir varsayım değildir (Wulfsohn ve Tsiatis, 1997; Kalbfleisch ve Prentice, 2011; McCrink ve diğ., 2013). Boylamsal değişkenler kesikli zaman noktalarında ölçülen ve genellikle ölçüm hatası içeren değişkenlerdir. Birleşik modeller, bu değişkenlerin sürekli olarak tahmin edilmesine olanak sağlar ve sağ kalım sürecinde boylamsal değişkenlerin ölçüm hatası içermeyen sürekli tahmin değerlerini kullanarak parametre tahminlerinde bulunur. Ölçüm hatasından arındırılmış sürekli tahmin edilen boylamsal değişkenler parametre

tahminlerindeki yanlılığı azaltır ve daha etkili tahminlerin elde edilmesini sağlar. Özellikle sağ kalım sürecinde boylamsal değişkenlerle bireyin ilgilenilen sağ kalım olayını yaşaması arasındaki ilişkinin incelendiği klinik çalışmalarda önemlidir çünkü ölçüm hatası içeren boylamsal değişkenlerin modelde kullanılması tehlike oranlarının yüksek tahmin edilmesine neden olur ve dolayısıyla tehlike oranlarının yüksek tahmin edilmesine yol açar. Birleşik modelde elde edilen tahminler Cox model ile karşılaştırıldığında daha küçük standart hataya sahiptirler (Ibrahim ve diğ., 2010; Lawrence Gould ve diğ., 2015). Konu ile ilgili olarak Birleşik model ve Cox modelden elde edilen tahminlerdeki yanlılık miktarlarının detaylı olarak incelendiği Ibrahim ve diğ. (2010) çalışmasına bakılabilir.

Zamana bağlı içsel değişkenlerin yer aldığı modelde parametre tahminleri için önerilen yaklaşımlardan diğeri Self ve Pawitan (1992) tarafından geliştirilen iki aşamalı yaklaşımdır. Bu yaklaşım parametre tahminlerindeki yanlılığı azaltır fakat sağ kalım modelinde ölçüm hatası içeren boylamsal ölçümleri kullandığından tamamen yansız bir yaklaşım değildir. Birleşik modelde, sağ kalım sürecinin tahmininde, modelde yer alan zamana bağlı değişkenlerin ölçüm hatası içermediği ve boylamsal ölçümlerin sürekli değerler olduğu varsayımı altında parametre tahminleri yapıldığından parametre tahminlerindeki yanlılık genişletilmiş Cox model ve iki aşamalı yaklaşım gibi alternatif yöntemlere göre daha azdır (Wulfsohn ve Tsiatis, 1997; Ibrahim ve diğ., 2010; Rizopoulos, 2012b; McCrink ve diğ., 2013). Birleşik model, genellikle boylamsal süreç için çok düzeyli model olarak da bilinen LME modeli (Laird ve Ware, 1982), sağ kalım süreci için Cox oransal tehlike modeli (Cox, 1972) olmak üzere iki alt modelden oluşur. Birleşik model, zamana bağlı içsel değişkenler ile sağ kalım süreci arasındaki ilişkinin ortaya çıkarılmasına olanak sağlar. Süreçler arasındaki ilişki, iki modelde de ortak parametre olan rastgele etki parametreleri ile sağlanır (Faucett ve Thomas, 1996; Wulfsohn ve Tsiatis, 1997; Tsiatis ve Davidian, 2004; Henderson ve diğ., 2000).

Birleşik model tanıtılırken birinci ve ikinci bölümde kullanılan notasyonlar kullanılacaktır. Sağ kalım sürecinde, T_i^* i . birey için gerçek olay zamanını göstermektedir. C_i sansürlenme zamanı ile T_i^* gerçek olay zamanının minimum değeri, $\min(T_i^*, C_i)$, T_i gözlenen olay zamanını ve $\delta_i = I(T_i^* \leq C_i)$ olay gösterge fonksiyonunu göstermektedir. Boylamsal süreçte, $y_i(t)$ i . birimin t zamanında

gözlenen içsel değişkenleri değerini göstermektedir. Burada, $y_i(t)$ içsel değişken değeri belirli bir t_{ij} zamanı için gözlenen değerdir. Gözlenen boylamsal ölçümler $y_{ij} = y_i(t_{ij}), j = 1, 2, \dots, n_i$ şeklinde gösterilebilir.

4.1.1 Sağ Kalım Alt Modeli

Birleşik modelde amaç ölçüm hatasından arındırılmış boylamsal ölçümler ile olay yaşama riski arasındaki ilişkiyi incelemektir. Bu amaçla, sağ kalım alt modelinde kullanılan Cox oransal tehlike modelinde $y_i(t)$ boylamsal ölçümlerin gözlenen değerlerinin yerine, Rizopoulos (2012b) tarafından önerilen $m_i(t)$ terimi kullanılır. $m_i(t)$ i . bireyin t zamanındaki gözlenemeyen gerçek değerini göstermektedir. Birleşik modelde, Cox oransal tehlike modelinde kullanılan $m_i(t)$ 'nin ölçüm hatası içermediği varsayıldığından genişletilmiş Cox modelinde kullanılan $y_i(t)$ 'den farklıdır

Birleşik modelde $m_i(t)$, gözlenemeyen gerçek değerler ile sağ kalım olayı yaşama riski arasındaki ilişkiyi inceleyebilmek için Cox oransal tehlike modeli,

$$\begin{aligned} h_i(t | M_i(t), w_i) &= \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T_i^* < t + \Delta t | T_i^* > t, M_i(t), w_i)}{\Delta t} \\ &= h_0(t) \exp(\gamma^T w_i + \alpha m_i(t)), \quad t > 0 \end{aligned} \quad (4.1)$$

şeklinde yazılabilir. Modelde $M_i(t) = (m_i(s), 0 \leq s < t)$, i . bireyin başlangıç zamanından t zamanına kadar olan boylamsal ölçümlerini, $h_0(\cdot)$ temel risk fonksiyonunu, w_i , cinsiyet, tedavi yöntemi gibi temel açıklayıcı değişkenleri, γ , temel açıklayıcı değişkenlere karşılık gelen regresyon parametresini göstermektedir. α boylamsal ölçümlerin ile sağ kalım olayı yaşama riski üzerindeki etkisini gösteren regresyon parametresidir. $\exp(\gamma)$ herhangi bir t zamanı için w_{ij} açıklayıcı değişkeninde meydana gelen bir birimlik değişim için tehlike oranını gösterir. $\exp(\alpha)$, herhangi bir t zaman noktasında, $m_i(t)$ değerindeki bir birimlik değişimin sonucu olarak aynı zaman noktasında meydana gelen sağ kalım olayı için göreceli artışı ifade eder.

Eşitlik 4.1'de verilen modelde Cox oransal tehlike modelinde, sağ kalım olayı yaşama riski sadece t zamanında alınan $m_i(t)$ boylamsal ölçüm değerine bağlıdır. t zamanına kadar olan $M_i(t)$ boylamsal ölçüm değerlerini de modele dahil etmek için Eşitlik 3.4'te verilen sağ kalım fonksiyonu ve birikimli tehlike fonksiyonu arasındaki ilişkiden yararlanarak,

$$S(t) = \exp\left(-\int_0^t h(s)ds\right)$$

şeklinde olmak üzere sağ kalım fonksiyonu

$$\begin{aligned} S_i(t | M_i(t), w_i) &= P(T_i^* > t | M_i(t), w_i) \\ &= \exp\left(-\int_0^t h_0(s) \exp \gamma^\top w_i + \alpha m_i(s) ds\right) \end{aligned} \quad (4.2)$$

şeklinde yazılabilir. Eşitlik 4.2'de verilen sağ kalım fonksiyonu, i . bireyin t zamanına kadar boylamsal ölçümlerine bağlıdır. (Rizopoulos, 2012b)

Temel tehlike fonksiyonu $h_0(\cdot)$ tanımsız olarak bırakılabilir fakat birleşik modelde tanımsız olarak bırakılması parametrelerin standart hata tahminlerinin küçük tahmin edilmesine yol açabilir. Bu nedenle $h_0(\cdot)$ temel tehlike fonksiyonu tanımsız olarak bırakıldığında parametrelerin standart hata tahminlerinin dayanıklı (bootstrap) tahmin yöntemleri ile elde edilmesi önerilmiştir (Hu ve diğ., 2009). $h_0(\cdot)$ temel tehlike fonksiyonu için Weibull, log-normal, gamma dağılımı gibi parametrik dağılımlar da kullanılabilir. Örneğin, λ şekil ρ ölçek parametresi olmak üzere Weibull dağılımına sahip temel tehlike fonksiyonu, $h_0(t) = \lambda \rho (\lambda t)^{\rho-1}$ şeklinde yazılabilir. $\rho > 1$ artan risk oranını, $\rho < 1$ azalan risk oranını, $\rho = 1$ sabit risk oranını göstermektedir. $\rho = 1$ durumu aynı zamanda üstel dağılıma karşılık gelmektedir. Temel tehlike fonksiyonu için parametrik ve esnek yöntemler de tercih edilebilir. Çalışmalarda, temel tehlike fonksiyonu için tercih edilen yöntemler genellikle parçalı sabit (piecewise constant) ve regresyon spline yaklaşımlarıdır (Rizopoulos, 2012b).

Parçalı sabit modelde temel tehlike fonksiyonu,

$$h_0(t) = \sum_{q=1}^Q \xi_q I(v_{q-1} < t < v_q) \quad (4.3)$$

biçiminde gösterilir. $0 = v_0 < v_1 < \dots < v_q$ zaman ölçeğinin bölümlerini gösterir. v_Q gözlenen en büyük zamandan daha büyüktür ve ξ_q , (v_{q-1}, v_q) zaman aralığındaki tehlike değerini gösterir. Düğüm sayısı arttıkça temel tehlike fonksiyonunun belirlenmesi daha esnek hale gelmektedir.

Regresyon Spline modelinde log temel tehlike fonksiyonu,

$$\log h_0(t) = \kappa_0 + \sum_{d=1}^m \kappa_d B_d(t, q) \quad (4.4)$$

şeklinde yazılabilir. $\kappa^T = (\kappa_0, \kappa_1, \dots, \kappa_m)$, spline katsayılarını, q , B spline fonksiyonu $B(\cdot)$ 'nin derecesini göstermektedir. \dot{m} düğüm sayısını göstermek üzere $m = \dot{m} + q + 1$ şeklinde ifade edilir. Parçalı sabit modelde olduğu gibi düğüm sayısı arttıkça temel tehlike fonksiyonu daha esnek olmaktadır (Rizopoulos, 2012b).

4.1.2 Boylamsal Alt Model

Sağ kalım alt modelinde tanımlanan $m_i(t)$, boylamsal ölçümlerin gözlenemeyen gerçek değerlerini belirtir. Boylamsal ölçümlerin sağ kalım olayı yaşama riski üzerindeki etkisini ölçebilmek için $m_i(t)$ 'nin tahminin elde edilmesi ve her bireyin t zamanına kadar olan ölçüm değerlerinin de modele dahil edilmesi gerekir. Bu nedenle boylamsal ölçümlerin normal dağıldığı varsayımıyla boylamsal alt model için LME modeli ,

$$\begin{aligned} y_i(t) &= m_i(t) + \varepsilon_i(t) \\ m_i(t) &= x_i^T(t)\beta + Z_i^T(t)b_i \\ b_i &\sim N(0, D) \\ \varepsilon_i(t) &\sim N(0, \sigma^2) \end{aligned} \quad (4.5)$$

şeklinde yeniden düzenlenebilir. Modelde, $x_i(t)$, sabit etkilere ait tasarım vektörünü, β sabit etkiler parametresini $z_i(t)$ rastgele etkiler tasarım vektörünü ve b_i rastgele etkileri göstermektedir. $\varepsilon_i(t)$ hata terimleri 0 ortalama ve σ^2 varyans ile normal dağılıma sahiptir. b_i rastgele etkiler, 0 ortalama ve D varyans kovaryans matrisi ile normal dağılıma sahip olduğu ve b_i rastgele etkiler ile ε_i hata terimlerinin bağımsız olduğu varsayılır (Rizopoulos, 2012b).

4.2 Birleşik Modelde Parametre Tahmin Yöntemleri

4.2.1 İki Aşamalı Yaklaşım

Boylamsal süreç ve sağ kalım sürecinin ilişkili olduğu durumlarda, bu iki süreci ayrı ayrı modellemek süreçler arası ilişki olduğu bilgisini göz ardı ettiğinden parametre tahminlerinde yanlı sonuçların elde edilmesine neden olabilir. Hem birleşik modellemede ML tahmin edicilerinin elde edilmesinde karşılaşılan hesaplama

zorluklarının hem de süreçlerin ayrı ayrı modellenmesinde karşılaşılan parametre tahminlerindeki yanlışlık probleminin üstesinden gelmek için Self ve Pawitan (1992) parametre tahminlerinde iki aşamalı yaklaşımı önermişlerdir. İki aşamalı yaklaşımın adımları şu şekildedir.

- 1.aşama: Ölçüm hatası içermeyen boylamsal değişkenlerin gerçek değerlerinin tahmini için veriye uygun model oluşturulur.
- 2.aşama: Birinci aşamada sağ kalım bilgisi içermeden tahmin edilen ölçüm hatası içermeyen boylamsal ölçüm değerleri sağ kalım modeline açıklayıcı değişken olarak eklenir. Sağ kalım modeli için parametre tahminleri kısmi olabilirlik yöntemi ile elde edilir.

İki aşamalı yaklaşım boylamsal ölçümlerin ölçüm hatası içermesinden kaynaklanan parametre tahminlerindeki yanlışlığı azaltır fakat tamamen ortadan kaldırmaz. Bu yaklaşımın standart yazılımlarla uygulanabilirliği kolay olsa da tamamen yansız tahmin ediciler elde edilmediği, Dafni ve Tsiatis (1998); Tsiatis ve Davidian (2001); Ye ve diğ. (2008); Sweeting ve Thompson (2011) tarafından simülasyon çalışmaları ile gösterilmiştir (Rizopoulos, 2012b; McCrink ve diğ., 2013). Birleşik modelde yansız parametre tahminlerinin elde edilebilmesi için birleşik olabilirlik fonksiyonu yardımıyla parametre tahminlerinin eş zamanlı olarak elde edilmesine olanak sağlayan ML yönteminden yararlanır.

4.3 Birleşik Modelde ML Yöntemi

Birleşik modellemede parametre tahminleri genellikle ML yöntemine dayanır. ML yönteminde, parametre tahminleri T_i, δ_i, y_i gözlenen çıktuların log olabilirlik fonksiyonunun model parametrelerine göre türevinin alınması ile elde edilir. Rastgele etki parametresi b_i iki süreç için de ortak parametredir. Birleşik olabilirlik fonksiyonu tanımlanırken b_i rastgele etki parametresinin boylamsal ve sağ kalım çıktıları arasındaki birlikteliği ve tekrarlı ölçümler arasındaki ilişkiyi açıkladığı varsayılır. $\theta = (\theta_t^\top, \theta_y^\top, \theta_b^\top)^\top$ tam parametre vektöründe yer alan $\theta_t = (\gamma^\top, \alpha)^\top$ sağ kalım modeli parametre vektörü, $\theta_y = (\beta^\top, \sigma_\epsilon^2)^\top$ boylamsal model parametre vektörü ve $\theta_b = D$ şeklinde olmak üzere T_i ve y_i 'nin koşullu bağımsızlığına dayalı olarak olabilirlik fonksiyonu,

$$\begin{aligned}
L(\theta) & \prod_{i=1}^n p(T_i, \delta_i, y_i | b_i, \theta) \\
& = \prod_{i=1}^n \int p(T_i, \delta_i | b_i, \theta_t) p(y_i | b_i, \theta_y) p(b_i; \theta_b) db_i
\end{aligned} \tag{4.6}$$

şeklinde tanımlanır (Rizopoulos, 2012b; Elashoff ve diğ., 2016).

Eşitlik 4.6'da verilen birleşik olabilirlik fonksiyonunda yer alan sağ kalım sürecinin koşullu yoğunluk fonksiyonu,

$$\begin{aligned}
p(T_i, \delta_i | b_i, \theta_t, \beta) & = \left(h_0(t_i) \exp \left\{ \gamma^\top w_i + \alpha m_i(t_i) \right\} \right)^{\delta_i} \\
& \times \exp \left(- \int_0^{t_i} h_0(s) \exp \left\{ \gamma^\top w_i + \alpha m_i(s) \right\} ds \right)
\end{aligned} \tag{4.7}$$

şeklindedir. Rastgele etkilerle birlikte boylamsal yanıtların yoğunluk fonksiyonu,

$$\begin{aligned}
p(y_i | b_i, \theta_y) p(b_i; \theta_t) & = \prod_{j=1}^{n_i} p(y_i(t_{ij}) | b_i, \theta_y) p(b_i; \theta_b) \\
& = (2\pi\sigma_\varepsilon^2)^{-\frac{n_i}{2}} \exp \left(- \|y_i - X_i\beta - Z_i b_i\|^2 / 2\sigma_\varepsilon^2 \right) \\
& \times (2\pi)^{-\frac{q_b}{2}} |D|^{-\frac{1}{2}} \exp \left(- \frac{1}{2} - b_i^\top D^{-1} b_i \right)
\end{aligned} \tag{4.8}$$

olarak yazılır. Eşitlik 4.8'te yer alan q_b rastgele etkiler vektörünün boyutunu göstermektedir (Rizopoulos, 2012b).

Parametre tahminlerinin elde edilebilmesi için 4.6'da verilen olabilirlik fonksiyonunun logaritması alınıp,

$$\ell(\theta) = \sum_{i=1}^n \left\{ \log p(T_i, \delta_i | b_i, \theta_t) + \log p(y_i | b_i, \theta_y) + \log p(b_i | \theta_t) \right\} \tag{4.9}$$

θ parametresine göre maksimizasyonu yapıldığında parametre tahminleri EM algoritması ya da Newton Raphson algoritması gibi iteratif yöntemler kullanılarak elde edilebilir (Wulfsohn ve Tsiatis, 1997).

4.3.1 Birleşik Modelde EM Algoritması

Birleşik modelde, birleşik olabilirlik fonksiyonunun çözümünde EM algoritmasının tercih edilmesinin nedeni, EM algoritmasının ikinci aşaması olan maksimizasyon

aşamasında bazı parametrelerin kapalı formlarının güncellenebilir olmasıdır. Kapalı formu bulunmayan parametrelerin tahmini için Newton Raphson yönteminden yararlanır (Wulfsohn ve Tsiatis, 1997). Birleşik modelde EM algoritmasının adımları şu şekilde gerçekleştirilir:

- E adımı: EM algoritmasının aşamaları gerçekleştirilirken b_i rastgele etkiler kayıp gözlem olarak varsayılır. Amaç gözlenen veriye ait log olabilirlik fonksiyonunu maksimize eden $\hat{\theta}$ değerlerini bulmaktır.

$$\begin{aligned} Q(\theta | \theta^{it}) &= \sum_{i=1}^n \int \log p(T_i, \delta_i, y_i, b_i; \theta) p(b_i | T_i, y_i, \delta_i; \theta^{it}) db_i \\ &= \sum_{i=1}^n \int \left\{ \log p(T_i, \delta_i | b_i, \theta_t) + \log p(y_i | b_i, \theta_y) + \log p(b_i | \theta_t) \right\} \\ &\quad \times p(b_i | T_i, \delta_i, y_i; \theta^{it}) db_i \end{aligned} \quad (4.10)$$

Eşitlik 4.10'da tanımlanan ifade, rastgele etkilere göre ve Eşitlik 4.7'de verilen sağ kalım fonksiyonunda yer alan ifadenin zamana göre integralini içerir. Bu integrallerin kapalı formda çözümleri bulunmadığından Gauss Hermite kareleme (quadrature) yöntemi ve Monte Carlo örnekleme gibi sayısal yöntemlerle integrallere yakınsama sağlanabilir (Rizopoulos ve Ghosh, 2011).

- M adımı: Parametrelerin güncellenmesi için gerekli olan kapalı formlar rastgele etkiler varyans kovaryans matrisi D ve ölçüm hatasının varyansı σ_ϵ^2 için mevcuttur. Boylamsal ve sağ kalım modelindeki regresyon parametreleri için Newton Raphson yönteminden yararlanarak M adımı gerçekleştirilebilir (Wulfsohn ve Tsiatis, 1997).

Ölçüm hatası varyansının tahmini

$$\begin{aligned} \hat{\sigma}^2 &= \frac{1}{\sum_{i=1}^n n_i} \sum_i \int (y_i - X_i\beta - Z_i b_i)^\top (y_i - X_i\beta - Z_i b_i) p(b_i | T_i, \delta_i, y_i; \theta) db_i \\ &= \frac{1}{\sum_{i=1}^n n_i} \sum_i (y_i - X_i\beta)^\top (y_i - X_i\beta - 2Z_i \tilde{b}_i) + \text{tr}(Z_i^\top Z_i \text{Var}(b_i | T_i, \delta_i, y_i; \theta^{it})) + \tilde{b}_i^\top Z_i^\top Z_i \tilde{b}_i \end{aligned} \quad (4.11)$$

şeklinde elde edilir. Eşitlikte yer alan $\tilde{b} = E(b_i | T_i, \delta_i, y_i; \theta^{it})$ koşullu beklenen değeri göstermektedir.

Rastgele etkilerin varyans kovaryans matrisi D 'nin tahmini

$$\hat{D} = \frac{1}{\sum_i n_i} \text{Var}(b_i | T_i, \delta_i, y_i; \theta^{(it)}) + \tilde{b} \tilde{b}^\top \quad (4.12)$$

eşitliği ile elde edilir.

