

Esin AVCI tarafından hazırlanan BAYESCİ SAĞKALIM ANALİZİ VE MEME KANSERİ VERİLERİ ÜZERİNE BİR UYGULAMA adlı bu tezin Doktora tezi olarak uygun olduğunu onaylarım.


Yrd. Doç. Dr. Meral YAY

Tez Yöneticisi

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

Başkan: : Prof. Dr. Nalan CİNEMRE

Üye : Yrd. Doç. Dr. Meral YAY

Üye : Prof. Dr. Nural BEKİROĞLU

Üye : Prof. Dr. Gülay KIROĞLU

Üye : Doç. Dr. Tuncay CAN

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

ÖZET

Bu çalışmada, açıklayıcı olan ve olmayan önsel bilgiye dayalı Bayesci Sağkalım Analizi tüm yönleriyle irdelenmiş ve yöntemin üstünlüğü klasik parametrik sağkalım analizi ile karşılaştırılarak gösterilmiştir.

Giriş niteliğini taşıyan birinci bölümde sağkalım analizi ile ilgili genel bilgiler ve tanımlar verilmiştir. Bayesci yaklaşımın tanıtıldığı giriş bölümü aynı zamanda Bayesci Sağkalım Analizi için de temel oluşturmaktadır.

Bayesci yaklaşımın ayrıntılı olarak ele alındığı ikinci bölümde, önsel dağılım kavramı ve Bayesci yaklaşımda kullanılan önsel dağılım türleri tanıtılmıştır. Ardından sonsal dağılım kavramı tanıtılarak, Markov Zinciri Monte Carlo (MCMC) ve Metropolis Hasting yöntemleri anlatılmıştır.

Sağkalım analizi ile ilgili temel kavramların tanıtıldığı üçüncü bölümde, parametrik ve yarı parametrik sağkalım modelleri anlatılmış; sağkalım süresinin Üstel, Weibull ve Lognormal dağılıma sahip olduğu parametrik modeller ayrıntısı ile irdelenmiştir. Bayesci yaklaşım kullanılarak üç farklı dağılıma sahip modeller için parametre tahminleri elde edilmiş ve model seçim kriterleri anlatılmıştır.

Tez çalışmasının uygulamasını oluşturan dördüncü bölümde ise, 9 yıl izlemlili meme kanserli 458 hastaya ilişkin veriler incelenmiş ve tüm prognostik faktörler gözönünde bulundurularak seçim yapılmıştır. Veriler üzerinde Klasik Parametrik Sağkalım Analizinin yanısıra açıklayıcı olan ve olmayan önsel bilgiye dayalı Bayesci Sağkalım Analizi uygulanmıştır. Anılan üç yönteme ilişkin parametre tahminleri elde edilmiş ve karşılaştırma yapılmıştır.

Tez çalışmasının son bölümünde, önsel bilgiye dayalı Bayesci yaklaşımla elde edilen sonuçların önemi vurgulanmış ve yapılan çalışmanın bundan sonraki araştırmalar için temel oluşturması hedeflenmiştir.

BAYESIAN SURVIVAL ANALYSIS AND AN APPLICATION ON BREAST CANCER DATA

ABSTRACT

In this study, Bayesian survival analysis with informative and noninformative prior examined in all aspects of and the advantage of the method is shown by compared with the frequentis analysis.

In the first chapter as an introduction, general information and definitions for survival analysis are given. At the same time, this chapter covers Bayesian approach which is the basis for Bayesian survival analysis.

Bayesian approach is discussed in the second chapter, the concept of prior distribution and types of prior distribution used in bayesian approach is introduced. And then the concept of posterior distribution is introduced, Markov Chain Monte Carlo (MCMC) and Metropolis Hasting methods described.

The third section introduced the basic concepts related to survival analysis. Parametric and semi-parametric models are explained. Parametric model of exponentially, weibull and lognormal distributed survival time examined with the details. Also bayesian parameter estimation approach for the three models are obtained and model selection criteria are explained.

The fourth chapter of the thesis that cover application. 9 year follow-up 458 patients with breast cancer data is examined and all the prognostic factors took place considering the elections. Frequentis and Bayesian survival analysis with informative and noninformative prior applied to the data. The parameter estimates are obtained for three method mentioned and are compared.

In the last chapter of the thesis, the importance of the result which obtained by Bayesian approach used prior highlighted and it is aimed to be the basis for the future research.

ÖNSÖZ

Çalışmanın ortaya çıkmasındaki çabası ve özellikle tez formatına dönüştürülmesindeki emeğinden dolayı danışman hocam Yrd. Doç. Dr. Meral YAY'a,

Çalışmanın daha iyi bir dil kullanılarak yazılması için verdiği uğraş ve yapıcı eleştirilerinden dolayı değerli bölüm başkanı Prof. Dr. Nalan CİNEMRE'ye,

Çalışmanın içeriğinin oluşmasında ve kullanılan kaynakların temininde yapıcı eleştirilerini ve desteğini esirgemeyen değerli hocam Prof. Dr. Nural BEKİROĞLU'na,

Beni bir an olsun yalnız bırakmayan ve desteğini esirgemeyen sevgili AİLEME'e sonsuz teşekkürler.

İÇİNDEKİLER DİZİNİ

ÖZET	i
ABSTRACT	ii
ÖNSÖZ	iii
İÇİNDEKİLER DİZİNİ	iv
ÇİZELGELER DİZİNİ	vi
ŞEKİLLER DİZİNİ	vii
KISALTMALAR DİZİNİ	viii
1.GİRİŞ	1
2. BAYESÇİ YAKLAŞIM	5
2.1. Bayes Teoremi	5
2.2. Bayesci Çıkarsama	8
2.2.1. Önsel Dağılım	9
2.2.1.1. Açıklayıcı Olmayan Önsel Dağılım	10
2.2.1.2. Açıklayıcı Önsel Dağılım.....	13
2.2.2. Sonsal Dağılım	13
2.2.3. Markov Zinciri Monte Carlo Yöntemi	15
2.2.3.1. Metropolis Hastings Algoritması	16
2.2.3.2. Gibbs Örnekleme.....	17
3. SAĞKALIM ANALİZİ	20
3.1. Parametrik Modeller.....	28
3.1.1. Olabilirlik Fonksiyonu	28
3.1.2. Üstel Dağılım	29
3.1.2.1. Üstel Dağılıma Sahip Sağkalım Verisine İlişkin Bayesci Yaklaşım	31
3.1.3. Weibull Dağılım.....	36
3.1.3.1. Weibull Dağılıma Sahip Sağkalım Verisine İlişkin Bayesci Yaklaşım	40
3.1.4. Lognormal Dağılım.....	42
3.1.4.1. Lognormal Dağılıma Sahip Sağkalım Verisine İlişkin Bayesci Yaklaşım	44

3.2. Yarı Parametrik Modeller	46
3.2.1. Kısmi Olabilirlik	47
3.2.2. Oransal Hazard Modeli (OHM)	47
3.2.2.1. Oransal Hazard Modeli için Bayesci Yaklaşım	48
3.3. Model Parametresinin Anlamlılığı	48
3.4. Model Seçim Yöntemleri	49
3.4.1. Sapma İstatistiği (-2logl)	49
3.4.2. Akaike Bilgi Kriteri (AIC)	50
3.4.3. Bayesci Bilgi Kriteri (BIC)	50
4. UYGULAMA.....	51
4.1. Kullanılan Programlar	51
4.2. Tanımlayıcı İstatistikler ve Klasik Parametrik Sağkalım Analizi.....	51
4.3. Bayesci Sağkalım Analizi (BSA).....	58
5. SONUÇ VE ÖNERİLER.....	72
KAYNAKLAR	76
ÖZGEÇMİŞ.....	82

ÇİZELGELER DİZİNİ

Çizelge 2.1. Farklı dağılımlar için Jeffrey önselleri.....	12
Çizelge 2.2. Farklı dağılımlar eşlenik önseller.....	13
Çizelge 4.1. Üç kategorili değişkene ait kukla değişkenler	52
Çizelge 4.2. Meme kanserli hastalara ilişkin değişkenlerin dağılımı.....	53
Çizelge 4.3. Lognormal dağılım için elde edilen uyum iyiliği test sonuçları	56
Çizelge 4.4. Logaritması alınan verilerin Kolmogorov-Smirnov Normallik testi sonuçlar	56
Çizelge 4.5. Meme kanserli hastaların sağkalım sürelerine ilişkin KPSA sonuçları.....	57
Çizelge 4.6. Anlamli bulunan parametrelere göre elde edilen KPSA sonuçları	58
Çizelge 4.7. Meme kanserli hastalara ilişkin bağımsız değişkenlerin önsel bilgileri	61
Çizelge 4.8. Meme kanserli hastaların sağkalım sürelerine ilişkin BSA-I sonuçları	62
Çizelge 4.9. Anlamli bulunan parametrelere için elde edilen BSA-I sonuçları	63
Çizelge 4.10. Meme kanserli hastaların sağkalım sürelerine ilişkin BSA-II sonuçları.....	66
Çizelge 4.11. Anlamli bulunan parametreler için BSA-II sonuçları.....	67
Çizelge 4.12. Meme kanserli hastaların sağkalım sürelerine ilişkin anlamli bulunan değişkenlerin KPSA, BSA-I ve BSA-II yöntemlerine göre elde edilen uyum istatistikleri.....	70

ŞEKİLLER DİZİNİ

Şekil 3.1. Farklı zamanlarda analize dahil olan hastaların çalışmaya giriş süreleri gösteren grafik.....	21
Şekil 3.2. Farklı zamanlarda analize dahil olan hastaların çalışmaya giriş sürelerinin başlama noktasına çekildiğini gösteren grafik.....	22
Şekil 3.3. Üstel dağılıma ait olasılık yoğunluk fonksiyonunun grafiği	30
Şekil 3.4. $\lambda = 1$ ve $\lambda = 1/2$ için, Üstel dağılımın sağkalım fonksiyonunun grafiği...	30
Şekil 3.5. Üstel dağılımın Hazard fonksiyonunun grafiği.....	31
Şekil 3.6. Farklı γ şekil ve $\alpha = 1$ ölçek parametrelili Weibull dağılımının olasılık yoğunluk fonksiyonunun grafiği.....	37
Şekil 3.7. $\alpha = 1$ Ölçek parametrelili Weibull dağılımının $\log S(t)$ eğimi.....	37
Şekil 3.8. $\alpha = 1$ için Weibull dağılımının Hazard fonksiyonunun grafiği.....	38
Şekil 3.9. Sırasıyla $\mu = 0$ için değişen σ^2 ve $\sigma^2 = 0.5$ için değişen μ değerleri için Lognormal olasılık yoğunluk fonksiyonlarının grafiği.....	42
Şekil 3.10. Lognormal dağılımın Hazard fonksiyonunun grafiği	43
Şekil 4.1 Ölen meme kanserli hastaların sağkalım sürelerine ilişkin dağılımın olasılık yoğunluk fonksiyonunun grafiği.....	54
Şekil 4.2. Ölen meme kanserli hastaların sağkalım Hazard fonksiyonunun grafiği ..	55
Şekil 4.3. BSA-I'e göre anlamlı bulunan parametrelere için elde edilen iz grafikleri	64
Şekil 4.4. BSA-I'e göre anlamlı bulunan parametreler için Kernel yoğunluk grafikleri	65
Şekil 4.5. BSA-II'ye göre anlamlı bulunan parametreler için iz grafikleri.....	68
Şekil 4.6. BSA-II'ye göre anlamlı bulunan parametreler için Kernel yoğunluk grafikleri.....	69
Şekil 4.7. KPSA, BSA-I ve BSA-II'nin parametre tahmini ve standart hatalarının grafiği.....	71

KISALTMALAR DİZİNİ

BSA	: Bayeſçi Saękalım Analizi
MCMC	: Markov Zinciri Monte Carlo
KPSA	: Klasik Parametrik Saękalım Analizi
EYSYA	: En Y¼ksek Sonsal Yoęunluk Aralıęı
MC	: Markov Zinciri
AIC	: Akaike Bilgi Kriteri
KM	: Kaplan-Meier
OHM	: Oransal Hazard Modelleri
OH	: Oransal Hazard
HF	: Hızlandırma Fakt¼r¼
BSA-I	: Açıklayıcı Önsel Bilgiye Dayalı Bayeſçi Saękalım Analizi
BSA-II	: Açıklayıcı Olmayan Önsel Bilgiye Dayalı Bayeſçi Saękalım Analizi

1.GİRİŞ

Sağkalım en kısa tanımıyla yaşamda kalma süresidir. Daha genel bir ifade ile belirli bir başlangıç değerinden (doğum, tedavi başlangıcı veya işverenin iş vermesinden) belirli bir olguya ulaşmaya (ölüm, hastalığın nüks etmesi veya işten çıkarılma) kadar geçen süre olarak tanımlanır. Sağkalım bir süreç olduğu için verilerin negatif değer almayacağı açıktır. Sağkalım verilerinin analizi ve modellenmesi, 17. yy da aktüerya ve demografinin gelişmesiyle şekillenmeye başlamıştır. Sağkalım analizi konusundaki ilk önemli gelişme, Kaplan ve Meier'in 1958 yılında yayımlanan "sağkalım eğrisinin tahmini" makalesiyle kendini göstermiştir (Kaplan ve Meier, 1958). Bu çalışmanın ardından, sağkalım süresine etki eden risk faktörlerinin belirlenmesi amacıyla geliştirilen yeni yöntemler araştırmacıların ilgisini çekmiş ve araştırmaların sayısı gün geçtikçe artmıştır. Son araştırmalar parametrik sağkalım analizinin yanı sıra yarı parametrik ve Bayesçi sağkalım analizi üzerinde odaklanmıştır.

Parametrik sağkalım analizi, sağkalım süresinin belirli bir dağılıma sahip olmasını gerektirir. Bu varsayımın sağlanması halinde sağkalıma etki eden faktörlere ait parametrelerin tahmini gerçekleştirilebilir. (Feigl ve Zelen, 1965). Sağkalım verisi modele uygun bir parametrik dağılıma sahip olduğunda; parametrik sağkalım analizi, yarı parametrik sağkalım analizine oranla avantajlı duruma gelir. (Richard ve diğ., 1998). Ayrıca parametrik sağkalım analizi, olgunun gerçekleşme hızına (sağkalım zamanının artma veya azalma oranı) ilişkin oran tahminininin yapılmasına da yardımcı olur.

1972 yılında David Cox tarafından ortaya atılan yarı parametrik sağkalım analizi, oransal riskin (sağkalım süresine etki eden prognostik faktörlerin, Cox regresyon yöntemi uygulanmasıyla elde edilen katsayıların birbirine göre oranlarıdır) hesaplanmasında kullanılan bir yöntemdir. Paket programlardaki uygulama kolaylığı ve sağkalım dağılımlarına uyma zorunluluğunun olmaması, Cox regresyonun kullanımını yaygınlaştırmıştır ancak uzun takip süreleri için oransal Hazard

varsayımını sağlanmaması durumunda uygulanması tavsiye edilmez (Valsecchi ve diğ., 1996). Yapılan son arařtırmalarda Cox regresyon kullanılarak elde edilen parametre tahminlerinin, parametrik sađkalım analizinden elde edilen tahminlere gre yetersiz olduđu da belirlenmiřtir.

Son yıllarda klasik yntemlerdeki kısıtlamalar Bayesci yaklařıma olan ilgiyi arttırmıřtır. Bayesci yaklařım subjektif dřncenin temel tařı olarak kabul edilen Bayes teoremine dayanarak geliřtirilmiřtir ve yaklařıma gre parametreler klasik yaklařımdaki gibi sabit olarak deđil, olasılıđa bađlı olarak tanımlanır. Dolayısı ile her bir parametreye iliřkin bir dađılım sz konusudur. Bu olasılıklar “kanaat derecesi (degrees of belief)” olarak tanımlamaktadır. Bir bařka deyiřle parametreler rasgele deđiřken olarak ele alınmaktadır. “Bayesci” kelimesi de, parametre tahminleri iin yapılan ıkarsamalarda Thomas Bayes’in teoreminden faydalanılmasından dođmuřtur (Ibrahim ve diğ., 2001; SAS Institute, 2006).

Bayesci yaklařım, karmařık veriyi modellemede nsel (prior) bilgiye bařvurma esnekliđi nedeniyle klasik yntemlere gre olduka avantajlıdır (Ibrahim ve diğ., 2001; Wong ve diğ., 2005). nsel bilginin elde edilmesi, Bayesci ıkarsamada nemli rol oynar. Bu bilgi, daha nce yapılmıř alıřmalardan elde edilen bilgiler, gemiř deneyimler ya da harici bir kanıt gibi bađımsız kaynaklar (aıklayıcı nsel bilgiler) ıřıđında, gzlenen verinin subjektif yorumuna dayanır ve buradan hareketle elde edilen yeni bilginin bileřimi olan sonsal dađılımla aıklanır (Congdon, 2006; Wong ve diğ., 2005).

2001 yılında yayımlanan Ibrahim, Chen ve Sinha’nın “Bayesian Survival Analysis” adlı kitabının ardından Bayesci Sađkalım Analizi (BSA) daha yaygın kullanılmaya bařlamıřtır. Bu konudaki en nemli rnek alıřmalar ařađıdaki gibi sıralanabilir:

- Wong ve arkadaşlarının, okul ncesi inli ocukların aktif diř rmesini durdurmada, silver diamin florid ve sodyum florid parlatıcılarının etkinliđini arařtırmaya ynelik yaptıkları klinik alıřmadan elde edilen ve diř rmesi gerekleřmiř veri seti zerindeki uygulaması (Wong ve diğ., 2005)

- Calle ve arkadaşlarının, yağsız ve tam yağlı yoğurt ürünlerinin raf ömürlerini tahmin etmede açıklayıcı olan ve olmayan önsel bilgiye dayanarak yaptıkları uygulama (Calle ve diğ., 2006).
- Yin ve İbrahim'in, cilt kanseri ile ilgili gerçek bir veri seti üzerindeki uygulaması. (Yin ve İbrahim, 2006)
- Lin, 2010 yılında yaptığı doktora tez açılışmasında log-lineer medyana sahip regresyon modeli için hem benzetim hem de gerçek veri seti üzerine uygulama yaparak, Cox modeline alternatif yarı parametrik model geliştirmiştir.
- Kurt ve arkadaşları, 2011 yılında Weibull regresyon modelinin, klasik ve Bayesci yaklaşımla karşılaştırmak amacıyla benzetim çalışması (Kurt ve diğ., 2011).
- Ahmed ve İbrahim, 2011 yılında Weibull dağılıma sahip sağkalım fonksiyonunu Jeffrey önsel dağılımı kullanılarak Bayesci ve En Çok Olabilirlik Tahminlerinin karşılaştırma çalışması (Ahmet ve diğ., 2011).
- Abu-Taleb, Smadi ve Alawneh, 2007 yılında Üstel dağılıma sahip sağkalım verilerinin Bayes parametre tahmini, marjinal sonsallar ve güvenilirlik aralığı eşlenik önsel dağılım kullanılarak elde edilen çalışması (Abu-Taleb ve diğ., 2007).

Bayesci analizde, karmaşık modeller önsel dağılımla birleştirildiğinde, işlem yükü fazla olan analizler ile karşılaşılır. Ancak Bayesci analizlerin uygulanması Markov Zinciri Monte Carlo (MCMC) yöntemleri yardımıyla oldukça kolaylaşmaktadır (Gilks ve diğ., 1996). MCMC, temelde Markov zinciri kullanan Monte Carlo integrasyonudur. Monte Carlo integrasyonu ilgilenilen dağılımlardan örneklem seçer ve örneklem ortalamasını elde eder ve bu noktada çeşitli özelliklere sahip Markov zincirini kullanır. Yöntem ücretsiz olarak indirilebilen WinBUGS paket programı yardımıyla kolaylıkla uygulanmaktadır. BUGS (Spiegelhalter ve diğ., 1996), özel olarak MCMC yönteminin tam olasılık modellerine uygulanması için

kullanılan bir programdır. Bu programda tüm parametreler rasgele deęişken olarak ele alınmaktadır. Markov Zinciri elde etmede öne çıkan yöntemlerden bazıları Metropolis Hastings algoritması ve Gibbs örneklemesidir.

Bayesci yaklaşımın sağkalım analizinin son yıllarda yaygın kullanılmasının en önemli sebeplerinden biri, sansürlü (censored) veriler için sonsal dağılımın hesaplamasında Gibbs örneklemesinin kullanılabilmesidir. Gibbs örneklemesi ilk olarak Brooks (1998), Geman ve Geman (1984) tarafından uygulanmıştır. Gibbs örneklemesinin en önemli avantajı, karmaşık hesaplamalara getirdiği kolaylıktır.

Bu çalışmada, açıklayıcı olan ve olmayan önsel bilgiye dayalı Bayesci Sağkalım Analizi (BSA) tüm yönleriyle irdelenmiş, ardından klasik yöntemle karşılaştırması yapılmıştır. Bu amaçla 9 yıl izlemlı meme kanseri verilerine etki eden prognostik faktörler incelenmiş ve açıklayıcı olmayan önsel bilgiye dayalı BSA'nin, klasik yöntemle elde edilen sonuçlarla benzerlik gösterdiği görülmüştür. Açıklayıcı önsel bilgiye dayanan sağkalım analizinin ise hem klasik hem de açıklayıcı olmayan önsel bilgiye dayalı Bayesci sağkalım analizinden daha duyarlı parametre tahminleri verdiği de yapılan çalışmada ortaya konulmuştur.

2. BAYESCI YAKLAŞIM

İstatistiksel analizlerde kullanılan birçok yaklaşım gibi, Bayes yaklaşımının kullanım alanını da gün geçtikçe genişlemektedir. 18. yy sonlarında Thomas Bayes tarafından ortaya atılan Bayes teoremi, bu yaklaşımın temelini oluşturmaktadır.

2.1. BAYES TEOREMİ

Olasılığın ilk subjektif tanımı, 18. yy sonlarında Thomas Bayes tarafından ortaya konmuştur. 19. yüzyılın sonu ve 20. yüzyılın ilk yarısında klasik istatistiğin altın çağını yaşaması, Bayes teoremini gölgede bırakarak gelişimini yavaşlatsa da 21. yüzyılın başlarından itibaren Bayesci yaklaşıma ait hesaplama tekniklerinin bilgisayar yazılımlarının geliştirilmesiyle, teorem tekrar yaygın olarak kullanılmaya başlanmıştır. Bayesci yaklaşımın klasik yaklaşımdan en önemli farkı, parametrelerin raslantı değişkeni olarak analize dahil edilmesidir. Raslantı değişkenler olarak tanımlanan bu önsel dağılımlar, önceki çalışmalardan ele edilmiş bilgiyi içermektedir. Ancak klasik yaklaşımı benimseyen araştırmacılar; parametreler için seçilen önsel dağılımların subjektifliği ve hatalı sonuçlara yol açtığı gerekçesiyle Bayesci yaklaşımdan uzak durmuşlardır.