Birleşik modelde yer alan sabit etkiler parametresi β ve $\theta_t = (\gamma, \alpha)$ sağ kalım modeli parametreleri için olabilirlik fonksiyonunun kapalı çözümleri bulunmamaktadır. Bu nedenle bu parametrelerin tahmin edicilerinin elde edilmesinde Newton Raphson yöntemi kullanılır (Wulfsohn ve Tsiatis, 1997; Rizopoulos ve diğ., 2009).

$$\beta^{(it+1)} = \beta^{(it)} - \left\{ \partial S(\hat{\beta}^{(it)}) / \partial \beta \right\} \quad (4.13)$$

$$\theta_t^{(it+1)} = \theta_t^{(it)} - \left\{ \partial S(\hat{\theta}_t^{(it)}) / \partial \theta_t \right\} \quad (4.14)$$

Burada $\beta^{(it)}$ ve $\theta_t^{(it)}$, β ve θ_t parametrelerinin güncel değerlerini ifade eder. β ve θ_t 'ye karşılık gelen skor vektörleri,

$$S(\beta) = \sum_i X_i^\top y_i - X_i \beta - Z_i b_i / \sigma^2 + \alpha \delta_i X_i(T_i) - \exp(\gamma^\top w_i) \left[\int_0^{T_i} h_0(s) \alpha x_i(s) \exp(\alpha x_i^\top(s) \beta + z_i^\top(s) b_i) \times p(b_i | T_i, \delta_i, y_i; \theta) ds db_i \right] \quad (4.15)$$

ve

$$S(\gamma) = \sum_i w_i \left[\delta_i - \exp(\gamma^\top w_i) \int_0^{T_i} h_0(s) \alpha x_i(s) \exp(\alpha x_i^\top(s) \beta + z_i^\top(s) b_i) \times p(b_i | T_i, \delta_i, y_i; \theta) ds db_i \right] \quad (4.16)$$

ve

$$S(\alpha) = \sum_i \delta_i (x_i^\top(T_i) \beta + Z_i^\top(T_i) \tilde{b}_i - \exp(\gamma^\top w_i) \int_0^{T_i} h_0(s) \alpha x_i(s) \exp(\alpha x_i^\top(s) \beta + z_i^\top(s) b_i) \times p(b_i | T_i, \delta_i, y_i; \theta) ds db_i) \quad (4.17)$$

biçimindedir (Rizopoulos, 2012b).

EM algoritmasında rastgele etkiler kayıp gözlem olarak kabul edilir. EM algoritmasında parametrelerin tahminlerinin elde edilebilmesi için rastgele etkilerin

de tahminine ihtiyaç vardır. Rastgele etkiler, rastgele değişkenler olduğu ve olasılık dağılımına sahip olduğu için Rizopoulos (2012b)'de yer aldığı gibi bu parametrelerin tahmin edicilerinin elde edilmesinde Bayesçi yaklaşım kullanılabilir. $p(b_i; \theta)$ rastgele etki parametrelerinin önsel dağılımı olarak kabul edilir. $p(T_i, \delta_i | b_i; \theta)$ ve $p(y_i | b_i; \theta)$ koşullu olabilirlik kısmından yararlanarak rastgele etkiler parametresinin sonsal dağılımı

$$\begin{aligned} p(b_i | T_i, \delta_i, y_i; \theta) &= \frac{p(T_i, \delta_i | b_i; \theta)p(y_i | b_i; \theta)p(b_i; \theta)}{p(T_i, \delta_i, y_i; \theta)} \\ &\propto p(T_i, \delta_i | b_i; \theta)p(y_i | b_i; \theta)p(b_i; \theta) \end{aligned} \quad (4.18)$$

şeklinde elde edilir. Rastgele etkilerin sonsal dağılımının kapalı formda çözümü bulunmamaktadır bu nedenle çözümünün sayısal yöntemlerle elde edilmesi gerekir. Boylamsal ölçümlerin sayısı arttıkça, $n \rightarrow \infty$, rastgele etkilerin sonsal dağılımı normal dağılıma yaklaşmaktadır.

$$p(b_i | T_i, \delta_i, y_i; \theta) \xrightarrow{P} N(\hat{b}_i, \bar{H}^{-1}) \quad (4.19)$$

Sonsal dağılımın modu ve ortalaması,

$$\begin{aligned} \bar{b}_i &= \int b_i p(b_i | T_i, \delta_i, y_i; \theta) \\ \hat{b}_i &= \arg \max_b \left\{ \log p(b_i | T_i, \delta_i, y_i; \theta) \right\} \end{aligned} \quad (4.20)$$

rastgele etkilerin sonsal dağılımının varyansı ya da Hessian matrisinin tersi,

$$\begin{aligned} \text{Var}(b_i) &= \int (b_i - \bar{b}_i) p(b_i | T_i, \delta_i, y_i; \theta) \\ H_i &= \left\{ \frac{\partial^2 \log p(b | T_i, \delta_i, y_i; \theta)}{\partial b^\top \partial b} \Big|_{b=\hat{b}_i} \right\}^{-1} \end{aligned} \quad (4.21)$$

eşitlikleri ile hesaplanabilir. Hessian matrisinin tersinin köşegen elemanları parametrelere ait varyans değerlerini vermektedir.

EM algoritmasında yakınsamanın kontrolü,

$$\max \{ |\theta^{(it)} - \theta^{(it-1)}| / (|\theta^{(it-1)}| + \epsilon_1) \} < \epsilon_2 \quad (4.22)$$

eşitliği ile yapılır. θ^{it} , it . aşamadaki parametre değerlerini ifade eder. Eğer tahmin edilen parametre değerleri ile hesaplanan eşitlik ε_2 gibi küçük bir değerden küçük ise yakınsama sağlanır. ε_1 ve ε_2 değerleri genellikle 10^{-3} ya da 10^{-4} gibi küçük değerleri ifade eder (Rizopoulos, 2012b).

4.3.2 Sayısal İntegrasyon Yöntemi

Birleşik modelde, parametre tahminlerinin elde edilebilmesi için Eşitlik 4.7’de verilen olabilirlik fonksiyonundaki integrallerin hesaplanması gerekir. Olabilirlik fonksiyonunda yer alan integrallerden biri sağ kalım modelinin yoğunluk fonksiyonunda yer alan integraldir. Bu integral tek boyutlu olduğu için 7 noktalı ya da 15 noktalı Gauss Kronrod kuralından (Vetterling ve Press, 1992) yararlanarak yakınsama sağlanabilir (Rizopoulos ve Ghosh, 2011). Eşitlik 4.7’de açık şekilde yazılan sağ kalım modeline ait yoğunluk fonksiyonu

$$p(T_i, \delta_i | b_i, \theta_i, \beta) = \left(h_0(T_i) \exp \left\{ \gamma^\top w_i + \alpha m_i(T_i) \right\} \right)^{\delta_i} \times \exp \left(- \int_0^{T_i} h_0(s) \exp \left\{ \gamma^\top w_i + \alpha m_i(s) \right\} ds \right) \quad (4.23)$$

şeklinindedir. Yoğunluk fonksiyonunda yer alan integrale 15 noktalı Gauss Kronrod kuralı kullanılarak,

$$\int_0^{T_i} h_0(s) \exp \gamma^\top w_i + \alpha m_i(s) ds \approx \frac{T_i}{2} \sum_{q=1}^{15} \pi_q H_i(w_i, \gamma, m_i) \left(\frac{T_i(1 + s_q)}{2} \right) \quad (4.24)$$

yakınsama sağlanabilir. π_q ve s_q sırasıyla ağırlıkları ve apsileri gösterir. T_i bireylerin son sağ kalım zamanlarını ifade etmektedir (Rizopoulos ve Ghosh, 2011).

Birleşik model tanımında yer alan diğer integralin hesaplanabilmesi için rastgele etkilere göre integralin alınması gerekir ve rastgele etkilerin boyutu arttıkça integral almak zorlaşır. Birleşik model literatüründe Wulfsohn ve Tsiatis (1997); Henderson ve diğ. (2000); Song ve diğ. (2002) tarafından kullanılan Gauss Hermit kareleme kuralı ya da Monte Carlo örnekleme genellikle tercih edilen yöntemlerdir. Rizopoulos ve diğ. (2009) ve Ye ve diğ. (2008) bu yöntemlere alternatif olarak birleşik modeldeki integrale Laplace yaklaşımı ile yakınsama sağlandığında, rastgele etkilerin boyutunun arttığı

durumlarda, Gauss Hermit kareleme yaklaşımı ve Monte Carlo örneklemesine göre daha etkili sonuçların elde edildiğini göstermişlerdir. Bu üç yaklaşıma alternatif olarak Rizopoulos (2012a) tarafından hesaplama zorluklarını azaltan bir yöntem önerilmiştir. Bu yaklaşımın gerçekleştirilebilmesi için öncelikle model parametreleri için skor fonksiyonunun aşağıdaki gibi tanımlanması gerekir.

$$\begin{aligned}
S(\theta) &= \sum_i \frac{\partial}{\partial \theta^\top} \log \int p(T_i, \delta_i | b_i; \theta) p(y_i | b_i; \theta) p(b_i; \theta) db_i \\
&= \sum_i \frac{1}{p(T_i, \delta_i, y_i; \theta)} \frac{\partial}{\partial \theta^\top} \int p(T_i, \delta_i | b_i; \theta) p(y_i | b_i; \theta) p(b_i; \theta) db_i \\
&= \sum_i \frac{1}{p(T_i, \delta_i, y_i; \theta)} \int \frac{\partial}{\partial \theta^\top} p(T_i, \delta_i | b_i; \theta) p(y_i | b_i; \theta) p(b_i; \theta) db_i \\
&= \sum_i \int \left[\frac{\partial}{\partial \theta^\top} \log p(T_i, \delta_i | b_i; \theta) p(y_i | b_i; \theta) p(b_i; \theta) \right] \\
&\quad \times \frac{p(T_i, \delta_i | b_i; \theta) p(y_i | b_i; \theta) p(b_i; \theta)}{p(T_i, \delta_i, y_i; \theta)} db_i \\
&= \sum_i \int A(b_i, \theta) p(b_i | T_i, \delta_i, y_i; \theta) db_i
\end{aligned} \tag{4.25}$$

$A(\cdot)$ tam verinin skor vektörü olmak üzere, $A(b_i; \theta) = \partial p(T_i, \delta_i | b_i; \theta) p(y_i | b_i; \theta) p(b_i; \theta) / \partial \theta^\top$ şeklinde gösterilir (Rizopoulos, 2012b). Rastgele etkilerin fonksiyonu $A(\cdot)$ 'nin herhangi bir formu için, skor vektöründe yer alan integrale önceden belirlenen apsis noktalarındaki integrand değerlerinin ağırlıklı toplamı ile yakınsama sağlandığı standart Gauss Hermit kuralı eşitlikteki gibi yazılabilir.

$$\begin{aligned}
E\{A(\theta, b_i) | T_i, \delta_i, y_i; \theta\} &= \int A(\theta, b_i) p(b_i | T_i, \delta_i, y_i; \theta) db_i \\
&\approx 2^{q_b/2} \sum_{t_1, \dots, t_q} \pi_t A(\theta, b_t \sqrt{2}) p(b_t \sqrt{2} | T_i, \delta_i, y_i; \theta) \exp(-\|b_t\|^2)
\end{aligned} \tag{4.26}$$

Eşitlik 4.26'de yer alan \sum_{t_1, \dots, t_q} ifadesi, $\sum_{t_1=1}^K, \dots, \sum_{t_q=1}^K$ şeklinde gösterilen K kareleme noktalarının kısa gösterimidir. $b_t^\top = (b_{t_1}, \dots, b_{t_q})$, π_q ağırlık noktalarına karşılık gelen apsisleri ifade eder. K kareleme noktalarının sayısı arttıkça yaklaşımın doğruluğu artmaktadır. Standart Gauss Hermit kuralında, integrand değerleri apsislerin kartezyen çarpımları üzerinden her bir rastgele etki için değerlendirildiğinden bu durum rastgele etkilerin boyutu q_b arttıkça hesaplama zorluklarına yol açmaktadır. Rastgele

etkilerin boyutundan kaynaklanan hesaplama zorluklarının üstesinden gelebilmek için Rizopoulos (2012a) tarafından önerilen yeni yaklaşım yapay uyarlanabilir (pseudo adaptive) Gauss Hermit kuralı $\tilde{r}_t = \tilde{b}_t + \sqrt{2}\tilde{B}_t^{-1}b_t$, $\tilde{b}_t = \arg \max_b \{\log p(y_t, b_t; \theta)\}$ ve \hat{B}_t , Hessian matrisinin Choleski faktörünü belirtmek üzere eşitlikteki gibi yazılabilir (Rizopoulos, 2012a).

$$\begin{aligned} E\{A(\theta, b_t) \mid T_t, \delta_t, y_t; \theta\} &= \int A(\theta, b_t) p(b_t \mid T_t, \delta_t, y_t; \theta) db_t \\ &\approx 2^{q_b/2} |\tilde{B}_t|^{-1} \sum_{t_1, \dots, t_q} \pi_t A(\theta, \tilde{r}_t) p(\tilde{r}_t \mid T_t, \delta_t, y_t; \theta) \exp(-\|b_t\|^2) \end{aligned} \quad (4.27)$$

Önerilen yeni yaklaşımda standart Gauss Hermite kuralına göre daha az nokta ile çalışılır ve daha az nokta ile çalışılmasına rağmen standart Gauss Hermite kuralına göre daha hızlı yakınsama sağlanır (Rizopoulos, 2012b). Konu ile ilgili detaylı bilgi için standart Gauss Hermite ve yapay uyarlanabilir Gauss Hermite yönteminin karşılaştırıldığı Rizopoulos (2012a) çalışması incelenebilir.

4.4 Birleşik Modelde Hipotez Testi

Birleşik modelde, $H_0 : \theta = \theta_0$ yokluk hipotezinin testi için olabilirlik oran testi, skor testi ve Wald testinden yararlanır.

- Olabilirlik Oran testi:

Olabilirlik oran testi için kullanılan test istatistiği,

$$LRT = -2\{\ell(\hat{\theta}_0) - \ell(\hat{\theta})\} \quad (4.28)$$

şeklinde gösterilir. $\hat{\theta}_0$ ve $\hat{\theta}$ yokluk ve alternatif hipotez altında ML tahmin edicileridir.

- Skor testi:

Skor testi için test istatistiği,

$$\begin{aligned} I(\hat{\theta}) &= \sum_{i=1}^n \frac{\partial S_i(\theta)}{\partial \theta} \Big|_{\theta=\hat{\theta}} \\ U &= S^T(\hat{\theta}_0) \{I(\hat{\theta}_0)\}^{-1} S(\hat{\theta}_0) \end{aligned} \quad (4.29)$$

$S(\cdot)$ skor fonksiyonunu ve $I(\cdot)$ bilgi matrisini göstermektedir.

- Wald testi:

Wald test istatistiği,

$$W = (\hat{\theta} - \theta_0)^T I(\hat{\theta})(\hat{\theta} - \theta_0) \quad (4.30)$$

şeklinde gösterilir.

Yokluk hipotezi altında, üç test istatistiğinin dağılımı asimptotik olarak p serbestlik dereceli χ^2 dağılımı gösterir. Skor testinde, model yokluk hipotezi altında ve Wald testinde model alternatif hipotez altında değerlendirilirken, olabilirlik oran testinde her iki hipotez altında da modelin değerlendirilmesi gerekir.

Olabilirlik oran, Skor ve Wald testleri sadece iki iç içe modelin(nested model) karşılaştırılmasında uygundur. Farklı modellerin karşılaştırılmak istendiğinde Akaike Bilgi Kriteri (Akaike Information Criteria, AIC) ve Bayesçi Bilgi Kriteri (Bayesian Information Criteria, BIC) gibi bilgi kriterlerinden yararlanılabilir.

$$\begin{aligned} AIC &= -2\ell(\hat{\theta}) + 2n_p \\ BIC &= -2\ell(\hat{\theta}) + n_p \log(n) \end{aligned} \quad (4.31)$$

Burada $\hat{\theta}$ ML tahmin edicisini, n_p modeldeki parametre sayısını ve n örneklem genişliğini göstermektedir. Aynı veri için iki model karşılaştırıldığında AIC ve BIC bilgi kriterlerinden daha küçük bilgi kriteri değerine sahip model, uygun model olarak kabul edilir (Rizopoulos, 2012b).

4.5 Birleşik Modelde Artık Analizi

LME modelinde, birey-özel artıklar ve marjinal artıklar olmak üzere iki tür artık incelenir. Birim özel artıklar Eşitlik 4.5'te verilen boylamsal alt model olan LME modelinin geçerliliğinin kontrolünde kullanılır.

$$\begin{aligned} y_i(t) &= m_i(t) + \varepsilon_i(t) \\ b_i &\sim N(0, D) \\ \varepsilon_i(t) &\sim N(0, \sigma^2) \end{aligned} \quad (4.32)$$

LME modeli için standartlaştırılmış artıklar $\hat{\beta}$ ve $\hat{\sigma}^2$ ML tahmin edicilerini ve \hat{b}_i Bayes tahmin edicisini göstermek üzere

$$r_i^{yss} = y_i(t) - x_i^T(t)\hat{\beta} - z_i^T(t)\hat{b}_i \quad (4.33)$$

ile hesaplanabilir. Marjinal artıklar için ise Eşitlik 4.34'te verilen marjinal model kullanılır.

$$\begin{aligned} y_i &= X_i\beta + \varepsilon_i^* \\ \varepsilon_i^* &\sim N(0, Z_i D Z_i^T + \sigma^2 I_{ni}) \end{aligned} \quad (4.34)$$

Marjinal modelde standartlaştırılmış artıklar $\hat{V}_i = Z_i D Z_i^T + \sigma^2 I_{ni}$ 'nin tahmin edilen varyans kovaryans matrisi olmak üzere

$$r_i^{ysm} = \hat{V}_i^{-1/2}(y_i - X_i\hat{\beta}) \quad (4.35)$$

şeklinde hesaplanır. Marjinal artıklar ve birey-özel artıklar birleşik modelde, boylamsal alt model varsayımlarının geçerliliğinin kontrolünde kullanılabilirler (Rizopoulos, 2012b).

Sağ kalım alt modelinde kullanılan Oransal Cox modelinin geçerliliğinin kontrolü için kullanılan artık türü Martingale artıklarıdır. Martingale artıkları eşitlikteki gibi hesaplanır.

$$\begin{aligned} r_i^{tm} &= N_i(t) - \int_0^t R_i(s)h(s | \hat{M}_i(s); \hat{\theta})ds \\ r_i^{tm} &= N_i(t) - \int_0^t R_i(s)\hat{h}_0(s) \exp\left\{\hat{\gamma}w_i + \hat{\alpha}\hat{m}_{i,s}\right\}ds \end{aligned} \quad (4.36)$$

$N_i(t)$, t zamanından önce olay yaşayan bireylerin sayısını, $R_i(t)$, t zamanında risk altındaki bireylerin sayısını gösterir. Martingale artıkları, i . birey için t zamanında gözlenen olay sayısı ile aynı zaman noktasında modelden elde edilen beklenen olay sayısı arasındaki fark alınarak hesaplanır.

Sağ kalım alt modeli için kullanılan diğer artık türü (Cox ve Snell, 1968) tarafından önerilen Cox-Snell artıklarıdır.

$$r_i^{tcs} = \int_0^{T_i} \hat{h}_0(s) \exp\{\hat{\gamma}w_i + \hat{\alpha}\hat{m}_{i,s}\}ds \quad (4.37)$$

Modelde, tahmin edilen parametre değerleri gerçek parametre değerlerine yakınsa r_i^{tcs} değerleri sansürlü örneklemeler gibi üstel dağılım gösterirler. r_i^{tcs} değerleri tahmin

edilen birikimli tehlike fonksiyonu $\hat{H}_i(t) = -\log \hat{S}_i(t)$ deęerlerine eřittir. r_i^{tcs} deęerleri üstel daęılıma sahip ise $\hat{H}_i(t)$ deęerlerinin daęılımı yaklaşık olarak üstel daęılıma eřit olur (Klein ve Moeschberger, 2006; Collett, 2015).



5 BİRLEŞİK MODELDE PARAMETRE TAHMİNLERİ İÇİN BAYESÇİ YAKLAŞIM

Birleşik modelde olabilirlik fonksiyonu karmaşık bir yapıya sahiptir. Olabilirlik fonksiyonunun maksimizasyonuna dayanan ML yöntemi ile parametre tahminleri hesaplanmak istendiğinde karşılaşılan en büyük zorluk, olabilirlik fonksiyonunun boylamsal ve sağ kalım süreçleri için ortak parametre olan rastgele etki parametrelerine göre integralinin alınmasıdır. Birleşik modelde, olabilirlik fonksiyonunda yer alan integralin kapalı yapıda çözümü yoktur ve bu integrale yakınsama sağlamak için genellikle Gauss Hermite yöntemi ya da Monte Carlo yaklaşımı gibi sayısal yöntemlerden yararlanır. Gözlem sayısı az olduğunda integrale Gauss Hermite çeyreklik kuralı ile yakınsama sağlamak kolay olsa da rastgele etkilerin boyutu arttıkça hesaplamalar daha karmaşık hale gelmektedir. Birleşik model gibi karmaşık yapıda olabilirlik fonksiyonuna sahip modellerde, Bayesçi yöntemlerle parametre tahminlerini elde etmenin en önemli avantajı olabilirlik fonksiyonundaki yüksek boyutlu integrallerin hesaplanmasında kolaylık sağlamasıdır. Bayesçi yöntemde Markov Zinciri Monte Carlo (MCMC) benzetim yöntemleri ile sonsal dağılım fonksiyonuna dayalı olarak parametrelere yakınsama sağlar. Son yıllardaki Bilgisayar yazılımlarındaki gelişmeler de Bayesçi yöntemlerin uygulanabilirliğini arttırmış ve karmaşık yapıdaki modellerin çözümünde tercih edilen bir yöntem olmasını sağlamıştır.

Birleşik modelde bayesçi yöntemin uygulandığı çalışmalarda sağ kalım alt modeli için genellikle Cox oransal tehlike modeli tercih edilirken boylamsal alt modelde kullanılan modeller farklılık göstermektedir. Faucett ve Thomas (1996), boylamsal alt model için rastgele etkiler modeli ve sağ kalım alt modeli için Cox oransal tehlike modelinde parametre tahminlerini Bayesçi yaklaşım ile elde etmişlerdir. ML tahminlerine yakın değerler elde edebilmek için bilgilendirici olmayan önseller kullanmışlar ve sabit ve rastgele etkiler parametreleri için parametre tahminlerini MCMC algoritması

yöntemlerinden Gibbs örneklemesi ile elde etmişlerdir. Wang ve Taylor (2001), Faucett ve Thomas (1996) gibi rastgele etkiler modelini kullanmışlardır. Boylamsal alt modele Ornstein-Uhlenbeck sürecini dahil ederek bireylerin kitle ortalaması etrafındaki rastgele değişimlerini esnek bir modelle açıklayamayı amaçlamışlardır. Önerilen model bireylerde, zaman boyunca meydana gelen hızlı ya da yavaş değişimleri açıklamada esneklik sağlasa da parametre sayısı arttıkça hesaplamalarda problemlere neden olmaktadır. (Lawless, 2011) R. Brown ve G. Ibrahim (2003) birleşik modelde, boylamsal alt model parametreleri için parametrik önsel dağılımlar yerine Dirichlet süreci önselini kullanılarak dağılım varsayımlarında esneklik sağlamayı amaçlamışlardır. Rizopoulos ve Ghosh (2011) çok değişkenli birleşik modelde bireylere ait boylamsal değişimleri açıklamak için boylamsal alt modelde spline tabanlı yaklaşım kullanarak bireylerin boylamsal ölçümlerinde meydana gelen değişimlerini esnek bir modelle açıklamayı amaçlamışlardır ve parametre tahminleri için Bayesçi yaklaşım kullanmışlardır. Huang ve diğ. (2011b) boylamsal yanıtların ve rastgele etkilerin çarpık normal dağılımdan geldiği durumlar için birleşik modelde parametre tahminlerinde bayesçi yöntemi kullanmışlardır.