Bayesci yaklaşımda, önsel dağılımdan ve veriden gelen bilgi birleştirilerek sonsal dağılım elde edilir. Sonsal bilginin elde edilmesindeki en önemli sorun, dağılımın hesaplama güçlüğüdür. Karmaşık dağılımların analitik olarak çözümünün oldukça zor olması, bilgisayar yazılımlarını daha hızlı gelişmeye itmiş ve bu şekilde çözüm zorluğunu ortadan kaldırmıştır. Analitik yöntemler ile elde edilemeyen bu tür problemlerin çözümünde kullanılan yakınsama tekniği, belirli problemlerde sonuca ulaşılmasını sağlasa da karmaşık problemlerin çözümünde yetersiz kaldığı gözlemlenmiştir. Ancak 1990 yılında Gelfand ve Smith Markov Zinciri Monte Carlo (MCMC) yöntemini ortaya koymaları ile geleneksel yöntemlerle tanımlanamayan ve çözülemeyen problemler ortadan kalkmış, ayrıca karmaşık modellerin tahmini yapılabilmektedir (Ntzoufras, 2009).

Bayes teoremi, koşullu olasılıkların hesaplanmasında kullanılan matematiksel bir formüle dayanmaktadır. Koşullu olasılık klasik olarak şu şekilde tanımlanabilir: A ve B iki olay olmak üzere; A olayının gerçekleşmesi durumunda B olayının meydana gelme olasılığı; B olayının koşullu olasılığı olarak isimlendirilir. B olayının oluşmasında ikiden fazla A olayının ($A = A_1 \cup \dots \cup A_n$ ve $A_i \cap A_j = \phi$) etkisinin olması durumunda B olayının koşullu olasılığı, $P(B|A_i)$ ile gösterilir. Aşağıdaki gibi hesaplanır.

$$P(B|A_i) = \frac{P(A_i \cap B)}{P(A_i)} \quad (2.1)$$

Burada $P(A_i)$, birbirinden bağımsız A_i olaylarının gerçekleşme olasılığı; $P(A_i \cap B)$, A_i ve B olaylarının birlikte gerçekleşme olasılığıdır. Klasik koşullu olasılığa göre; A_i olaylarının hangi olasılıklar ile gerçekleştiği bilinmekte, A_i olaylarının hangi olasılıklar ile B olayını gerçekleştirdiği bilinmemekle birlikte A_i olaylarının gerçekleşmesi durumunda B olayının gerçekleşme olasılığı hesaplanmak istenmektedir. Klasik tanımdan farklı olarak Bayesci yaklaşımda ise, ikiden daha fazla A olayının olması durumunda, B olayının koşullu olasılığı;

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{P(B)} = \frac{P(B|A_i)P(A_i)}{\sum_{i=1}^n P(B|A_i)P(A_i)} \quad (2.2)$$

ile verilmektedir. Burada $P(B|A_i)$ ve $P(A_i)$, yukarıda ifade edildiği gibidir. $P(B)$, B olayının gerçekleşme olasılığıdır. Klasik tanımdan farklı olarak (2.2) eşitliğindeki $P(A_i|B)$, bir etki olasılığını ifade etmektedir. Bu etki olasılığı, her bir A_i olayının hangi olasılık ile B olayını gerçekleştirdiğini gösterir (Ntzoufras, 2009). Daha açık bir ifade ile Bayesci yaklaşımda koşullu olasılık, A_i olaylarının hangi olasılıklar ile gerçekleştiği ve A_i olaylarının hangi olasılıklar ile B olayını gerçekleştirdiği bilinmekte, buna bağlı olarak B olayının gerçekleşmesini sağlayan A_i olaylarının hangi olasılıklar ile B olayına etki ettiği hesaplanmak istenmektedir.

Klasik yaklaşımda, verilerin gerekli varsayımları sağlaması halinde istatistiksel model kurulur ve bu modele ait parametre tahminleri yapılır. Önsel bilgiyi kullanan Bayesci yaklaşımda ise, klasik yaklaşımın aksine; model kurulumu, varsayım gerektirmeden (2.2)'de verilen koşullu olasılığa dayandırılır. Bayesci yaklaşıma göre model açıklandığında; $P(A_i)$, A olayının önsel dağılımı; $P(B|A_i)$ ise olayın olabilirlik fonksiyonu olarak tanımlanır. Bu iki bilginin birleştirilmesi sonucunda oluşan $P(A_i|B)$ ise sonsal dağılımı ifade eder ve parametre tahmini sonsal dağılım yardımıyla yapılır. A ve B olaylarının istatistiksel bir çalışmada karşılığı sırasıyla y ve θ olarak ifade edildiğinde, Bayes yaklaşımı aşağıdaki gibi ifade edilir:

$$f(\theta|y) = \frac{f(y|\theta)f(\theta)}{f(y)} \quad (2.3)$$

Burada $f(\theta)$ önsel dağılımı, $f(y|\theta)$ olabilirlik fonksiyonunu ifade etmektedir. $f(y)$, y 'nin marjinal dağılımıdır ve θ 'ya göre integral alındığı için bu parametreyi içermez. Aynı zamanda $f(y)$, sonsal dağılımı (sürekli veriler için integral, kesikli veriler için toplam alınarak) 1'e eşitleyen "normalleştirme" katsayısı olarak adlandırılır. Bayesci yaklaşımda amaç, θ 'nın sonsal dağılımını elde etmek olduğundan bu terim çoğu kez ihmal edilebilir. Bu durumda sonsal dağılım, yaklaşık olarak aşağıdaki gibi ifade edilebilir:

$$f(\theta|y) \propto f(y|\theta)f(\theta) \quad (2.4)$$

Çalışmadan elde edilen y_i ($i=1, \dots, n$) gözlemlerine ait olabilirlik fonksiyonu,

$$f(y|\theta) = \prod_{i=1}^n f(y_i|\theta) \quad (2.5)$$

eşitliği ile ifade edilir. Klasik yaklaşımda parametre uzayında bir sabit olan ancak değeri bilinmeyen θ 'nın tahminine, (2.5)'te verilen olabilirlik fonksiyonunun maksimize edilmesiyle ulaşılır. Bir başka deyişle bu tahmine, veriyi özetlemede kullanılan örneklem ortalaması, varyansı, basıklık ve çarpıklık katsayısı gibi basit

istatistiksel ölçüler yardımıyla ulaşmak mümkün değildir. Ayrıca birçok araştırmada zaman ve maliyet kısıtları nedeniyle küçük örneklerle çalışma zorunluluğu olduğundan, parametre tahminleri için model kurma gerekliliği ortaya çıkar (Menteşe, 2011). Gözlem sayısının yetersiz olduğu veri kümesinden daha duyarlı sonuçlar elde edebilmek amacıyla, bilinmeyen parametre değerine ait var olan tüm bilgi bir olasılık modeline dayandırılır ve tahmin sürecinde kullanılır. Bu düşünce Bayesci yaklaşımın doğmasına ve gelişmesine neden olmuştur.

2.2. BAYESCİ ÇIKARSAMA

Bayesci yaklaşıma göre model kurmadaki amaç, bilinmeyen parametre değerlerinin tahmin edilmesidir. Jaynes (1985)'e göre; Bayesci parametre tahmininde hem önsel hem de sonsal dağılım, parametrenin herhangi bir ölçülebilir özelliğinin yanı sıra, parametre hakkında kişilerin sahip olduğu bilgilerin düzeyini temsil etmektedir. Bu noktada Bayesci istatistikte bilinmeyen parametre değerlerinin sabit değerler olarak değil, birer raslantı değişkeni olarak ele alındığını belirtmek gerekir. Parametrelere olan bu bakış farklılığı, klasik ve Bayesci yaklaşımlar arasındaki temel ayırımın en önemli nedenidir. Bayesci yaklaşımda bilgiler olasılığa dayalı olarak ifade edildikleri için, her bir parametreye ilişkin bir dağılımın varlığı söz konusudur. Veriler gözlemlenmeden önce kişinin, özne bilgisi dahilinde parametreye atadığı dağılım önsel dağılım olarak adlandırılır ve bu dağılımdan hareketle sonsal dağılıma ulaşılır. Çıkarsama sonunda parametreye ait tüm bilgiler, sonsal dağılım yardımıyla elde edilir.

Bayesci yaklaşımın klasik yaklaşımdan daha kullanışlı ve üstün olduğu bazı durumlar kısaca aşağıdaki gibi özetlenebilir.

- Klasik yaklaşımda birçok teorem bulunmaktadır. Bayesci yaklaşım ise sadece Bayes teoremine dayanır. Dolayısıyla ile Bayesci yaklaşım çıkarsama yapmak için daha kolay bir yol izler.
- Klasik yaklaşımda parametre tahmini için önsel bilgi ihmal edilirken Bayes yaklaşımda tahmin önsel bilgiye dayanarak yapılır. Bu noktada Bayesci yaklaşım üstünlük sağlamaktadır.
- Klasik yaklaşımda parametre için bir aralık tahmin yapılırken, Bayesci yaklaşımda sonsal dağılım yardımıyla parametre tahmininin olasılığını elde

edilir. Bayesci yaklaşımda tahminin bir aralık olarak değil doğrudan hesaplanabilmesi uygulamada avantaj sağlar.

- Klasik yaklaşımda karmaşık modellerin parametrelerini tahmin etmek oldukça zordur. Bayesci yaklaşım ise parametre tahmininde marjinal sonsal dağılımı kullandığından, parametre tahminlerini elde etmek daha kolaydır.
- Klasik yaklaşımda maliyet ve zaman tasarrufu sağlamak için küçük örneklem tercih edilir ve bu nedenle parametre tahminlerinin güvenilirliği yitirilebilir. Ancak Bayesci yaklaşımda küçük örneklem ile güvenilir sonuçlar elde etmek mümkündür.
- Klasik yaklaşımda, parametre tahminin yansızlık ve tutarlılık gibi özellikleri sağlaması gerekmektedir. Bayesci yaklaşımda ise parametre tahmini bir olasılık olduğundan bu özellikleri sağlamasına gerek yoktur.
- Bayesci yaklaşımda uygun önsel dağılımlar kullanılarak elde edilen varyans, klasik yaklaşımla elde edilenden daha küçük ve aralık tahminleri de daha dardır. Sonuç olarak Bayesci yaklaşımla daha etkin tahminlere ulaşmak mümkündür (Press, 1989; Demirhan, 2004).
- Bayesci yaklaşımda farklı zaman dilimlerinde toplanmış veriler, parametrenin sonsal dağılımının oluşturulmasında kullanılabilir. Bu durum yaklaşımın esnekliği konusunda araştırmacıya yardımcı olur.

2.2.1. Önsel Dağılım

Bayesci modeller oluşturulurken yapılacak ilk iş önsel dağılımın belirlenmesidir (Gelman ve diğ., 2004). Önsel dağılım, benzer çalışmalardan elde edilmiş veya uzman görüşleri doğrultusunda oluşturulan dağılımlardır. Önsel dağılımın seçiminde yapılacak hata yanlış çıkarımlara neden olabilir. Ayrıca aynı model için seçilen farklı önsel dağılımların farklı sonsal dağılımlara yol açacağı aşikardır. Birçok durumda hangi önsel dağılımın daha uygun olacağını kestirmek zordur. Böyle bir durumda önsel dağılımın sonuçları ne kadar değiştireceği model seçim kriterleri ile tespit edilmelidir.

Bayesci yaklaşımın en genel özelliği, örneklem büyüklüğü arttıkça önsel dağılımın sonsal dağılım üzerindeki etkisinin azalmasıdır. Bu durumda olabilirlik

fonksiyonunun dağılım özelliği, önsel dağılımın özelliğini bastırır. Örnek vermek gerekirse, olabilirlik fonksiyonunun dağılımı sivri ve önsel dağılımı daha basık ise, sonsal dağılım sivri özellik gösterecektir. Bu nedenle önsel dağılımın sonsal dağılıma katkısı olması gerekenden daha az olacaktır. (Box ve Tiao, 1973).

Önsel dağılımlar, açıklayıcı olan (informative) ve açıklayıcı olmayan (noninformative) olmak üzere iki temel gruba ayrılır. 1973 yılında Box ve Tiao açıklayıcı olan önsel dağılımları, veri seti hakkında bilgi veren, açıklayıcı olmayan önsel dağılımları ise veri seti hakkında çok az bilgi veren önsel dağılımlar olarak tanımlamışlardır (Congdon, 2003). Model parametresi hakkında herhangi bir bilginin olmaması durumunda sonsal dağılım üzerinde en az etkiyi yapacak olan açıklayıcı olmayan önsel dağılımların kullanılması tercih edilmelidir.

Önsel dağılımlara ait diğer bir ayırım ise uygun (Proper) ve uygun olmayan (improper) önsel dağılım sınıflamasıdır. Tanım aralığında integral ya da toplamı alındığında “1” değerini veren sürekli ya da kesikli bir önsel bir dağılım uygun önsel dağılım, vermiyorsa uygun olmayan (improper) önsel dağılım olarak adlandırılır. Uygulamada kullanılan önseller genellikle uygun olmayan önsel dağılımlardır. Önsel dağılımlar ile ilgili dikkat edilmesi gereken en önemli nokta, kullanılan önsel dağılım uygun olmayan önsel olması halinde bile sonsal dağılımın mutlaka uygun olması gerektiğidir.

2.2.1.1. Açıklayıcı Olmayan Önsel Dağılım

Açıklayıcı olmayan önsel dağılım, çalışmalarda en yaygın kullanılan önsel dağılım türüdür ve parametrenin tanımlı olduğu aralık bilgisi dışında herhangi bir bilginin olmaması durumunda kullanılır. Bayesci yaklaşımda açıklayıcı olmayan önsel dağılım kullanıldığında, sonuçlar klasik yaklaşımla elde edilen sonuçlarla örtüşmektedir. Bunun nedeni, parametre tahmininde elde edilen verinin sağladığı bilgi dışında farklı bir bilgi kullanmamasıdır.

Açıklayıcı olmayan önsel dağılımın aralığının sınırlarından en az biri sonsuz olursa, normalleştirme katsayısı aşağıdaki gibi belirsiz bir integrale eşit olur ve hesaplanamaz.

$$f(y) = \int f(\theta) d\theta = \infty \quad (2.6)$$

Bu durumda açıklayıcı olmayan önsel aynı zamanda da uygun olmayan bir önsel dağılım olmaktadır. Başlıca açıklayıcı olmayan önsel dağılımlar, yaygın/dağınık (diffuse), zayıf (weak), etkisiz (vague), düzgün (flat) ve eşlenik (Conjugate) gibi adlarla nitelendirilen önsel dağılımlardır (Box ve Tiao, 1973; Yardımcı, 1992; Gelman ve diğ., 2004; Demirhan, 2004; Christensen ve diğ., 2011).

Uygulamalarda yoğun olarak kullanılan açıklayıcı olmayan önsel dağılımlar aşağıda verilmiştir.

Düzcün Önsel Dağılım

Düzcün önsel dağılım, en çok kullanılan açıklayıcı olmayan önsel dağılımdır ve Bayes-Laplace'ın "yetersiz neden prensibi"ne dayanır. Bu prensip elde kanıt olmadığında tüm olaylara aynı önsel olasılık değerini verir (Congdon, 2003). Örneğin bir olay için K tane durum söz konusu ise, bu prensibe göre her bir durumun önsel olasılığı düzcün dağılım ile $\frac{1}{K}$ olur.

Jeffrey Önsel Dağılımı

1961 yılında Jeffrey tarafından ortaya atılan ve açıklayıcı olmayan önsel dağılım seçimi için kullanılan bir kuralı içerir. Bu kurala göre önsel dağılım orantısal olarak, Fisher bilgi matrisinin determinantının kareköküne eşittir ve aşağıdaki şekilde elde edilir.

$$f(\theta) \propto |J(\theta)|^{1/2} \quad (2.7)$$

Bir tahmine ait en çok olabilirlik tahmin edicisinin duyarlılığını ölçer ve aşağıdaki Fisher bilgi matrisine dayanılarak açıklanır:

$$J(\theta) = -\int \left[\frac{d^2 \log f(y|\theta)}{d\theta^2} \right] f(y|\theta) dy \quad (2.8)$$

Jeffrey önseli yardımıyla herhangi parametrik bir model için açıklayıcı olmayan önsel dağılım elde edilebilir. Farklı dağılımlar için elde edilen Jeffrey oranları aşağıdaki çizelgede verilmiştir (Gill, 2008).

Çizelge 2.1. Farklı dağılımlar için Jeffrey önselleri

<i>Olabilirlik Fonksiyonu</i>	<i>Parametre(ler)</i>	<i>Önsel Dağılım</i>
Bernolli	$\theta = p$	$(p(1-p))^{-1/2}$
Binom	$\theta = p$	$(p(1-p))^{-1/2}$
Poisson	$\theta = \lambda$	$\lambda^{-1/2}$
Normal (σ^2 biliniyor)	$\theta = \mu$	1
Normal (μ biliniyor)	$\theta = \sigma^2$	σ^{-1}
Normal	$\theta = \mu, \sigma^2$	σ^{-1}

Eşlenik (Conjugate) Önsel Dağılım

Yapılan araştırmalarda sonsal dağılımın analitik olarak elde edilmesi hedeflenir. Bu durumda önsel dağılım ile sonsal dağılımın aynı dağılım ailesinden (eşlenik) olması araştırmacılara uygulama kolaylığı sağlar. Eşlenik önsel dağılımlar aynı zamanda ek veri olarak da algılanabilir (Gelman ve diğ., 2004). Eşlenik dağılımlara ilişkin çizelge aşağıdaki gibidir:

Çizelge 2.2. Farklı dağılımlar eşlenik önseller

<i>Olabilirlik Fonksiyonu</i>	<i>Parametre(ler)</i>	<i>Önsel Dağılım</i>	<i>Sonsal Dağılım</i>
Bernolli	p	Beta	Beta
Binom	p	Beta	Beta
Poisson	λ	Gamma	Gamma
Üstel	λ	Gamma	Gamma
Normal (σ^2 biliniyor)	μ	Normal	Normal
Normal (μ biliniyor)	σ^2	Ters Gamma	Ters Gamma
Normal	μ, σ^2	Normal-Gamma	Normal-Gamma
Tek biçimli	a,b	Pareto	Pareto
Gamma	α, β	Gamma	Gamma

Uygulamalarda karmaşık modeller için eşlenik önsel dağılım kullanımı mümkün olmayabilir. Bu tür durumlarda eşlenik olmayan (nonconjugate) önseller kullanılır (Gelman ve diğ., 2004).

2.2.1.2. Açıklayıcı Önsel Dağılım

Açıklayıcı önsel dağılım, olabilirlik fonksiyonu tarafından etkisi azaltılmayan önsel dağılımdır. Bu tip önseller deneyimlerden, benzer geçmiş çalışmalardan ve uzman görüşlerinden elde edilen bilgi çerçevesinde belirlenirler. Sonsal dağılım üzerinde oldukça etkili oldukları için önsel dağılım belirlenirken çok dikkatli olunmalıdır. Ayrıca açıklayıcı önsel dağılımın matematiksel yapısının da çok karmaşık olmaması gerekir. Aksi halde elde edilen tahminler yanıltıcı olabilir. Veri yapısına uygun ve gerçeği yansıtan önsel dağılım ile çalışıldığında ise Bayesci yöntemin performansı oldukça artmaktadır. Önceki çalışmaların meta analizinin yapılması, verinin histogramının çizilmesi veya kesikli önsel dağılımın kullanılması yardımıyla açıklayıcı önsel dağılımları belirlemek mümkündür.

2.2.2. Sonsal Dağılım

Tahmin edilecek parametre hakkında tüm bilgiyi içeren sonsal karakteristiklerini ortaya koyan istatistikler ortalama, mod, medyan olarak sıralanabilir. Sonsal

ortalama, sonsal mod ve sonsal medyan hesaplanması sırasıyla aşağıdaki eşitlikler ile elde edilir:

$$\bar{\theta} = \int \theta f(\theta|y) d\theta \quad (2.9)$$

$$\hat{\theta} = \arg \max_{\theta} f(\theta|y) \quad (2.10)$$

$$\text{medyan} = \int_{-\infty}^{\tilde{\theta}} f(\theta|y) d\theta = 0.5 \text{ integralini sağlayan } \tilde{\theta} \text{ dır.} \quad (2.11)$$

Bayesci çıkarsama süreci, $f(\theta)$ önsel dağılımından $f(\theta|y)$ sonsal dağılımına geçiş sürecidir. Bayesci yaklaşımda sonsal dağılımın özetlenmesinde kullanılan diğer bir istatistik ise Bayes güven aralığıdır. Klasik yaklaşımda güven aralığı, aralığın parametreyi içermesi olasılığı olarak yorumlanırken, Bayesci yaklaşımda bu yorum parametrenin aralığa düşmesi olasılığı olarak değişmektedir. Bayesci yaklaşımda aralığın içinde bulunan her noktanın sahip olduğu olasılık yoğunluk değeri, aralık dışında bulunanlardan daha büyük olduğunda elde edilen aralığa” En Yüksek Sonsal Yoğunluk Aralığı” (EYSYA) (Highest Posterior Density Region) adı verilir (Demirhan, 2004; Box ve Tiao, 1973; Pres, 1989; Gelman ve diğ., 2004).

Bayesci uygulamalarda temel amaç ilgilenilen parametre için tahminde bulunmaktır. Ancak parametre sayısının birden fazla olduğu durumlarda, eşitlik (2.4)’ün uygulanmasıyla bileşik sonsal dağılım elde edilmesi gerekir. Bu durumda bileşik sonsal dağılımın ilgilenilmeyen parametreler üzerinden integrali alınarak, ilgilenilen parametre için marjinal dağılım hesaplanır. Bu işlemle, ilgilenilen parametre üzerinden diğer parametrelerin etkileri yok edilir. Buna göre marjinal sonsal dağılım, ilgilenilen parametreye ait, diğer önemsiz parametre etkilerinden arındırılmış bilgiyi içeren bir dağılım şeklini alır. Örneğin parametre vektörü, θ_1 ve θ_2 gibi iki bileşenden meydana geldiğinde ve θ_1 parametresi tahmin edilmek istendiğinde θ_1 ’in marjinal sonsal dağılımı aşağıdaki şekilde elde edilir:

$$f(\theta_1|y) = \int_{R\theta_2} f(\theta_1, \theta_2|y) d\theta_2 \quad (2.12)$$

Ancak bu bileşenler arasında bir koşul yapısı söz konusu ise aşağıdaki eşitliğin kullanılması tavsiye edilir: (θ_1, θ_2 'ye koşullu olması durumunda)

$$f(\theta_1|\theta_2^*, y) = \frac{f(\theta_1, \theta_2 = \theta_2^*|y)}{\int f(\theta_1, \theta_2 = \theta_2^*|y)d\theta_1} \quad (2.13)$$

Bayesci analizde, (2.4) eşitliği ile verilen model bir çok parametrenin varlığı söz konusu olduğunda karmaşık ve zor bir şekil alır. Bu gibi durumlarda çözüm Markov Zinciri Monte Carlo (MCMC) yöntemi ile gerçekleştirilmektedir (Gilks ve diğ., 1996).

2.2.3. Markov Zinciri Monte Carlo Yöntemi

Bayesci yaklaşımda sonsal dağılımın karmaşık bir yapıya sahip olması nedeniyle parametre tahmini Markov Zinciri Monte Carlo (MCMC) benzetim yöntemi ile yapılır. Oluşturulan örneklemelerden yola çıkan benzetim yöntemi, dağılımdaki parametrelerin diğer parametrelere göre koşullu olasılıklarının elde edilmesiyle uygulanır (Gilks ve diğ., 1996). Markov zinciri ve örnekleme yöntemine dayanan Monte Carlo yaklaşımı, çok boyutlu ve karmaşık integral problemlerine yaklaşık çözüm getiren stokastik bir yöntemdir. MCMC ile Monte Carlo benzetim yöntemleri arasındaki temel fark, MCMC'de üretilen değerlerin Monte Carlo benzetimin aksine birbirinden bağımsız olmamasıdır. MCMC'de üretilen rasgele sayılar birbirine oldukça bağımlı yapı gösterir ve her bir benzetim değerinin bir önceki değere bağlı olduğu zincir değeri üretir. Eğer bu zincir yeterince uzun çalışırsa, ilgilenilen son dağılım istenilen şekle gelmiş demektir. Rasgele yürüyüş (random walk) yöntemine göre tesadüfi olarak çekilen örneklemelerden özet istatistikler elde edilir ve bu şekilde son dağılımın ortalamasına, meydanına ve en yüksek sonsal yoğunluk aralığına ulaşılabilir.

MCMC yönteminin en önemli özelliği, dağılıma en uygun benzetim algoritmasının seçilmesi durumunda, zincirin başlangıç değerine bakılmaksızın sonsal dağılımın yakınsamasını garanti etmesidir. Zincir çok uzun bir benzetim sürecinde çalıştırılırsa sonsal dağılım daha yüksek doğrulukta elde edebilir (Kurt, 2008). İstenen sonuca kaç

iterasyon sonucunda yakınsayacağı; belirlenen başlangıç değerlerine, zincir sayısına ve modelin etkin bir şekilde işleyip işlememesine bağlıdır. Bu konuda geliştirilmiş tanımsal yöntemler vardır ve iz grafikleri bunlardan biridir. Son yıllarda tercih edilen bir program olan WinBUGS yardımıyla iterasyonun takibi yapılarak iz grafikleri (Trace plot) elde edilebilmektedir. İz grafiklerinde salınımın veya titreşimin fazla olması, benzetim algoritmasının sonsal dağılımı hızlı bir biçimde yaklaştığını ve Markov Zincirinin (MC) oldukça hızlı bir şekilde yeni alanlara geçiş yaptığını göstermektedir. İterasyon sayısının belirlenmesinde ise MC Hatasına (MC Error) bakılmaktadır. MC Hatası, Markov Zinciri algoritması ile yapılan tahminin standart hatasını gösterir. Bu hatanın, parametre tahmine ilişkin standart hataya oranının mümkün oldukça küçük olması istenir (Genellikle 0.05'ten küçük). MC Hata değerine bakarak zincir sayısına da karar verilir. Parametre sayısı fazla olması durumunda, MCMC uzun çalıştırılmalıdır. Bu durumda iterasyon sayısı artacaktır.

Metropolis Hastings ve Gibbs örnekleme, MCMC'de en çok kullanılan algoritmalarıdır. Bu yöntemler ile karmaşık sonsal dağılımlardan örneklem çekmek ve sonsal istatistikleri elde etmek mümkündür.

2.2.3.1. Metropolis Hastings Algoritması

Sonsal dağılımdan parametre tahmini, Metropolis Hastings algoritması kullanılarak aşağıdaki adımlar ile gerçekleştirilir (O'Neill, 2002).

Adım 1: Başlangıç değerleri θ_0 seçilir.

Adım 2: Olabilirlik fonksiyonu ve önsel dağılımın çarpımından elde edilen $h(\theta)$ fonksiyonu hesaplanır.

Adım 3: Öneri dağılımından ($g(\theta)$) üretilen θ_n değerleri kullanılarak

Adım 4: $p_{kabul} = \frac{h(\theta)g(\theta_n|\theta)}{h(\theta_n)g(\theta|\theta_n)}$ kabul olasılığı hesaplanır.

Adım 5: Düzgün dağılımdan bir sayı üretilerek (u) kabul olasılığı karşılaştırılır.

Adım 6: $p_{kabul} < u$ ise θ eski durumunda kalır, aksi takdirde yeni üretilen değerler θ 'nın yeni değerleri olarak alınır.

Görüldüğü gibi $g(\theta)$ dağılımı, rasgele sayı üretilmesi kolay bir dağılım olmalı ve aynı zamanda, θ 'ya bağlı olmayan bir sabit dışında tamamen bilinmelidir.

$g(\theta)$ öneri dağılımının simetrik olması durumunda $g(\theta_n|\theta) = g(\theta|\theta_n)$ olacağından, kabul olasılığı sadece iki noktanın sonsal dağılım oranlarına bağlı olacaktır. Bu durumda:

- i. θ_n aday noktası, sonsal dağılım oranını arttırsa (>1) kabul edilir ve tekrar bir aday noktası seçilir.
- ii. θ_n aday noktası, sonsal dağılım oranını azaltırsa (<1) red edilir, zincir hareket etmez.

2.2.3.2. Gibbs Örneklemesi

Gibbs örnekleme Markov zincirini kullanan bir güncelleme yöntemidir ve sonsal dağılımın, tam koşullu dağılım serisi tarafından açıklandığı durumda kullanılır. Metropolis Hastings algoritmasının özel bir durumu olan Gibbs örneklemesindeki temel amaç koşullu ilişkiler üzerinden tahmin sürecini periyodik olarak işleterek, parametrelerin bileşik dağılımlarına ulaşabilmektir (Gill, 2008).

$(\theta = \theta_1, \theta_2, \dots, \theta_k)$ olmak üzere ilgilenilen parametreye ait dağılımın $\pi(\theta)$ olduğunu varsayalım. Ayrıca θ vektörüne ait tüm tam koşullu dağılımların kümesi de Θ ile gösterelim. Burada θ 'nın k tane bileşeninin her biri tek ya da çok boyutlu olabilir. Bu durumda θ için tam koşullu dağılımlar kümesi aşağıdaki gibi ifade edilir:

$$\pi(\Theta) = \pi(\theta_i|\theta_{-i}) \quad (2.14)$$

Burada θ_{-i} tam koşullu dağılımlar kümesinin θ_i 'yi içermediği halidir. Gibbs örneklemesinde tanımlanan dağılımlardan örneklemler çekilebilmesi için, ilgilenilen parametrelere ait tüm tam koşullu dağılımların analitik olarak tanımlanabilir formda ve bu dağılımların tamamen birbiriyle bağımlı olması gerekmektedir. Bu şekilde tam

Gibbs örnekleme yöntemi yeterince uzun çalıştırıldığında, algoritmada var olan tüm döngüler koşullu dağılımları kullanarak, θ 'nın bütün bileşenleri için birer örneklem üretecektir. Gibbs örneklemesinin en etkileyici yanı, koşullu dağılımların bileşik dağılımdan örneklem elde etmek için yeterli tüm bilgiyi içermesidir.

1984'te Geman tarafından verilen Gibbs örneklemesinin, θ parametrelerinin tüm koşullu dağılımların kolayca elde edilebilir olmadığı durumda çalışmadığını belirtmek gerekir. Bu durumda parametreler için Metropolis Hastings algoritması kullanılarak güncelleştirme yapılır (Gilks ve diğ., 1996).

Gibbs örneklemesinden farklı olarak Metropolis Hastings algoritması, her iterasyonda hareket etmeyi gerektirmez. Önerilen dağılımın simetrik olması durumunda $g(\theta_n|\theta) = g(\theta|\theta_n)$ olacağından, kabul olasılığı sadece iki noktanın sonsal dağılım değerlerine bağlı olacaktır.

3. SAĞKALIM ANALİZİ

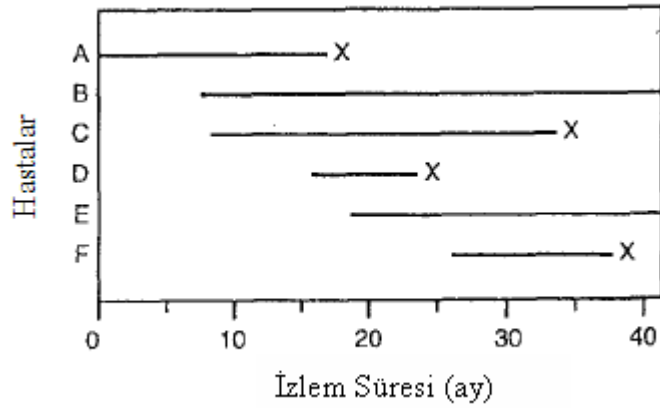
Sağkalım analizi; ölümcül hastalığa yakalanan bir bireye tedavi uygulandıktan sonra, bireyin daha ne kadar yaşayabileceğini ya da hastalığın ne kadar süre sonra tekrarlanabileceğini tahmin etmek amacıyla geliştirilmiş bir yöntemdir. Analiz ayrıca tedavi tiplerinin sağkalım süresine etkilerini incelemeye veya bir ürünün raf ömrünü tahmin etmeye de kullanılmaktadır (Kurt, 2008).

İleriye yönelik bir analiz olması, prognostik faktörler içermesi sebebiyle, sağkalım analizi son yıllarda çok yaygın bir biçimde kullanılmaktadır. Sağkalım verilerinin analizinde son yıllarda farklı yöntemler uygulanmaya başlamıştır. Bunlar içerisinde en yaygın bilinenleri; Yaşam Tablosu (Life Table, LT), Kaplan-Meier (KM) Yöntemi, Cox Regresyon Yöntemi (Oransal Hazard Modelleri, OHM) ve benzetim yöntemlerinin gelişmesi ile öne çıkan Bayesci Sağkalım Analizi (BSA) şeklinde saymak mümkündür. Özellikle BSA'nın önsel olasılıkları dikkate alarak uygulanması, anılan yöntemlerle daha etkin sonuçlar elde etmeyi mümkün kılmaktadır (Kurt, 2008).

Bayesci sağkalım analizini açıklamadan önce sağkalım verilerinde karşılaşılabilecek bazı kavramları tanımlamakta yarar görülmüştür. Bunlardan ilki izlem süresi ya da bir diğer adıyla sağkalım süresidir. Sağkalım süresi, bir bireyin ölümcül bir hastalığa yakalandıktan sonra iyileşmesine, hastalığın tekrarlanmasına ya da ölümüne kadar geçen süredir. Sağkalım süresi, analizde bağımlı değişken olarak kullanılır ve genellikle Üstel, Weibull ya da Lognormal dağılıma sahip bir değişkendir. Sağkalım analizinde karşılaşılabilecek bir diğer kavram ise bağımsız değişkenler yani prognostik faktörlerdir. Bu faktörler sağkalım süresi üzerinde etkili olan değişkenlerdir ve dağılım şekli konusunda herhangi bir kısıtlama yoktur. Sansürlü (censored) veri, sağkalım analizinde karşılaşılabilecek kavramlardan bir diğeridir ve önceden belirlenen izlem süresi sonunda, ölüm ya da beklenen bir olaya erişemeyen gözlemlerden oluşur. Oysa ki analize dahil edilen tüm hastaların, ölüm ya da belirlenen başka bir olaya erişmesi beklenir. Sansürlü veri kullanımı, bu olaya ulaşma süresini bir

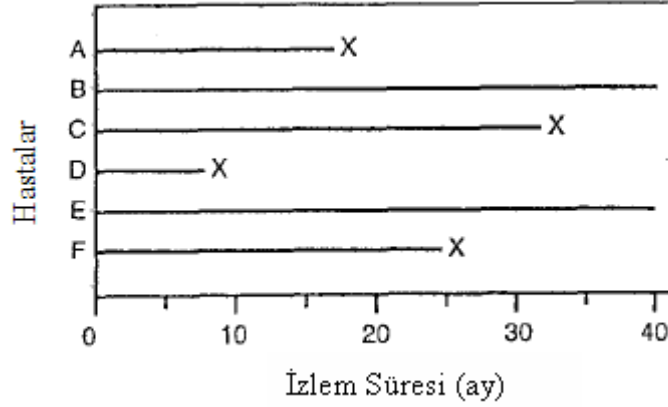
başka deyişle hangi tedavinin ya da yöntemin daha iyi sonuç verdiğini görme süresini uzatacağı için sağkalım analizinde yaygın bir şekilde kullanılmaktadır. Sağkalım süresi içinde bazı birimlerin ölmemesi ya da beklenen olgunun gözlenmemesi (administrative censoring), bazı gözlem birimlerinin çeşitli nedenlerle kaybolması (lost of follow up) ya da gözlem birimlerinin izlem dışı bırakılması (withdrawing) halinde sansürlü veri ile karşılaşmak mümkündür.

Sağkalım analizinde hastaların izlenmeye başladığı tarihi farklı biçimlerde belirlemek mümkündür. Bu tarih, tedaviye başlanan tarih veya cerrahi müdahale tarihi olarak tanımlanabilir. Başlangıç noktasının yani hastayı izlemeye başlanan tarihin, tüm hastalar için aynı zamanda olması beklenemez. Dolayısı ile veriler çalışmaya farklı zamanlarda dahil olabilirler. Bu durumda gözlemler “aşamalı (dereceli) sansürlü” olarak adlandırılırlar ve izlem süresi bittiğinde aşamalı sansürlü olan gözlemlerin tümü, çalışmanın başlama noktasına çekilerek analiz edilirler. Şekil 3.1’de 40 ay izlem süresine sahip, aşamalı sansürlü verileri açık bir biçimde görmek mümkündür ve incelendiğinde hastaların izlenmeye başladıkları sürelerin farklı olduğu görülür. “X” ile gösterilen hastalar ölen hastalar iken, B ve E hastaları sansürlü gözlemleri oluşturmaktadır.



Şekil 3.1. Farklı zamanlarda analize dahil olan hastaların çalışmaya giriş süreleri gösteren grafik

Şekil 3.2.’de ise; farklı zamanlarda izlenmeye başlanan hastalar, çalışmanın başlama noktasına çekilerek analiz edilmişlerdir (Dişçi, 2008).



Şekil 3.2. Farklı zamanlarda analize dahil olan hastaların çalışmaya giriş sürelerinin başlama noktasına çekildiğini gösteren grafik

Sağkalım verisinin karakteristik özelliklerinin belirlenmesinde ve sağkalım süresine etki eden prognostik faktörlerin incelenmesine olanak veren sağkalım ve Hazard fonksiyonları tanımlanması bu noktada önem kazanmaktadır. Sağkalım fonksiyonu, sağkalım analizinin temelini oluşturmaktadır ve belirli bir t zamanına kadar veya t zamanından daha fazla yaşayan bireylerin olasılığını ifade etmektedir. Fonksiyon aşağıdaki gibi ifade edilir:

$$S(t) = 1 - F(t) = P(T > t) \quad (3.1)$$

Burada $F(t) = P(T \leq t) = \int_0^t f(u)du$ olarak tanımlanır. (3.1)'de dikkat edilmesi gereken $S(t)$ 'nin, t'nin "0" ve " ∞ " değerleri için $S(0) = 1$ ve $S(\infty) = \lim_{t \rightarrow \infty} S(t) = 0$ olduğu, ayrıca monoton azalan eğilim gösterdiği, yani bir hastanın izlemeye alındığında yani başlama noktasında sağkalım olasılığı "1" iken, izleme süresi sonsuza giderken hastanın sağkalım olasılığı "0" olmaktadır.

Sağkalım verisinin karakteristik özelliğini ortaya çıkaran ve sağkalım süresine etki eden prognostik faktörlerin incelenmesine olanak veren diğer bir fonksiyon da Hazard fonksiyonudur. Belirlenen bir olguya ulaşma fonksiyonu ya da oranı olarak da anılan Hazard (risk) fonksiyonu, t zamanındaki ani ölüm ya da hastalığa yakalanma riski olarak tanımlanabilir. Endüstride bir ürünün bozulmasını ya da raf

ömrünün sona ermesini, tıpta bir bireyin hastalığa yakalanmasını ya da ölüm riskini ifade etmek için kullanılır. Hazard fonksiyonu aşağıdaki eşitlikteki gibi ifade edilir:

$$h(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t < T \leq t + \Delta t | T > t)}{\Delta t} = \frac{f(t)}{S(t)} \quad (3.2)$$

$h(t)$ 'den hareketle $f(t) = -\frac{d}{dt}S(t)$ şeklinde ve $S(t) = \exp(-\int_0^t h(u)du)$ olarak elde edilir.

(3.2) eşitliğinden de görüldüğü gibi; $f(t)$ olasılık yoğunluk fonksiyonu, $S(t)$ sağkalım fonksiyonu ve $h(t)$ Hazard fonksiyonlarından herhangi ikisinin bilinmesi durumunda bilinmeyen üçüncü fonksiyon kolaylıkla elde edilmektedir.

Prognostik faktörlerin, Hazard ve sağkalım fonksiyonlarına olan etkileri, sırasıyla Oransal Hazard (OH) ve Hızlandırma faktörü (HF) yardımıyla incelenebilir. OH, Hazard'ın karşılaştırılmasına olanak verirken; HF, sağkalım zamanlarının karşılaştırılmasında kullanılır. Karşılaştırılmak istenen prognostik faktörlere ait Hazard fonksiyon oranlarının, sağkalım zamanı boyunca sabit olmasına Oransal Hazard varsayımı denilmektedir. Karşılaştırılmak istenen prognostik faktörler X^* ve X ile olarak tanımlandığında OH'ın matematiksel ifadesi,

$$\frac{\hat{h}(t, X^*)}{\hat{h}(t, X)} = \hat{\theta} \quad (3.3)$$

ile ifade edilir.

Oransal Hazard varsayımının değerlendirilmesinde kullanılan yöntemleri şu şekilde sıralamak mümkündür: grafik yöntemleri (log(-log)) sağkalım eğrileri (Kalbfleisch Prentice, 1980)), Arjas grafikleri (Arjas, 1988), Schoenfeld artıkları ile sağkalım süresinin rankının korelasyon testi (Schoenfeld, 1982). Sözü edilen yöntemler aşağıda verilmiştir.

- Log (-log) Sağkalım Eğrileri

Log(-log) sağkalım eğrileri yöntemi en yaygın kullanılan grafiksel yöntemdir. İncelenen prognostik faktörün farklı kategorileri üzerinden $-\ln(-\ln)$ sağkalım eğrilerinin tahmininin karşılaştırmasını içermektedir. Elde edilen paralel eğriler orantılı Hazard varsayımının sağlandığını göstermektedir (Kalbfleisch ve Prentice, 1980; Therneau ve Grambsch, 2000).

Sağkalım eğrisi tahmini $[0,1]$ aralığında değer alırken, $-\ln(-\ln \hat{S})$ eğrisi $(-\infty, \infty)$ aralığında değer almaktadır.

Log (-log) sağkalım eğrisi, sağkalım eğrisi tahminine uygulanan basit bir dönüşümdür. Sağkalım olasılığı tahminin iki kez doğal logaritması alınarak elde edilmektedir.

Sağkalım fonksiyonu $[0,1]$ aralığında değerler aldığından $\ln S(t, x)$ ve $\ln S_0(t)$ negatif olmaktadır. Negatif bir sayının logaritması alınmadığından, $\ln S(t, x)$ 'in ikinci kez logaritmasını alabilmek için bu fonksiyonun $-\ln S(t, x)$ olarak kullanılması gerekmektedir. İkinci logaritma pozitif ya da negatif olabilir. Bu nedenle ikinci kez negatifinin alınmasına gerek yoktur. Tutarlılık için, $\ln(-\ln)$ ifadesini elde etmek için ikinci logaritmanın önüne eksi isareti konulmaktadır.

Açıklanan bu grafiksel yöntemin bazı zayıf yönleri vardır. Paralelligin nasıl belirleneceğine karar vermek öznel olabilir. Log(-log) sağkalım eğrilerinin paralel olmadığına dair güçlü bir kanıt yoksa, orantılı Hazard varsayımının sağlandığı varsayılmaktadır. Sürekli prognostik faktörlerin nasıl sınıflandırılacağıının belirlenmesi önemlidir. Farklı sınıflandırmalar farklı grafiksel çizimleri ortaya koymaktadır. Sürekli prognostik faktörler sınıflandırılırken, sınıf sayısının az alınması önerilmektedir. Sınıfların seçimi mümkün olduğu kadar anlamlı olmalı ve uygun sayı dengesinin sağlanması gerekmektedir. Log(-log) sağkalım eğrilerinin kullanılması ile ilgili bir sorun da birçok prognostik faktör için orantılı Hazard varsayımının eş zamanlı olarak nasıl değerlendirileceğidir. Eş zamanlı karşılaştırmalar için, bütün prognostik faktörler ayrı ayrı sınıflandırılmalı, sınıfların

farklı birleşimleri oluşturulmalı ve daha sonra aynı grafik üzerinde bütün log(-log) sağkalım eğrileri karşılaştırılmalıdır. Herbir birleştirilmiş sınıf için yeterli sayı olsa bile, paralel olmama durumuna neden olan prognostik faktörlere karar vermek zordur (Kleinbaum ve Klein, 1996).

- Arjas Grafikleri

Orantılı Hazard modelinde uyum iyiliğinin test edilmesi için kullanılan grafiksel bir yöntemdir. Alternatif modellerin tahminine gereksinim duymadığından ve sadece parametre tahmininde kullanılan kısmi olabilirlik ifadesindeki niceliklere benzer nicelikleri içerdiğinden kullanılması kolay bir yöntemdir (Arjas, 1988). Grafiksel yöntemler içinde en uygun sonucu veren yöntemdir (Persson, 2002).

Grafiklerin değerlendirilmesi öznelidir. Gözlenen ve beklenen (modelden tahmin edilen) başarısızlık sıklıkları arasında doğrudan bir karşılaştırma yapmaktadır.

Doğru belirlenmiş bir modelde bu sıklıkların arasında yaklaşık bir dengenin olması beklenmektedir.

Arjas grafikleri yönteminde bireyler tabakalara ayrılır ve herbir tabaka için bir grafik çizilir. Cox regresyon modeli veriye uygunsa, grafiğin eğiminin bire yakın ve yaklaşık olarak doğrusal olması beklenmektedir. Model yanlış belirlendiği takdirde grafik üzerinde gözle görülebilir farklı şekiller gözlemlenmektedir (Arjas, 1988).

- Schoenfeld Artıkları ile Sağkalım Sürelerinin Rankının Korelasyon Testi Yöntemi

Cox regresyon modelinde kullanılan Cox-Snell, değiştirilmiş Cox-Snell, Martingale ve Sapma artıklarının iki dezavantajı vardır:

- Ağırlıklı olarak gözlenen sağkalım zamanına bağlı olmaları,
- Birikimli Hazard fonksiyonunun tahminine ihtiyaç duymalarıdır.