Bu bölümde, Bayesçi yaklaşım ile birleşik modelde parametre tahminlerinin elde edilmesi sürecinden bahsedilecektir. Bu amaçla, öncelikle bayesçi yaklaşım ve Bayesçi yaklaşımda sonsal dağılımdan örneklem üretilmesini sağlayan Monte Carlo Markov Chain (MCMC) benzetim yöntemlerinden Gibbs örneklemesi ve MH algoritması anlatılacaktır. Birleşik modelde Bayesçi yaklaşımın uygulanabilmesi için seçilen önsel dağılımlar ve birleşik olabilirlik fonksiyonu yardımıyla model parametrelerinin tam koşullu dağılımları elde edilecektir. Birleşik model parametrelerine ait sonsal dağılımı yardımıyla parametreler için örneklem üretilmesinde kullanılacak olan algoritmanın aşamalarından bahsedilecektir. Önerilen Bayesçi modelde, parametreler için yakınsamanın kontrolünde kullanılacak olan Geweke tanısı, Gelman ve Rubin tanısı ve Heidelberger ve Welch tanıları anlatılacaktır.

5.1 Bayesçi Yaklaşım

Bayesçi yaklaşım, model parametrelerinin dağılımlarından yararlanarak olasılık modelinin tahmin edilme sürecidir. Bayesçi yaklaşımda, model parametrelerinin rastgele değişken olduğu varsayılır ve parametreler hakkındaki bilgi olasılık dağılımına

dayalı olarak ifade edilir. Parametreler için belirlenen dağılımlar arařtırmacının parametre hakkındaki kesin olmayan bilgisine dayanarak belirlenen dağılımlardır ve önsel dağılım olarak adlandırılırlar.

Bayesçi yaklaşımda, önsel dağılımdan ve veriden gelen bilgi birleřtirilerek parametreler için sonsal dağılımlar elde edilir. $y = \{y_1, y_2, \dots, y_n\}$ gözlenen veri bilindiğinde θ parametresinin kořullu sonsal dağılımının ifade edilebilmesi için öncelikle θ ve y 'nin birleřik olabilirlik fonksiyonunun belirlenmesi gerekir. Birleřik olabilirlik fonksiyonu, $p(\theta)$, θ parametresinin önsel dağılımını, $p(y | \theta)$ örneklem modelinin olabilirlik fonksiyonunu belirtmek üzere,

$$p(\theta, y) = p(y | \theta)p(\theta) \quad (5.1)$$

biçiminde yazılabilir. (Gelman ve diğ., 2013) Birleřik olabilirlik fonksiyonu $p(\theta, y)$ belirlendikten sonra y 'ler bilindiğinde θ 'nın kořullu sonsal dağılımı $p(\theta | y)$, Bayes yaklaşımının temelini oluřturan ve Thomas Bayes (1763) tarafından önerilen Bayes kuralından yararlanarak yazılabilir.

$$p(\theta | y) = \frac{p(\theta, y)}{p(y)} = \frac{p(\theta)p(y | \theta)}{p(y)} \quad (5.2)$$

Kesikli veriler için $p(y) = \sum_{\theta} p(\theta)p(y | \theta)$ ve sürekli veriler için $p(y) = \int_{\theta} p(\theta)p(y | \theta)d\theta$ şeklinde hesaplanır. Eřitlik 5.2'de paydada yer alan $p(y)$, y 'nin marjinal dağılımını ifade eder ve sonsal dağılımın normalleřtirme sabiti olarak adlandırılır. y 'nin marjinal dağılımı $p(y)$, önsel dağılım ile olabilirlik fonksiyonunun θ parametresine göre integralinin alınması ile elde edilir. İntegral sonlu olduđu sürece, paydadaki integralin sonucu sonsal dağılım hakkında ek bir bilgi sađlamaz. Böylelikle sonsal dağılım, önsel dağılım ve olabilirlik fonksiyonuna orantılı olarak,

$$p(\theta | y) \propto p(y | \theta)p(\theta) \quad (5.3)$$

Sonsal dağılım \propto Olabilirlik \times Önsel dağılım

şeklinde yazılabilir (Gelman ve diğ., 2013; Hoff, 2009). Eřitlik 5.2'de verilen ifade parametreler için elde edilebilecek çıkarımların çekirdek fonksiyonunu oluřturmaktadır. θ parametresi ile ilgili tüm çıkarımlar sonsal dağılım $p(\theta | y)$ 'den yararlanılarak yapılır.

Önsel dağılım, arařtırmacı tarafından belirlenen, ilgilenilen parametre hakkındaki bilgileri içeren olasılık dağılımıdır. Önsel dağılım ve verinin olasılık dağılımı birleřtirilerek elde edilen sonsal dağılım sayesinde ilgilenilen parametre hakkındaki bilgiler güncellenir. Örnekleme geniřliđinin küçük olduđu durumlarda uygun önsel seçimi önemlidir ancak örnekleme geniřliđi arttıkça önsel dağılımın sonsal dağılım üzerindeki etkisi azalır. Önsel dağılım bilgilendirici (informative) önsel ve bilgilendirici olmayan (noninformative) önsel olarak iki gruba incelenebilir. Bilgilendirici olmayan önseller aynı zamanda dađımık (vague), düz (flat), belirsiz (diffuse) önsel olarak da bilinir (Gelman ve diđ., 2006).

- *Bilgilendirici Önseller*

Önsel dağılım, arařtırmacının geçmiř deneyimlerinden ve önceki çalıřmalardan edindiđi bilgilere dayanarak belirlenebilir. Bu tür önseller genellikle subjektif önsel ya da bilgilendirici önsel dağılım olarak bilinir. Bilgilendirici önseller olabilirlik fonksiyonu tarafından etkisi azaltılmamıř sonsal dağılımın belirlenmesinde etkisi olan önsellerdir. Gerçek uygulamalarda veri yapısına uygun önsel dağılımın belirlenmesi önemlidir çünkü önceki çalıřmalardan ve geçmiř deneyimlerden elde edilen bilginin güncel bilgi ile birleřtirildiđi uygun bir önsel kullanımı Bayeřçi yaklařımın gücünü arttırmaktadır (Gelman ve diđ., 2013).

- *Bilgilendirici Olmayan Önseller*

Bilgilendirici olmayan önseller sonsal dağılım üzerinde etkisi az olan önsellerdir. Bilgilendirici olmayan önseller arařtırmacı tarafından veri ile ilgili herhangi bir önsel bilgi belirtilmeden objektif bir řekilde seçilen önsellerdir. Bilgilendirici olmayan önseller kullanıldıđı durumda sonsal dağılım genellikle olasılık yoğunluk fonksiyonuna benzer ve veri ile ilgili ek bir bilgi sađlanmadıđından genellikle Bayeřçi yaklařımdan elde edilen parametre tahminleri ML tahminleri ile benzerlik gösterir (Gelman ve diđ., 2013; Carlin ve Louis, 2010).

- *Eřlenik(Conjugate) Önseller*

θ parametresinin önsel dağılımı, $p(\theta)$ ile $p(\theta | y)$ sonsal dağılım aynı dağılım ailesinden ise θ parametresinin önsel dağılımına eřlenik önsel denir.

$$P(\theta) \in F \Rightarrow p(\theta | y) \in F \quad (5.4)$$

Örneğin, $p(y | \theta)$ örneklem dağılımı binom dağılımı ise ve θ parametresinin önsel dağılımı beta dağılımı olarak belirlenmiş ise θ parametresinin sonsal dağılımı $p(\theta | y)$ beta dağılımı olur. Eşlenik önseller uygulama açısından kolaylık sağlar (Gelman ve diğ., 2013).

- *Jeffrey Önsel Dağılım*

Jeffreys (1961) tarafından önerilen Jeffrey önseli, θ tek parametrelili olduğu durumda $p(\theta | y)$ olabilirlik fonksiyonuna ait Fisher bilgi matrisinin kareköküne eşittir.

$$p(\theta) \propto \sqrt{I(\theta)} \quad (5.5)$$

θ parametre vektörü iki ya da daha fazla parametre içerdiği durumda Jeffrey önseli eşitlikteki gibi ifade edilir.

$$p(\theta) \propto |I(\theta)|^{-1/2} \quad (5.6)$$

Jeffrey önseli herhangi bir parametrik model için bilgi içermeyen önselin elde edilmesini sağlar (Hoff, 2009).

5.2 Markov Zinciri Monte Carlo Yöntemi

MCMC yöntemi, θ parametresine ait $p(\theta | y)$ sonsal dağılım yardımıyla θ parametresi için $\theta^{(1)}, \dots, \theta^{(s)}$ şeklinde sıralı örneklemelerin elde edildiği benzetim yöntemidir ve karmaşık model yapılarının çözümüne pratik bir yol sağlar. Bu yöntemde $\theta^{(s)}$ örneklem değeri, bir önceki $\theta^{(s-1)}$ güncel örneklem değerine bağlı olarak üretilir ve bu durum Markov zinciri olarak bilinen zincir yapısını oluşturur. MCMC yönteminin en önemli özelliği algortima iyi bir şekilde yazıldığında ve yeterince uzun çalıştırıldığında başlangıç değerine bakılmaksızın ilgilenilen sonsal dağılıma yakınsama sağlanmasıdır. Uygulamalarda genellikle parametre tahminleri ML yöntemi ile elde edilebiliyorsa, ML tahminleri başlangıç değeri olarak seçilir. MCMC yönteminde, Monte Carlo yaklaşımında olduğu gibi θ parametresinin $E[g(\theta)]$

beklenen değerine, $\{g(\theta^{(1)}), \dots, g(\theta^{(S)})\}$ örneklem dizisinin ortalaması ile yakınsama sağlanır.

$$\frac{1}{S} \sum_{s=1}^S g(\theta^{(s)}) \rightarrow E[g(\theta)] = \int g(\theta) p(\theta) d\theta \quad (5.7)$$

Burada, $p(\theta)$, θ parametresinin önsel dağılımını ifade etmektedir. Yakınsama Markov zincirleri ile sağlandığından bu yaklaşıma Markov Zinciri Monte Carlo yaklaşımı denir (Hoff, 2009; Gelman ve diğ., 2013). Monte Carlo yöntemi ile MCMC yöntemi arasındaki temel fark, MCMC yönteminde her iterasyonda örneklem parametrelerin bir önceki güncel değerlerine koşullu olarak üretildiğinden örneklem birbiriyle bağımlı olmasıdır.

Bayesçi yöntemde, parametrelere ait sonsal dağılımdan parametrelerin elde edilmesi analitik olarak zor olduğundan sonsal dağılımlardan örneklem üretme süreci için genellikle MCMC benzetim yöntemlerinden Gibbs örnekleme ve Metropolis Hasting(MH) algoritmasından yararlanır.

5.2.1 Gibbs Örnekleme

Parametrelerin sonsal dağılımı bilinen bir dağılım yapısına benzerlik gösterdiğinde Gibbs örnekleme ile parametreler için örneklem sonsal dağılımdan elde edilebilir. Gibbs örnekleme, uygulama açısından kolay olması nedeniyle çok boyutlu problemlerin çözümünde en çok tercih edilen örnekleme yöntemlerinden biridir.

Gibbs örneklemesinin her bir aşamasında ilgili parametre için örneklem üretilirken parametrelerin tam (full) koşullu dağılımlarından yararlanır. İlgili parametre hariç diğer tüm parametrelere ve gözlenen veriye koşullu olarak yazıldığından tam koşullu dağılım olarak bahsedilir.

$\theta = (\theta_1, \theta_2, \dots, \theta_p)$ parametre vektörü şeklinde p alt parametreden oluşsun. $p(\theta) = p(\theta_1, \theta_2, \dots, \theta_p)$, θ parametresinin önsel dağılımını ve $p(\theta | y)$ birleşik sonsal dağılımını ifade etsin. θ parametre vektörünün alt parametreleri için Gibbs örneklemesinin aşamaları şu şekilde gerçekleştirilir;

1. $s, s = 1, \dots, S$ iterasyon sayısını göstermek üzere $s = 0$ iterasyonu için herhangi bir başlangıç değeri seçilir, $\theta^{(0)} = \theta_1^{(0)}, \theta_2^{(0)} \dots, \theta_p^{(0)}$.

2. $s+1$. iterasyon için θ parametre vektörünün her bir alt parametre bileşeni üretilir,

$$\begin{aligned} \theta_1 \text{ parametresi için } (s+1). \text{ örneklem, } & \theta_1^{(s+1)}, p(\theta_1^{(s+1)} | \theta_2^{(s-1)}, \theta_3^{(s-1)}, \dots, \theta_p^{(s-1)}, y) \\ \theta_2 \text{ parametresi için } (s+1). \text{ örneklem, } & \theta_2^{(s+1)}, p(\theta_2^{(s+1)} | \theta_1^{(s-1)}, \theta_3^{(s-1)}, \dots, \theta_p^{(s-1)}, y) \\ & \vdots \\ \theta_p \text{ parametresi için } (s+1). \text{ örneklem, } & \theta_p^{(s+1)}, p(\theta_p^{(s+1)} | \theta_1^{(s-1)}, \theta_2^{(s-1)}, \theta_3^{(s-1)}, \dots, \theta_{p-1}^{(s-1)}, y) \end{aligned} \quad (5.8)$$

3. $s < S$ ise 2.adıma tekrar dönülür. Parametreler için önceden belirlenen S maksimum iterasyon sayısına ulaşıldığında işlem durdurulur.

Gibbs örnekleme bağımlı vektörler dizisi üretir.

$$\begin{aligned} \theta^{(1)} &= \{\theta_1^{(1)}, \theta_2^{(1)}, \dots, \theta_p^{(1)}\} \\ \theta^{(2)} &= \{\theta_1^{(2)}, \theta_2^{(2)}, \dots, \theta_p^{(2)}\} \\ &\vdots \\ \theta^{(S)} &= \{\theta_1^{(S)}, \theta_2^{(S)}, \dots, \theta_p^{(S)}\} \end{aligned}$$

Gibbs örneklemesinde $\theta^{(s)}$ parametre vektöründe yer alan örneklemler, $\theta^{(s-1)}$ güncel değerlerine bağlıdır. Her bir iterasyonda bir önceki örnekleme değerine bağlı olarak üretildiğinden Markov zinciri oluşturur.

5.2.2 Metropolis Hasting (MH) Algoritması

θ parametresinin sonsal dağılımı $p(\theta | y)$ için her zaman bilinen bir dağılım yapısı elde edilemeyebilir. Karmaşık ve çok parametrelili modellerde sonsal dağılım için bilinen bir dağılım yapısı elde edilemediğinde sonsal dağılımdan örneklemler üretmek için Metropolis ve diğ. (1953) tarafından önerilen MH algoritması kullanılır. MH algoritmasında amaç $p(\theta^* | \theta^{(s)})$ simetrik öneri (proposal) dağılımı yardımıyla θ parametresinin güncel değeri $\theta^{(s)}$ 'e yakın θ^* öneri örnekleme üretmektir. Öneri dağılımı olarak genellikle θ^s bilindiğinde θ^* elde etmenin olasılığı ile θ^* bilindiğinde θ^s elde etmenin olasılığı eşit olan $p(\theta^* | \theta^{(s)}) = p(\theta^s | \theta^{(*)})$ simetrik bir dağılım seçilir. MH algoritmasında en sık kullanılan simetrik dağılım normal dağılımdır.

MH algoritmasının aşamaları şu şekilde gerçekleştirilir:

1. θ^* için öneri dağılımından yeni bir değer üretilir, $\theta^* \sim p(\theta | \theta^{(s)})$
2. r , kabul oranı hesaplanır.

$$r = \min \left\{ \frac{p(\theta^* | y)}{p(\theta^{(s)} | y)}, 1 \right\}$$

3. $U \sim \text{uniform}(0, 1)$ düzgün dağılımdan U için bir rastgele sayı üretilir.
4. 2. adımda hesaplanan r kabul oranı ile U örnekleme karşılaştırılır. $r < U$ ise $\theta^{(s+1)} = \theta^*$ olarak kabul edilir, diğer durumda ise $\theta^{(s+1)} = \theta^{(s)}$ olarak kabul edilir.

İlgili parametre için öneri dağılımı simetrik bir dağılım olarak belirlenemediği durumda algoritmanın ikinci aşamasında kullanılan kabul oranı yerine,

$$r = \min \left\{ \frac{p(\theta^* | y)/p(\theta^* | \theta^{(s)})}{p(\theta^{(s)} | y)/p(\theta^{(s)} | \theta^*)}, 1 \right\}$$

eşitliği kullanılır (Gelman ve diğ., 2013; Hoff, 2009).

5.3 Birleşik Modelde Olabilirlik Fonksiyonu ve Önsel Dağılımlar

Bölüm 4'te bahsedilen birleşik modelde boylamsal alt model

$$\begin{aligned} y_i(t) &= m_i(t) + \varepsilon_i \\ m_i(t) &= X_i(t)\beta + Z_i b_i \end{aligned} \tag{5.9}$$

ve sağ kalım alt modeli,

$$h_i(t | M_i(t), w_i) = h_0(t) \exp(\gamma^\top w_i + \alpha m_i(t)) \tag{5.10}$$

için birleşik olabilirlik fonksiyonu, $\theta = (\theta_t, \theta_y, \theta_b)^\top$ tam parametre vektörünü ve $\theta_t = (\gamma^\top, \alpha)^\top$, $\theta_y = (\beta^\top, \sigma_\varepsilon^2)^\top$ ve $\theta_b = (D)$ boylamsal ve sağ kalım alt modeline ait parametreleri içeren alt parametre vektörleri olmak üzere birleşik olabilirlik fonksiyonu,

$$\begin{aligned}
p(T_i, \delta_i, y_i | b_i, \theta) &= \int p(T_i; \delta_i | b_i, \theta_t, \beta) p(y_i | b_i; \theta_y) p(b_i; \theta_t) db_i \\
&= \left(h_0(t) \exp(\gamma^\top w_i + \alpha m_i(t)) \right)^{\delta_i} \times \left(- \int_0^{T_i} h_0(s) \exp\{\gamma^\top w_i + \alpha m_i(s)\} ds \right) \\
&\times (2\pi\sigma_\varepsilon^2)^{-n_i/2} \exp\left(-\|y_i - X_i\beta - Z_i b_i\|^2 / 2\sigma_\varepsilon^2\right)
\end{aligned} \tag{5.11}$$

şeklindedir. Bayesçi yaklaşımda, birleşik olabilirlik fonksiyonu belirlendikten sonra birleşik sonsal dağılımın belirlenebilmesi için parametreler için önsel seçimi yapılması gerekmektedir. Bu çalışmada, parametreler için belirlenen önsel dağılımlar bilgilendirici olmayan önsellerdir. Önsel dağılımlar, literatürde bayesçi yaklaşımda regresyon parametreleri ve varyans parametreleri için yapılan önsel seçimleri referans alınarak belirlenmiştir. Birleşik modelde, boylamsal ve sağ kalım alt modelinde yer alan regresyon parametreleri, β, γ ve α için önseller normal dağılım olarak belirlenmiştir. Boylamsal alt modelde ölçüm hatasının varyansı σ_ε^2 için düzgün dağılım ve rastgele etkilere ait varyans kovaryans matrisi D için ters Wishtart (IW , inverse Wishart) dağılımı önsel dağılım olarak belirlenmiştir (Hoff, 2009; Gelman ve diğ., 2013; Rizopoulos ve Ghosh, 2011).

Boylamsal alt modele ait parametrelerin önsel dağılımları,

$$\begin{aligned}
p(\beta) &\sim MVN(\mu_\beta, \Sigma_\beta) \\
p(\sigma_\varepsilon^2) &\sim Uniform(a, b)
\end{aligned} \tag{5.12}$$

Sağ kalım alt modeline ait parametrelerin önsel dağılımları,

$$\begin{aligned}
p(\gamma) &\sim MVN(\mu_\gamma, \Sigma_\gamma) \\
p(\alpha) &\sim N(\mu_\alpha, \sigma_\alpha^2)
\end{aligned} \tag{5.13}$$

Varyans kovaryans matrisi D için seçilen önsel dağılım eşlenik önsel dağılımdır.

$$p(D) \sim IW(v, S)$$

IW dağılımında, $v > q_b + 1$ serbestlik derecesini ve S , $(q_b \times q_b)$ boyutlu simetrik pozitif tanımlı ölçek matrisini ifade etmektedir.