Bu dezavantajlar, Schoenfeld (1982) tarafından önerilen artıklar ile ortadan kaldırılmaktadır. Bu artıklar skor artık olarak da adlandırılmaktadır. Tanımlanan diğer artık türlerinden farklı olan bir diğer özelliği de herbir birey için artığın tek bir değeri yoktur. Cox regresyon modelinde kullanılan herbir birey, her açıklayıcı değişkeni için bir artık değerine sahiptir (Collet, 2003).

p tane prognostik faktör olduğunu ve n tane bağımsız birey için sağkalım zamanının, prognostik faktörlerin ve durum değişkenininin (t_i, x_i, v_i) ile gösterildiği varsayalım. $i=1, \dots, n$ olmak üzere olguya ulaşmış gözlemler için $v_i = 1$ ve sansürlü gözlemler için $v_i = 0$ dir. Schoenfeld'in (1982), orantılı Hazard modeli ile kullanılması için önerdiği artıklar, log kısmi olabilirliğin türevine bireysel katkıya dayanmaktadır. j. Prognostik faktör için türev,

$$\frac{\partial L_p(\beta)}{\partial \beta_j} = \sum_{i=1}^n v_i \left[x_{ji} - \frac{\sum_{l \in R(t_i)} x_{jl} \exp(\beta' x_l)}{\sum_{l \in R(t_i)} \exp(\beta' x_l)} \right] = \sum_{i=1}^n v_i \{x_{ji} - a_{ji}\} \quad (3.4a)$$

biçimindedir. Burada

$$a_{ji} = \frac{\sum_{l \in R(t_i)} x_{jl} \exp(\beta' x_l)}{\sum_{l \in R(t_i)} \exp(\beta' x_l)} \quad (3.4b)$$

olarak tanımlanmaktadır. Eşitlik (3.4a)'da x_{ji} çalışmadaki i. birey için j. ($j=1, 2, \dots, p$) prognostik faktörünün değeri ve $R(t_i)$ ise t_i zamanında riskte olan tüm bireylerin kümesidir.

j. prognostik faktör için i. bireyin Schoenfeld artığının tahmin edicisi, Eşitlik (3.4)'ten

$$\hat{r}_{S_{ji}} = v_i (x_{ji} - \hat{a}_{ji}) \quad (3.5)$$

biçiminde elde edilir. Burada \hat{a}_{ji} , eşitlik (3.4b)'de verildiği gibidir (Schoenfeld, 1982). Schoenfeld artıklarının sıfırdan farklı değerleri sadece olguya ulaşmış gözlemler için açığa çıkmaktadır (Hosmer ve Lemeshow, 1999)

Orantılı Hazard varsayımını incelemek için Schoenfeld artıklarına dayanan bir test geliştirilmiştir (Grambsch ve Therneau, 1994). Belirli bir prognostik faktör için Schoenfeld artıkları ile bireylerin sağkalım zamanlarının rankı arasındaki korelasyon kullanılarak orantılı Hazard varsayımı incelenmiştir. Bu teste göre, orantılı Hazard varsayımının sağlanması için korelasyonun sıfıra yakın olması beklenmektedir. Yokluk hipotezi “orantılı Hazard varsayımı sağlanmaktadır” biçimindedir. Yokluk hipotezi hiçbir zaman tam anlamıyla kanıtlanamamaktadır. Ancak yokluk hipotezini reddedecek yeterli kanıtın olmadığı söylenebilir.

Bu test istatistiği, orantılı Hazard varsayımının incelenmesi için kullanılan grafiksel yöntemlere göre daha nesnel bir kriter sağlamaktadır. Grafiksel yöntemler ise daha öznel-dir. İlgilenilen olguya erken bir zamanda karşılaşan bireyler için artıklar pozitif olma eğilimi gösterirse ve olguya geç bir zamanda karşılaşan bireyler için artıklar negatif olma eğilimi gösterirse, Hazard oranı zaman eksenini boyunca sabit olmaz, bu durumda orantılı Hazard varsayımı sağlanmaz.

Prognostik faktörlerin sağkalım fonksiyonuna olan etkilerinin belirlenmesinde HF'den yararlanılır. HF'nün, sağkalım fonksiyonlarının oranlarının sağkalım zamanına göre sabit olma varsayımını sağlaması şart değildir. Ancak HF'nin kullanılabilmesi için sağkalım zamanının, bir dağılımının ve şekil parametresinin olması gerekmektedir. Karşılaştırılmak istenen prognostik faktörlere ait sağkalım fonksiyonları sırasıyla $S_1(t)$ ve $S_2(t)$ olarak tanımlandığında HF, $t \geq 0$ için sağkalım fonksiyonu ile ilişkiyi $S_2(t) = S_1(\gamma t)$ şeklinde ifade edilir. Burada γ sabitine *Hızlandırma Faktörü* (HF) adı verilir. γ , prognostik faktörlere ait katsayıların tahmin değerlerinin (α) üstel'inin alınmasıyla elde edilir ($\gamma = \exp(\alpha)$). Bu durumda HF, $t \geq 0$ için $S_2(t) = S_1(\exp(\alpha)t)$ veya benzer şekilde $S_2(t) = S_1(\exp(-\alpha)t)$ olarak gösterilebilir.

3.1. PARAMETRİK MODELLER

Sağkalım süresine ait veriler negatif değer almaz ve dağılımları sağa çarpıktır. Ayrıca sağkalım verilerinin dağılımı, üstel dağılım ailesindedir (Üstel, Weibull, Lognormal, vb.). Sağkalım verisinin uyduğu dağılıma ait varsayımlar sağlandığı zaman, sağkalım ve Hazard fonksiyonları açık olarak tanımlanabilir ve buradan hareketle parametre tahminleri kolayca elde edilebilir. Sağkalım verilerinin hangi dağılıma uygun olduğuna karar vermek için dağılımlara ait bazı karakteristiklerin bilinmesi gerekir. Bu karakteristikler olasılık yoğunluk fonksiyonu, sağkalım fonksiyonu ve Hazard fonksiyon grafikleridir. Bayesci sağkalım analizinde, parametrik modeller önemli bir rol oynamaktadır ve parametrik modellerden elde edilen sağkalım tahminlerine ait grafikler, yarı parametrik sağkalım eğrilerinden (Kaplan-Meier eğrilerinden) daha tutarlı sonuçlar verir. Parametrik modellerin daha çok tercih edilmesinin en önemli sebepleri kullanım kolaylığı ve ilgilenilen olgunun uyduğu dağılımın şeklinin bilinmesidir (Kleinbaum ve diğ., 1996).

Birçok parametrik sağkalım modellerinde oransal Hazard varsayımı sağlanamamaktadır. Sadece Üstel ve Weibull dağılımlarına sahip modeller OH varsayımını sağlamakta yeterlilik göstermektedir (Kleinbaum ve diğ., 1996). Ancak tüm parametrik modellerde sağkalım zamanına ait dağılım ve şekil parametresi elde edilebildiğinden, prognostik faktörlerin etkisinin incelenmesinde HF yeterli olmaktadır.

3.1.1. Olabilirlik Fonksiyonu

Sağkalım analizinde, verilerinin belirli bir dağılıma sahip olduğu varsayımı altında, dağılıma ilişkin parametrelerin tahmininde olabilirlik fonksiyonundan yararlanılır. Sağkalım veri kümesinin n tane gözlemden meydana geldiğini varsayalım. Veriye ilişkin sağkalım zamanının (t); hem belirli bir olguya ulaşmış ve hem de sansürlü verilerden oluştuğunu düşünelim. Olguya ulaşmış ve sansürlü zamanları ayırt etmede durum değişkeni olarak “ v ” yi kullanalım ve $v_i = 1$ ile olguya ulaşan, $v_i = 0$ ile sansürlü sağkalım zamanını gösterelim. Ayrıca sağkalım verilerinin, $f(t, \theta)$ yoğunluk fonksiyonuna ve $S(t, \theta)$ sağkalım fonksiyonuna sahip olduğu varsayalım.

Burada $\theta = (\theta_1, \dots, \theta_p)$, dağılıma ait parametreyi ifade etmektedir (örneğin Üstel dağılım için $p=1$ ve $\theta = \lambda$ dır). Olguya ulaşmış veriler tamamlanmış veri niteliğini taşıdığı için, uydukları dağılım belirlenebilmekte ve olabilirlik fonksiyonu, yoğunluk fonksiyonlarının çarpımı ile elde edilebilmektedir. Ancak sansürlü veriler henüz tamamlanmamış oldukları için, belirli bir t zamanına kadar veya t zamanından daha fazla yaşayan bireylerin sağkalım olasılığını, sağkalım fonksiyonundan yararlanarak elde etmek mümkündür. Burada sağkalım fonksiyonu, gerçekleşen olgunun dağılımına dayanarak elde edilir.

Olguya ulaşmış ve sansürlü veriler bağımsız oldukları için; ortak olabilirlik fonksiyonu, olabilirliklerinin çarpımına eşittir (Lee ve Wang, 2003 ve Collet, 2003). Ortak olabilirlik fonksiyonunu aşağıdaki gibi ifade etmek mümkündür:

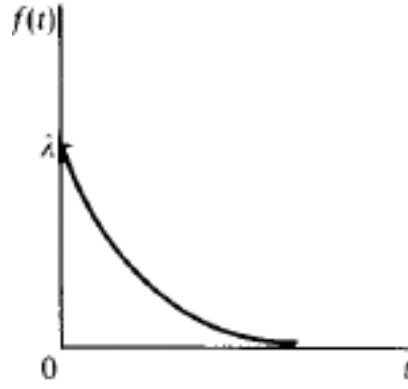
$$L(\theta) = \prod_{i=1}^n f(t_i, \theta)^{v_i} S(t_i, \theta)^{(1-v_i)} \quad (3.6)$$

3.1.2. Üstel Dağılım

Üstel dağılım, sağkalım analizinde kullanılan başlıca parametrik dağılımlardan biridir. Dağılımın tek bir parametresi vardır ve iki veya daha fazla parametrelili dağılımların temelini oluşturmaktadır. Üstel dağılımda, belli bir zaman aralığında olgunun gerçekleşme olasılığı sabit kalmaktadır. Ayrıca üstel dağılım, iki olgunun gerçekleşmesi arasında geçen zamanın dağılımını da vermektedir. (Evans ve diğ., 2000). Üstel dağılımda olgunun gerçekleşme olasılığı sabittir. Hesaplamalarda kolaylık sağlasa da uygulamalarda değişim gösteren olgunun gerçekleşme olasılığı ile karşılaşılmaktadır (Türkan, 2007).

Üstel dağılımın olasılık yoğunluk fonksiyonu aşağıdaki eşitlikte, fonksiyona ait grafik ise Şekil 3.3'te verilmiştir:

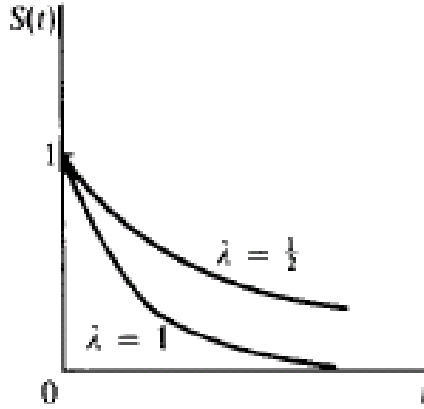
$$f(t) = \lambda \exp(-\lambda t) \quad (3.7)$$



Şekil 3.3. Üstel dağılıma ait olasılık yoğunluk fonksiyonunun grafiği

Üstel dağılımın sağkalım fonksiyonu eşitlik (3.8)'de, fonksiyona ilişkin grafik ise Şekil 3.4.'te verilmiştir.

$$S(t) = 1 - F(t) = \exp(-\lambda t) \quad (3.8)$$

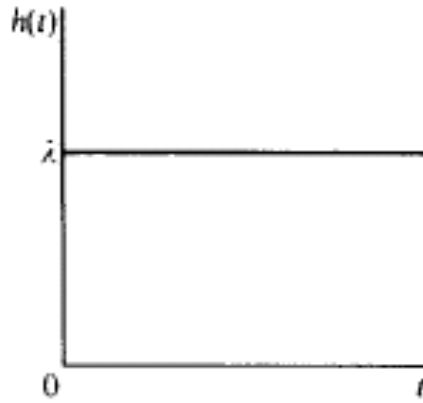


Şekil 3.4. $\lambda = 1$ ve $\lambda = 1/2$ için, Üstel dağılımın sağkalım fonksiyonunun grafiği

Üstel dağılımın Hazard fonksiyonu (Hazard oranı);

$$h(t) = \frac{f(t)}{S(t)} = \lambda \quad (3.9)$$

ile ifade edilir. Üstel dağılımın Hazard fonksiyonunun grafiği ise aşağıdaki gibidir:



Şekil 3.5. Üstel dağılımın Hazard fonksiyonunun grafiği

Üstel dağılıma sahip bir sağkalım verisinde; yüksek λ değeri yüksek riski bir başka deyişle düşük sağkalımı ifade ederken, düşük λ değeri düşük risk ve yüksek sağkalımı temsil eder. Ayrıca üstel dağılıma sahip modelde; OH ve HF de elde edilen katsayılar, birbirinin çarpmaya göre tersidir. Bu durum sadece üstel dağılım için geçerlidir.

Sağkalım çalışmasında; $t = (t_1, \dots, t_n)'$ 'nin sağkalım zamanını, $v = (v_1, \dots, v_n)'$ 'nin sağkalım zamanına ilişkin durum değişkenini, $D = (n, y, v)'$ 'nin ise gözlenen verileri gösterdiğini düşünelim. Sağkalım zamanının üstel dağılıma sahip olduğu varsayımı altında, λ 'ya ait olabirlik fonksiyonu;

$$L(\lambda|D) = \prod_{i=1}^n f(t_i|\lambda)^{v_i} S(t_i|\lambda)^{(1-v_i)} = \lambda^d \exp(-\lambda \sum_{i=1}^n t_i) \quad (3.10)$$

şeklinde ifade edilir. Burada $d = \sum_{i=1}^n v_i$ dir.

3.1.2.1. Üstel Dağılıma Sahip Sağkalım Verisine İlişkin Bayesci Yaklaşım

Bayesçi yaklaşımda λ parametre tahmini, olabirlik fonksiyonu ve $f(\lambda)$ önsel dağılımın birleştirilmesiyle oluşan sonsal dağılım yardımıyla elde etmek mümkündür. λ 'nın sonsal dağılımı,

$$f(\lambda|D) \propto f(D|\lambda)f(\lambda) \quad (3.11)$$

ile ifade edilir. Buradan hareketle λ parametresi için üstel dağılımın eşlenik dağılımlarından, Gamma dağılımına sahip olduğunu varsayalım ve dağılımı $\zeta(\alpha_0, \lambda_0)$ ile gösterelim. Bu durumda λ parametresi için Gamma dağılımı aşağıdaki eşitlik ile tanımlanabilir:

$$f(\lambda|\alpha_0, \lambda_0) \propto \lambda^{\alpha_0-1} \exp(-\lambda_0 \lambda) \quad (3.12)$$

λ parametresine ait sonsal dağılımı elde etmek amacıyla (3.10) ve (3.12) eşitlikleri birleştirilir ve sonsal dağılım,

$$\begin{aligned} f(\lambda|D) &\propto L(\lambda|D)f(\lambda|\alpha_0, \lambda_0) \propto (\lambda^{\sum_{i=1}^n v_i} \exp\{-\lambda \sum_{i=1}^n t_i\})(\lambda^{\alpha_0-1} \exp(-\lambda_0 \lambda)) \\ &= \lambda^{\alpha_0+d-1} \exp\{-\lambda(\lambda_0 + \sum_{i=1}^n t_i)\} \end{aligned} \quad (3.13)$$

ile elde edilir.

λ parametresine ait sonsal ortalama ve varyans;

$$E(\lambda|D) = \frac{\alpha_0 + d}{\lambda_0 + \sum_{i=1}^n t_i} \quad (3.14)$$

$$Var(\lambda|D) = \frac{\alpha_0 + d}{(\lambda_0 + \sum_{i=1}^n t_i)^2} \quad (3.15)$$

olarak elde edilir.

t_f gelecekte gerçekleşen olgu zamanını ifade ettiğinde, bu zamanın tahmini için elde edilen sonsal dağılım;

$$\begin{aligned}
\pi(t_f|D) &= \int_0^{\infty} f(t_f|\lambda) f(\lambda|D) d\lambda \propto \int_0^{\infty} \lambda^{\alpha_0+d+1-1} \exp\{-\lambda(t_f + \lambda_0 + \sum_{i=1}^n t_i)\} d\lambda \\
&= \Gamma(\alpha_0 + d + 1) (\lambda_0 + \sum_{i=1}^n t_i + t_f) \propto (\lambda_0 + \sum_{i=1}^n t_i + t_f)^{-(d+\alpha_0+1)}
\end{aligned} \tag{3.16}$$

şeklinde ifade edilir. Sonsal tahmin ediciye ait dağılımın normalleştirilmesi ise aşağıdaki eşitlik yardımıyla yapılabilir:

$$f(t_f|D) = \left\{ \begin{array}{l} \frac{(d + \alpha_0)(\lambda_0 + \sum_{i=1}^n t_i)^{(\alpha_0+d)}}{(\lambda_0 + \sum_{i=1}^n t_i + t_f)^{(d+\alpha_0+1)}}; y_f > 0 \\ 0; d.d \end{array} \right\} \tag{3.17}$$

Üstel dağılıma sahip sağkalım sürelerine etki eden prognostik faktörlerin tahmin edilmesi için; λ parametresi, px1 boyutlu x'_i prognostik faktörler vektörü ile $\lambda_i = \varphi(x'_i\beta)$ şeklinde bir ilişki içindedir. Burada β , px1 prognostik faktörlere ait katsayıları ve $\varphi(\cdot)$ sağkalım zamanına ait dağılımın parametresini göstermektedir. $\varphi(\cdot)$ genellikle; $\varphi(x'_i\beta) = \exp(x'_i\beta)$ veya $\varphi(x'_i\beta) = (x'_i\beta)^{-1}$ olarak seçilir. $\varphi(x'_i\beta) = \exp(x'_i\beta)$ şeklinde seçildiğinde olabirlik fonksiyonu;

$$\begin{aligned}
L(\beta|D) &= \prod_{i=1}^n f(t_i|\lambda_i)^{v_i} S(t_i|\lambda_i)^{(1-v_i)} \\
&= \prod_{i=1}^n [\exp(x'_i\beta) \exp(-t_i \exp(x'_i\beta))]^{v_i} [\exp(-t_i \exp(x'_i\beta))]^{(1-v_i)} \\
&= \exp\left\{ \sum_{i=1}^n v_i x'_i \beta \right\} \exp\left\{ - \sum_{i=1}^n t_i \exp(x'_i \beta) \right\}
\end{aligned} \tag{3.18}$$

olarak ifade edilir. Burada $D = (x, y, X, v)$ ve X , i. satırı x'_i 'lerden oluşan nxp boyutlu matristir.

Bayesçi yaklaşımda rasgele değişken olarak tanımlanan β parametresi için önerilen önsel dağılımlar, uygun olmayan (tanım aralığında dağılımın integrali alındığında 1

değerini vermeyen) düzgün dağılım ($\pi(\beta) \propto 1$) ve normal önsel dağılımdır. Genel olarak β parametresinin, p boyutlu normal önsel dağılıma ($N_p(\mu_0, \Sigma_0)$) sahip olduğu varsayılmaktadır. Burada μ_0 önsel ortalamayı ve Σ_0 önsel kovaryans matrisini ifade etmektedir. Bu durumda β için sonsal dağılım;

$$f(\beta|D) \propto L(\beta|D)f(\beta|\mu_0, \Sigma_0) \quad (3.19)$$

ile elde edilir. Eşitlikte $f(\beta|\mu_0, \Sigma_0)$; ortalaması μ_0 ve varyansı Σ_0 olan çok değişkenli normal dağılımdır. (3.19)'de verilen sonsal dağılımın bilinen bir dağılım yapısı olmadığından MCMC yöntemine gereksinim duyulur.

MCMC, Gibbs örnekleme yöntemi yardımıyla sonsal dağılımdan örneklem çeken bir yöntemdir. Ancak Gibbs örnekleme yönteminin uygulanabilmesi için, tam koşullu olasılığın yazılabilmesi gerekir. β_j , β 'ya ait j. elemanı ve $\beta^{(-j)}$, j. eleman çıkarıldıktan sonraki β vektörünü temsil ettiğinde; j=1,2,...,p için j. tam koşul aşağıdaki gibi ifade edilir:

$$f(\beta_j|D, \beta^{(-j)}) \propto L(\beta_j, \beta^{(-j)}|D)f(\beta_j, \beta^{(-j)}|\mu_0, \Sigma_0) \quad (3.20)$$

Burada dikkat edilmesi gereken; elde edilen tam koşullu olasılığın, her bir p tam koşullu dağılımın olabilirliliği ile önselinin çarpımı olarak elde edildiğidir. Ayrıca her bir gözlem için β_j ve $\beta^{(-j)}$ rasgele değişkenlerinin sabit olarak kabul edildiği de dikkat edilmesi gereken bir başka noktadır. Koşullu dağılımdan örneklem çekmede, Metropolis-Hastings ve Gibbs algoritmalarının da temelini oluşturan RED örnekleme algoritması yaygın olarak kullanılmaktadır. RED örnekleme algoritması, istenen dağılımdan gözlemler türetmek için rasgele örneklem çeme yöntemidir. İstenen dağılıma sahip olmayan gözlemleri “red” ettiği için bu şekilde adlandırılmaktadır. RED örnekleme algoritmasını basit bir örnekle açıklamak gerekirse, (-1,1) aralığında düzgün dağılıma sahip x ve y bağımsız değişkenlerinin oluşturduğu kareden, birim çember özelliğini sağlayan bir veri kümesi elde etmek isteyelim. Bu durumda sadece $x^2 + y^2 \leq 1$ eşitliğini sağlayan x ve y noktaları veri kümesine alınır, geri kalan

veriler red edilir. Red örneklem algoritmasının uygulanabilmesi için dağılımlar log-konkav olmalıdır çünkü log-konkav olan bir dağılımın koşullu olasılıkları da log-konkav olmaktadır. Bir fonksiyon aşağıdaki eşitsizliği sağlıyorsa log-konkav olduğu söylenir:

$$f(\theta x + (1-\theta)y) \geq f(x)^\theta f(y)^{1-\theta} \quad (3.21)$$

Üstel dağılım ailesi genellikle log-konkavlık özelliğini sağlamaktadır. Log-konkavlık özelliğini sağlayan dağılımlardan benzetim yöntemleri yardımıyla kolaylıkla örneklem çekilebilir (Gilks ve Wild., 1992). Bir fonksiyonun log-konkavlık özelliğini sağlayıp sağlamadığı, fonksiyonun logaritmasının parametreye göre ikinci türevinin negatif değer alıp almadığı ile anlaşılır. Üstel dağılım için (3.20)'de tanımlanan tam koşullu dağılımın log-konkav olduğu (3.22) ile görülmektedir.

$$\frac{\partial^2}{\partial \beta_j^2} \log(f(\beta_j | D, \beta^{(-j)})) \leq 0 \quad (3.22)$$

Kernel (Çekirdek) düzlemesi kullanılarak sonsal dağılımın logaritması

$$\log(f(\beta | D)) = \sum_{i=1}^n v_i x_i' \beta - \sum_{i=1}^n t_i \exp(x_i' \beta) + \log(f(\beta | \mu_0, \Sigma_0)) \quad (3.23)$$

ile elde edilir. (3.23) eşitliğinin log-konkav olup olmadığı

$$\begin{aligned} \frac{\partial^2}{\partial \beta_j^2} \log(f(\beta | D)) &= -\sum_{i=1}^n t_i x_{ij}^2 \exp(x_i' \beta) + \frac{\partial^2}{\partial \beta_j^2} \log(f(\beta | \mu_0, \Sigma_0)) \\ &= -\sum_{i=1}^n t_i x_{ij}^2 \exp(x_i' \beta) - \sigma_0^{(jj)} \leq 0 \end{aligned} \quad (3.24)$$

ile görülür. Burada $\sigma_0^{(jj)}$, Σ_0^{-1} 'in j. köşegen elemanını ifade etmektedir. Buradan hareketle $f(\beta | D)$ 'de yer alan her bir β elemanının log-konkav olduğunu söylemek mümkündür.

β için normal önsel dağılım kullanıldığında, önsel dağılımın β parametresine göre log-konkav olduğu aşağıdaki eşitliklerden görülebilir:

$$\frac{\partial}{\partial \beta} \log(f(\beta | \mu_0, \Sigma_0)) = -\Sigma_0^{-1}(\beta - \mu_0) \quad (3.25)$$

$$\frac{\partial^2}{\partial \beta \partial \beta'} \log(f(\beta | \mu_0, \Sigma_0)) = -\Sigma_0^{-1} = -\sigma_0^{(jj)} \quad (3.26)$$

Sonuç olarak önsel dağılım log-konkav olduğu sürece sonsal dağılım da log-konkav olacaktır (Ibrahim ve diğ., 2001). Log-konkav olan sonsal dağılımdan benzetim yöntemleri yardımıyla örneklemeler çekilerek β parametresine ait tahminler yapılır.

Üstel dağılıma sahip sağkalım verisinin, sağkalım fonksiyonunu $S(t) = \exp(-\exp(x'\beta)t)$ ile ifade edilir. Dolayısıyla Bayesci yaklaşımla sağkalım fonksiyonu hakkında bilgi edinmek için β 'nin sonsal dağılımından örneklemeler elde etmek yeterlidir.

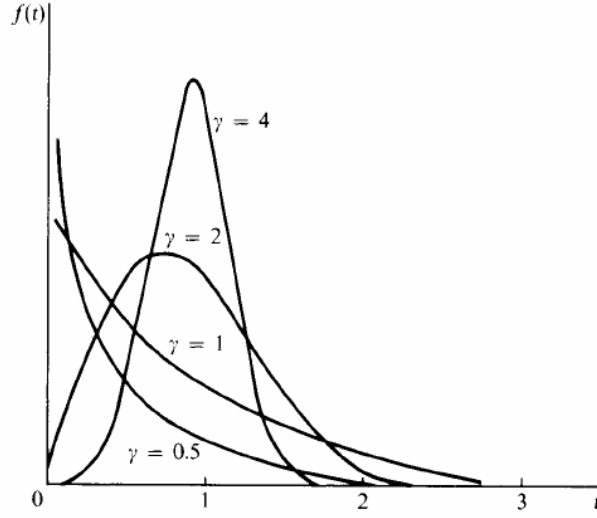
3.1.3. Weibull Dağılım

Parametrik sağkalım dağılımları içinde en yaygın kullanılan dağılımlardan biri olan Weibull dağılımı, üstel dağılımın genelleştirilmiş halidir. Ancak üstel dağılımın aksine sabit Hazard oranına sahip olmadığı için, çok geniş uygulama alanına sahiptir. Güvenirlilik, hastalık ve ölüm ile ilgili hesaplamalarda sıkça kullanılmaktadır (Türkan, 2007).

Weibull dağılımının γ ve λ ile gösterilen iki parametresi vardır. γ şekil parametresi, λ ise ölçek parametresi olarak tanımlanmaktadır. Weibull dağılımın olasılık yoğunluk fonksiyonu aşağıdaki gibi ifade edilir.

$$f(t) = \lambda \gamma (\lambda t)^{\gamma-1} \exp(-\lambda t)^\gamma \quad (3.27)$$

Dağılımın olasılık yoğunluk fonksiyonuna ait grafik ise Şekil 3.6.'daki gibidir.

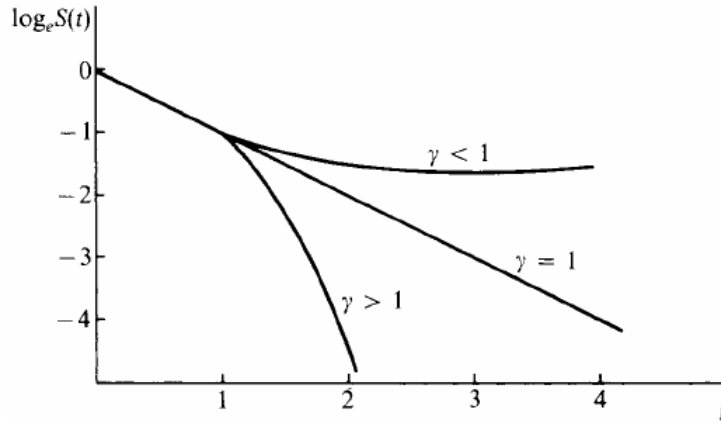


Şekil 3.6. Farklı γ şekil ve $\lambda=1$ ölçek parametrelili Weibull dağılımının olasılık yoğunluk Fonksiyonunun grafiği

Weibull dağılımına ilişkin sağkalım fonksiyonu ise aşağıdaki gibi ifade edilir.

$$S(t) = 1 - F(t) = \exp(-(\lambda t)^\gamma) \quad (3.28)$$

$\lambda=1$ ölçek parametrelili Weibull dağılımının farklı şekil parametreleri için $\log S(t)$ sağkalım zamanına göre değişimi Şekil 3.7’de verilmektedir.



Şekil 3.7. $\lambda=1$ Ölçek parametrelili Weibull dağılımının $\log S(t)$ eğimi

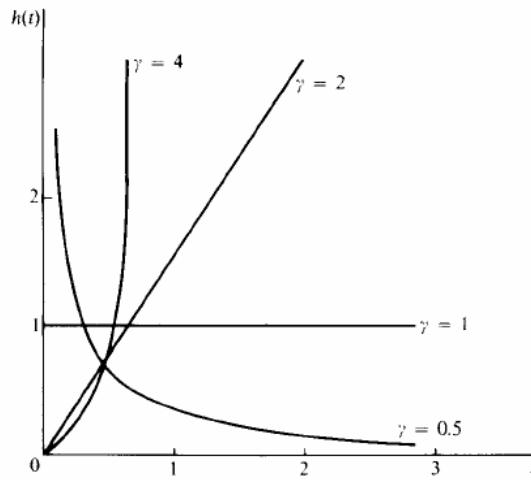
Şekil 3.7’de görüldüğü gibi izlem süresi uzadığında, $\gamma=1$ için sağkalım fonksiyonu doğrusal olarak azalırken, $\gamma < 1$ için sağkalım fonksiyonu yavaşça azalmakta ve sabit

bir değere yaklaşmaktadır. Ayrıca $\gamma > 1$ için sağkalım fonksiyonu hızlı bir şekilde sifira yaklaşmaktadır.

Weibull dağılımının Hazard fonksiyonu (Hazard oranı);

$$h(t) = \alpha \gamma t^{\gamma-1} \quad (3.29)$$

ile ifade edilir ve dağılımın Hazard fonksiyonuna ait grafik aşağıdaki şekilde verilmiştir.



Şekil 3.8. $\lambda=1$ için Weibull dağılımının Hazard fonksiyonunun grafiği

Şekilde $\lambda=1$ ve $\gamma = 0.5, 1, 2, 4$ değerleri için, γ 'nın sağkalım zamanı ile olan ilişkisi gösterilmektedir. Üstel dağılımda olduğu gibi izlem süresi uzadıkça, $\gamma=1$ için Hazard oranı sabit kalmaktadır. Zaman ilerlediğinde Hazard oranı $\gamma > 1$ için yine artış gösterirken, $\gamma < 1$ için azalmaktadır. Bu nedenle Weibull dağılımı hem azalan hem artan hem de sabit Hazard oranlı sağkalım dağılımlarını modellemede kullanılabilir. Azalan ve artan Hazard oranı için en belirgin örnekler sırasıyla akciğer kanserli hastaların ve başarılı ameliyat geçiren hastaların dağılımıdır. Weibull dağılımından elde edilen model, hem HF hem de OH özelliğinin sağlanması için γ şekil parametresinin, prognostik faktörlerin farklı düzeyleri için değişmemesi gerekmektedir. Weibull dağılımında HF ve OH özelliklerinin incelenmesinde, $\ln[-\ln \hat{S}(t)] - \ln(t)$ grafiğinden yararlanır. $\ln[-\ln \hat{S}(t)] - \ln(t)$ grafiğinin iki veya

daha fazla prognostik faktörün farklı düzeyleri için çizilmesi sonucunda karşılaşılan farklı durumlar aşağıdaki gibi açıklanabilir:

- Paralel veya düz olarak elde edilen doğru; dağılımın Weibull olduğunu ayrıca OH ve HF'nin da sağlandığını gösterir.
- Paralel ve "1" eğimine sahip düz doğru; dağılımın üstel olduğunu ayrıca OH ve HF'nin sağladığını gösterir.
- Paralel ancak düz olmayan doğru; dağılımın Weibull olduğunu, HF'nin sağlanmadığını ancak OH'nin sağlandığını gösterir. Dolayısıyla yarı parametrik model kullanılabilir.
- Paralel ve düz olmayan doğru; dağılımın Weibull olduğunu ve OH'nin ihlal edildiğini gösterir.
- Paralel olmayan düz doğru; dağılımın Weibull ancak OH ve HF'nin ihlal edildiğini gösterir. Ayrıca farklı prognostik faktör düzeyleri için sabit olmayan γ 'lere sahip olduğu anlamına gelir.

Weibull dağılımına dayanarak elde edilen modelin diğer bir özelliği ise; logaritması alınmış sağkalım süresine göre doğrusal olmasıdır. Sağkalım çalışmasında, $t = (t_1, \dots, t_n)'$ 'nin sağkalım zamanını, $v = (v_1, \dots, v_n)'$ 'nin durum değişkenini (olgunun gerçekleşmesi durumunda $v_i = 1$ ve sansürlü olması durumunda $v_i = 0$ değerlerini alır) ve $D = (n, y, v)'$ 'nin ise gözlenen verileri gösterdiğini varsayalım. Sağkalım zamanının (λ, γ) parametrelili Weibull dağılıma sahip olduğu varsayımı altında (λ, γ) 'ya ait olabilirlik fonksiyonu;

$$L(\lambda, \gamma | D) = \prod_{i=1}^n f(t_i | \lambda, \gamma)^{v_i} S(t_i | \lambda, \gamma)^{(1-v_i)} \quad (3.30)$$

$$= \lambda^d \exp \left\{ d\gamma + \sum_{i=1}^n (v_i (\lambda - 1) \log(t_i) - \exp(\gamma) t_i^\lambda) \right\}$$

ile elde edilir. Burada $d = \sum_{i=1}^n v_i$ dır.

3.1.3.1. Weibull Dağılıma Sahip Sağkalım Verisine İlişkin Bayesci Yaklaşım

Bayesçi yaklaşımda (λ, γ) parametre tahminleri, sağkalım verisinden elde edilen olabilirlik fonksiyonu ile (λ, γ) parametresine ait $f(\lambda, \gamma)$ önsel dağılımın birleştirilmesiyle oluşan sonsal dağılım yardımıyla yapılır. (λ, γ) 'nın sonsal dağılımı,

$$f(\lambda, \gamma | D) \propto f(D | \lambda, \gamma) f(\lambda, \gamma) \quad (3.31)$$

elde edilir. (λ, γ) 'nın parametreleri bilinmediği varsayıldığında ortak bilgi verici olmayan önsel dağılımlardan, eşlenik önsel dağılım elde edilemez. Bu durumda ortak önsel tanımlanması için λ ve γ 'nın bağımsız olduğu varsayılır. Burada λ 'nın α_0 ölçek ve k_0 şekil parametrelili $\zeta(\alpha_0, k_0)$ Gamma dağılımına ve γ 'nın da μ_0 ortalamalı ve σ_0^2 varyanslı Normal dağılıma $N(\mu_0, \sigma_0^2)$ sahip olduğu varsayılır. Bu durumda (λ, γ) parametreleri için eşlenik önsel dağılımlar aşağıdaki eşitlik ile tanımlanır.

$$f(\lambda | \alpha_0, k_0) \propto \alpha^{\alpha_0 - 1} \exp(-k_0 \alpha) \quad (3.32a)$$

$$f(\gamma | \mu_0, \sigma_0^2) \propto -\frac{1}{2\sigma_0^2} \exp(\lambda - \mu_0)^2 \quad (3.32b)$$

Yukarıda (3.30) eşitliği ile verilen olabilirlik ve (3.32) eşitliği ile verilen önsel dağılımlar birleştirildiğinde, (λ, γ) parametresine ait sonsal dağılım;

$$\begin{aligned} f(\lambda, \gamma | D) &\propto L(\lambda, \gamma | D) f(\lambda | \alpha_0, k_0) f(\gamma | \mu_0, \sigma_0^2) \\ &\propto \left[\prod_{i=1}^n f(t_i | \lambda, \gamma)^{v_i} S(t_i | \lambda, \gamma)^{(1-v_i)} \right] f(\lambda | \alpha_0, k_0) f(\gamma | \mu_0, \sigma_0^2) \\ &= \lambda^{\alpha_0 + d - 1} \exp \left\{ d\gamma + \sum_{i=1}^n (v_i (\lambda - 1) \log(t_i) - \exp(\gamma) t_i^\alpha) - k_0 \lambda - \frac{1}{2\sigma_0^2} (\gamma - \mu_0)^2 \right\} \end{aligned} \quad (3.33)$$

şeklinde elde edilir. Burada $d = \sum_{i=1}^n v_i$ olarak ifade edilir. (λ, γ) ortak sonsal dağılımının λ ve γ parametrelerine göre marjinal dağılımlarının integral hesaplamalarındaki karmaşıklık nedeniyle $[\lambda|\gamma, D]$ ve $[\gamma|\lambda, D]$ koşullu dağılımlar benzetim yöntemleri ile çözülmektedir. $[\lambda|\gamma, D]$ ve $[\gamma|\lambda, D]$ 'nin log-konkav olması benzetim yöntemlerinden olan Gibbs örneklemesinin uygulanabilirliğini artırmaktadır.

Weibull dağılımına sahip sağkalım zamanına etki eden prognostik faktörlerin tahmin edilmesi amacıyla, γ parametresi px1 boyutlu x'_i prognostik faktörler vektörü ile $\gamma_i = (x'_i\beta)$ şeklinde bir ilişki içinde bulunduğunu belirtmek gerekir. Burada β , px1 prognostik faktörlere ait katsayıları göstermektedir. Bayesçi yaklaşımda her bir parametre rasgele değişken olarak kabul edildiğinden, β parametresi için yaygın olarak kullanılan önsel dağılımların varlığından söz etmek gerekir. Dağılımlar; uygun olmayan düzgün dağılım $\pi(\beta) \propto 1$ ve normal önsel dağılımdır. β 'nin $N_p(\mu_0, \Sigma_0)$ önsel dağılıma ve α 'nın ise α_0 ölçek ve k_0 şekil parametrelili $\zeta(\alpha_0, k_0)$ Gamma dağılımına sahip olduğu varsayıldığında ortak sonsal dağılım;

$$f(\beta, \lambda|D) \propto \lambda^{\alpha_0+d-1} \exp\left\{\sum_{i=1}^n v_i x'_i \beta + v_i (\lambda - 1) \log(t_i) - t_i^\lambda \exp(x'_i \beta)\right\} - k_0 \lambda - \frac{1}{2} (\beta - \mu_0) \Sigma_0^{-1} (\beta - \mu_0) \} \quad (3.34)$$

olarak ifade edilir. Yukarıdaki eşitlikte $D = (x, y, X, v)$ ve X , i. satırı x'_i 'lerden oluşan nxp boyutlu bağımsız değişkene ait matris olarak tanımlanır.

β 'nin sonsal dağılımının elde edilmesi karmaşık integral hesaplamaları gerektirdiğinden benzetim yöntemi uygulanır. $[\lambda|\beta, D]$ ve $[\beta|\lambda, D]$ koşullu dağılımlarının log-konkavlık özelliği sağlamasından dolayı benzetim yöntemlerinden Gibbs örnekleme kullanılır.

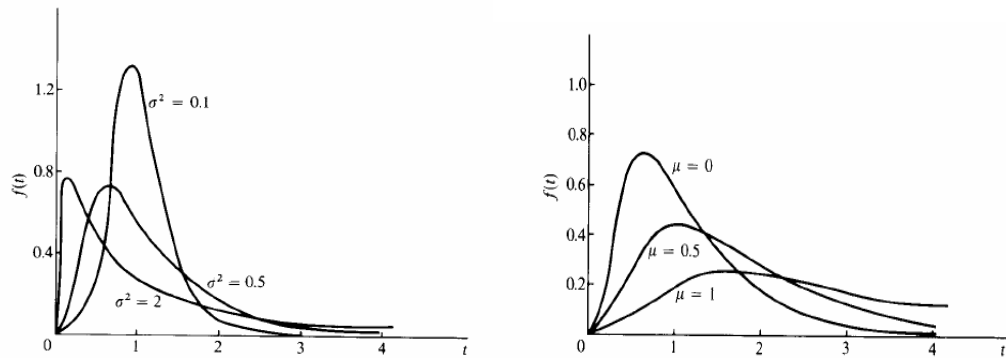
3.1.4. Lognormal Dağılım

Yaygın kullanılan sağkalım dağılımlarından biri de Lognormal dağılımdır. Dağılımın sıkça tercih edilmesinin en önemli nedeni; logaritması alınan rasgele değişkenin kümülatif değerinin, standart normal dağılımdan elde edilebilmesidir. Buradan hareketle dağılıma ait olasılıklar kolayca bulunabilmektedir. Lognormal dağılım, sabit olmayan Hazard oranı gösteren veriler için uygun bir yapı gösterir. Bu tür veriler için yapılan logaritmik dönüşüm yardımıyla, dağılımın yaklaşık olarak simetrik olması sağlanır (Evans ve diğ., 2000). Burada dikkat edilmesi gereken nokta dönüşümün uygulanabilmesi için rasgele değişkeninin pozitif değerler alması gerektiğidir (Türkan, 2007).

Ölçek parametresi $m = e^\mu > 0$, şekil parametresi $\sigma > 0$ ve rasgele değişken $0 \leq t < +\infty$ aralığında tanımlanmak üzere olmak üzere Lognormal dağılımın olasılık yoğunluk fonksiyonu;

$$f(t) = (2\pi)^{-1/2} (t\sigma)^{-1} \exp\left\{-\frac{1}{2\sigma^2} (\log(t) - \mu)^2\right\} \quad (3.35)$$

olarak tanımlanır. Dağılıma ilişkin olasılık yoğunluk fonksiyonunun grafiği ise aşağıdaki gibidir.



Şekil 3.9. Sırasıyla $\mu=0$ için değişen σ^2 ve $\sigma^2=0.5$ için değişen μ değerleri için Lognormal olasılık yoğunluk fonksiyonlarının grafiği

Lognormal dağılımın dağılım fonksiyonu ise (3.36)'da verildiği gibidir (Evans ve diğ., 2000).

$$F(t) = \Phi\left(\frac{\ln t - \mu}{\sigma}\right) \quad (3.36)$$

Eşitlikte $\Phi(\cdot)$, standart normal dağılımın birikimli dağılım fonksiyonudur ve aşağıdaki gibi ifade edilir:

$$\Phi(t) = \frac{1}{\sqrt{2\pi}} \int_0^t e^{-\frac{t^2}{2}} dt \quad (3.37)$$

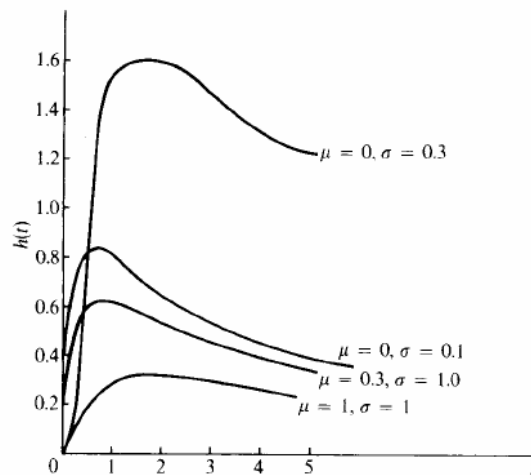
Lognormal dağılıma ait sağkalım fonksiyonu,

$$S(t) = 1 - \Phi\left(\frac{\log(t) - \mu}{\sigma}\right) \quad (3.38)$$

olarak, risk (Hazard) fonksiyonu ise,

$$h(t) = \frac{f(t)}{S(t)} = \frac{(2\pi)^{-1/2} (t\sigma)^{-1} \exp\left\{-\frac{1}{2\sigma^2} (\log(t) - \mu)^2\right\}}{1 - \Phi\left(\frac{\log(t) - \mu}{\sigma}\right)} \quad (3.39)$$

olarak ifade edilir. Dağılımın Hazard fonksiyonuna ait grafik Şekil 3.10'da verildiği gibidir (Evans ve diğ., 2000).



Şekil 3.10. Lognormal dağılımın Hazard fonksiyonunun grafiği

Grafikten de görüleceği gibi, Lognormal dağılımın Hazard fonksiyonu $t=0$ için sıfır değerini alır, maksimum noktaya ulaştıktan sonra azalan bir eğim gösterir. İzlem süresi uzadıkça sifira yaklaşır.

Lognormal dağılımına sahip sağkalım verisine ilişkin model, HF'yi sağlarken, OH'yi sağlamamaktadır. Bu durum göz önünde bulundurularak bir sağkalım çalışmasında; $t = (t_1, \dots, t_n)'$ 'nin sağkalım zamanını, $v = (v_1, \dots, v_n)'$ 'nin durum değişkenini ve $D = (n, y, v)'$ 'nin gözlenen verileri gösterdiğini varsayalım. Sağkalım zamanının (μ, σ^2) parametrelili ve Lognormal dağılıma sahip olduğu varsayımı altında, (μ, σ^2) 'ya ait olabilirlik fonksiyonu;

$$\begin{aligned}
 L(\mu, \sigma^2 | D) &= \prod_{i=1}^n f(t_i | \mu, \sigma)^{v_i} S(t_i | \mu, \sigma)^{(1-v_i)} \\
 &= (2\pi\sigma^2)^{-d/2} \exp\left\{-\frac{1}{2\sigma^2} \sum_{i=1}^n v_i (\log(t_i) - \mu)^2\right\} \\
 &\times \prod_{i=1}^n t_i^{-v_i} \left(1 - \Phi\left(\frac{\log(t_i) - \mu}{\sigma}\right)\right)^{(1-v_i)}
 \end{aligned} \tag{3.40}$$

ile ifade edilir.