5.4 Birleşik Modelde Sonsal Dağılımlar

Bayesçi yaklaşımın ikinci aşaması olan önsel dağılımlar belirlendikten sonra parametrelerin tam koşullu dağılımlarının elde edilmesinde kullanılan sonsal dağılım fonksiyonu Eşitlik 5.2' de verilen Bayes kuralından yararlanılarak,

$$p(\beta, \gamma, \alpha, b_i, \sigma_\varepsilon^2, D | T_i, \delta_i, y_i) \propto p(T_i, \delta_i, y_i | b_i, \theta) \times p(\theta) \quad (5.14)$$

şeklinde yazılabilir. Olabilirlik fonksiyonu ve parametreler için belirlenen önsel dağılımlar açık olarak yazıldığında sonsal dağılım aşağıdaki gibi elde edilir:

$$\begin{aligned} p(\beta, \gamma, \alpha, b_i, \sigma_\varepsilon^2, D | T_i, \delta_i, y_i) &\propto \prod_{i=1}^n \left(h_0(t) \exp(\gamma^\top w_i + \alpha m_i(t)) \right)^{\delta_i} \\ &\times \left(- \int_0^{T_i} h_0(s) \exp\{\gamma^\top w_i + \alpha m_i(s)\} ds \right) \\ &\times \prod_{i=1}^n \prod_{j=1}^{n_i} (2\pi\sigma_\varepsilon^2)^{-1/2} \exp \left\{ - \left(y_i(t_{ij}) - X_i^\top(t_{ij})\beta - Z_i^\top(t_{ij})b_i \right)^2 / 2\sigma_\varepsilon^2 \right\} \\ &\times (2\pi)^{-q_b/2} |D|^{-1/2} \exp(b_i^\top D^{-1} b_i) \\ &\times |\Sigma_\beta|^{-1/2} \exp \left\{ - \frac{1}{2} (\beta - \mu_\beta)^\top |\Sigma_\beta|^{-1} (\beta - \mu_\beta) \right\} \\ &\times |\Sigma_\gamma|^{-1/2} \exp \left\{ - \frac{1}{2} (\gamma - \mu_\gamma)^\top |\Sigma_\gamma|^{-1} (\gamma - \mu_\gamma) \right\} \\ &\times (\sigma_\alpha^2)^{-1/2} \exp \left\{ - \frac{(\alpha - \mu_\alpha)^2}{2\sigma_\alpha^2} \right\} \\ &\times |D|^{-(v+q_b+1)/2} \exp \left\{ - \frac{1}{2} \text{tr}(SD^{-1}) \right\} \\ &\times p(\sigma_\varepsilon^2) \end{aligned} \quad (5.15)$$

Birleşik modelde yer alan sabit etkiler parametresi β , rastgele etkiler parametresi b_i , temel değişkenlere ait regresyon parametresi γ , ilişki parametresi α ve varyans parametreleri σ_ε^2 ve D için koşullu sonsal dağılımları Eşitlik 5.15 yardımıyla bulunabilir.

5.4.1 Boylamsal Alt Model Parametreleri İçin Tam Koşullu Dağılımlar

Boylamsal alt modelde β sabit etkiler parametresinin tam koşullu dağılımı;

$$\begin{aligned}
p(\beta \mid T_i, \delta_i, y_i, b_i, \gamma, \alpha, \sigma_\varepsilon^2, D) &\propto \prod_{i=1}^n \left(h_0(t) \exp(\gamma^\top w_i + \alpha m_i(t)) \right)^{\delta_i} \\
&\times \left(- \int_0^{T_i} h_0(s) \exp\{\gamma^\top w_i + \alpha m_i(s)\} ds \right) \\
&\prod_{j=1}^{n_i} (2\pi\sigma_\varepsilon^2)^{-1/2} \exp \left\{ \left(y_i(t_{ij}) - X_i^\top(t_{ij})\beta - Z_i^\top(t_{ij})b_i \right)^2 / 2\sigma_\varepsilon^2 \right\} \\
&\times |\Sigma_\beta|^{-1/2} \exp \left\{ - \frac{1}{2} (\beta - \mu_\beta)^\top |\Sigma_\beta|^{-1} (\beta - \mu_\beta) \right\}
\end{aligned} \tag{5.16}$$

şeklinde elde edilir. μ_β ve Σ_β , β parametresinin önsel dağılım parametreleridir, sırasıyla ortalama vektörü ve varyans kovaryans matrisini ifade eder.

Varyans parametresi için genellikle ters gamma (IG, inverse gamma) dağılımı önsel dağılım olarak belirlenir. Önsel dağılım, ters gamma dağılımı olarak belirlendiğinde eşlenik koşullu sonsal dağılım da ters gamma dağılımı olacağından sonsal dağılım Gibbs örneklemesi ile kolayca güncellenebilir. Ters gamma dağılımı ile çalışmasının zorluğu, ters gamma dağılımında, $IG(\varepsilon, \varepsilon)$ ε parametresinin seçimi 1,0.01,0.001 gibi küçük değerler seçildiğinde yani $\varepsilon \rightarrow 0$ iken koşullu eşlenik dağılım bilgilendirici özelliğini kaybederek uygun olmayan sonsal dağılıma neden olmasıdır. Bu nedenle Gelman ve diğ. (2006) varyans parametresi önsel dağılımı bilgilendirici olmayan düzgün önsel dağılımla çalışılmasını önermişlerdir. Düzgün önsel dağılımda, belirlenen $[a, b]$ aralığında ($U(0, 100)$ gibi geniş bir aralık) parametrenin sonsal dağılımı için aynı olasılık değerleri belirlenir ve $b \rightarrow \infty$ giderken önsel dağılımın sonsal dağılım üzerindeki etkisi azalır. Hiyerarşik modellerde grup sayısı küçük olduğunda, düzgün önsel dağılımının seçilmesi varyans parametresinin büyük tahmin edilmesi problemine yol açacağından düzgün dağılım ile çalışılırken grup sayısının $J > 3$ olmasına dikkat edilmedilir. Bu çalışmada, Gelman ve diğ. (2006) çalışması referans alınarak σ_ε^2 varyans parametresinin önsel dağılımı için düzgün dağılım belirlenmiştir.

Boylamsal alt modelde σ_ε^2 varyans parametresinin tam koşullu dağılımı;

$$\begin{aligned}
p(\sigma_\varepsilon^2 \mid T_i, \delta_i, y_i, b_i, \beta, \gamma, \alpha, D) &\propto \prod_{i=1}^n \prod_{j=1}^{n_i} (2\pi\sigma_\varepsilon^2)^{-1/2} \exp \left\{ \left(y_i(t_{ij}) - X_i^\top(t_{ij})\beta - Z_i^\top(t_{ij})b_i \right)^2 / 2\sigma_\varepsilon^2 \right\} \\
&\times p(\sigma_\varepsilon^2)
\end{aligned} \tag{5.17}$$

b_i rastgele etkiler parametresinin tam koşullu dağılımı;

$$\begin{aligned}
p(b_i | T_i, \delta_i, y_i, \beta, \gamma, \alpha, \sigma_\varepsilon^2, D) &\propto \left(h_0(t) \exp(\gamma^\top w_i + \alpha m_i(t)) \right)^{\delta_i} \\
&\times \left(- \int_0^{T_i} h_0(s) \exp\{\gamma^\top w_i + \alpha m_i(s)\} ds \right) \\
&\prod_{j=1}^{n_i} (2\pi\sigma_\varepsilon^2)^{-1/2} \exp \left\{ \left(y_i(t_{ij}) - X_i^\top(t_{ij})\beta - Z_i^\top(t_{ij})b_i \right)^2 / 2\sigma_\varepsilon^2 \right\} \\
&\times (2\pi)^{-q_b/2} |D|^{-1/2} \exp(b_i^\top D^{-1} b_i)
\end{aligned} \tag{5.18}$$

şeklindedir.

Rastgele etki parametresine ait varyans kovaryans matrisi D 'nin tam koşullu dağılımı;

$$\begin{aligned}
p(D | T_i, \delta_i, y_i, b_i, \beta, \gamma, \alpha, \sigma_\varepsilon^2) &\propto \prod_{i=1}^n |D|^{-1/2} \exp(b_i^\top D^{-1} b_i) \\
&\times |D|^{-(v+q_b+1)/2} \exp \left\{ -\frac{1}{2} \text{tr}(SD^{-1}) \right\}
\end{aligned} \tag{5.19}$$

biçiminde gösterilir. v ve S parametreleri sırasıyla serbestlik derecesini ve ölçek matrisini ifade etmek üzere IW dağılımının parametreleridir. q_b , rastgele etki parametresinin boyutunu göstermektedir. Eşitlik 5.19'in sağ tarafında yer alan $b_i^\top D^{-1} b_i$ ifadesi

$$(b_i^\top D^{-1} b_i) = \text{tr}(b_i^\top D^{-1} b_i) = \text{tr}(D^{-1} (\sum_{i=1}^n b_i b_i^\top)) \tag{5.20}$$

olarak yazılabilir. (Nydic, 2012) Eşitlik 5.19 tekrar düzenlendiğinde D varyans kovaryans matrisinin tam koşullu dağılımı,

$$\begin{aligned}
p(D | T_i, \delta_i, y_i, b_i, \beta, \gamma, \alpha, \sigma_\varepsilon^2) &\propto |D|^{-\frac{n}{2}} |D|^{-\frac{v+q_b+1}{2}} \exp \left\{ -\frac{1}{2} \text{tr}(\sum_{i=1}^n b_i b_i^\top D^{-1}) \right\} \times \exp \left\{ -\frac{1}{2} \text{tr}(SD^{-1}) \right\} \\
&\propto |D|^{-\frac{v+n+q_b+1}{2}} \exp \left(-\frac{1}{2} \text{tr} \left[\left(\sum_{i=1}^n b_i b_i^\top + S \right) D^{-1} \right] \right)
\end{aligned} \tag{5.21}$$

şeklindedir. Tam koşullu dağılım düzenlendiğinde D varyans kovaryans matrisi ters Wishart dağılır. D varyans kovaryans matrisi için koşullu sonsal dağılım ters Wishart dağılımı ile güncellenebilir.

$$D \sim IW \left(\mathbf{v} + n, \sum_{i=1}^n b_i b_i^T + S \right) \quad (5.22)$$

5.4.2 Sağ Kalım Alt Modeli Parametreleri İçin Tam Koşullu Dağılımlar

Sağ kalım alt modelinde temel değişkenlere karşılık gelen regresyon parametresi γ 'nın tam koşullu dağılımı aşağıdaki gibi elde edilir.

$$\begin{aligned} p(\gamma | T_i, \delta_i, y_i, b_i, \beta, \alpha, \sigma_\varepsilon^2, D) &\propto \prod_{i=1}^n \left(h_0(t) \exp(\gamma^T w_i + \alpha m_i(t)) \right)^{\delta_i} \\ &\times \left(- \int_0^{T_i} h_0(s) \exp\{\gamma^T w_i + \alpha m_i(s)\} ds \right) \\ &\times |\Sigma_\gamma|^{-1/2} \exp \left(- \frac{1}{2} (\gamma - \mu_\gamma)^T |\Sigma_\gamma|^{-1} (\gamma - \mu_\gamma) \right) \end{aligned} \quad (5.23)$$

Sağ kalım süreci ile boylamsal süreç arasındaki ilişkiyi açıklayan, sağ kalım alt modeline ait ilişki parametresi α 'nın tam koşullu dağılımı,

$$\begin{aligned} p(\alpha | T_i, \delta_i, y_i, b_i, \beta, \alpha, \sigma_\varepsilon^2, D) &\propto \prod_{i=1}^n \left(h_0(t) \exp(\gamma^T w_i + \alpha m_i(t)) \right)^{\delta_i} \\ &\times \left(- \int_0^{T_i} h_0(s) \exp\{\gamma^T w_i + \alpha m_i(s)\} ds \right) \\ &\times (\sigma_\alpha^2)^{-1/2} \exp \left(- \frac{(\alpha - \mu_\alpha)^2}{2\sigma_\alpha^2} \right) \end{aligned} \quad (5.24)$$

şeklinde elde edilir.

Birleşik modelde yer alan tüm parametreler için tam koşullu sonsal dağılımlar elde edildikten sonra model parametreleri için örneklem üretme sürecine geçilebilir. Bunun için MCMC yöntemlerinden Gibbs örnekleme ve MH algoritmasının birlikte kullanıldığı algoritmadan yararlanılacaktır.

5.5 Markov Zincirinde Yakınsamanın Değerlendirilmesi

MCMC yönteminde temel düşünce ilgilenilen sonsal dağılımla aynı durağan dağılıma sahip bir Markov süreci oluşturmaktır. MCMC yönteminde ilgili sonsal

dağılıma yakınsamanın kontrolü iz grafikleri (trace plot), otokorelasyon grafikleri ile belirlenebileceği gibi Geweke testi (Geweke ve diğ., 1991), Heidelberger ve Welch testi (Heidelberger ve Welch, 1983) gibi testlerle de yakınsamanın kontrolü sağlanabilir. Bu testler Markov zincirinin tek zincire sahip olduğu durumlarda kullanılan testlerdir. İki ya da daha fazla zincir kullanıldığı durumlarda, parametreler için yakınsamanın kontrolünde Geweke ve Rubin (Gelman ve diğ., 1992) tanısında yararlanır.

- *İz Grafikleri*

İz grafikleri, her parametre için zincirin durağan bir dağılıma yakınsayıp yakınsamadığı hakkında bilgi verir. Modeldeki her parametre için çizdirilirler ve yakınsamanın kontrolünde en çok kullanılan yöntemlerden biridir. Zincir devam ederken, dağılım noktası belirli bir değer üzerindeyse zincirin durağanlığa ulaştığı söylenebilir. İz grafikleriyle ilk kaç gözlemin atılacağına da karar verilebilir. Eğer grafik üzerinde büyük dalgalanmalar var ise zincir devam ederken örneklem hedef dağılımdan gittikçe uzaklaşıyorsa bu yakınsamanın gerçekleşmediğini gösterir.

- *Otokorelasyon*

Otokorelasyon grafikleri ile yakınsamanın kontrolü direkt olarak yapılamasa da bu grafikler yakınsamanın değerlendirilmesine yardımcı olur. MCMC algoritması yüksek korelasyonlu örneklem üretiyorsa bu durum Markov Zincirinin yavaş yakınsadığını ve iterasyon sayısının artırılması gerektiğini ifade eder. Örneklem kovaryansı,

$$\hat{\gamma}(h) = \frac{1}{n-h} \sum_{t=1}^{n-h} (\theta_i^{t+h} - \bar{\theta}_i)(\theta_i^t - \bar{\theta}_i), 0 \leq h < b \quad (5.25)$$

şeklinde olmak üzere h gecikmenin örneklem otokorelasyonu,

$$\hat{\rho}(h) = \frac{\hat{\gamma}(h)}{\hat{\sigma}^2}, |h| < 0 \quad (5.26)$$

şeklinde hesaplanır (Institute Inc, 2008).

- *Geweke Tanısı*

Geweke testinde, zincir ilk ve son iterasyon olmak üzere iki parçaya ayrılarak sonsal ortalamalar üzerinden yakınsamanın kontrolü yapılır. θ^t Markov zinciri, $\theta_1^t : t = 1, \dots, n_1$ ve $\theta_2^t : t = n_a, \dots, n$ iki alt kümeye ayrılır. $1 < n_1 < n_a < n$ ve $n_2 = n - n_a + 1$ olmak üzere ilk ve son iterasyonun sonsal ortalamaları;

$$\bar{\theta}_1 = \frac{1}{n_1} \sum_{t=1}^{n_1} \theta^t \quad \bar{\theta}_2 = \frac{1}{n_2} \sum_{t=1-n_a}^n \theta^t \quad (5.27)$$

şeklinde tanımlanır. Geweke testi, z istatistiğine dayanır ve aşağıdaki gibi hesaplanır.

$$Z_n = \frac{\bar{\theta}_1 - \bar{\theta}_2}{\sqrt{\frac{\hat{s}_1(0)}{n_1} + \frac{\hat{s}_2(0)}{n_2}}} \quad (5.28)$$

Eşitlikte asimptotik varyanslar spektral yoğunluk tahminleri üzerinden tanımlanmaktadır. Zincirde yakınsama sağlandığında, iterasyon sayısı arttıkça $n \rightarrow \infty$, Geweke testi 0 ortalama ve 1 varyans ile standart normal dağılıma yakınsar (Sinharay, 2003).

- *Heidelberger ve Welch Tanısı*

Heidelberger ve Welch durağanlık testi (Heidelberger ve Welch, 1983) Markov zincirinin kovaryans durağan süreçten geldiği hipotezini test ederek Markov zincirinin durağanlığını değerlendirir. Eğer yokluk hipotezi reddedilirse, zincirdeki ilk %10'luk kısmı atılır ve kalan %90 üzerinden tekrar test edilir. Bu süreç durağanlık testi geçilene kadar ya da en az %50 iterasyon atılana kadar devam edilir (Sinharay, 2003).

- *Gelman ve Rubin Tanısı*

MCMC algoritmasında iki ya da daha fazla zincir kullanıldığında, Gelman ve diğ. (1992) yakınsama tanısından yararlanarak yakınsamanın kontrolü yapılabilir. Bu yakınsama tanıları, varyans analizinde olduğu gibi, her bir model parametresi için zincir içi varyansı ve zincirler arası varyansı karşılaştırılarak değerlendirilir. Bu iki varyans arasındaki büyük fark yakınsama sağlanmadığını ifade eder (Carlin ve Louis, 2010).

Gelman ve Rubin tanısında n , $\{t = 1, \dots, n\}$ iterasyon uzunluğunu ve M , $\{m = 1, \dots, M\}$ zincir sayısını ve $\hat{\theta}_m$ m . zincirin sonsal ortalamasını ve $\hat{\sigma}_m^2$ m . zincirin varyansını belirtmek üzere,

$$\begin{aligned}\bar{\theta}_m &= \frac{1}{n} \sum_{t=1}^n \theta_m^t & \bar{\theta} &= \frac{1}{M} \sum_{m=1}^M \bar{\theta}_m \\ \hat{\sigma}_m^2 &= \sum_{t=1}^n (\theta_m^t - \bar{\theta}_m)^2\end{aligned}\tag{5.29}$$

B ve W zincirler arası varyans ve zincir içi varyans ve \hat{V} varyans tahmini,

$$\begin{aligned}B &= \frac{n}{M-1} \sum_{m=1}^M (\bar{\theta}_m - \bar{\theta})^2 \\ W &= \frac{1}{M} \sum_{m=1}^M \hat{\sigma}_m^2 \\ \hat{V} &= \frac{n-1}{n} + \frac{m+1}{mn} \frac{B}{W}\end{aligned}\tag{5.30}$$

şeklinde hesaplanır. Paralel çalıştırılan M zincir hedef dağılıma ulaştıysa sonsal varyans grup içi varyansı W 'ye yakın bir değer olarak tahmin edilir. Bu durumda \hat{V}/W oranının 1'e yakın bir değer olması beklenir (Carlin ve Louis, 2010). \hat{V}/W oranının karekökü potansiyel ölçek küçültme faktörü (PSRF, potential scale reduction factor) olarak adlandırılır ve eşitlikteki gibi hesaplanır.

$$\hat{R}_c = \sqrt{\frac{df + 3 \hat{V}}{df + 1 W}}\tag{5.31}$$

df serbestlik derecesi $df \approx 2\hat{V}/var(\hat{V})$ şeklinde hesaplanır. Büyük PSFR değerleri zincirler arası varyansın, zincir içi varyanstan büyük olduğunu ve daha uzun simülasyonlara ihtiyaç olduğunu belirtir. PSFR oranı 1'e yakın bir değer olması her bir zincirin durağan olduğunu ve hedef sonsal dağılıma yakınsamanın sağlandığını ifade eder (Gelman ve diğ., 2013).

6 UYGULAMA

Bu bölümde birleşik modellemede Bayesçi yaklaşım ile parametre tahminlerinin elde edilmesi amacıyla. Memorial Sloan Kettering Kanser Merkezinden alınan 341 tiroid kanseri hastasına ait veri seti üzerinde bir uygulama yapılmıştır. Veri seti, 1998 – 2015 yılları boyunca takip edilen hastaların boylamsal ölçümleri ile birlikte sağ kalım verilerini içermektedir. Hastaların %58.7’i kadınlardan ve %41.3’i erkeklerden oluşmaktadır. Tiroid kanseriyle en çok ilişkili olduğu düşünülen tiroglobulin değerleri bireylerden ziyaret zamanlarında tekrarlı olarak ölçülmüştür. Hastalardan farklı ölçüm zamanlarında alınan tekrarlı tiroglobulin değerleri ölçümlerinin sayısı ortalama 12.56’dır, hastalara ait en az 5 ölçüm en fazla 83 tekrarlı ölçüm kaydedilmiştir. Hastalardan farklı zamanlarda ölçümler alındığından alınan ölçüm sayıları hastadan hastaya farklılık göstermektedir ve ziyaret süreleri eşit değildir. 341 hastadan farklı ziyaret zamanlarında toplam 5484 tekrarlı tiroglobulin ölçümü alınmıştır. Bu çalışma için ilgilenilen sağ kalım olayı bireylerde tümörün tekrarlama olarak belirlenmiştir. Bireylerde, tümörün tekrarlama kadar geçen süre bireylerin sağ kalım süreleri olarak alınmıştır ve sağdan sansürlü veri ile çalışılmıştır. İlk ölçümün alındığı tarihten önce tümörün tekrarladığı hastalar analize dahil edilmemiştir. Bireyler çalışmada en uzun 17 yıl kalmıştır ortalama sağ kalım süresi 7 yıldır (medyan=5.98). Tümörün tekrarlamadığı ya da ölüm gibi nedenlerle çalışmadan ayrılan hastalar sansürlü olarak tanımlanmıştır. Çalışmada yer alan hastaların izlenme süresi sona erdiğinde, 276 (%80.9) hastanın ilgilenilen sağ kalım olayını yaşamadığı ve 65 (%19.1) hastada ise tümörün tekrarladığı görülmüştür. Tümörün tekrarının görüldüğü 65 hastadan %56.9’u kadın ve %43.1’i erkek hastalardan oluşmaktadır. Uygulamada zamandan bağımsız temel açıklayıcı değişken olarak cinsiyet değişkeni ve temel yaş değişkeni alınmıştır. Temel yaş değişkeni hastanın çalışmaya dahil olduğu andaki yaşını ifade etmektedir.

Bayesçi yöntemle parametre tahminlerine geçmeden önce R programında yer alan ve ML yöntemine dayanan JM paketiyle (Rizopoulos, 2010) analiz yapılacaktır. Birleşik model analizi ile

- Tiroglobulin değerlerindeki değişim ile tümörün tekrarlaması arasındaki ilişki var mıdır?
- Aynı yaşa ve aynı cinsiyete sahip iki bireye ait tiroglobulin değerlerindeki farklılık bireylerde tümör tekrarının görülme riski nasıl etkilemektedir?
- Aynı cinsiyete sahip bireylerdeki başlangıç yaşındaki farklılık, bireylerde tümör tekrarı yaşama riski üzerinde etkili midir?

soruları cevaplanmaya çalışılacaktır.

6.1 Klasik Yaklaşım ile Parametre Tahminleri

Birleşik model ile tiroid veri setinin analizi için boylamsal alt model,

$$\log(\text{Tg}_{ij}) = \log(\text{Tg}_{ij}^*) + \varepsilon_{ij} = \beta_0 + \beta_1 \text{Zaman}_{ij} + \beta_2 \text{Temel Yaş}_i + b_{0i} + b_{1i} \text{Zaman}_{ij} + \varepsilon_{ij} \quad (6.1)$$

Sağ kalım alt modeli,

$$h_i(t) = h_0(t) \exp(\gamma_1 \text{Cinsiyet} + \gamma_2 \text{Temel Yaş} + \alpha \{\log(\text{Tg}_{ij}^*)\}) \quad (6.2)$$

şeklinindedir. Modelde yer alan $\log(\text{Tg}_{ij}^*)$ değişkeni tiroglobulin değerleri ölçüm hatasından arındırılmış ve gözlenemeyen gerçek değerlerini göstermektedir.