3.1.4.1. Lognormal Dağılıma Sahip Sağkalım Verisine İlişkin Bayesci Yaklaşım

Bayesci yaklaşıma göre (μ, σ^2) parametrelerinin tahminleri; olabilirlik fonksiyonu ve (μ, σ^2) parametresine ait $f(\mu, \sigma^2)$ önsel dağılımın birleştirilmesi ile meydana gelen sonsal dağılım yardımıyla elde edilir. (μ, σ^2) 'nin sonsal dağılımı,

$$f(\mu, \sigma^2 | D) \propto L(D | \mu, \sigma^2) f(\mu, \sigma^2) \tag{3.41}$$

şeklinde elde edilir. Bilinmeyen (μ, τ) parametreleri için $\tau = 1/\sigma^2$ dönüşümü yapılarak bilinmeyen (μ, τ) parametreleri için, açıklayıcı olmayan önsel dağılımlardan hareketle eşlenik önsel dağılımı elde etmek mümkün değildir. Bu durumda ortak bir önsel tanımlaması yapılır, ayrıca μ ve τ 'nin bağımsız olduğu

varsayılr. μ 'nın μ_0 ortalamalı, $\frac{1}{\tau\tau_0}$ varyanslı normal dağılıma $N(\mu_0, \frac{1}{\tau\tau_0})$ ve τ 'nin $\alpha_0/2$ ölçek parametrelili ve $\lambda_0/2$ şekil parametrelili Gamma $\tau \sim \zeta(\alpha_0/2, \lambda_0/2)$ sahip olduđu varsayılr. Buradan hareketle (μ, τ) parametreleri için eşlenik önsel dağılımlar aşağıdaki gibi ifade edilir:

$$f(\mu|\mu_0, \frac{1}{\tau\tau_0}) \propto -\frac{1}{2\tau\tau_0} \exp(\mu - \mu_0)^2 \quad (3.42a)$$

$$f(\tau|\alpha_0/2, \lambda_0/2) \propto \tau^{(\alpha_0/2)-1} \exp(-(\lambda_0/2)\tau) \quad (3.42b)$$

(3.40) ile verilen olabilirlik fonksiyonu ile (3.42) ile verilen önsel dağılımları birleştirildiğinde, (μ, τ) parametresine ait sonsal dağılım;

$$\begin{aligned} f(\mu, \tau|D) &\propto L(\mu, \sigma|D) f(\mu, \tau|\mu_0, \tau_0, \alpha_0, \lambda_0) \\ &\propto \tau^{\frac{\alpha_0+d}{2}-1} \exp\left\{-\frac{\tau}{2} \left[\sum_{i=1}^n v_i (\log(t_i) - \mu)^2 + \tau_0 (\mu - \mu_0)^2 + \lambda_0 \right]\right\} \\ &\times \prod_{i=1}^n t_i^{-v_i} (1 - \Phi(\tau^{1/2} (\log(t_i) - \mu)))^{(1-v_i)} \end{aligned} \quad (3.43)$$

şeklinde elde edilir.

(μ, τ) ortak sonsal dağılımı karmaşık bir fonksiyonel yapıya sahip olduğundan daha önce verilen diğer parametrik modellerde olduğu gibi $f(\mu, \tau|D)$ 'den $[\mu|\tau, D]$ ve $[\tau|\mu, D]$ koşullu dağılımların elde edilmesi zorlaşır. Bu nedenle benzetim yöntemlerine başvurulur. Ancak burada dikkat edilmesi gereken nokta $[\mu|\tau, D]$ koşullu sonsal dağılımı log-konkav olup, $[\tau|\mu, D]$ koşullu sonsal dağılımının genellikle log-konkav olmamasıdır. Koşullu dağılımlardan birinin log-konkavlık özelliğini sağlamaması, araştırmacıyı benzetim yöntemlerinden Metropolis algoritmasının kullanımına yönlendirir.

Lognormal dağılımlı Sağkalım zamanına etki eden prognostik faktörlerin etkilerinin tahmin edilmesi amacıyla, μ parametresi $p \times 1$ boyutlu x_i' prognostik faktörler vektörü ile $\mu_i = (x_i' \beta)$ şeklinde bir ilişki içinde tanımlanır. Burada β , $p \times 1$ prognostik faktörlere ait katsayılarıdır. Bayesçi yaklaşımda her bir parametre rasgele değişken olarak kabul edildiğinden, β parametresi için yaygın olarak kullanılan önsel dağılımların varlığından söz etmek gerekir. Bu önsel dağılımlar; uygun olmayan düzgün dağılım $\pi(\beta) \propto 1$ ve normal önsel dağılımdır. β 'nin $\beta|\tau \sim N_p(\mu_0, \tau^{-1}\Sigma_0)$ normal önsel dağıldığı varsayıldığında ortak sonsal dağılım;

$$f(\beta, \tau|D) \propto \tau^{\frac{\alpha_0+d}{2}-1} \exp\left\{-\frac{\tau}{2}\left[\sum_{i=1}^n v_i (\log(t_i) - x_i' \beta)^2 + (\beta - \mu_0)' \Sigma_0^{-1} (\beta - \mu_0) + \lambda_0\right]\right\} \times \prod_{i=1}^n t_i^{-v_i} (1 - \Phi(\tau^{-1/2} (\log(t_i) - x_i' \beta)))^{(1-v_i)} \quad (3.44)$$

ile verilir. Burada $D = (x, y, X, v)$ olarak ve X , i . x_i' 'lerden oluşan $n \times p$ boyutlu bağımsız değişkene ait matris olarak tanımlanır.

(3.44) eşitliğinden β 'ya ait sonsal dağılımın elde edilmesi karmaşık integral hesaplamaları gerektirdiğinden nümerik integrasyon veya MCMC yönteminin kullanılmasına ihtiyaç duyulur. Lognormal regresyon modeli için $[\beta|\tau, D]$ koşullu sonsal dağılımı log-konkav özelliğini sağladığından Gibbs örnekleme kullanılır. Ancak $[\tau|\beta, D]$ log-konkavlık özelliğini sağlamadığından örneklem çekmek için Metropolis algoritmasının kullanılması daha uygundur.

3.2. YARI PARAMETRİK MODELLER

1972 yılında David Cox tarafından ortaya atılan yarı parametrik sağkalım analizi, olguya ulaşılmış sağkalım süresine etki eden prognostik faktörlerin elde edilmesi için kullanılan bir yöntemdir.

3.2.1. Kısmi Olabilirlik

Sağkalım süresine etki eden prognostik faktörlerin tahminlerinde kullanılan Cox model, Hazard fonksiyonuna dayanarak elde edilir (Van Belle ve diğ., 2004) ve aşağıdaki gibi tanımlanır.

$$h(t) = h_0(t)e^{x_i\beta} \quad (3.45)$$

Burada $x = (x_1, \dots, x_n)$ prognostik faktörlere ait vektör ve $h_0(t)$, temel Hazard fonksiyonu olarak adlandırılır. $h_0(t)$ 'ye "temel" denilmesinin nedeni $x = 0$ olduğunda $h(t) = h_0(t)$ olmasıdır. Yarı parametrik bir model olan Cox regresyon modeli, $h_0(t)$ 'nin herhangi bir dağılıma uygunluğunu öngörmez (Cox, 1972, 1975). Bu nedenle β parametre tahmininde $h_0(t)$ 'ye bağlı olmayan bir olabilirlik fonksiyonunun tanımlanması gerekir. Tanımlanan bu olabilirlik fonksiyonuna kısmi olabilirlik fonksiyonu adı verilir. Kısmi olabilirlik fonksiyonu; Hazard'ı, sıralanan sağkalım süreleri içerisinde olguya ulaşan gözlemler için tanımlar. Kısmi olabilirlik fonksiyonun kullanılabilmesi için, bir sağkalım zamanında ikiden fazla olgu gerçekleşmemelidir. Yani veride herhangi bir bağımlılık olmamalıdır. Bağımlılık olması durumunda Breslow veya Efron yaklaşımları kullanılabilir (BIOST, 2004). Buna göre kısmi olabilirlik aşağıdaki şekilde tanımlanır.

$$PL(\beta) = \prod_{i=1}^n \left(\frac{e^{x_i'\beta}}{\sum_{l \in R_i} e^{x_l'\beta}} \right)^{v_i} \quad (3.46)$$

(İbrahim ve ark, 2001). Burada R_i riski, i . olgunun gerçekleşme zamanında, olguya ulaşma riski altında olan bireylerin oluşturduğu risk setini göstermektedir.

3.2.2. Oransal Hazard Modeli (OHM)

Sağkalım süresine etki eden prognostik faktörlerin Hazard fonksiyonuna dayanarak yazılan Cox model, farklı gözlemlerin Hazard fonksiyonlarının birbirine oranının sabit olması halinde orantısal Hazard modeli (OHM) olarak tanımlanır.

3.2.2.1. Oransal Hazard Modeli için Bayesci Yaklaşım

Bayesçi yaklaşımda β parametre tahminleri, kısmi olabilirlik fonksiyonu ve β parametresine ait $f(\beta)$ önsel dağılımın birleştirilmesiyle oluşan sonsal dağılım yardımıyla elde edilir. β 'nın sonsal dağılımı,

$$f(\beta|D) \propto f(D|\beta)f(\beta) \quad (3.47)$$

ile elde edilir. Yukarıdaki eşitlikte $D = (x, y, X, v)$ ve X , i . satırı x_i 'lerden oluşan $n \times p$ boyutlu bağımsız değişkene ait matristir. β parametresinin tahmininde en yaygın kullanılan önsel dağılımlar uygun olmayan düzgün dağılım $\pi(\beta) \propto 1$ ve normal önsel dağılımdır. β 'nın μ_0 ortalamalı ve σ_0^2 varyanslı normal dağılıma $N(\mu_0, \sigma_0^2)$ sahip olduğu varsayılırsa, β parametresine ait eşlenik önsel dağılımlar aşağıdaki gibi ifade edilir:

$$f(\beta|\mu_0, \sigma_0^2) \propto -\frac{1}{2\sigma_0^2} \exp(\lambda - \mu_0)^2 \quad (3.48)$$

(3.46) ile verilen kısmi olabilirlik ve (3.48) ile verilen önsel dağılım birleştirilerek, β parametresine ait sonsal dağılım;

$$f(\beta|D) \propto PL(\beta|D)f(\beta|\mu_0, \sigma_0^2) \quad (3.49)$$

şeklinde elde edilir. β 'ya ait sonsal dağılımın çözümünde ise MCMC yöntemi kullanılır.

3.3. MODEL PARAMETRESİNİN ANLAMLILIĞI

Parametrik veya yarı parametrik Bayesci sağkalım analizinde, prognostik faktörlere ait katsayıların anlamlılığı; α anlamlılık seviyesinde $H_0: \beta = 0$ sıfır hipotezine karşılık $H_1: \beta \neq 0$ alternatif hipotezi ile test edilir. Bayesci yaklaşımda önsel

dağılım sürekli olduğunda buna bağlı olarak sonsal dağılım da sürekli olacağı, ayrıca sürekli bir dağılımın herhangi bir noktadaki integrali “sıfır” olacağı için katsayılarla ilişkin hipotez test edilemez. Bu nedenle β için hipotez testi $(1-\alpha)\%100$ güvenilirlik aralığından yararlanılarak yapılır. Bayesci yaklaşımda hipotez testinin değerlendirmesi, güven aralığının “sıfır” değerini içerip içermemesine göre yapılır. Aralık “sıfır”ı içermiyorsa sıfır hipotezi “RED” aksi durumda kabul edilir. Güven aralıkları çekilen örneklemelerin dağılımı yardımıyla yapıldığından, elde edilen olasılıklar önsel bilgi niteliğindedir. Dolayısı ile Bayesci güven aralığında, gözlenen veriler ve önsel dağılımın birleştirilmesiyle elde edilen sonsal dağılım bir önbilginin sonucudur. (Bolstad, 2007).

3.4. MODEL SEÇİM YÖNTEMLERİ

Sağkalım modellerinde uygun olan modele karar verebilmek için sapma istatistiği ($-2\log L$), Akaike bilgi kriteri (Akaike’s information criterion, AIC) ve Bayeci bilgi kriteri (BIC) kullanılmaktadır. En küçük sapma istatistiğine, AIC veya BIC değerine sahip model en uygun model olarak kabul edilir (Klein ve diğ., 1997).

3.4.1. Sapma İstatistiği ($-2\log L$)

Uyum iyiliği testi olarak kullanılan sapma istatistiği, olabilirlik oran istatistiğini kullanmaktadır. Bu yöntemde amaç; prognostik faktörler ile kurulan modelin hiçbir prognostik faktör içermeyen modelden ne kadar iyi olduğunu ortaya koyabilmektir. Sapma istatistiği, p sayıda prognostik faktörün bulunduğu modelden elde edilen (indirgenmiş) olabilirlik fonksiyonu ($L[indirgenmiş]$) ile prognostik faktör içermeyen modeli ($L[tam]$) karşılaştırır. Sapma;

$$Sapma = -2 \ln \left[\frac{L[indirgenmiş]}{L[tam]} \right] \quad (3.50)$$

şeklinde elde edilir. Bayesci yaklaşımda sapma istatistiği, sonsal yoğunluklardan yararlanır ve en iyi sonsal tahmin değerlendirmesi aşağıdaki eşitlik kullanılarak yapılır (Congdon, 2006).

$$D(\bar{\beta}) = -2\log(p(y|\bar{\beta})) \quad (3.51)$$

Burada $\bar{\beta}$ sonsal ortalama vektörüdür ve $\bar{\beta} = \frac{1}{M} \sum_{m=1}^M \beta^m$ ile hesaplanır.

3.4.2. Akaike Bilgi Kriteri (AIC)

Bir model seçim kriteri olan Akaike bilgi kriteri (AIC), model değerlendirmede oldukça yaygın olarak kullanılan bir ölçüdür (Congdon, 2006). Parametrik yaşam modelleri için Akaike bilgi kriteri,

$$AIC = -2\log(L(\hat{\beta})) + 2(p + 2 + k) \quad (3.52)$$

eşitliği ile ifade edilir. Burada p, parametre sayısını; $\hat{\beta}$, β parametresinin en çok olabilirlik tahmini ve $L(\hat{\beta})$, olabilirlik fonksiyonunu ifade etmektedir.

AIC değerleri hesaplanırken üstel regresyon modeli için k=0; Weibull, log-lojistik ve log-normal regresyon modelleri için k =1 ve Gamma regresyon modeli için k = 2 olarak alınır (Klein ve diğ., 1997).

3.4.3. Bayesci Bilgi Kriteri (BIC)

Yaygın kullanılan model seçim yöntemlerinden bir diğeri de Bayeci bilgi kriteri (BIC) dir. Bu kriterdeki log-olabilirlik; parametre sayısına (p) ve olguya ulaşmış toplam gözlem sayısına dayanmaktadır. BIC aşağıdaki gibi tanımlanmaktadır: (Congdon, 2006)

$$BIC = -2\log(L(\hat{\beta})) + p \log(N_e) \quad (3.53)$$

Burada p, parametre sayısını ve N_e ise olguya ulaşmış gözlemlerin sayısını göstermektedir.

4. UYGULAMA

Çalışmanın uygulama bölümünde; Klasik parametrik sağkalım analizi (KPSA) ile açıklayıcı ve açıklayıcı olmayan önsel dağılım bilgisine dayanan Bayesci sağkalım analizi yöntemlerinin karşılaştırılması yapılarak, parametre tahminlerinin elde edilmesi amaçlanmıştır. Bu amaçla 1993-2002 yılları arasında İstanbul Üniversitesi Tıp Fakültesi Meme Cerrahi Kliniğinde tedavi gören 458 meme kanseri hastasının sağkalım süresine etki eden prognostik faktörler incelenmiş, KPSA ve Bayesci yaklaşımlara göre parametre tahminleri elde edilmiş ve sonuçlar değerlendirilmiştir.

4.1. KULLANILAN PROGRAMLAR

Analizler, EasyFit 5.0 (EasyFit 2008), SPlus 2000 paket ve WinBUGS programlarında yapılmıştır. Olguya ulaşmış (Ölen) 69 meme kanserli hastalara ait sağkalım sürelerinin dağılımının tanımlanması için (olasılık yoğunluk, Hazard grafiklerinin elde edilmesi ve test istatistiklerinin elde edilmesi) EasyFit 5.0 (EasyFit 2008), belirtici istatistiklerin hesaplanması ve klasik parametrik sağkalım analizi için SPlus 2000, Bayesci sağkalım analizi için ise WinBUGS dilinde macro yazılmıştır.

4.2. TANIMLAYICI İSTATİSTİKLER VE KLASİK PARAMETRİK SAĞKALIM ANALİZİ

Çalışmada ele alınan meme kanserli hastaların 9 yıllık izlem süresi sona erdiğinde, 458 hastadan 389'sının (%84) hayatta kaldığı (sağdan sansürlü) ve 69'ünün (%16) öldüğü gözlenmiştir. Ayrıca hastaların sağkalım sürelerini etkileyen prognostik faktörler; yaş, aile hikayesi, menepoz durumu, tümör büyüklüğü, TNM evresi, histolojik tip, östrojen reseptör durumu, progesteron reseptör durumu, tedavi türü (radyoterapi ve hormon terapisi) ve aksiller lenf nodu tutulumu olarak belirlenmiştir. Çalışmada ele alınan prognostik faktörler içinden yaş ve tümör büyüklüğü sürekli değişken olduğundan veride herhangi bir işleme tabi tutulmadan yer almıştır. Menepoz durumu (var, yok), histolojik tip (diğer, duktal), östrojen reseptör durumu

(negatif, pozitif), progesteron reseptör durumu (negatif, pozitif), tedavi türü (radyoterapi ve hormon terapisi), aksiller lenf nodu tutulumu (negatif, pozitif) ise dikotom yani iki kategorili değişkenler olarak gözlenmiştir. Değişkenlerin referans kategorisine (seçilen kategoriye) 1, diğer kategoriye 0 değeri verilerek kodlama yapılmıştır. İki'den fazla kategorili prognostik faktörlerden aile hikayesi (yok, meme kanseri, diğer kanserler) ve TNM evresi (0, IIA, IIB) için referans kategori olarak ilk kategori (genellikle ya ilk ya da sonuncu kategori seçilir) seçilmiş ve diğer kategorilere göre karşılaştırma yapılmıştır. Değişkenlere verilen kategorik değerlerin sayısal herhangi bir anlamı olmadığı için kukla değişken kullanılmıştır. Kukla değişkenler, 0-1 değerini alan değişkenlerdir. Kukla değişken olarak tanımlanacak değişkenin, ilgilenilen kategorisine 1 aksine 0 değeri verilerek tanımlanır ve sayısı, kategori sayısından bir eksik olarak belirlenir. Aile hikayesi ve TNM evresi değişkenlerinin de yukarıda ifade edildiği gibi üç kategorisi bulunmakta ve üç kategorili bir değişkene ait kukla değişken gösterimi aşağıdaki gibi ifade edilmektedir. Çalışmamızda referans kategorileri, daha önce yapılan meme kanseri çalışmalarında ele alınan referans kategorileri baz alınarak belirlenmiştir.

Çizelge 4.1. Üç kategorili değişkene ait kukla değişkenler

<i>Kategoriler</i>	<i>Kukla Değişkenler</i>	
	<i>K1</i>	<i>K2</i>
A (Referans Kategori)	0	0
B	1	0
C	0	1

Öncelikle meme kanserli hastalara ilişkin değişkenlerin dağılımı ile ilgili genel bilgiler aşağıdaki çizelgede özetlenmiştir.

Çizelge 4.2. Meme kanserli hastalara ilişkin değişkenlerin dağılımı

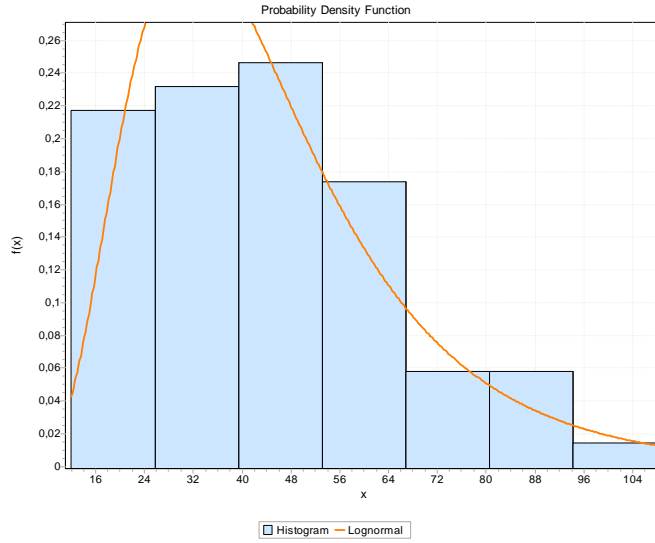
<i>Bağımsız Değişkenler</i>	<i>Hasta Sayısı</i>	<i>Yüzde</i>	<i>Ölen Hasta Sayısı</i>	<i>Sağkalan Hasta Sayısı</i>
Yaş (≤ 50)	232	51	44	188
Yaş (>50)	226	49	25	201
Aile Hikayesi (Yok)	370	80	58	312
Aile Hikayesi (Meme kanseri)	55	12	7	48
Aile Hikayesi (Diğer Kanser)	33	8	4	29
Menepoz Durumu (yok)	329	71	55	274
Menepoz Durumu (var)	129	29	14	115
Tümör Büyüklüğü (≤ 24)	243	53	24	219
Tümör Büyüklüğü (> 24)	215	47	45	170
Histolojik Tip (Diğer)	95	21	21	74
Histolojik Tip (Duktal)	363	79	48	315
TNM Evresi (0)	123	26	18	105
TNM Evresi (IIA)	188	41	24	164
TNM Evresi (IIB)	147	33	27	120
Östrojen Reseptör Durumu (Negatif)	176	38	48	128
Östrojen Reseptör Durumu (pozitif)	282	62	21	261
Progesteron Reseptör Durumu (Negatif)	248	54	54	194
Progesteron Reseptör Durumu (pozitif)	210	46	15	195
Tedavi TÜRÜ (MKC)	133	29	23	110
Tedavi TÜRÜ (MRM)	325	71	46	279
Aksiller Lenf Nodu Tutulumu (Negatif)	238	52	52	186
Aksiller Lenf Nodu Tutulumu (Pozitif)	220	48	17	203

* MKC: Meme koruyucu cerrahi, MRM: Modifiye Radikal mastektomi

Meme kanserli hastalara ait veride sadece yaş ve tümör büyüklüğü sürekli değişken, diğer değişkenler kategorik değişkenler olduğundan sadece bu değişkenlere ait ortalama ve standart sapma gibi istatistiklerin verilmesi uygun görülmüştür. Buna göre yaş değişkeninin ortalaması 50.832 ve standart sapması 13.441 dir. Ayrıca meme kanserli hastaların %51'i 50 yaş altında iken, %49'si 50 yaş üstündedir.

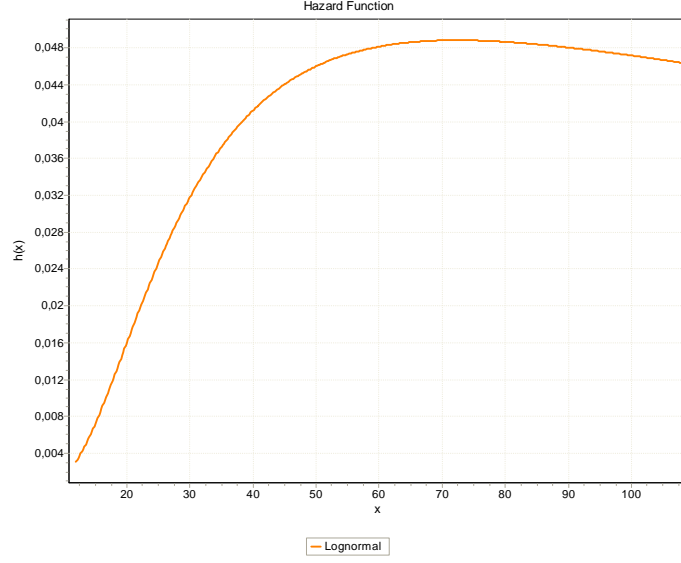
Tümör büyüklüğü değişkeninin ortalaması 24.055 ve standart sapması 10.335 dir. Çizelge 4.2’de de ifade edildiği gibi ölen hastaların %65’nin tümör büyüklüğünün, 24 mm’den büyük olduğu gözlenmiştir. Ölen hastaların yaklaşık %40’nın TNM evresi IIA ve IIB’ye sahip olduğu gözlenmiştir. Hastaların %71’ine modifiye radikal mastektomi (MRM) uygulandığı, %29’una ise meme koruyucu cerrahi (MKC) uygulandığı söylenebilir.