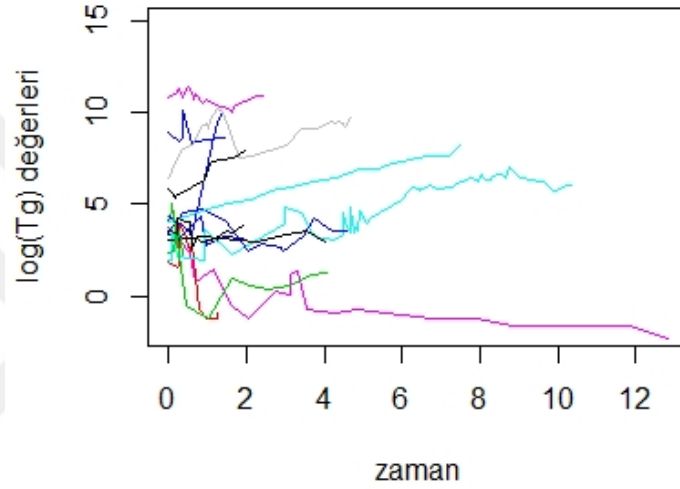
Modelde yer alan değişkenlerin açıklamaları Çizelge 6.1’de verilmiştir.

R(v3.5.0) programı JM paketiyle Rizopoulos (2010) ile yapılan analiz sonucunda elde edilen çıktılar Çizelge 6.2’de verilmiştir. JM paketinde parametre tahminleri, ML yöntemine dayanmaktadır.

ML yöntemi ile elde edilen sonuçlara göre boylamsal alt modelde, zaman değişkenine karşılık gelen parametre değeri pozitif bir değer olarak tahmin edilmiştir. Zaman

Çizelge 6.1: Tiroid veri setinde kullanılan değişkenler ve açıklamaları

Değişkenler	Tanımı
Tg	Bireylerden alınan tekrarlı tiroglobulin değerleri
Cinsiyet	Cinsiyet=1 ise kadın Cinsiyet=2 ise erkek
Temel Yaş	Bireylerin ilk ölçümün alındığı zamandaki yaşı
Zaman _{ij}	Bireylerden alınan tiroglobulin değerlerinin ölçüm zamanı

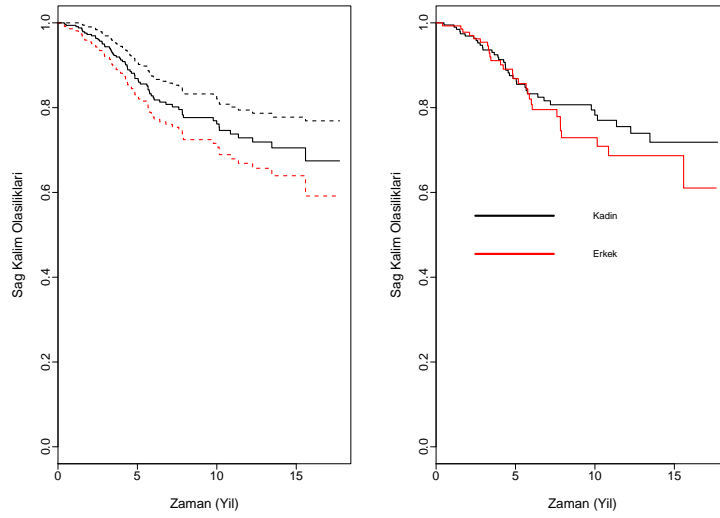


Şekil 6.1: Tiroglobulin ölçümlerinin zamana boyunca değişimi Rastgele seçilen 15 hastaya ait tiroglobulin ölçümlerinin zaman boyunca değişimini göstermektedir.

Çizelge 6.2: Birleşik Model Analiz Sonuçları

Model	Değişken	Parametre	Tahmin	Standart Hata	% 95 GA	p değeri
Boylamsal alt model	Sabit terim	β_0	1.1270	0.5358	(0.07687, 2.1772)	0.0354
	Zaman _{ij}	β_1	0.1436	0.0363	(0.0726, 0.2147)	<0.001
	Temel Yaş	β_2	0.0444	0.0092	(0.0264, 0.0624)	<0.001
Sağ kalım alt modeli	Cinsiyet	γ_1	-0.1976	0.2641	(-0.7188, 0.3201)	0.4544
	Temel Yaş	γ_2	0.0186	0.0095	(0.00003, 0.03710)	0.0496
	log(Tg*)	α	0.2379	0.0336	(0.1722, 0.3037)	<0.001

değişkeninde meydana gelen bir yıllık artış tiroglobulin değerlerinde 0.14 birimlik bir artışa karşılık gelmektedir. Aynı şekilde temel yaş değişkeni tiroglobulin değerlerindeki değişim üzerinde etkili bir değişken olarak tahmin edilmiştir. Başlangıç yaşındaki bir birimlik artış tiroglobulin değerlerini 0.0444 birim arttırdığı söylenebilir.



Şekil 6.2: Tüm hastalara ve cinsiyete göre Kaplan-Meire Eğrisi. Sol tarafta yer alan grafik tüm hastalar için tahmin edilen sağ kalım olasılıklarının Kaplan-Meire eğrisini göstermektedir. Sağ tarafta yer alan grafik Cinsiyete göre tahmin edilen sağ kalım olasılıklarının Kaplan-Meire eğrisini göstermektedir.

Birleşik model ile Tiroglubin değerlerinde meydana gelen değişim ile bireylerde tümörün tekrarlaması arasındaki ilişkiyi gösteren, $\log(Tg^*)$ değişkenine ait α parametresi 0.23 olarak tahmin edilmiştir. Bu durum, tiroglobulin değerlerindeki bir birimlik artışın bireyde tümörün tekrarlaması riskini 1.26 ($\exp(0.2378)$) kat arttırdığı şeklinde yorumlanabilir. Tiroglobulin değerlerindeki değişim ile bireylerde tümör tekrarının görülmesi arasında güçlü bir ilişki vardır. Başlangıç yaşı ile tümör tekrarının görülmesi arasındaki ilişki anlamlıdır. Bu durum başlangıç yaşındaki bir birimlik artışın bireylerde tümör tekrarının görülme riskini 1.02 ($\exp(0.0185)$) kat arttırdığı şeklinde yorumlanabilir. Bu çalışmada yer alan bireyler için cinsiyet değişkeni ile bireylerde tümörün tekrarlama riski arasında ilişki anlamlı değildir ($p=0.5793$).

6.2 Bayesçi Yaklaşım İle Parametre Tahminleri

6.2.1 Parametrelerin Önsel Seçimleri

Bayesçi yaklaşımın gerçekleştirilebilmesi için öncelikle birleşik modelde yer alan tüm parametreler için belirlenen önsel dağılımların ve önsel dağılım hiperparametre değerlerinin belirlenmesi gerekir. Önsel dağılımların belirlenmesinde Bölüm 5'te bahsedildiği gibi bilgilendirici önsel ya da bilgilendirici olmayan önseller seçilebilir.

Bu çalışma için önsel seçiminde bilgilendirici olmayan önseller kullanılmıştır. Regresyon parametreleri için normal dağılım ve ölçüm hatasının varyansı ve rastgele etki parametrelerinin varyans kovaryans matrisi için ise sırasıyla düzgün dağılım ve ters Wishart dağılımı olarak belirlenmiştir. Boylamsal alt modelde sabit etki parametresi β , sağ kalım alt modelinde temel değişkenlere karşılık gelen γ parametresi ve α ilişki parametresinin önsel dağılımları 0 ortalama 100 varyansa sahip normal dağılım olarak belirlenmiştir. Ölçüm hatasının varyansı σ_ϵ^2 için önsel dağılım $U(0, 100)$ düzgün dağılım olarak belirlenmiştir. D varyans kovaryans matrisinin önsel dağılımı ölçek parametresi 2×2 boyutlu olarak birim matris olmak üzere ters Wishart dağılımı $IW(2, \text{diag}(1))$ belirlenmiştir.

6.2.2 Gibbs Örnekleme ve MH Algoritması

Birleşik modelde, boylamsal alt modelde sabit etkiler parametresi β ve sağ kalım alt modelinde temel değişkenlere karşılık gelen regresyon parametresi γ ve ilişki parametresi α ve b_i rastgele etkiler parametresi ve σ_ϵ^2 ölçüm hatasının varyansı için tam koşullu sonsal dağılımlar bilinen bir dağılım yapısına dönüştürülemediğinden bu parametrelerin sonsal dağılımların güncellenebilmesi için MH algoritmasından yararlanılmıştır. Varyans kovaryans matrisi D için ise bilinen bir dağılım yapısına dönüştürülebildiğinden sonsal dağılım Gibbs örnekleme ile güncellenmiştir. Bu nedenle oluşturulan algoritma Gibbs örnekleme ve MH algoritması birleştirilerek oluşturulmuştur. Parametrelerin sonsal dağılımlarının güncellenebilmesi için oluşturulan algoritma şu şekilde gerçekleşmektedir;

1. $s=0$ için $\beta^{(0)}, \gamma^{(0)}, \alpha^{(0)}, \sigma_\epsilon^{2(0)}, b_i^{(0)}, D^{(0)}$ başlangıç değerleri olarak belirlenir.
2. $s=1, \dots, S$ için aşağıdaki iterasyonlar tekrarlanır.

- (a) Boylamsal alt modelde sabit etkiler parametresi β için

$$\beta^{(s)} \sim p(\beta \mid \gamma^{(s-1)}, \alpha^{(s-1)}, \sigma_\epsilon^{2(s-1)}, b_i^{(s-1)}, D^{(s-1)}, T_i, \delta_i, y_i)$$

MH algoritması ile sonsal dağılımdan örneklemler üretilir.

- (b) Rastgele etkiler parametresi b_i için

$$b_i^{(s)} \sim p(b_i \mid \beta^{(s)}, \gamma^{(s-1)}, \alpha^{(s-1)}, \sigma_\epsilon^{2(s-1)}, D^{(s-1)}, T_i, \delta_i, y_i)$$

sonsal dağılımından MH algoritması ile örneklemeler üretilir.

(c) σ_ε^2 ölçüm hatasının varyansı için

$$\sigma_\varepsilon^{2(s)} \sim p(\sigma_\varepsilon | \beta^{(s)} \gamma^{(s-1)}, \alpha^{(s-1)}, D^{(s-1)}, T_i, \delta_i, y_i)$$

sonsal dağılımından MH algoritması ile örneklemeler elde edilir.

(d) D varyans kovaryans matrisi için örneklemeler Gibbs örnekleme ile elde edilir.

$$D^{(s)} \sim p(D | \beta^{(s)}, \gamma^{(s-1)}, \alpha^{(s-1)}, \sigma_\varepsilon^{2(s)}, b_i^{(s)}, T_i, \delta_i, y_i)$$

(e) Sağ kalım alt modeli temel değişkenlere regresyon parametresi γ için

$$\gamma^{(s)} \sim p(\gamma | \beta^{(s)}, \alpha^{(s-1)}, \sigma_\varepsilon^{2(s)}, b_i^{(s)} D^{(s)}, T_i, \delta_i, y_i)$$

MH algoritması ile sonsal dağılımdan örneklemeler elde edilir.

(f) Sağ kalım süreci ile boylamsal süreç arasındaki ilişkiyi açıklayan parametre α için

$$\alpha^{(s)} \sim p(\alpha | \beta^{(s)}, \gamma^{(s)}, \sigma_\varepsilon^{2(s)}, b_i^{(s)} D^{(s)}, T_i, \delta_i, y_i)$$

MH algoritması ile sonsal dağılımdan örneklemeler elde edilir.

3. Belirlenen iterasyon sayısına ulaşıldığında işlem durdurulur.

Bayesçi yaklaşım için WinBugs, Stata, SAS gibi yazılımlar mevcuttur. Bu çalışma için R yazılımı tercih edilmiştir R (v.3.5.0) programında yazılan algoritma iki zincir ve her zincir 10000 iterasyon olarak çalıştırılmıştır. MCMC algoritmasının ilk aşaması olan başlangıç değerleri birinci zincirde regresyon parametreleri için 0 ve varyans parametreleri için 1 olarak belirlendi. İkinci zincirde ise başlangıç değerleri parametrelerin ML tahminleri olarak belirlenmiştir. Zincirlerde ilk 2000 iterasyona yakma (burnin) işlemi uygulandı. Bu işlem parametrelere ait sonsal çıktılarla ilgili özetleyici bilgiler (ortalama, standart sapma vb.) elde edilirken başlangıç değerlerinin etkisini azaltmak amacıyla yapılmaktadır. Özellikle MH algoritmasında örneklemeler arası otokorelasyonu azaltmak için kullanılan inceltme (thinning) yöntemi uygulandıktan sonra elde edilen parametre değerleri ve otokorelasyon grafiklerinde büyük bir farklılık olmadığı için örneklemelere inceltme uygulanmadan, yakma işlemi uygulandıktan sonraki örneklemelerin tamamı parametre tahminleri ve yakınsama

tanıları için kullanılmıştır. Yakma işlemi sonucunda yakınsama, her bir zincirde 8000 örneklem olmak üzere toplam 16000 sonsal örneklem ile elde edilmiştir. Bölüm 6.2’te bahsedildiği gibi parametrelerin sonsal dağılımları belirlenirken sağ kalım modeli parametreleri γ , α , boylamsal alt model parametresi β ve varyans parametresi σ_ϵ^2 için bilinen bir dağılım yapısı elde edilemediğinden bu parametrelere ait sonsal dağılımlar güncellenirken MH algoritmasından yararlanılmıştır. MH algoritmasının gerçekleştirilebilmesi için parametreler için öneri dağılımlarının belirlenmesi gerekir. Öneri dağılımları genellikle simetrik bir dağılım olarak belirlenir ve genellikle normal dağılım seçilir. Bu çalışmada β , γ , α , σ_ϵ^2 parametreleri için öneri dağılımı literatürde de MH algoritmasında öneri dağılımı için sıklıkla kullanılan normal dağılım belirlenmiştir. Algoritmada normal dağılımın varyansı boylamsal ve sağ kalım alt modelinin ayrı ayrı modellenmesinden elde edilen varyanslar olarak belirlendi ve her iterasyonda güncel örneklem değerleri ortalama değeri olarak güncellenmiştir.

6.2.3 Parametre Tahminleri

Birleşik modelde boylamsal ve sağ kalım alt modellerine ait parametrelerin MCMC benzetim yöntemi ile sonsal dağılımlar güncellenerek elde edilen parametrelerin sonsal ortalama değerleri, standart sapma ve HPD güven aralıkları Çizelge 6.3’da verilmiştir.

Çizelge 6.3: Bayesçi Yaklaşım Parametre Tahminleri

Model	Değişken	Parametre	Sonsal Ortalama	Standart Sapma	HPD Aralığı %95
Boylamsal alt model	Sabit terim	β_0	1.1356	0.067	(0.9985, 1.2678)
	Zaman _{ij}	β_1	0.1435	0.0052	(0.1327, 0.1530)
	Temel Yaş	β_2	0.0444	0.0092	(0.0421, 0.0464)
	D11	σ_{11}^2	6.8992	0.5373	(5.8790, 7.9654)
	D12	σ_{12}	0.1831	0.0845	(0.0132, 0.3454)
	D22	σ_{22}^2	0.2921	0.0263	(0.2421, 0.3438)
			σ_ϵ	0.9351	0.0093
Sağ kalım alt modeli	Cinsiyet	γ_1	-0.2362	0.2877	(-0.7870, 0.3233)
	Temel Yaş	γ_2	0.1856	0.0092	(0.0002, 0.0368)
	$\log(Tg^*)$	α	0.2137	0.0305	(0.1585, 0.2754)

Bayesçi yaklaşım ile elde edilen sonuçlara göre boylamsal alt modelde Zaman ve Temel yaş değişkeni, sağ kalım alt modelinde temel yaş değişkeni ve iki süreç arasındaki ilişkiyi gösteren $\log(Tg^*)$ değişkeni anlamlı değişken olarak bulunmuştur. Boylamsal alt modelde, zaman değişkeni pozitif bir değer olarak tahmin edilmiştir

ve bu durum zaman ile tiroglobulin deęerleri arasında pozitif bir iliřki olduęunu gsterir. Zaman deęiřkeninde meydana gelen bir yıllık artıř tiroglobulin deęerlerinde 0.14 birimlik artıřa karřılık gelmektedir řeklinde yorumlanabilir. Aynı řekilde temel yař deęiřkeni tiroglobulin deęerlerindeki deęiřim zerinde etkili bir deęiřken olarak tahmin edilmiřtir. Temel yař deęiřkeninde meydana gelen bir birimlik artıř bireylerin tiroglobulin deęerlerinde 0.0435 birimlik artıřa karřılık gelmektedir.

Birleřik model ile Tiroglobulin deęerlerinde meydan gelen deęiřim ile bireylerde tmrn tekrarlaması arasındaki iliřkiyi gsteren, $\log(\text{Tg}^*)_{ij}$ deęiřkenine ait α parametresi 0.2137 olarak tahmin edilmiřtir. Tiroglobulin deęerlerindeki bir birimlik artıřın bireyde tmrn tekrarlaması riskini 1.24 ($\exp(0.2137)$) kat arttırdıęı řeklinde yorumlanabilir. Bayeřci yaklařım ile elde edilen sonulara gre bařlangı yařı ile bireyde tmrn tekrarlaması arasındaki iliřki anlamlıdır. Bu durum bařlangı yařındaki bir birimlik artıřın bireylerde tmr tekrarının grlme riskini 1.02 kat arttıracıęı řeklinde yorumlanabilir. Bu alıřmada yer alan bireyler iin cinsiyet deęiřkeni ile bireylerde tmrn tekrarlama riski arasındaki iliřki HPD aralıęı sıfır deęerini ierdięinden anlamlı deęildir.

Bayeřci yntemde parametreler iin bilgilendirici olmayan nseller kullanılmıřtır. nsel seimlerinde bilgilendirici olmayan nseller kullanıldıęı iin Bayeřci yntem ile elde edilen parametre tahminleri ve ML tahminlerinde yakın sonular elde edilmiřtir. Gven aralıkları karřılařtırıldıęında HPD gven aralıklarının β_0 , β_1 ve β_2 parametrelerinin asimptotik gven aralıklarına gre daha dar bir aralıęına sahip olduęu grlmektedir. Bu parametreler iin bilgilendirici olmayan nseller kullanılmasına raęmen nsel bilgi parametreler iin sonsal daęılımlar gncellenerek elde edilen rneklemelerde u deęerlerin yer almasının nne geerek parametrelerin gven aralıklarını sınırlandırmıřtır. Boylamsal ve saę kalım alt modelinde yer alan tm deęiřkenler iin ML ynteminde anlamlı bulunan deęiřkenler Bayeřci yntemde de anlamlı deęiřkenler olarak bulunmuřtur.

6.2.4 Yakınsama Tanıları

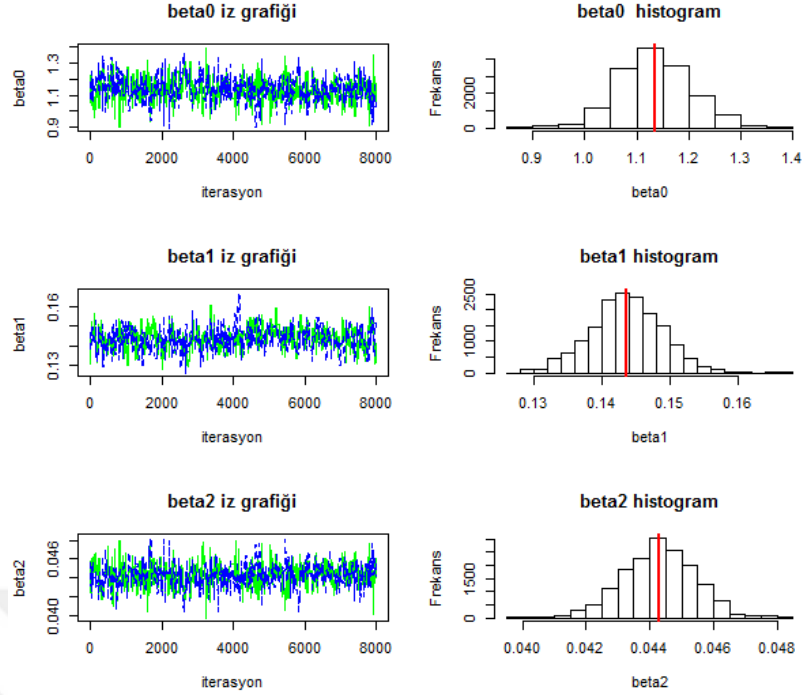
MCMC zincirleri Blm 6.2.2' de verilen algoritma yardımıyla oluřturulmuřtur. Tm parametreler iin yakınsamanın kontrolnn yapılabilmesi amacıyla iz grafiklerinden, kmlatif grafiklerinden, otokorelasyon grafiklerinden ve Gelman ve Rubin tanısından

(Gelman ve diğ., 1992) ve Geweke (Geweke ve diğ., 1991) tanısından yararlanılmıştır. Tüm grafikler MCMC benzetiminde 10000 iterasyon uygulandıktan ve başlangıç parametresinin etkisini en aza indirmek amacıyla ilk 2000 örnekleme yakma işlemi uygulandıktan sonra elde edilen 8000 sonsal örneklem üzerinden çizdirilmiştir. MCMC benzetimi ile elde edilen sonsal çıktılarının grafiklerle gösterilmesi ve özetlenmesi aynı zamanda yakınsama tanılarındaki kullanılan testlerden de yararlanmak amacıyla R yazılımında bulunan CODA (Plummer ve diğ., 2006) paketinden yararlanılmıştır. Gelman ve Rubin tanısında ölçek azaltma faktörünün 1'e yakın değerler elde edildiğinde parametreler için yakınsamanın gerçekleştiği söylenebilir. Geweke tanısında ise her parametre için sonsal örneklemelerden hesaplanan Z skorlarının $|Z| < 1.96$ değerinden küçük olması yakınsamanın sağlandığının göstergesidir. Gelman ve Rubin tanısı ve Geweke tanısı için elde edilen yakınsama sonuçları Çizelge 6.4'te verilmiştir. Çizelge 6.4'ten de görüldüğü gibi Gelman ve Rubin tanısı ve Geweke tanısı için yakınsama koşulları sağlandığından modeldeki parametreler için yakınsamanın gerçekleştiği söylenebilir.