Ölen hastalara ait sağkalım sürelerinin, olasılık yoğunluk, sağkalım ve Hazard grafikleri EasyFit 5.0 paket programı (EasyFit2008) yardımıyla tüm dağılım türleri için incelenmiş ve en çok gözlenen dağılım şeklinin Lognormal olduğu belirlenmiştir. Ayrıca ölen hastaların sağkalım süresi dağılımına ait olasılık yoğunluk fonksiyonunun grafiği Şekil 4.1’de verilmiştir. Bu grafiğe göre verilerin, Lognormal dağılımın önemli bir özelliği olan sağa çarpıklığı sağladığı gözlemlenmektedir (EasyFit 2008).



Şekil 4.1 Ölen meme kanserli hastaların sağkalım sürelerine ilişkin dağılımın olasılık yoğunluk fonksiyonunun grafiği

Sağkalım süresine ait olasılık yoğunluk fonksiyonu, Lognormal dağılıma uygunluk gösterse de, dağılımının daha belirgin tanımlanabilmesi için Hazard fonksiyon grafiğinin de incelenmesi gerekir. Çünkü Hazard fonksiyonu grafiği, ilgilenilen dağılımın tanımlanması ve diğer dağılımlardan ayırt edilmesi için önem taşıyan bir grafikdir. Sağkalım süresine ait Hazard fonksiyonunun grafiği Şekil 4.2.’de verildiği gibidir (EasyFit 2008).



Şekil 4.2. Ölen meme kanserli hastaların sağkalım Hazard fonksiyonunun grafiği

Sağkalım süresine ait olasılık yoğunluk ve Hazard fonksiyon grafiklerinden elde edilen sonuçlar birleştirildiğinde, tanı konduktan sonra ölen meme kanserli hastalara ait sağkalım süresi verilerinin Lognormal dağılıma uyduğu söylenebilir. Bu durum yani ölen hastaların sağkalım süresine ait verinin Lognormal dağılım gösterip göstermediği bir kez de hipotez kurularak test edilmiştir. Tüm hesaplamalarda hata payı (α) 0.05 olarak alınmıştır. Bu analizler için kurulan hipotez aşağıdaki gibi ifade edilir:

H_0 : Tanı konduktan sonra ölen hastaların, öldükleri zamana kadar geçen süreleri (sağkalım süresi) Lognormal dağılmaktadır.

H_1 : Tanı konduktan sonra ölen hastaların, öldükleri zamana kadar geçen süreleri (sağkalım süresi) Lognormal dağılmamaktadır.

Hipotezin testinde sürekli dağılımlar için uyum iyiliği testlerinden Kolmogorov-Smirnov ve Anderson-Darling testleri kullanılmıştır. Çizelge 4.3.'ten görüldüğü gibi, uyum iyiliği testleri sonucu elde edilen p değerleri her iki test için de 0.05'ten büyük olduğundan, Lognormallik için kurulan hipotez kabul edilmiş ve ölen hastalara ait sağkalım sürelerinin dağılımının Lognormal olduğu bir kez daha gösterilmiştir.

Çizelge 4.3. Lognormal dağılım için elde edilen uyum iyiliği test sonuçları

<i>Kolmogorov-Smirnov</i>		α
Örneklem büyüklüğü	69	$p > \alpha$
Test istatistiği	0,09551	
p değeri	0.414	
<i>Anderson-Darling</i>		α
Örneklem büyüklüğü	69	$p > \alpha$
Test istatistiği	0,70384	
p değeri	0.324	

Lognormal dağılımda, verilere uygulanan logaritmik dönüşümün ardından dağılımın yaklaşık olarak simetrik olması gerekir. Bu nedenle verilerin Lognormal dağılım gösterdiğini ve simetrik olduğunu vurgulamak için, ölen hastaların sağkalım süresi verilerinin logaritması alınarak normallik testi yapılmıştır. Kolmogorov-Smirnov testi yardımıyla logaritmik dönüşüm yapılan ölen hastaların sağkalım süresi verilerinin normal dağılıma uygun olup olmadığı test edilmiştir. Ölen hastaların sağkalım süresi verilerine ait normallik testi sonuçları Çizelge 4.4.'de özetlendiği gibidir.

Çizelge 4.4. Logaritması alınan verilerin Kolmogorov-Smirnov Normallik testi sonuçları

<i>Kolmogorov-Smirnov</i>	<i>Kolmogorov-Smirnov test istatistiği</i>	<i>p çift yönlü</i>
Ln takip süresi	0,664	0,770

Elde edilen p değerinin 0.05'ten büyük olması, hipotezin kabul edilmesine ve logaritması alınan ölen hastaların sağkalım süresi verilerinin normal dağılım gösterdiği şeklinde yorum yapılmasına neden olur. Normallik testinin ardından, prognostik faktörlerin sağkalım süresine olan etkilerinin belirlenmesi için KPSA'ya göre analiz sonuçları elde edilmiş ve sonuçlar aşağıdaki çizelgede özetlenmiştir.

Çizelge 4.5. Meme kanserli hastaların sağkalım sürelerine ilişkin KPSA sonuçları

<i>Bağımsız Değişkenler</i>	$\hat{\beta}$	$\hat{\sigma}_{\hat{\beta}}$	<i>Z</i>	<i>p</i>	<i>Hızlandırma Faktörü (HF)</i>
Yaş (Yıl)	-0.0029	0.0060	-0.4739	0.6360	0.9971
Aile Hikayesi (Meme kanseri)	-0.0377	0.2150	-0.1754	0.8610	0.9630
Aile Hikayesi (Diğer kanser)	0.3860	0.2937	1.3145	0.1890	1.4711
Menepoz Durumu	0.2267	0.1849	1.2259	0.2200	1.2545
Tümör Büyüklüğü	-0.0253	0.0077	-3.2726	0.001	0.9750
TNM Evresi IIA	0.6284	0.2140	2.9367	0.003	1.8746
TNM Evresi IIB	0.7648	0.2603	2.9381	0.003	2.1486
Histolojik Tip (Duktal)	-0.0047	0.1621	-0.0288	0.977	0.9953
Östrojen Reseptör Durumu (Negatif)	0.4731	0.1770	2.6737	0.0075	1.605
Progesteron Reseptör Durumu (Negatif)	0.2158	0.1732	1.2462	0.2130	1.2408
Tedavi TÜRÜ (MKC,MRM)	0.0358	0.1618	0.2210	0.8250	1.0364
Aksiller Lenf Nodu Tutulumu (Pozitif)	0.8022	0.1869	4.2922	0.000	2.2304

Sonuçlar incelendiğinde, meme kanserli hastaların sağkalım süresi üzerinde etkili olmayan prognostik faktörlerin; yaş, aile hikayesi, menepoz durumu, histolojik tip, progsteron reseptör durumu ve tedavi türü değişkenleri olduğu görülmüştür ($p>0.05$). Diğer taraftan, tümör büyüklüğü ($p<0.001$), TNM evresi IIA ($p<0.003$), TNM evresi IIB ($p<0.003$), östrojen reseptör durumu ($p<0.008$) ve aksiller lenf nodu tutulumu ($p<0.000$) değişkenlerinin sağkalım süresi üzerinde önemli etkilerinin olduğu belirlenmiştir. Anlamlı bulunan prognostik faktörlere göre KPSA yeniden ele alındığında, sonuçlar aşağıdaki gibi özetlenebilir.

Çizelge 4.6. Anlamlı bulunan parametrelere göre elde edilen KPSA sonuçları

<i>Bağımsız Değişkenler</i>	$\hat{\beta}$	$\hat{\sigma}_{\beta}$	<i>Z</i>	<i>p</i>	<i>HF</i>
Tümör Büyüklüğü	-0.026	0.008	-3.279	0.001	0.974
TNM Evresi IIA	0.664	0.216	3.079	0.002	1.943
TNM Evresi IIB	0.808	0.258	3.131	0.001	2.243
Östrojen Reseptör Durumu (Negatif)	0.580	0.153	3.785	0.000	1.786
aksiller Lenf Nodu Tutulumu (Pozitif)	0.849	0.186	4.565	0.000	2.337

Çizelge incelendiğinde, tümör büyüklüğü ($p<0.001$), TNM evresi IIA ($p<0.002$), TNM evresi IIB ($p<0.001$), östrojen reseptör durumu ($p<0.000$) ve aksiller lenf nodu tutulumu ($p<0.000$) değişkenlerinin sağkalım süresi üzerindeki etkilerinin önemli olduğu görülmektedir. Çizelgeden elde edilen bir başka önemli bilgi, sağkalım zamanının azaldığı durumun açıklanmasına yardımcı olur. Sağkalım zamanı, tümör büyüklüğü arttıkça 0.974 oranında, TNM evre IIA'nın evre (0)'a göre 1.943 oranında, TNM evre IIB'nın evre (0)'a göre 2.243 oranında, östrojen reseptör durumu negatif olanların pozitif olanlara göre 1.786 oranında ve aksiller lenf nodu tutulumu pozitif olanların negatif olanlara göre 2.337 oranında azalmıştır.

4.3. BAYESCI SAĞKALIM ANALİZİ (BSA)

Açıklayıcı olan ve olmayan önsel dağılım bilgisine dayanarak elde edilen Bayesci sağkalım analiz sonuçları, WinBUGS programında macro yazılarak elde edilmiştir. Bu noktada WinBUGS programı uygulanırken dikkat edilmesi gereken noktalar hakkında bilgi verilmesi uygun görülmüştür.

WinBUGS yazılımının temeli BUGS'a (Bayesian inference Using Gibbs Sampling) yani "Gibbs Örneklemesi Kullanılarak Bayesci Sonuç Çıkarma" ya dayanmaktadır. BUGS, ilk kez 1996'da uygulanmaya başlamasının ardından geliştirilerek ayrıca grafik arayüzlü hale dönüştürülerek, 2003 yılında yeni bir isimle yani "WinBUGS" olarak kullanıma sunulmuştur. WinBUGS yazılımının komut satırları daha kısadır, ve değişkenlerin model dışında ayrıca bildirilmesine gerek yoktur. Yazılımda ayrıca dizilerin boyutu sınırlandırılmamış ve deyimlerin noktalı virgülle sonlandırılmasına gerek görülmemiştir. Yazılımda MCMC kullanılarak, karmaşık istatistiksel

modellerin Bayesci yaklaşım ile çözümüne ulaşmak da mümkün hale gelmiştir. Uygulamada WinBUGS yazılımının tercih edilmesinin nedenlerinin başında, Bayesci modeli oluşturabilmek ve modele ilişkin çıkarsama yapabilmektir. Bayesci yaklaşım için yazılan çok sayıda program olmasına rağmen; WinBUGS, içlerinde en çok tercih edilen yazılımdır. Bunun en önemli nedenlerinden biri de, amacı programlama olmayıp istatistiksel çalışma yapmak isteyen bir kullanıcılar için uygun bir yazılım olmasıdır. WinBUGS uygulanırken dikkat edilmesi gereken noktalar aşağıdaki gibi sıralanabilir:

- WinBUGS’da istenen sonsal dağılımın elde edilebilmesi ve başlangıç değerlerinin etkisinin en aza indirilmesi için, elde edilen Markov zincirinde iterasyonun başlangıç kısmını işlem dışı bırakmak gerekmektedir. İşlem dışı bırakmada kullanılan iterasyon sayısına “burn-in uzunluğu” denir. Zincir ne kadar yavaş yakınsarsa, burn-in periyodunun seçiminde o derece dikkatli olmak gerekir. Zincirin yakınsaklığını sağlayan en iyi başlangıç değerinden başlatmak periyodun hızlı çalışmasına neden olur. Burn-in periyodunun uzunluğunu hesaplamak için evrensel, sistematik ve kesin bir yöntem bulunmamaktadır. Burn-in uzunluğu, yakınsama hızına bağlı olarak değişir. Paralel zincirler çalıştırarak, zincirlerin durağan dağılımları ile ilgili farklılıklar daha kolay gözlemlenebilir.
- WinBUGS’da iterasyon sayısı, burn-in uzunluğundan sonraki iterasyon sayısıdır.
- WinBUGS’da başlangıç değerinin belirlenmesinde iki yol vardır: birincisi başlangıç değerinin araştırmacı tarafından atanması, ikincisi ise en çok olabilirlik yönteminden elde edilen parametre tahminlerinin başlangıç değeri olarak kullanılmasıdır. Uygulamada ikinci yol tercih edilir.
- WinBUGS sonuç sekmesinde yer alan iz grafikleri; Markov zincirindeki yakınsamayı göstermenin en kolay yöntemlerinden biridir. Grafik; Markov zincirinden elde edilen değerlerin iterasyon sayısına göre çizilir. Tüm zincir değerleri için elde edilen iz grafiklerine “Geçmiş (History)” grafiği adı verilir. Markov zinciri yakınsama gösterirse; grafikte salınımlar (titreşimler) artar aksi durumda ise azalır. Yakınsamanın az olması, her bir iterasyonda çok az bir değişimin meydana geldiğini ve doğru sonsal dağılımın elde edilmesi için iterasyon sayısının artırılması gerektiğini ifade eder.

- WinBUGS özet çizelgesinde yer alan MC hatası, sonsal istatistiklerinden elde edilen ve yeterli iterasyon sayısının belirlenmesinde kullanılan bir yöntemdir. Bu değer aynı zamanda sonsal istatistiklerde oluşan belirsizliği göstermektedir ve örneklem hacmi arttıkça azalış gösterir. Sonsal istatistiklerin kesinliğinden bahsedebilmek için; MC hata oranının, elde edilen sonsal standart hatanın %5 veya daha azına denk gelmesi gerekir.
- WinBUGS sonuç sekmesinde yer alan yoğunluk grafiği için Kernel düzleme yöntemi kullanılır. Grafiğin çan biçiminde olması, sonsal dağılım arzu edilen dağılıma uygunluk gösterdiği anlamına gelir.

Açıklayıcı önsel dağılıma dayalı Bayesci sağkalım analizinin (BSA-I) uygulaması sırasında, sağkalım süresine etki eden prognostik faktörler için geçmiş çalışmalardan elde edilen önsel bilginin kullanılması tavsiye edilmektedir. Ancak yapılan kaynak taraması sonucunda, çalışmadaki bazı değişkenler için bu tür bir bilgiye ulaşılamamıştır. Bu nedenle uygulamada bu değişkenler için, açıklayıcı olmayan normal önsel dağılıma dayalı ($N(0.0, 0,0001)$) bilgi kullanılmıştır. Kaynak taraması sonucunda çalışmadaki değişkenler için elde edilen ($\hat{\beta}$) ve ($\hat{\sigma}_{\beta}$) önsel bilgileri aşağıda özetlenmiştir. Burada yer alan kategorik değişkenlerin referans dışında yer alan kategorileri dikkate alınmıştır.

Çizelge 4.7. Meme kanserli hastalara ilişkin bağımsız değişkenlerin önsel bilgileri

<i>Bağımsız Değişkenler</i>	$\hat{\beta}$	$\hat{\sigma}_{\hat{\beta}}$	<i>Kaynak</i>
Yaş (Yıl)	-0,02	0,01	Chapman ve diğ. (2008)
Aile Hikayesi	0.1009	0.2288	Raman ve Venkatesan (2012)
Menepoz Durumu	0,075	0,0997	Raman ve Venkatesan (2012)
Tümör Büyüklüğü	-0.20	0,04	Sun ve diğ. (2002)
TNM Evresi-1	-	-	-
TNM Evresi-2	-	-	-
Histolojik Tip (Duktal)	-	-	-
Östrojen Reseptör Durumu (Negatif)	0,14	0,05	Sun ve diğ. (2002)
Progesteron Reseptör Durumu (Negatif)	-	-	-
Tedavi TÜRÜ (MKC,MRM)	-	-	-
Aksiller Lenf Nodu Tutulumu (Pozitif)	-0.26	0,08	Sun ve diğ. (2002)

Lognormal dağılım gösteren ölen hastaların sağkalım süresine etki eden prognostik faktörlerin belirlenmesinde WinBUGS'ta yazılan macro yardımıyla elde edilen sonuçlar (BSA-I) aşağıdaki gibidir.

Çizelge 4.8. Meme kanserli hastaların sağkalım sürelerine ilişkin BSA-I sonuçları

<i>Bağımsız Değişkenler</i>	$\hat{\beta}$	$\hat{\sigma}_{\hat{\beta}}$	<i>MC Hata</i>	<i>%2.5</i>	<i>%97.5</i>	<i>HF</i>
Yaş (Yıl)	-0.0005	0.0057	0.0002	-0.0106	0.0120	0.9995
Aile Hikayesi (Meme kanseri)	-0.0124	0.2062	0.0025	-0.408	0.4058	0.9877
Aile Hikayesi (Diğer kanser)	0.4015	0.2835	0.0036	-0.1329	0.9855	1.4941
Menepoz Durumu	0.1906	0.177	0.0032	-0.1557	0.5416	1.2099
Tümör Büyüklüğü	-0.0228	0.0072	0.0002	-0.0369	-0.0089	0.9775
TNM Evresi IIA	0.6533	0.207	0.0053	0.2623	1.071	1.9219
TNM Evresi IIB	0.7805	0.2526	0.0069	0.2972	1.286	2.1826
Histolojik Tip (Duktal)	-0.0219	0.1549	0.0028	-0.2828	0.3253	1.0221
Östrojen Reseptör Durumu (Negatif)	0.4565	0.1668	0.0035	0.1389	0.7913	1.5785
Progesteron Reseptör Durumu (Negatif)	0.2227	0.1672	0.0027	-0.1031	0.5539	1.2494
Tedavi TÜRÜ (MKC,MRM)	0.0386	0.1559	0.0029	-0.267	0.3411	1.0393
Aksiller Lenf Nodu Tutulumu (Pozitif)	0.8137	0.1786	0.0043	0.4779	1.174	2.2562

Daha önce de ifade edildiği gibi, sonsal dağılımın yakınsama sağlaması ve başlangıç değerlerinin etkisinin en aza indirilmesi için, Markov zincirinde elde edilen örneklemin başlangıç kısmı çıkarılır. Uygulamada 50000 iterasyonun ilk 2500 iterasyonu çıkarılmış ve elde edilen BSA-I sonuçları Çizelge 4.8’de verilmiştir. Çizelgeden hareketle MC hatanın, sonsal standart hataya oranı %5’ten küçük olduğu için, yeterli iterasyon sayısına ulaşıldığı söylenir. Ayrıca %2.5 ve %97.5 güven aralıkları incelendiğinde, “0” değerini içeren aralığa sahip olan katsayıların sağkalım süresi üzerine bir etkisi olmadığı söylenir. Buradan hareketle, meme kanserli hastaların sağkalım süresi üzerinde; yaş, aile hikayesi, menepoz durumu, histolojik

tip, progsteron reseptör durumu ve tedavi türü değişkenlerinin önemli etkilerinin olmadığı söylenebilir.

Tümör büyüklüğü, TNM evresi IIA, TNM evresi IIB, östrojen reseptör durumu ve aksiller lenf nodu tutulumu değişkenlerinin ise sağkalım süresi üzerinde önemli etkilerinin olduğu ifade edilir. Anlamlı bulunan prognostik faktörler için BSA-I tekrar uygulandığında elde edilen sonuçlar aşağıdaki gibi özetlenebilir. Çizelgedeki değerlerin, 50000 iterasyondan ilk 2500 iterasyonunun çıkarılarak elde edilen BSA-I sonuçları olduğunun altını çizmek bu noktada önem taşımaktadır.

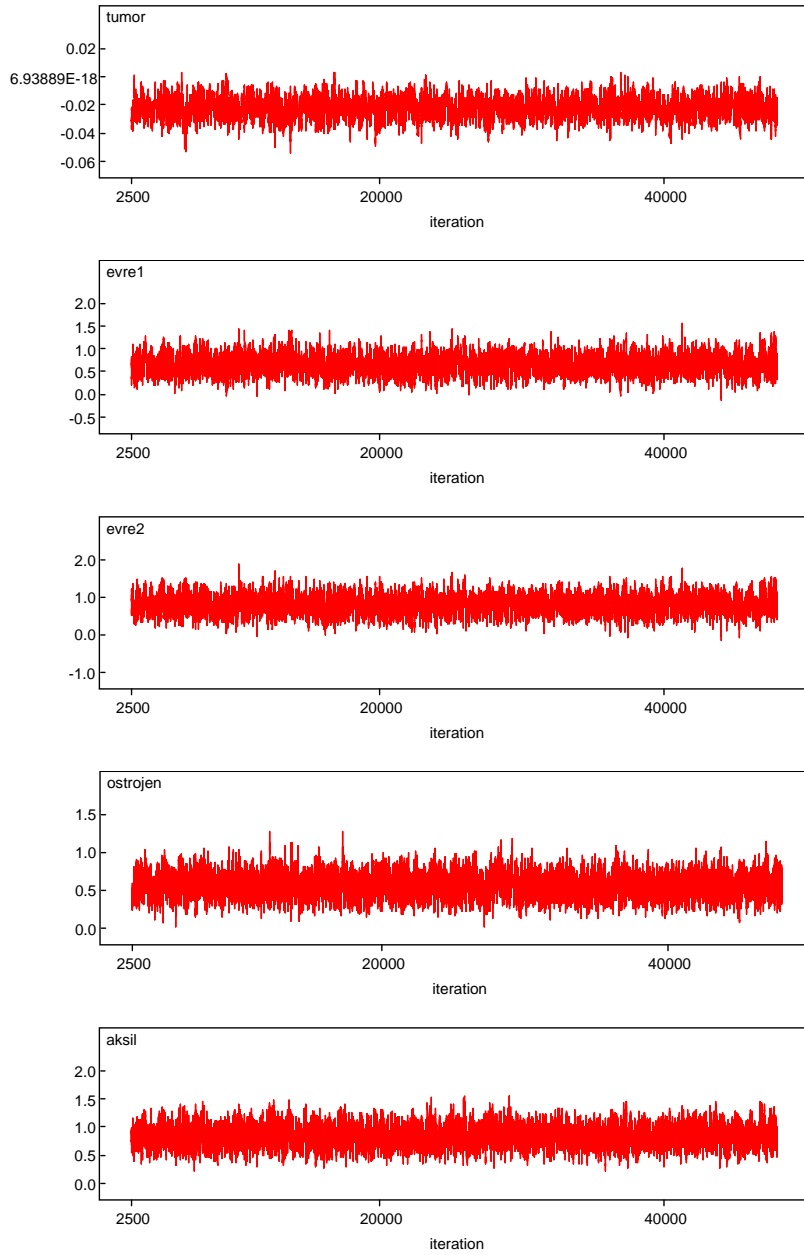
Çizelge 4.9. Anlamlı bulunan parametrelere için elde edilen BSA-I sonuçları

<i>Bağımsız Değişkenler</i>	$\hat{\beta}$	$\hat{\sigma}_{\hat{\beta}}$	<i>MC</i>	<i>%2.5</i>	<i>%97.5</i>	<i>HF</i>
			<i>Hata</i>			
Tümör Büyüklüğü	-0.0224	0.0069	0.0002	-0.0358	-0.0088	0.9778
TNM Evresi IIA	0.659	0.1931	0.0045	0.2936	1.051	1.9329
TNM Evresi IIB	0.7948	0.23	0.0052	0.3501	1.256	2.2140
Östrojen Reseptör Durumu (Negatif)	0.5626	0.1402	0.0030	0.2994	0.8517	1.7552
Aksiler Lenf Nodu Tutulumu (Pozitif)	0.8366	0.1686	0.0036	0.5185	1.183	2.3085

Çizelgeden hareketle MC hatanın, sonsal standart hataya oranı %5'ten küçük olduğundan yeterli iterasyonun sayısının elde edildiği söylenebilir. Ayrıca tümör büyüklüğü, TNM evresi IIA, TNM evresi IIB, östrojen reseptör durumu ve aksiller lenf nodu tutulumu değişkenlerinin sağkalım süresi üzerindeki etkilerinin önemli olduğunun da altı çizilmelidir. Sağkalım süresinin azalmasında tümör büyüklüğünün 0.9778, TNM evre IIA'nın evre (0)'e göre 1.9329, TNM evre IIB'nın evre (0)'a göre 2.2140, östrojen reseptör durumu negatif olanların pozitif olanlara göre 1.7552 ve aksiller lenf nodu tutulumu pozitif olanların negatif olanlara 2.3085 oranında etkili olduğu söylenebilir.