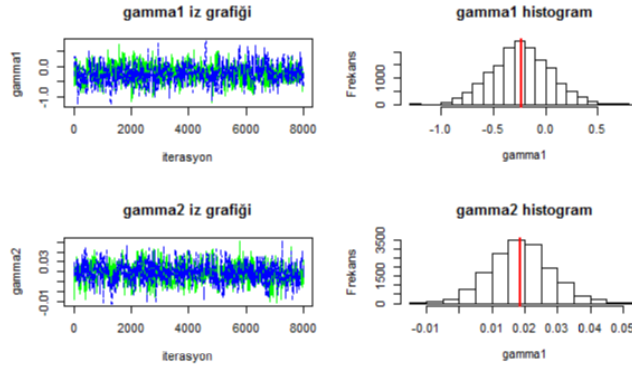
Çizelge 6.4: Yakınsama Tanıları

Değişken	Parametre	Gelman ve Rubin		Geweke
		Nokta Tahmini	Üst Sınır	
Sabit terim	β_0	1.01	1.03	0.077
Zaman	β_1	1.00	1.01	0.1185
Temel Yaş	β_2	1.00	1.00	1.4824
D11	σ_{11}^2	1.00	1.00	1.05
D12	σ_{12}^2	1.00	1.00	1.05
D22	σ_{22}^2	1.00	1.00	0.3093
Sigma	σ_ϵ	1.03	1.03	0.0796
Cinsiyet	γ_1	1.02	1.04	0.992
Temel Yaş	γ_2	1.00	1.01	0.4627
Tg*	α	1.01	1.02	0.7024

Şekil 6.3, Şekil 6.4, Şekil 6.5 ve Şekil 6.6 iz grafikleri ve sonsal örneklemelerin histogramları verilmiştir. Şekil 6.3, Şekil 6.4, Şekil 6.5 ve Şekil 6.6'de yer alan örneklemelerin histogram grafiklerinden tüm parametreler için sonsal örneklemelerin örneklem ortalaması etrafında dağıldığı görülmektedir. β , γ , α , σ_ϵ ve D varyans kovaryans parametreleri için çizdirilen iz grafiklerinden MCMC zincirlerinin durağan olduğu ve aynı zamanda iyi bir karışıma sahip olduklarını görülmektedir.

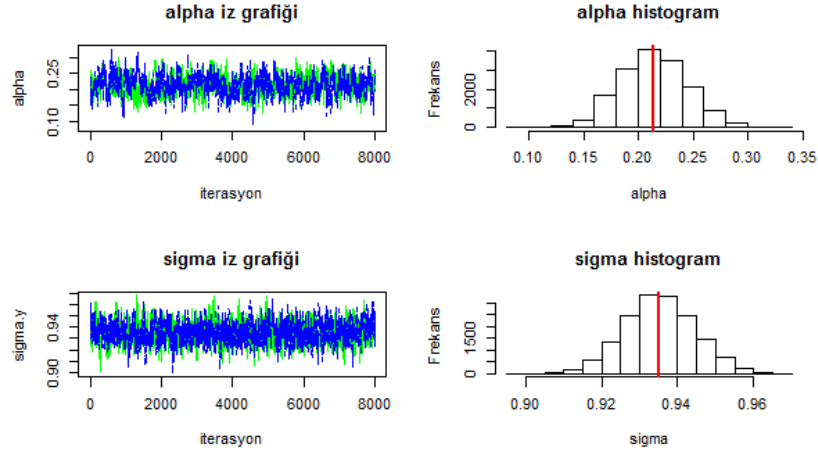


Şekil 6.3: β_0 , β_1 ve β_2 parametrelerine ait iz grafikleri ve sonsal dağılım histogram grafikleri.

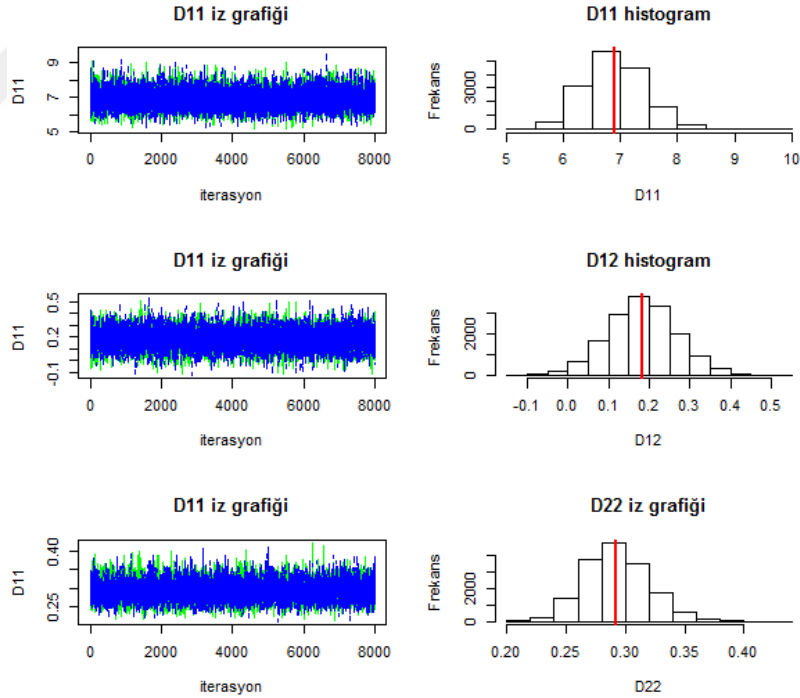


Şekil 6.4: γ_1 ve γ_2 parametrelerine ait iz grafikleri ve sonsal dağılım histogram grafikleri.

Tüm parametreler için otokorelasyon grafikleri Şekil 6.7’da verilmiştir. Şekil 6.7’den 25. gecikmeden sonra tüm parametreler için otokorelasyonun sifıra yaklaştığı görülmektedir. Bu durum gecikme sayısı arttıkça MCMC örneklemeleri arasındaki ilişkinin azaldığını göstermektedir. Grafikler incelendiğinde yakınsama β_0 , β_1 , β_2 ve α parametreleri için daha yavaş gerçekleşirken γ_1 , γ_2 ve σ_ϵ parametreleri için daha hızlı gerçekleşmektedir. D_{11} , D_{12} ve D_{22} parametreleri için örneklemeler Gibbs örnekleme ile elde edildiğinden MCMC örneklemeleri arasındaki otokorelasyon 3. gecikmeden sonra sifıra yaklaşmaktadır.

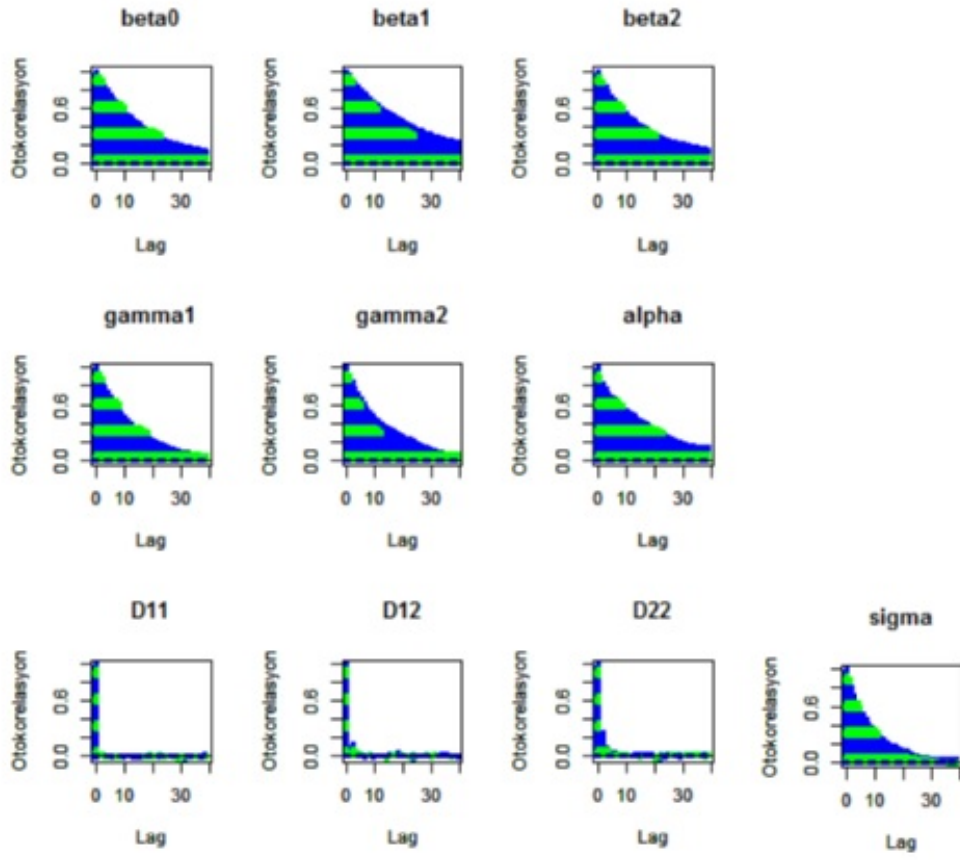


Şekil 6.5: α ve σ_{ε} parametrelerine ait iz grafikleri ve sonsal dağılım histogram grafikleri.



Şekil 6.6: D11, D12 ve D22 varyans kovaryans parametrelerine ait iz grafikleri ve sonsal dağılım histogram grafikleri.

Bayesçi yaklaşımla elde edilen parametre tahminleri ML yöntemi ile elde edilen parametre tahminlerine yakın olarak tahmin edilmiştir. Bunun nedeni bilgilendirici olmayan önseller kullanılmasından kaynaklanmaktadır.



Şekil 6.7: Zincir I ve Zincir II için Otokorelasyon grafikleri.

7 SİMÜLASYON

7.1 Birleşik Verinin Üretilmesi

Bu bölümde birleşik modelde ML ve Bayesçi yöntemin performanslarının değerlendirilmesi amacıyla simülasyon çalışması yapılmıştır. Simülasyon çalışmasında, uygulama bölümünde kullanılan tiroid veri setine benzer olarak boylamsal ve sağ kalım verisi üretilmiştir. Simülasyon çalışmasında, 7 yıl boyunca gözlenen hastalardan eşit sayıda ve 8 boylamsal ölçüm değeri alındığı varsayılmıştır. Değişkenler üretilirken R programında bulunan mvtnorm ve LaplaceDemon paketlerinden yararlanılmıştır.

Boylamsal alt modelde gerçek boylamsal yanıtlar,

$$m_i(t) = \beta_0 + \beta_1 Zaman_{ij} + \beta_2 TemelYa_i2 + b_{0i} + b_{1i} Zaman_{ij}$$

denkleminde yararlanılarak üretilmiştir. Boylamsal ölçümlerin alındığı Zaman değişkeni için tekdüze dağılımdan yararlanılmıştır. Böylelikle gerçek ver setine uygun olarak hastaların ziyaret zamanlarının birbirinden farklı olması sağlanmıştır. Modelde yer alan Temel yaş değişkeni ve Cinsiyet değişkeni sırasıyla normal dağılım ve binom dağılımından üretilmiştir. Tahmin yöntemlerinin performanslarının değerlendirilmesi amacıyla yanıt değişkenleri üretilirken uygulama bölümünde elde edilen varyans değeri $\sigma_1^2 = 0.96$, $\sigma_2^2 = 0.1$ ve $\sigma_3^2 = 10$ olmak üzere üç farklı varyans göz önünde bulundurulmuştur.

Birleşik modelde, iki süreç için ortak parametre olan rastgele etkiler parametrelerine ait varyans kovaryans matrisi,

$$D = \begin{bmatrix} \sigma_{b0}^2 & \sigma_{b0,b1} \\ \sigma_{b0,b1} & \sigma_{b1}^2 \end{bmatrix}$$

şeklindedir. Rastgele kesim parametresinin varyansı $\sigma_{b0}^2 = 6.98$ rastgele eğim parametresinin varyansı ise $\sigma_{b1}^2 = 0.33$ olarak belirlenmiştir. Kovaryans değeri ise

$\sigma_{b_1b_0} = 0.26$ 'dır. Doğrusal karma modelde bireylere ait değişimleri gösteren rastgele etki parametreleri 0 ortalama ve D varyans kovaryans matrisi ile normal dağılımdan üretilmiştir, $b_i \sim N(0, D)$. Boylamsal alt modeli için kullanılan doğrusal karma model parametrelerinin gerçek değerleri $\beta_0 = 1.127$, $\beta_1 = 0.1436$, $\beta_2 = 0.0443$ şeklindedir.

Modelde temel tehlike fonksiyonunun weibull dağıldığı varsayılmıştır. Sağ kalım alt modeli,

$$h_i(t) = \rho t^{\rho-1} \left\{ \gamma^T w_i + \alpha m_i(t) \right\}$$

şeklindedir. Sağ kalım alt modelinde yer alan zamandan bağımsız cinsiyet ve temel yaş değişkeninin gerçek değerleri sırasıyla $\gamma_1 = -0.197$ ve $\gamma_2 = 0.186$ 'dır. Birleşik modelde, boylamsal ve sağ kalım süreçleri arasındaki ilişkiyi gösteren α parametresinin değeri 0.23'tür.

Simülasyon çalışmasında, ML yöntemiyle parametre tahminlerinin elde edilmesi R programında açık kaynak olarak bulunan "JM" paketindeki jointmodel fonksiyonundan yararlanılarak elde edilmiştir. Bu pakette doğrusal karma model için "nlme" R paketindeki lme fonksiyonundan, Oransal Cox model için "survival" R paketinde bulunan coxph fonksiyonundan yararlanılmıştır. Bayesçi yaklaşımda parametre tahminleri R programında yazılan kodlarla elde edilmiştir. MCMC algoritması için Gibbs örnekleme ve Metropolis Hasting algoritmasını kullanan algoritma oluşturulmuştur.

Bayesçi yöntemde parametrelerin önsel dağılımları uygulama bölümünde olduğu gibi regresyon parametreleri için 0 ortalama ve 10 standart sapma ile normal dağılımdır. Rastgele etkilere ait varyans kovaryans matrisi önsel dağılımı $D \sim IW(2, diag(1))$, ölçüm hatasının varyansı için önsel dağılım $\sigma^2 \sim U(0, 100)$ olarak belirlenmiştir.

7.2 Simülasyon Sonuçları

Simülasyon zaman kısıtı nedeniyle 500 iterasyon için çalıştırılmıştır. MCMC örneklemesinde iterasyon sayısı 1200 olarak belirlenmiştir. İlk 200 örnekleme yakma işlemi uygulanmış ve sonuçlar 1000 örneklem üzerinden elde edilmiştir. Örneklem genişliği $n = 100$, $n = 300$ ve $n = 600$ ve varyans $\sigma^2 = (0.1, 0.96, 10)$ olmak üzere dokuz farklı simülasyon senaryosu üzerinden parametre tahminleri elde edilmiştir.

ML yöntemi ve Bayesçi yöntem ile elde edilen simülasyon sonuçları hata kareler ortalaması (MSE) ve standart hatalar üzerinden karşılaştırılmıştır. Varyans değerleri $\sigma^2 = (0.1, 0.96, 10)$ ve örneklem genişlikleri $n = (100, 300, 600)$ için simülasyon sonuçları sırasıyla Çizelge 7.1, Çizelge 7.2 ve Çizelge 7.3'te özetlenmiştir.

Çizelge 7.1: $\sigma_1^2 = 0.1$ ve farklı örneklem genişlikleri (100,300,600) için ML ve Bayesçi yöntem için simülasyon sonuçlarının karşılaştırılması

Yöntem		Parametre	β_0	β_1	β_2	γ_1	γ_2	α	σ_{11}	σ_{12}	σ_{22}
		Gerçek değer	1.127	0.1436	0.0443	-0.197	0.0186	0.2380	2.64	0.26	0.58
$n = 100$	ML	Ortalama	1.121	0.1464	0.0443	-0.2376	0.0212	0.2606	2.63	0.26	0.58
		Yan	-0.006	0.0028	0.0000	-0.0406	0.0026	0.0226	-0.01	0.00	0.00
		MSE	0.8566	0.0035	0.0002	0.3520	0.0004	0.0081	0.9241	0.0274	0.0022
		SH	0.0413	0.0026	0.0007	0.0264	0.0008	0.0037	0.0428	0.0074	0.0021
$n = 100$	Bayes	Ortalama	1.1257	0.1481	0.0443	-0.2306	0.0201	0.2682	2.64	0.26	0.63
		Yan	-0.0013	0.0045	0.0000	-0.0336	0.0015	0.0309	0.00	0.00	0.05
		MSE	0.8400	0.0036	0.0001	0.4665	0.0005	0.0107	0.9338	0.0276	0.0063
		SH	0.0409	0.0027	0.0006	0.0305	0.0010	0.0044	0.0224	0.0074	0.0017
$n = 300$	ML	Ortalama	1.1096	0.1430	0.0444	-0.1825	0.0185	0.2451	2.63	0.26	0.58
		Yan	-0.0174	-0.0006	-0.0001	0.0145	-0.001	0.0071	-0.01	0.00	0.00
		MSE	0.3186	0.0012	0.0001	0.0992	0.0001	0.0017	0.3737	0.0091	0.0007
		SH	0.0252	0.0016	0.0004	0.0004	0.0003	0.0018	0.0272	0.0042	0.0012
$n = 300$	Bayes	Ortalama	1.1145	0.1440	0.0443	-0.1840	0.0190	0.2391	2.63	0.26	0.60
		Yan	-0.0125	0.0004	0.0000	0.0130	0.0004	0.0011	-0.01	0.00	-0.02
		MSE	0.3167	0.0012	0.0001	0.1350	0.0001	0.0018	0.3739	0.0090	0.0012
		SH	0.0252	0.0016	0.0004	0.0164	0.0004	0.0019	0.0273	0.0043	0.0011
$n = 600$	ML	Ortalama	1.1209	0.1417	0.0441	-0.1983	0.0180	0.2415	2.64	0.26	0.58
		Yan	-0.0061	-0.0019	-0.0002	-0.0013	-0.0006	0.0035	0.00	0.00	0.00
		MSE	0.1470	0.0006	0.0001	0.00004	0.00005	0.0007	0.1603	0.0046	0.0004
		SH	0.0171	0.0011	0.0003	0.0085	0.0003	0.0012	0.0178	0.0030	0.0009
$n = 600$	Bayes	Ortalama	1.1262	0.1427	0.0440	-0.2121	0.0183	0.2331	2.64	0.26	0.59
		Yan	-0.0008	-0.0009	0.0003	-0.015	-0.0003	-0.004	0.00	0.00	0.01
		MSE	0.1467	0.0005	0.0001	0.052	0.0001	0.0008	0.1613	0.0046	0.0005
		SH	0.0171	0.0011	0.0003	0.0101	0.0003	0.0013	0.0178	0.0030	0.0008

ML(Maximum Likelihood): En Çok Olabilirlik, MSE:Hata Kareler Ortalaması, SH:Standart Hata.

Küçük hata varyanslarının kullanıldığı simülasyon senaryosu için ML ve Bayesçi yöntem iyi performans göstermiştir. Her iki yöntemde de tahmin edilen parametre değerleri için yanlılık ve standart hata değerleri düşüktür. Çizelge 7.1 incelendiğinde örneklem büyüklüğü arttıkça her iki yöntemde de parametrelerin gerçek değerine yakınsamanın gerçekleştiği ve MSE değerlerinin ve standart hatalar azaldığı görülmektedir. Çizelge 7.1'de özetlenen simülasyon sonuçlarına göre Bayesçi yöntemde, parametrelerin sonsal dağılımları güncellenerek elde edilen sonsal ortalama tahminlerinin, örneklem genişliği arttırıldığında gerçek parametre değerlerine

Çizelge 7.2: $\sigma_1^2 = 0.96$ ve farklı örneklem genişlikleri (100,300,600) için ML ve Bayeşçi yöntem için simülasyon sonuçlarının karşılaştırılması

Yöntem	Parametre	β_0	β_1	β_2	γ_1	γ_2	α	σ_{11}	σ_{12}	σ_{22}
	Gerçek değer	1.127	0.1436	0.0443	-0.197	0.0186	0.2380	2.64	0.26	0.58
$n = 100$	ML Ortalama	1.090	0.1380	0.0443	-0.192	0.021	0.2616	2.61	0.50	0.58
	ML Yan	-0.0368	-0.0056	0.000	0.005	0.002	0.4996	0.03	-0.01	0.00
	ML MSE	1.0216	0.0042	0.0003	0.3642	0.0004	0.0083	1.039	0.0321	0.003
	ML SH	0.0451	0.003	0.001	0.026	0.0008	0.0039	0.044	0.0079	0.0024
$n = 100$	Bayes Ortalama	1.092	0.1396	0.0443	-0.2055	0.0179	0.2616	2.62	0.24	0.65
	Bayes Yan	-0.035	-0.004	0.0000	-0.0083	-0.0007	0.4996	0.03	-0.02	0.07
	Bayes MSE	1.0045	0.0042	0.00005	0.4772	0.0005	0.0099	1.052	0.0318	0.009
	Bayes SH	0.0448	0.003	0.001	0.030	0.001	0.0043	0.045	0.0079	0.0019
$n = 300$	ML Ortalama	1.1141	0.1403	0.0440	-0.1813	0.0184	0.2463	2.64	0.26	0.58
	ML Yan	-0.0129	-0.0033	-0.003	0.0157	-0.0002	0.0083	0.00	0.00	0.00
	ML MSE	0.3272	0.0013	0.0001	0.0863	0.0001	0.0018	0.3875	0.0097	0.0010
	ML SH	0.0255	0.0016	0.0004	0.0131	0.0004	0.0019	0.0278	0.0044	0.0013
$n = 300$	Bayes Ortalama	1.1250	0.1402	0.0440	-0.1866	0.0197	0.2403	2.64	0.25	0.60
	Bayes Yan	-0.002	-0.0032	-0.003	0.0104	0.000	0.0023	0.00	0.01	0.02
	Bayes MSE	0.3256	0.0013	0.00004	0.1315	0.0001	0.0020	0.3895	0.0096	0.0014
	Bayes SH	0.0255	0.0016	0.0004	0.0162	0.0005	0.0020	0.0225	0.0043	0.0013
$n = 600$	ML Ortalama	1.1099	0.1437	0.0443	-0.1863	0.0182	0.2433	2.64	0.26	0.58
	ML Yan	-0.0170	-0.0001	0.0000	0.0107	-0.0004	0.0053	0.00	0.00	0.00
	ML MSE	0.1671	0.0006	0.0001	0.0429	0.0001	0.0008	0.1827	0.0055	0.0005
	ML SH	0.01826	0.0011	0.0003	0.0092	0.0003	0.0013	0.0190	0.0033	0.0010
$n = 600$	Bayes Ortalama	1.1251	0.1433	0.0441	-0.1923	0.0180	0.2386	2.64	0.26	0.59
	Bayes Yan	-0.0018	-0.0003	0.0002	0.0047	-0.0006	0.0086	0.00	0.00	0.01
	Bayes MSE	0.1661	0.0006	0.0001	0.0601	0.0001	0.0008	0.1862	0.0055	0.0007
	Bayes SH	0.01822	0.0011	0.0003	0.0109	0.0003	0.0013	0.0190	0.0033	0.0009

ML(Maximum Likelihood): En Çok Olabilirlik, MSE:Hata Kareler Ortalaması, SH:Standart Hata.

yakınsadığı ve küçük varyans değerleri için iyi bir performans gösterdiği söylenebilir. Hata varyansı $\sigma_1^2 = 0.1$ ve $\sigma_2^2 = 0.96$ değerleri için β_0 , β_1, γ_2 ve α parametreleri için gerçek değerlere yakın sonuçlar elde edilmiştir ve parametrelerdeki yanlılık ML tahminlerine göre azdır. Ayrıca bu parametreler için standart hata ve MSE değerleri de ML yöntemine göre küçüktür. Bayeşçi yöntemde, birleşik modelde önemli bir parametre olan ve iki süreç arasındaki ilişkiyi gösteren α parametresi için tahminler her simülasyon senaryosu için gerçek değere daha yakın sonuçlar elde edilmiştir.

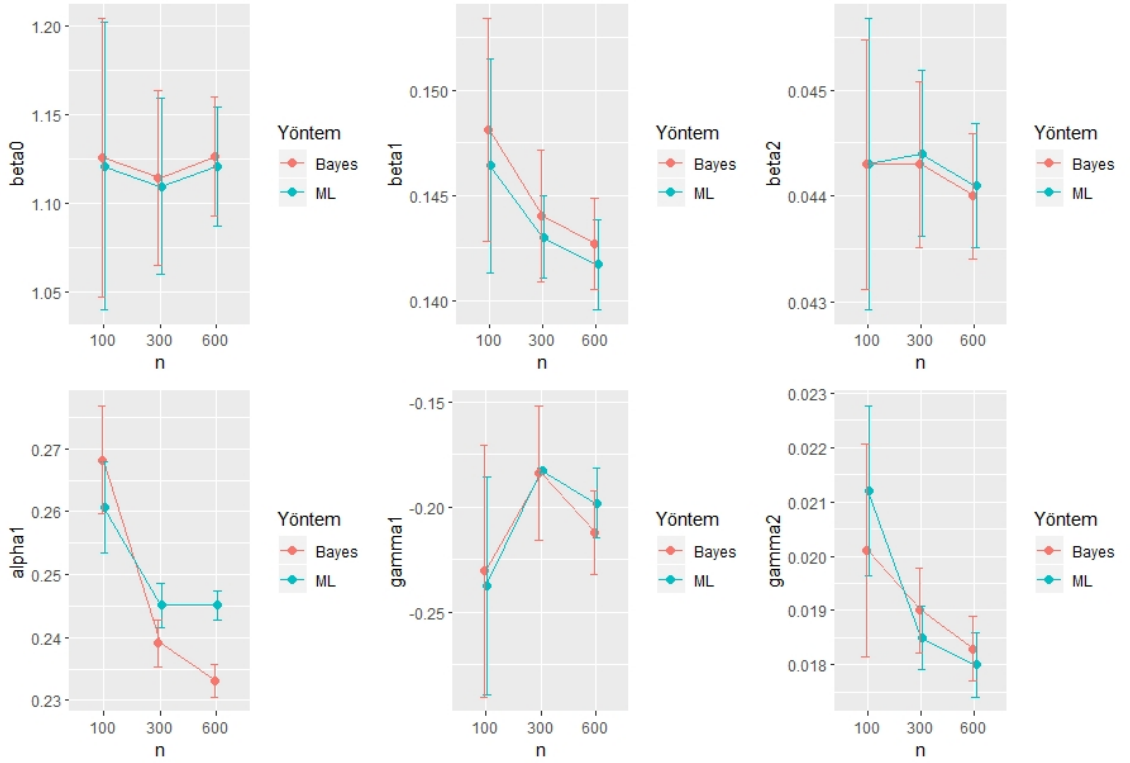
Çizelge 7.3'te de görüldüğü gibi hata varyansı $\sigma^2 = 10$ gibi büyük bir değer alındığında ve bilgilendirici olmayan önseller kullanıldığında Bayeşçi yöntemde parametrelere yakınsamada iyi performans göstermediği ve parametrelerin gerçek değerlerine yakınsamada ML yöntemi Bayeşçi yönteme göre daha iyi sonuçlar verdiği görülmüştür.

Çizelge 7.3: $\sigma_1^2 = 10$ ve farklı örneklem genişlikleri (100,300,600) için ML ve Bayesçi yöntem için simülasyon sonuçlarının karşılaştırılması

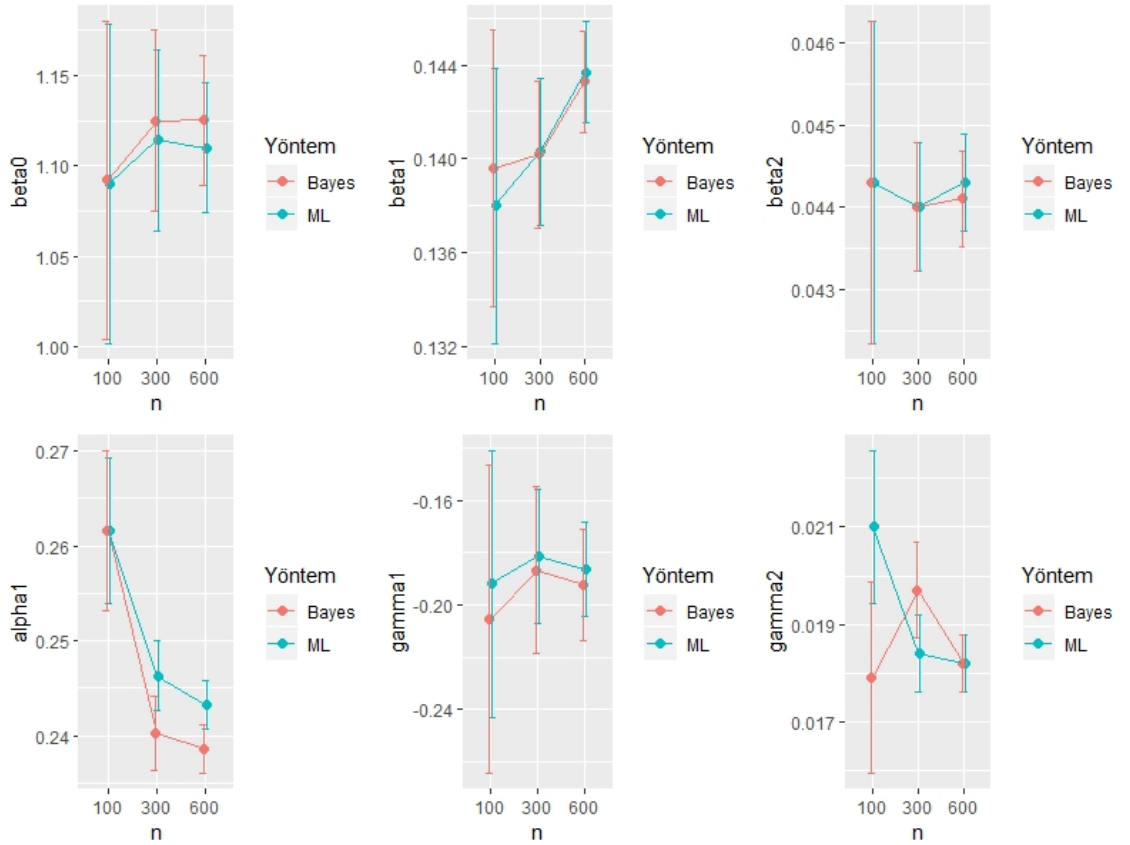
Yöntem	Parametre	β_0	β_1	β_2	γ_1	γ_2	α	σ_{11}	σ_{12}	σ_{22}	
	Gerçek değer	1.127	0.1436	0.0443	-0.197	0.0186	0.2380	2.64	0.26	0.58	
$n = 100$	ML	Ortalama	1.0935	0.1433	0.0446	-0.1895	0.0195	0.2646	2.61	0.25	0.57
		Yan	-0.0335	-0.0003	0.0003	0.0075	0.0009	0.0273	-0.01	-0.01	-0.01
		MSE	1.3999	0.0091	0.0004	0.3668	0.0003	0.0122	2.5917	0.0973	0.0126
		SH	0.0527	0.0042	0.0009	0.0270	0.0008	0.0048	0.0713	0.0139	0.0049
$n = 100$	Bayes	Ortalama	1.123	0.1467	0.0445	-0.2156	0.0200	0.2965	2.60	0.01	1.33
		Yan	-0.004	0.0031	0.0002	-0.0186	-0.0014	0.0585	-0.04	-0.25	0.75
		MSE	1.3862	0.0260	0.0001	1.0607	0.0017	0.0529	4.8019	5.5441	70.426
		SH	0.0524	0.0071	0.0009	0.0459	0.0018	0.0099	0.0971	0.1044	0.3741
$n = 300$	ML	Ortalama	1.1483	0.1391	0.0437	-0.1922	0.0182	0.2448	2.63	0.26	0.58
		Yan	0.0213	-0.0045	0.0004	0.0048	-0.0004	0.0068	-0.01	0.00	0.00
		MSE	0.4912	0.0027	0.0001	0.0927	0.0001	0.0024	0.8388	0.0370	0.0047
		SH	0.0312	0.0023	0.0005	0.0136	0.0004	0.0022	0.0407	0.0085	0.0031
$n = 300$	Bayes	Ortalama	1.1946	0.1248	0.0435	-0.1535	0.0185	0.2425	2.62	0.16	0.67
		Yan	0.0068	-0.0188	0.0002	0.0435	-0.0001	0.0045	-0.02	-0.10	0.09
		MSE	0.4828	0.0030	0.0001	0.1298	0.0001	0.0035	1.2125	0.0394	0.0823
		SH	0.0311	0.0023	0.0005	0.0159	0.0005	0.0026	0.0489	0.0076	0.0117
$n = 600$	ML	Ortalama	1.1196	0.1422	0.0439	-0.1835	0.0181	0.2436	2.63	0.26	0.58
		Yan	-0.007	-0.0014	-0.00008	0.0135	-0.0005	0.0064	-0.01	0.00	0.00
		MSE	0.2167	0.0015	0.0001	0.0419	0.00004	0.0013	0.4296	0.017	0.0024
		SH	0.0208	0.0017	0.0004	0.0091	0.0003	0.0015	0.0293	0.0058	0.0022
$n = 600$	Bayes	Ortalama	1.1695	0.1262	0.0437	-0.2081	0.0203	0.2350	2.64	0.21	0.61
		Yan	0.0425	-0.0173	-0.0003	-0.0111	0.0017	-0.0022	0.00	-0.05	0.03
		MSE	0.2171	0.0017	0.0001	0.0606	0.0001	0.0015	0.4404	0.015	0.0032
		SH	0.0208	0.0017	0.0004	0.0110	0.0003	0.0017	0.0296	0.0051	0.0015

ML(Maximum Likelihood): En Çok Olabilirlik, MSE:Hata Kareler Ortalaması, SH:Standart Hata.

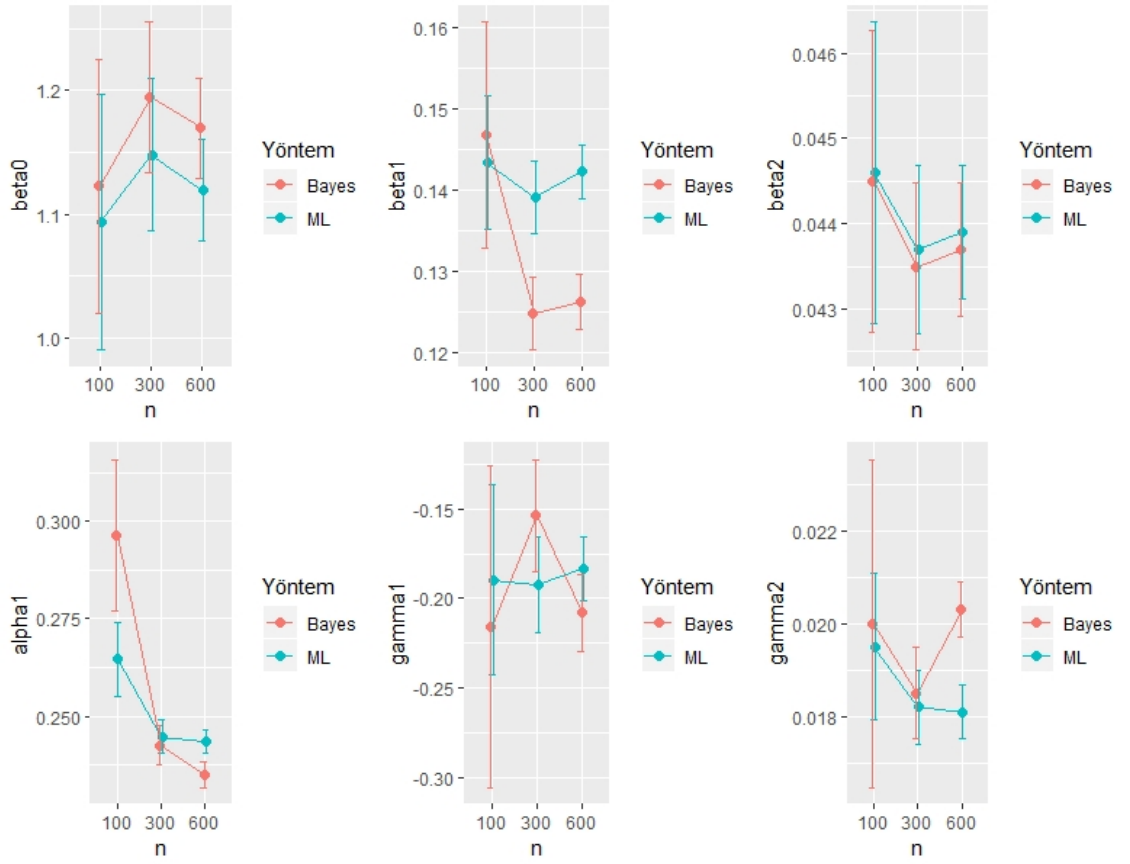
Genel olarak bir değerlendirme yapılacak olursa farklı örneklem genişlikleri ve küçük hata varyansları için ML yöntemi ve Bayesçi yöntem karşılaştırıldığında parametrelerin tahminleri ve standart hatalar benzer sonuçlar vermiştir. İki tahmin yönteminde de örneklem genişliği arttıkça MSE ve standart hatalar azaldığı gözlenmiştir. Tüm simülasyon senaryoları için parametrelere ait güven aralıklarını gösteren grafikler Şekil 7.1, Şekil 7.2 ve Şekil 7.3'te verilmiştir.



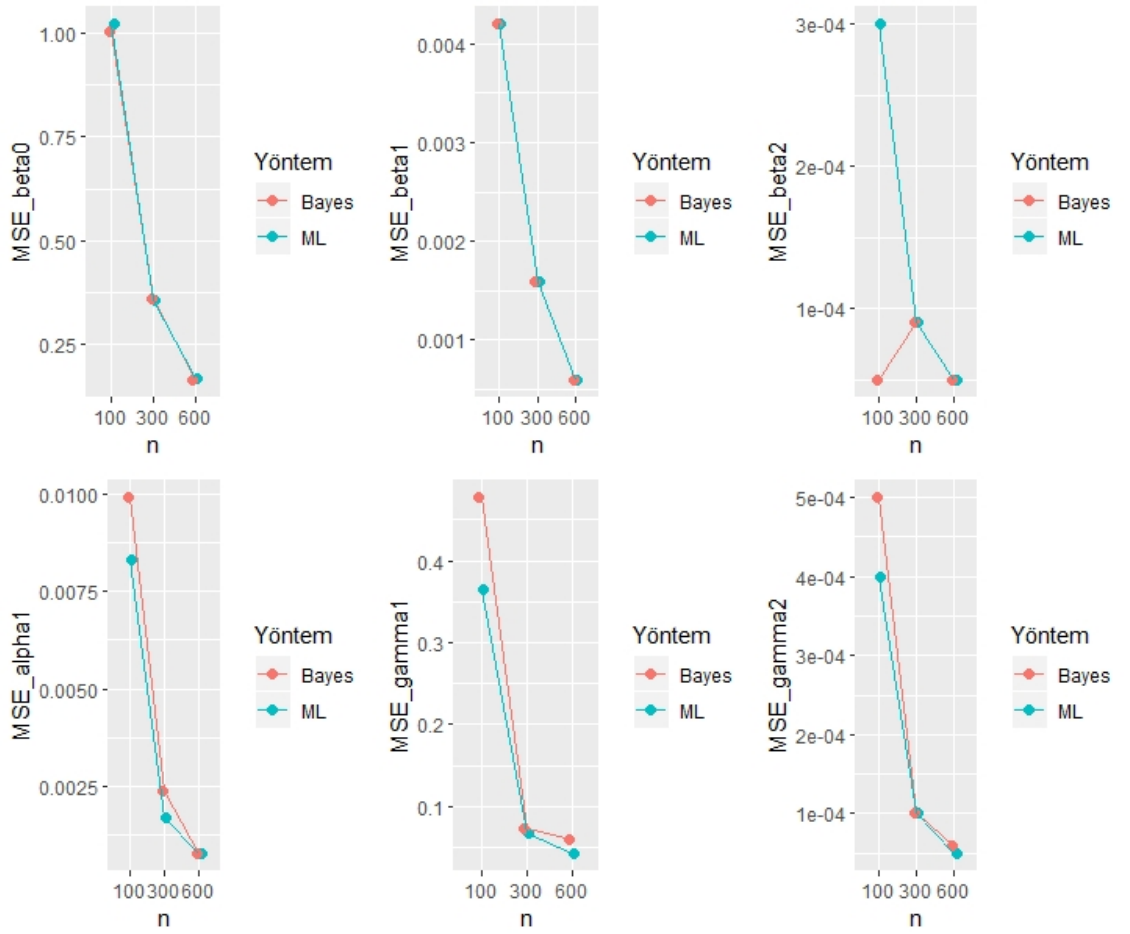
Şekil 7.1: $\sigma^2 = 0.1$ ve örneklem genişliği $n = (100, 300, 600)$ için Güven Aralıkları



Şekil 7.2: $\sigma^2 = 0.96$ ve örneklem genişliği $n = (100, 300, 600)$ için Güven Aralıkları



Şekil 7.3: $\sigma^2 = 10$ ve örneklem genişliği $n = (100, 300, 600)$ için Güven Aralıkları



Şekil 7.4: $\sigma^2 = 0.1$ ve farklı örneklem genişlikleri $n = (100, 300, 600)$ MSE değerlerinin değişimi

8 SONUÇ VE ÖNERİLER

Birleşik modelleme ile alternatif yöntemlere göre etkin ve yansız tahmin ediciler elde edildiğinden son yıllarda sıklıkla kullanılan bir yöntemdir. Birleşik modellemede boylamsal ve sağ kalım süreci arasında ilişki olduğu varsayılır ve bu süreçler arasındaki ilişki, iki süreç için ortak parametre olan rastgele etki parametreleri ile sağlanır. Birleşik modellemede parametre tahminleri genellikle ML yöntemine dayanmaktadır. ML yönteminde, olabilirlik fonksiyonunda yer alan integralin kapalı formda çözümü yoktur ve integralin alınması rastgele etkilerin boyutu arttıkça karmaşık hale gelmektedir. Bu integrale Gauss Hermite, Uyarlanabilir Gauss Hermit Laplace yaklaşımı ve Monte Carlo gibi sayısal yöntemlerle yakınsama sağlanabilir.

Bayesçi yöntemde, karmaşık olabilirlik yapıları önsel dağılımlarla birleştirildiğinde sonsal dağılımlardan parametre tahminlerinin elde edilmesi analitik olarak genellikle zordur ve parametre sayısı arttıkça işlem yükü de artmaktadır fakat MCMC benzetim yöntemleri ile parametrelere yakınsama sağlanabilir. Bu çalışmada ML yönteminde parametre tahminlerinde karşılaşılan hesaplama zorluklarının azaltılması ve parametre tahminlerinin elde edilmesinde pratik bir yol sağlaması açısından Bayesçi yöntem benimsenmiş ve ML yöntemine alternatif yöntem olarak MCMC benzetim yöntemi ile parametre tahminlerinin yapılması önerilmiştir.

Bayesçi yöntemin, ML yönteminden farkı gözlenen veriden elde edilen bilgi ile araştırmacının geçmiş deneyimlerine dayananan önsel bilginin birleştirilmesine dayalı bir yöntem olmasıdır. Bayesçi yöntemde, bilgilendirici önseller ya da bilgilendirici olmayan önseller kullanılabilir. Bilgilendirici önsel kullanmanın Bayesçi yöntemin gücünü arttırmaktadır fakat bu çalışmada parametreler için önsel dağılımlar belirlenirken bilgilendirici olmayan önseller kullanılmıştır.

Çalışmada analizlerin yapılabilmesi için 341 tiroid kanseri hastasına ait veri seti kullanılmıştır. Bireylerde tümör tekrarının görülmesi ve tiroglobulin değerleri

arasındaki ilişkinin ortaya çıkarılması ve bireylerde tümör tekrarının görülmesi üzerinde etkili olan faktörlerin ortaya çıkarılması amacıyla birleşik modellemeden yararlanılmıştır. Birleşik modellemede karşılaştırma yapabilmek amacıyla öncelikle ML yöntemiyle parametre tahminleri yapılmıştır.

Modelde regresyon parametreleri için normal dağılım, ölçüm hatasının varyansı için düzgün dağılım kullanılmıştır. Ölçüm hatasının varyansı için başlangıçta ters gamma dağılımı belirlenmiş fakat ters gamma dağılımında seçilen ölçek parametresi ile parametrelere yakınsamada sağlanamadığından ölçüm hatasının varyansı için önsel dağılım düzgün dağılım olarak belirlenmiştir. Önsel dağılım düzgün dağılım olarak belirlendiğinde parametreye yakınsama sağlanmıştır. Rastgele etki parametresine ait varyans kovaryans parametresi için önsel dağılım ters Wishart dağılımı olarak belirlenmiştir.

Parametrelerin tahminleri MCMC benzetim yönteminde sıklıkla kullanılan Gibbs örnekleme ve MH algoritması birleştirilerek elde edilmiştir. Parametre tahminleri elde edildikten sonra Bayesçi yöntemde kullanılan yakınsama tanılarıyla parametreler için yakınsama kontrolleri yapılmıştır. Bayesçi yöntemde yakınsamanın kontrolü için Geweke, Gelman ve Rubin tanıları kullanılmıştır. Aynı zamanda iz grafikleri ve otokorelasyon grafikleriyle de her parametre için yakınsamanın kontrolü yapılmıştır. Geweke, Gelman ve Rubin tanılarına göre parametrelere yakınsamanın sağlandığı görülmüştür. Çizdirilen iz grafikleri zincirlerin iyi bir karışıma sahip olduğunu göstermektedir.

Bayesçi yöntem ve ML yöntemi ile elde edilen parametre tahminleri karşılaştırıldığında, Bayesçi ve ML tahmin değerleri birbirine yakın değerler olarak tahmin edilmiştir. Bunun nedeni parametreler için önsel dağılımlar belirlenirken bilgilendirici olmayan önsel seçiminden kaynaklanmaktadır. Bayesçi yöntemde bilgilendirici olmayan önseller kullanıldığında sonsal dağılımlar genellikle olabilirlik fonksiyonuna benzemektedir ve ML yöntemi ile elde edilen parametre tahminleri ile Bayesçi yöntemler benzer sonuçlar vermektedir. Bu durum birleşik modellemede yüksek boyutlu veri setleri ile çalışıldığında ve ML yöntemi ile parametre tahminlerinin elde edilmesinin zor olduğu durumlarda bilgilendirici olmayan önseller ile MCMC benzetim yöntemleriyle elde edilen parametre tahminlerinin ML yöntemiyle elde edilen parametre tahminlerine yakın sonuçlar vereceğini göstermektedir. Ayrıca

bilgilendirici olmayan önsel dağılımların hiperparametre değerleri belirlenirken varyans değerleri için büyük değerler belirlenmiştir. Bu durum önsel dağılımın sonsal dağılım üzerindeki etkisini azaltmış ve parametre tahminleri klasik yöntemle benzer sonuçlar göstermiştir.

Son olarak oluşturulan algoritmanın performansının değerlendirilmesi ve uygulamada elde edilen sonuçların desteklenebilmesi için simülasyon çalışması yapılmıştır. Simülasyon sonuçlarına göre, örneklem büyüklüğü arttıkça Bayesçi yöntemde parametrelerin gerçek değerine yakınsadığı görülmektedir. Aynı zamanda klasik yöntem ve Bayesçi yöntemin performanslarının karşılaştırılabilmesi için R programında bulunan JM paketinden yararlanarak parametre tahminleri elde edilmiştir. JM paketinde parametre tahminleri ML yöntemi ile yapılmaktadır. Elde edilen simülasyon sonuçlarına göre her iki yöntemde de parametre tahminleri yakın sonuçlar vermektedir. Karmaşık yapıda olabilirlik fonksiyonuna sahip modellerde işlem kolaylığı açısından Bayesçi yöntem ML yöntemine alternatif bir yöntem olarak kullanılabilir.

Birleşik modelde, boylamsal ve sağ kalım alt modelinde farklı yapıda modeller kullanılabilir. Alt modellerin farklı modellere genişletilmesi durumunda ML yöntemiyle tahminler yapılmak istendiğinde yüksek boyutlu integrallerin alınması gerekebilir ve bu durum araştırmacılar için işlem yükünü arttırabilir. Bayesçi yöntemde yüksek boyutlu integrallerin alınmasına gerek yoktur çünkü Bayesçi yöntem, oluşturulan algoritmalar yardımıyla sonsal dağılımların güncellenerek parametre tahminlerinin elde edilmesine olanak sağlar.

Bayesçi yaklaşım karmaşık yapıdaki modellerde parametre tahminleri için hesaplama yükü açısından kolaylıklar sağlasa da çok parametrelili modellerde parametrelere yakınsamada sorunlar yaşanabilir. Bu nedenle Bayesçi yöntemde önsel dağılımlar belirlenirken parametreler uygun önsel dağılımların belirlenmesi önemlidir ve parametrelerin önsel dağılımlarının belirlenmesinde bilgilendirici önsel kullanılması Bayesçi yöntemin gücünü de arttırmaktadır.

Bu çalışmada, boylamsal ve sağ kalım süreçleri için ortak parametre olan rastgele etki parametrelerinin normal dağılımdan geldiği varsayılmıştır. Bu parametrelerin farklı dağılımlardan gelmesi ve modelde aykırı gözlemlerin bulunması durumunun birleşik modelde parametre tahminleri üzerine etkisinin incelenmesi ileriye yönelik planlanan

alıřmalar arasındadır. Aynı zamanda alıřmada, boylamsal alt modelin dođrusal karma model olduđu varsayılmıřtır. Ancak veri yapısı her zaman dođrusal karma model kullanımı iin uygun olmayabilir. Gibbs rneklemesi ve MH algoritması gibi MCMC benzetim yntemleri boylamsal alt modelin farklı modellere geniřletilmesinde esneklik sađladıđından birleřik modellemede boylamsal alt modelin farklı modeller kullanılarak esnetilmesi de ileriye ynelik planlanan alıřmalar arasındadır.



KAYNAKLAR

- Brown, E. R. (2009). Assessing the association between trends in a biomarker and risk of event with an application in pediatric hiv/aids. *The annals of applied statistics*, 3(3):1163.
- Brown, E. R., Ibrahim, J. G., ve DeGruttola, V. (2005). A flexible b-spline model for multiple longitudinal biomarkers and survival. *Biometrics*, 61(1):64–73.
- Carlin, B. P. ve Louis, T. A. (2010). *Bayes and empirical Bayes methods for data analysis*. Chapman and Hall/CRC.
- Collett, D. (2015). *Modelling survival data in medical research*. CRC press.
- Cox, D. R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society: Series B (Methodological)*, 34(2):187–202.
- Cox, D. R. ve Snell, E. J. (1968). A general definition of residuals. *Journal of the Royal Statistical Society: Series B (Methodological)*, 30(2):248–265.
- Dafni, U. G. ve Tsiatis, A. A. (1998). Evaluating surrogate markers of clinical outcome when measured with error. *Biometrics*, pages 1445–1462.
- De Gruttola, V. ve Tu, X. M. (1994). Modelling progression of cd4-lymphocyte count and its relationship to survival time. *Biometrics*, pages 1003–1014.
- Dempster, A. P., Laird, N. M., ve Rubin, D. B. (1977). Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)*, 39(1):1–22.
- Diggle, P., Diggle, P. J., Heagerty, P., Liang, K.-Y., Heagerty, P. J., Zeger, S., ve diğ. (2002). *Analysis of longitudinal data*. Oxford University Press.

- Elashoff, R., Li, N., ve diğ. (2016). *Joint modeling of longitudinal and time-to-event data*. CRC Press.
- Elashoff, R. M., Li, G., ve Li, N. (2008). A joint model for longitudinal measurements and survival data in the presence of multiple failure types. *Biometrics*, 64(3):762–771.
- Faraway, J. J. (2016). *Extending the linear model with R: generalized linear, mixed effects and nonparametric regression models*. CRC press.
- Faucett, C. L. ve Thomas, D. C. (1996). Simultaneously modelling censored survival data and repeatedly measured covariates: a gibbs sampling approach. *Statistics in medicine*, 15(15):1663–1685.
- Fitzmaurice, G. M., Laird, N. M., ve Ware, J. H. (2012). *Applied longitudinal analysis*, volume 998. John Wiley & Sons.
- Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., ve Rubin, D. B. (2013). *Bayesian data analysis*. CRC press.
- Gelman, A. ve diğ. (2006). Prior distributions for variance parameters in hierarchical models (comment on article by browne and draper). *Bayesian analysis*, 1(3):515–534.
- Gelman, A., Rubin, D. B., ve diğ. (1992). Inference from iterative simulation using multiple sequences. *Statistical science*, 7(4):457–472.
- Geweke, J. ve diğ. (1991). *Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments*, volume 196. Federal Reserve Bank of Minneapolis, Research Department Minneapolis, MN.
- Harville, D. A. (1974). Bayesian inference for variance components using only error contrasts. *Biometrika*, 61(2):383–385.
- Heidelberger, P. ve Welch, P. D. (1983). Simulation run length control in the presence of an initial transient. *Operations Research*, 31(6):1109–1144.
- Henderson, C. R., Kempthorne, O., Searle, S. R., ve Von Krosigk, C. (1959). The estimation of environmental and genetic trends from records subject to culling. *Biometrics*, 15(2):192–218.

- Henderson, R., Diggle, P., ve Dobson, A. (2000). Joint modelling of longitudinal measurements and event time data. *Biostatistics*, 1(4):465–480.
- Hickey, G. L., Philipson, P., Jorgensen, A., ve Kolamunnage-Dona, R. (2016). Joint modelling of time-to-event and multivariate longitudinal outcomes: recent developments and issues. *BMC medical research methodology*, 16(1):117.
- Hoff, P. D. (2009). *A first course in Bayesian statistical methods*, volume 580. Springer.
- Hu, W., Li, G., ve Li, N. (2009). A bayesian approach to joint analysis of longitudinal measurements and competing risks failure time data. *Statistics in medicine*, 28(11):1601–1619.
- Huang, X., Li, G., Elashoff, R. M., ve Pan, J. (2011a). A general joint model for longitudinal measurements and competing risks survival data with heterogeneous random effects. *Lifetime data analysis*, 17(1):80–100.
- Huang, X., Stefanski, L. A., ve Davidian, M. (2009). Latent-model robustness in joint models for a primary endpoint and a longitudinal process. *Biometrics*, 65(3):719–727.
- Huang, Y., Dagne, G., ve Wu, L. (2011b). Bayesian inference on joint models of hiv dynamics for time-to-event and longitudinal data with skewness and covariate measurement errors. *Statistics in Medicine*, 30(24):2930–2946.
- Ibrahim, J. G., Chu, H., ve Chen, L. M. (2010). Basic concepts and methods for joint models of longitudinal and survival data. *Journal of Clinical Oncology*, 28(16):2796.
- Institute Inc, S. (2008). *Sas/stat® 9.2 user's guide*.
- Jeffreys, H. (1961). *Theory of probability*, clarendon.
- Kalbfleisch, J. D. ve Prentice, R. L. (2011). *The statistical analysis of failure time data*, volume 360. John Wiley & Sons.
- Klein, J. P. ve Moeschberger, M. L. (2006). *Survival analysis: techniques for censored and truncated data*. Springer Science & Business Media.

- Kleinbaum, D. G. (1998). Survival analysis, a self-learning text. *Biometrical Journal: Journal of Mathematical Methods in Biosciences*, 40(1):107–108.
- Konar, N. M., Karaismailođlu, E., Portakal, O., Pinar, A., Dikmen, Z. G., ve Karađaođlu, A. E. (2018). Birleřik modelleme yaklařımı ile tekrarlı c-reaktif protein ölçümlerinin tanısal performansını etkileyen faktörlerin incelenmesi. *Türkiye Klinikleri Journal of Biostatistics*, 10(1).
- Laird, N. M. ve Ware, J. H. (1982). Random-effects models for longitudinal data. *Biometrics*, pages 963–974.
- Lavalley, M. P. ve Degruittola, V. (1996). Models for empirical bayes estimators of longitudinal cd4 counts. *Statistics in Medicine*, 15(21):2289–2305.
- Lawless, J. F. (2011). *Statistical models and methods for lifetime data*, volume 362. John Wiley & Sons.
- Lawrence Gould, A., Boye, M. E., Crowther, M. J., Ibrahim, J. G., Quartey, G., Micallef, S., ve Bois, F. Y. (2015). Joint modeling of survival and longitudinal non-survival data: current methods and issues. report of the dia bayesian joint modeling working group. *Statistics in medicine*, 34(14):2181–2195.
- Lee, E. T. ve Wang, J. (2003). *Statistical methods for survival data analysis*, volume 476. John Wiley & Sons.
- Li, N., Elashoff, R. M., Li, G., ve Tseng, C.-H. (2012). Joint analysis of bivariate longitudinal ordinal outcomes and competing risks survival times with nonparametric distributions for random effects. *Statistics in medicine*, 31(16):1707–1721.
- McCrink, L. M., Marshall, A. H., ve Cairns, K. J. (2013). Advances in joint modelling: a review of recent developments with application to the survival of end stage renal disease patients. *International Statistical Review*, 81(2):249–269.
- McCulloch, C. E. ve Neuhaus, J. M. (2005). Generalized linear mixed models. *Encyclopedia of biostatistics*, 4.

- Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., ve Teller, E. (1953). Equation of state calculations by fast computing machines. *The journal of chemical physics*, 21(6):1087–1092.
- Nydick, S. W. (2012). The wishart and inverse wishart distributions. *Electronic Journal of Statistics*, 6:1–19.
- Pantazis, N. ve Touloumi, G. (2007). Robustness of a parametric model for informatively censored bivariate longitudinal data under misspecification of its distributional assumptions: A simulation study. *Statistics in medicine*, 26(30):5473–5485.
- Plummer, M., Best, N., Cowles, K., ve Vines, K. (2006). Coda: convergence diagnosis and output analysis for mcmc. *R news*, 6(1):7–11.
- Prentice, R. L. (1982). Covariate measurement errors and parameter estimation in a failure time regression model. *Biometrika*, 69(2):331–342.
- R. Brown, E. ve G. Ibrahim, J. (2003). A bayesian semiparametric joint hierarchical model for longitudinal and survival data. *Biometrics*, 59(2):221–228.
- Rizopoulos, D. (2010). Jm: An r package for the joint modelling of longitudinal and time-to-event data. *Journal of Statistical Software (Online)*, 35(9):1–33.
- Rizopoulos, D. (2012a). Fast fitting of joint models for longitudinal and event time data using a pseudo-adaptive gaussian quadrature rule. *Computational Statistics & Data Analysis*, 56(3):491–501.
- Rizopoulos, D. (2012b). *Joint models for longitudinal and time-to-event data: With applications in R*. CRC press.
- Rizopoulos, D. ve Ghosh, P. (2011). A bayesian semiparametric multivariate joint model for multiple longitudinal outcomes and a time-to-event. *Statistics in medicine*, 30(12):1366–1380.
- Rizopoulos, D., Verbeke, G., ve Lesaffre, E. (2009). Fully exponential laplace approximations for the joint modelling of survival and longitudinal data. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 71(3):637–654.

- Rizopoulos, D., Verbeke, G., ve Molenberghs, G. (2008). Shared parameter models under random effects misspecification. *Biometrika*, 95(1):63–74.
- Searle, S. R., Casella, G., ve McCulloch, C. E. (2009). *Variance components*, volume 391. John Wiley & Sons.
- Self, S. ve Pawitan, Y. (1992). Modeling a marker of disease progression and onset of disease. In *AIDS epidemiology*, pages 231–255. Springer.
- Sinharay, S. (2003). Assessing convergence of the markov chain monte carlo algorithms: A review. *ETS Research Report Series*, 2003(1):i–52.
- Song, X., Davidian, M., ve Tsiatis, A. A. (2002). A semiparametric likelihood approach to joint modeling of longitudinal and time-to-event data. *Biometrics*, 58(4):742–753.
- Sweeting, M. J. ve Thompson, S. G. (2011). Joint modelling of longitudinal and time-to-event data with application to predicting abdominal aortic aneurysm growth and rupture. *Biometrical Journal*, 53(5):750–763.
- Tang, A.-M. ve Tang, N.-S. (2015). Semiparametric bayesian inference on skew-normal joint modeling of multivariate longitudinal and survival data. *Statistics in medicine*, 34(5):824–843.
- Tsiatis, A. A. ve Davidian, M. (2001). A semiparametric estimator for the proportional hazards model with longitudinal covariates measured with error. *Biometrika*, 88(2):447–458.
- Tsiatis, A. A. ve Davidian, M. (2004). Joint modeling of longitudinal and time-to-event data: an overview. *Statistica Sinica*, pages 809–834.
- Tsiatis, A. A., Degruottola, V., ve Wulfsohn, M. S. (1995). Modeling the relationship of survival to longitudinal data measured with error. applications to survival and cd4 counts in patients with aids. *Journal of the American statistical association*, 90(429):27–37.
- Ülgen, E. ve Asar, Ö. (2017). Tekrarlı gözlem ve sağkalım verilerinin bileşik modellenmesi. *Türkiye Klinikleri Biyoistatistik*, 9(2):156–166.

- Vetterling, W. T. ve Press, W. H. (1992). *Numerical recipes in Fortran: the art of scientific computing*, volume 1. Cambridge University Press.
- Wang, Y. ve Taylor, J. M. G. (2001). Jointly modeling longitudinal and event time data with application to acquired immunodeficiency syndrome. *Journal of the American Statistical Association*, 96(455):895–905.
- West, B. T., Welch, K. B., ve Galecki, A. T. (2014). *Linear mixed models: a practical guide using statistical software*. CRC Press.
- Williamson, P. R., Kolamunnage-Dona, R., Philipson, P., ve Marson, A. G. (2008). Joint modelling of longitudinal and competing risks data. *Statistics in medicine*, 27(30):6426–6438.
- Wu, L., Liu, W., ve Hu, X. (2010). Joint inference on hiv viral dynamics and immune suppression in presence of measurement errors. *Biometrics*, 66(2):327–335.
- Wulfsohn, M. S. ve Tsiatis, A. A. (1997). A joint model for survival and longitudinal data measured with error. *Biometrics*, pages 330–339.
- Ye, W., Lin, X., ve Taylor, J. M. (2008). Semiparametric modeling of longitudinal measurements and time-to-event data—a two-stage regression calibration approach. *Biometrics*, 64(4):1238–1246.
- Yu, M., Law, N. J., Taylor, J. M., ve Sandler, H. M. (2004). Joint longitudinal-survival-cure models and their application to prostate cancer. *Statistica Sinica*, pages 835–862.

EKLER

EK A : Matris İşlemleri

EK B : Önsel Seçimde Kullanılan Dağılımların Olasılık Yoğunluk Fonksiyonları



10 Matris İşlemleri

10.1 Matrisin Tersinin Türevi

A matrisi $n \times n$ boyutlu tekil olmayan kare bir matris olmak üzere, A matrisinin elemanları t skalerlerinin bir fonksiyonu olsun. A^{-1} matrisinin t skalerinine göre birinci dereceden türevi,

$$\frac{\partial A^{-1}}{\partial t} = -A^{-1} \frac{\partial A}{\partial t} A^{-1} \quad (10.1)$$

şeklinde elde edilir (McCulloch ve Neuhaus, 2005; Searle ve diğ., 2009).

10.2 Matris Determinantının Logaritmasının Türevi

A matrisi $n \times n$ boyutlu tekil olmayan kare bir matris olmak üzere, A matrisinin elemanları t skalerlerinin bir fonksiyonu olsun. A matrisinin determinantının logaritmasının t skalerinine göre birinci dereceden türevi;

$$\frac{\partial \ln|A|}{\partial t} = \text{tr} \left(A^{-1} \frac{\partial A}{\partial t} \right) \quad (10.2)$$

şeklinde elde edilir (McCulloch ve Neuhaus, 2005; Searle ve diğ., 2009).

11 Önsel Seçiminde Kullanılan Dağılımlara Ait Yoğunluk Fonksiyonları

11.1 Çok Değişkenli Normal Dağılım

μ , $p \times 1$ boyutlu ortalama vektörü ve Σ $p \times p$ boyutlu varyans kovaryans matrisi olmak üzere X rastgele değişkeni çok değişkenli normal dağılımı sahip ise, $X \sim MVN(\mu, \Sigma)$ X rastgele değişkeninin olasılık yoğunluk fonksiyonu,

$$p(x | \mu, \Sigma) = (2\pi)^{-p/2} |\Sigma|^{-1/2} \exp \left\{ -\frac{1}{2} (x - \mu)^\top \Sigma^{-1} (x - \mu) \right\}$$

biçimindedir.

11.2 Wishart ve Ters Wishart Dağılımı

M , $p \times p$ boyutlu simetrik pozitif bir matrisi ve $\nu \geq p$ serbestlik derecesini göstermek üzere, $p \times p$ boyutlu simetrik pozitif tanımlı X matrisi Wishart dağılımına sahip ise, $X \sim W(\nu, M)$ X matrisinin olasılık yoğunluk fonksiyonu,

$$p(X | \nu, S) = \frac{1}{2^{\nu p/2} \Gamma_p(\nu/2)} |S|^{-1/2} \times |X|^{(\nu-p-1)/2} \exp \left(-\frac{1}{2} X S^{-1} \right)$$

şeklinde yazılır.

$p \times p$ boyutlu X matrisi ters Wishart dağılımına sahip ise X matrisinin tersi Wishart dağılımına sahiptir. X matrisi, $X \sim IW(\nu, S)$ olasılık yoğunluk fonksiyonu eşitlikteki gibidir.

$$p(X | \nu, S) = \frac{1}{2^{\nu p/2} \Gamma_p(\nu/2)} |S|^{-1/2} \times |X|^{-(\nu+p+1)/2} \exp \left(-\frac{1}{2} X^{-1} S \right)$$

11.3 Tek Düze Dağılım

X sürekli rastlantı değişkeni tekdüze dağılıma sahip ise, X 'in olasılık yoğunluk fonksiyonu aşağıdaki gibidir.

$$f(x) = \frac{1}{b-a}, a \leq x \leq b \\ = 0, \text{diğer durumlar için}$$

ÖZGEÇMİŞ

Ad Soyad: Zeynep ATLI

Doğum Yeri ve Tarihi: Samsun 05/09/1985

E-Posta: zeynep.atli@msgsu.edu.tr

Lisans: İnönü Üniversitesi Ekonometri Bölümü

Y. Lisans: Sinop Üniversitesi İstatistik Programı