BSA-I'den elde edilen parametrelerin sonsal dağılımlarına ait Markov zincirlerinin yakınsaklıklarını iz ve Kernel yoğunluk grafikleri ile görmek mümkündür. Buna göre

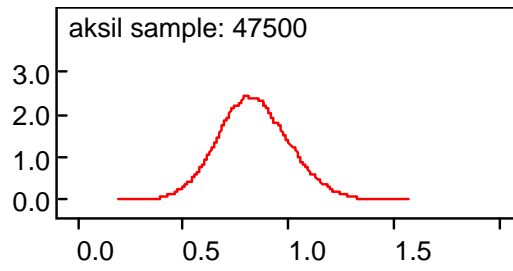
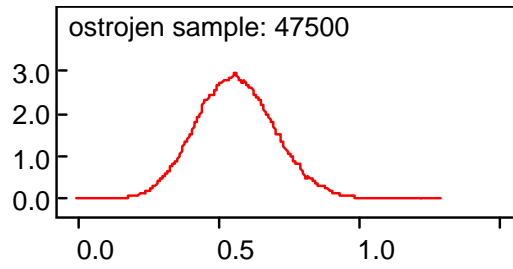
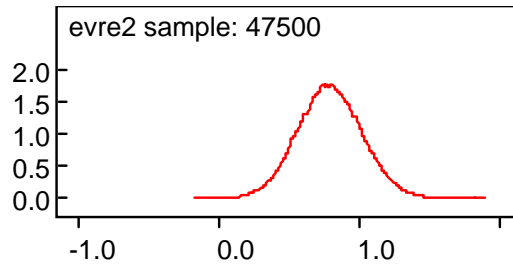
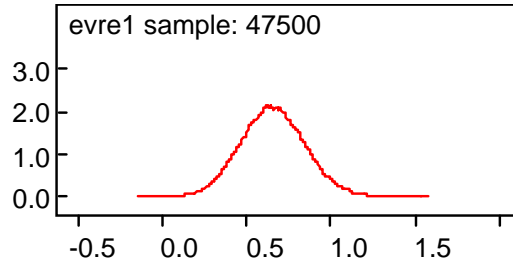
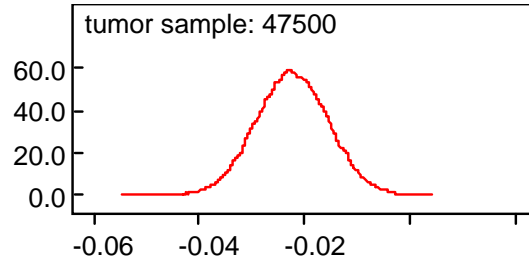
BSA-I’de anlamlı bulunan deęişkenler için çizilen iz ve Kernel yoğunluk grafikleri aőađıda verilmiőtir.



Őekil 4.3. BSA-I’e gre anlamlı bulunan parametrelere iin elde edilen iz grafikleri

Parametrelerin iz grafiklerinden oldukça salınım gsterdikleri (titreőim) ve sonsal dađılıma yakınsamanın hızlı biimde gerekleőtiđi sylenebilir.

Anlamlı bulunan parametreler iin çizilen Kernel yoğunluk grafikleri aőađıdaki gibidir:



Şekil 4.4. BSA-I'e göre anlamlı bulunan parametreler için Kernel yoğunluk grafikleri

Sonsal dağılıma ilişkin Kernel yoğunluk grafikeri incelendiğinde, parametrelerin çan eğrisi şeklinde olup dağılıma uygunluk gösterdiği bir başka deyişle ilgilenilen sonsal dağılıma ulaşıldığı söylenebilir.

Açıklayıcı olmayan önsel dağılım bilgisine dayalı Bayesci sağkalım analizi (BSA-II), sağkalım süresine etki eden prognostik faktörler için geçmişte yapılan herhangi bir çalışmaya ulaşılamadığı takdirde uygulanan bir analizdir. Bu noktada analizde prognostik faktörlerin her biri için açıklayıcı olmayan normal önsel dağılıma dayalı (N(0.0,0,0001)) bilginin kullanıldığını belirtmek gerekir. Ayrıca Lognormal dağılıma prognostik faktörlerin belirlenmesi amacıyla macro yazılmıştır. Sonsal dağılımın istenen dağılımına yakınsaması ve başlangıç değerlerinin etkisinin en aza indirilmesi için, Markov zincirinde örnekleminin başlangıç değerlerinin bir kısmı çıkarılmıştır. Bu çalışmada 50000 iterasyonun ilk 2500 adedi çıkartılmış ve elde edilen BSA-II sonuçları Çizelge 4.10'da verilmiştir.

Çizelge 4.10. Meme kanserli hastaların sağkalım sürelerine ilişkin BSA-II sonuçları

<i>Bağımsız Değişkenler</i>	$\hat{\beta}$	$\hat{\sigma}_{\hat{\beta}}$	<i>MC Hata</i>	%2.5	%97.5	<i>HF</i>
Yaş (Yıl)	-0.0032	0.0073	0.0002	-0.0171	0.0108	0.9968
Aile Hikayesi (Meme kanseri)	-0.0382	0.2557	0.0031	-0.532	0.4748	0.9625
Aile Hikayesi (Diğer kanser)	0.4635	0.3511	0.0041	-0.1954	1.187	1.5896
Menepoz Durumu	0.2483	0.2162	0.0048	-0.1705	0.6832	1.2818
Tümör Büyüklüğü	-0.0287	0.0092	0.0002	-0.0477	-0.0114	0.9717
TNM Evresi IIA	0.7231	0.2547	0.0074	0.2494	1.25	2.0608
TNM Evresi IIB	0.8578	0.309	0.0091	0.2715	1.499	2.3580
Histolojik Tip (Duktal)	-0.0029	0.1961	0.0042	-0.3884	0.3896	1.0029
Östrojen Reseptör Durumu (Negatif)	0.5627	0.2099	0.0047	0.1692	0.9937	1.7554
Progesteron Reseptör Durumu (Negatif)	0.25	0.2045	0.0034	-0.1471	0.6595	1.2840
Tedavi TÜRÜ (MKC,MRM)	0.0288	0.1904	0.0035	-0.3533	0.3999	1.0292
Aksiller Lenf Nodu Tutulumu (Pozitif)	0.9011	0.2257	0.0054	0.4771	1.372	2.4623

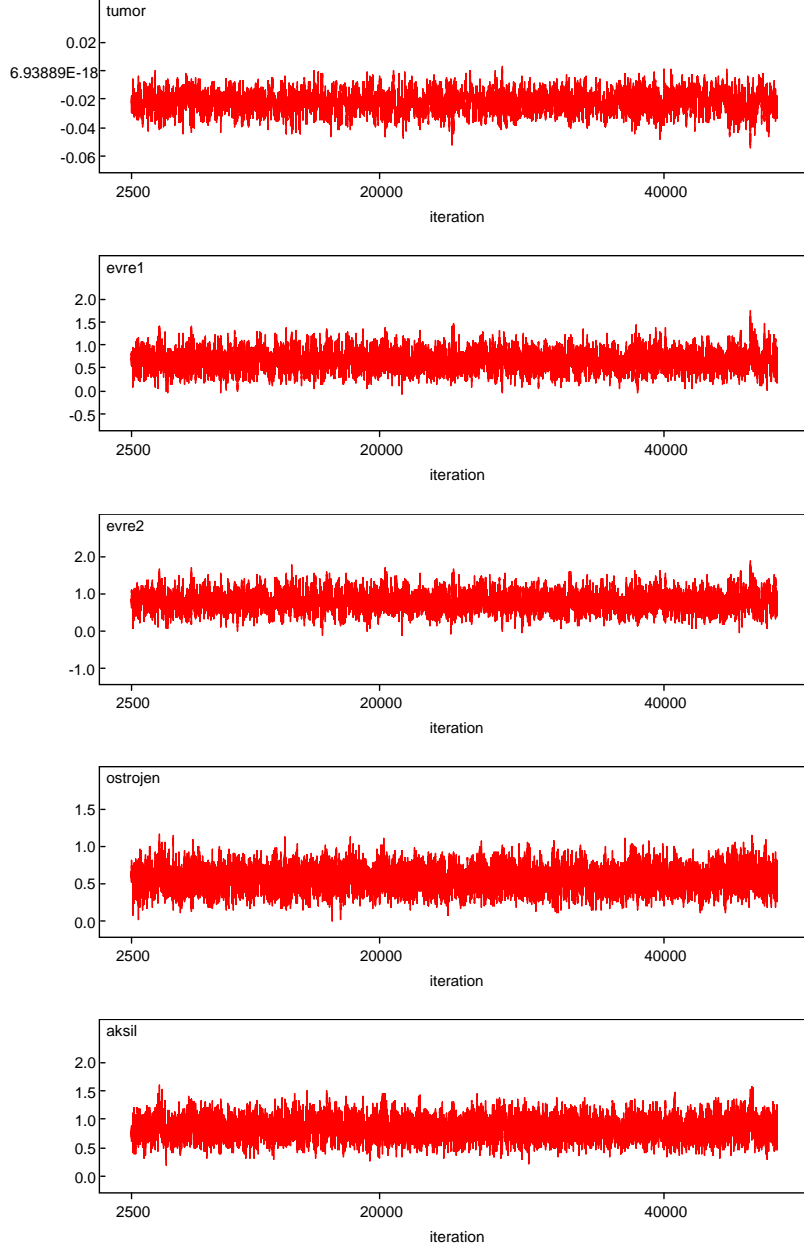
Çizelgeden de görüldüğü üzere, MC hatanın sonsal standart hataya oranı %5'ten küçük olduğu için yeterli iterasyon sayısı elde edilebilmiştir. %2.5 ve %97.5 güven aralıkları arasından “0” değerini içeren aralıklara karşılık gelen katsayıların sağkalım süresi üzerine herhangi bir etkisi olmadığı söylenir. Buradan hareketle meme kanserli hastaların sağkalım süresi üzerinde yaş, aile hikayesi, menepoz durumu, histolojik tip, progsteron reseptör durumu ve tedavi türü değişkenlerinin önemli etkilerinin olmadığı ifade edilir. Diğer taraftan tümör büyüklüğü, TNM evresi IIA, TNM evresi IIB, östrojen reseptör durumu ve aksiller lenf nodu tutulumu değişkenlerinin sağkalım süresi üzerindeki etkilerinin önemli olduğu söylenir. Anlamli bulunan prognostik faktörler için, BSA-II tekrar uygulandığında elde edilen sonuçlar aşağıdaki gibi özetlenebilir.

Çizelge 4.11. Anlamli bulunan parametreler için BSA-II sonuçları

<i>Bağımsız Değişkenler</i>	$\hat{\beta}$	$\hat{\sigma}_{\beta}$	<i>MC</i>	<i>%2.5</i>	<i>%97.5</i>	<i>HF</i>
			<i>Hata</i>			
Tümör Büyüklüğü	-0.0266	0.0083	0.0002	-0.0434	-0.0108	0.9738
TNM Evresi IIA	0.708	0.2312	0.0062	0.2752	1.186	2.0299
TNM Evresi IIB	0.8548	0.2788	0.0075	0.322	1.42	2.3509
Östrojen Reseptör Durumu (Negatif)	0.6181	0.1699	0.0034	0.3064	0.9739	1.8553
Aksiler Lenf Nodu Tutulumu (Pozitif)	0.8967	0.203	0.0047	0.5196	1.321	2.4515

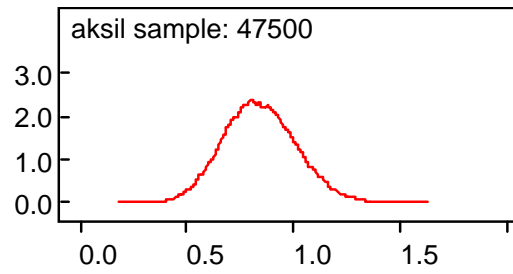
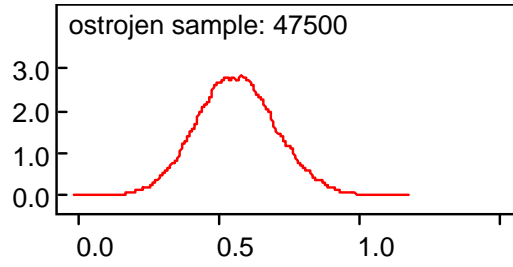
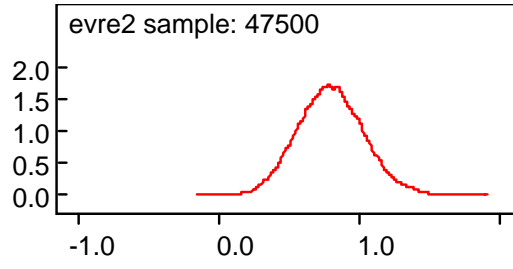
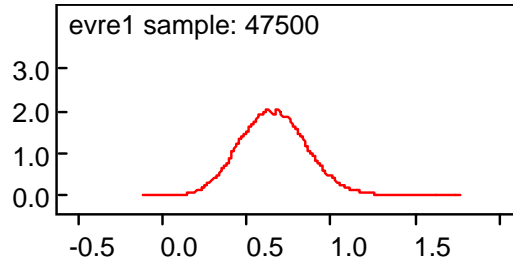
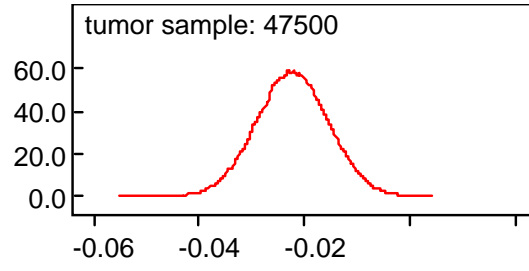
Çizelgedeki sonuçlar doğrultusunda tümör büyüklüğü, TNM evresi IIA, TNM evresi IIB, östrojen reseptör durumu ve aksiller lenf nodu tutulumu değişkenlerinin sağkalım süresi üzerindeki etkilerinin önemli olduğuna karar verilir. BSA-II'ye göre, tümör büyüklüğü arttıkça sağkalım süresinin 0.9738 oranında, TNM evre IIA'nın, evre (0)'e göre 2.0299 oranında, TNM evre IIB'nın, evre (0)'a göre 2.3509 oranında, östrojen reseptör durumu negatif olanların pozitif olanlara göre 1.8553 oranında ve aksiller lenf nodu tutulumu pozitif olanların negatif olanlara göre 2.4515 oranında azaldığı söylenebilir.

Parametrelerin sonsal dağılımından örneklemeler çeken Markov zincirinin yakınsaklığı iz ve Kernel yoğunluk grafikleri ile incelenir. BSA-II analizinde anlamlı bulunan parametrelere ait iz ve Kernel yoğunluk grafikleri sırasıyla aşağıda verilmektedir.



Şekil 4.5. BSA-II'ye göre anlamlı bulunan parametreler için iz grafikleri

Parametrelerin iz grafikleri incelendiğinde, grafiklerin salınım gösterdikleri ve sonsal dağılıma yakınsama hızının yüksek olduğu görülmektedir.



Şekil 4.6. BSA-II'ye göre anlamlı bulunan parametreler için Kernel yoğunluk grafikleri

Sonsal dağılıma ilişkin Kernel yoğunluk grafikeri incelendiğinde, parametrelerin çan eğrisine uygunluk gösterdiği, bir başka deyişle ilgilenilen sonsal dağılıma yaklaştığı söylenebilir.

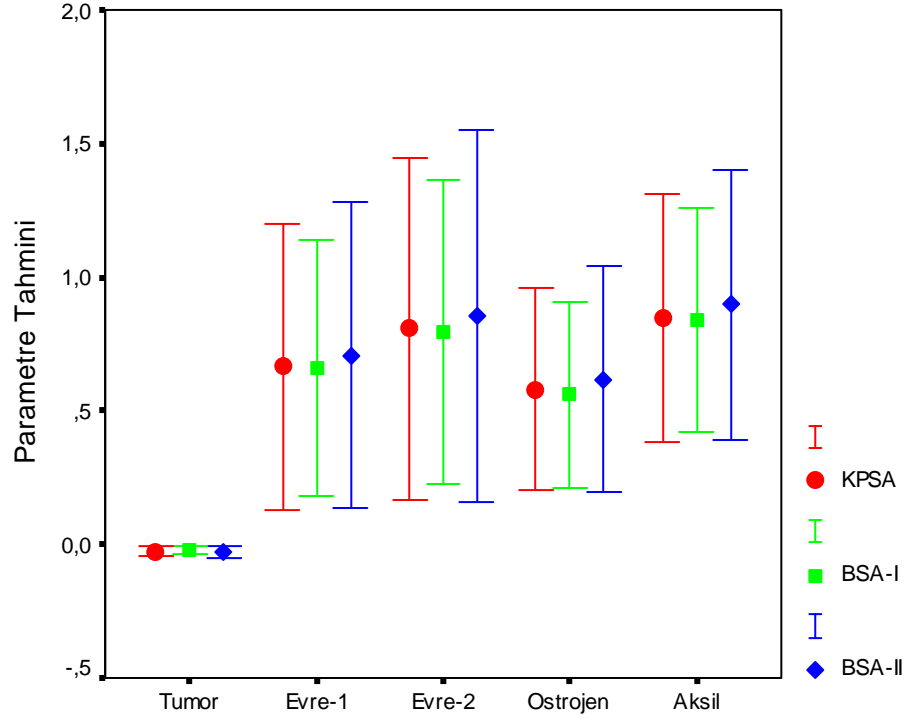
Meme kanserli hastaların sürelerine etki eden anlamlı prognostik faktörleri inceleyen KPSA, BSA-I ve BSA-II yöntemlerinin karşılaştırılması için uyum iyiliği istatistikleri aşağıdaki çizelge elde edilir.

Çizelge 4.12. Meme kanserli hastaların sağkalım sürelerine ilişkin anlamlı bulunan değişkenlerin KPSA, BSA-I ve BSA-II yöntemlerine göre elde edilen uyum istatistikleri

<i>Uyum istatistiği</i>	<i>KPSA</i>	<i>BSA-I</i>	<i>BSA-II</i>
-2xLogaritmik	884.213	867.603	885,045
olabilirlik			
AIC	900.213	883.603	901.045

Yukarıdaki çizelgede üç yöntem için elde edilen sonuçlar incelendiğinde, BSA-I yönteminin diğer yöntemlere nazaran en düşük olabilirliğe sahip olduğu, ayrıca en düşük AIC değerini verdiği görülmüştür. Bu nedenle diğer yöntemlere tercih edilebileceği söylenebilir.

KPSA, BSA-I ve BSA-II yöntemlerine göre anlamlı bulunan tümör büyüklüğü, TNM evresi IIA, TNM evresi IIB, östrojen reseptör durumu ve aksiller lenf nodu tutulumu değişkenlerinin parametre tahminleri ve standart hataları aşağıdaki şekilde olduğu gibi bir görünüme sahiptir.



Şekil 4.7. KPSA, BSA-I ve BSA-II'nin parametre tahmini ve standart hatalarının grafiği

Grafik incelendiğinde, BSA-I yöntemine ait anlamlı bulunan parametre tahminlerin standart hatalarının, KPSA ve BSA-II için elde edilen parametre tahminlerinin standart hatalarından daha küçük olduğu gözlemlenmiştir.

5. SONUÇ VE ÖNERİLER

Sağlık alanında sağkalım sürelerine ilişkin verilerin incelenmesi amacıyla kullanılan yöntemler her zaman güncelliğini korumaktadır. Sağkalım analizi, prognostik öğelerin ve tedavi başarılarının ortaya konması açısından önem taşımakta ve tıbbın çeşitli alanlarında (bireyin sağ kalma süresi, protezin dayanma süresi gibi) kullanılmaktadır. Lloyd D. Fisher ve Gerald van Belle “Biostatistics, A Methodology for the Health Science” isimli kitaplarında analize “Sağkalım Analizi” denmesinin nedenini şöyle vurgularlar: “Ölümden bahsetmek üzücü ancak sağkalımdan söz etmek daha memnun edicidir”.

Belirli bir dağılıma sahip sağkalım süresi üzerinde etkili olan faktörlerin belirlenmesinde kullanılan KPSA, birçok istatistiksel paket programlarda yer alan bir yöntem olduğu için araştırmacılar tarafından tercih edilmektedir. KPSA’ya alternatif olarak geliştirilmiş olan BSA ise son yıllarda oldukça sık kullanılmaya başlanmıştır. BSA, gözlenen veriden elde edilen yeni bilgi ile önceki benzer çalışmalardan elde edilen bilgilerin birleştirilmesine dayanan bir yöntemdir. Dolayısıyla önsel bilginin elde edilmesi, Bayesci çıkarsamada önemli bir rol oynar. Bayesci analiz; önceki bilgiler ve geçmiş deneyimler gibi açıklayıcı önsel bilgiler ile gözlenen verinin birleştirilerek subjektif olarak yorumlanmasına dayanır (Congdon, P., 2006; Wong ve ark., 2005). Bayesci ve klasik yöntemler arasındaki en önemli fark, BSA’nın önsel bilgiyi hesaba katması ve dolayısı ile araştırmacının subjektif kanaatlerini de yansıtmasıdır.

İncelenen olayla ilgili geçmişte herhangi bir çalışma yapılmamış ya da uygun önsel bilgiye, yapılan çalışmaların sonuçlarının güvenilir olmaması nedeniyle ulaşılamamış olabilir. Veri yapısına uygun ve gerçeği yansıtan açıklayıcı önsel bilgiye ulaşılamadığı durumda, açıklayıcı olmayan önsel bilgi kullanılmasına rağmen, Calle ve ark. (2006) ve İbrahim ve ark. (2001), açıklayıcı olmayan önsel bilgiye dayalı Bayesci ve klasik yaklaşımların benzer sonuçlar verdiği görülmüştür.

İbrahim ve ark. (2001), açıklayıcı olmayan ya da açıklayıcı olup da uygun olmayan önsel bilgiyi kullanmanın, değişkenlerin anlamlılığının belirlenmesinde faydalı olabileceğini fakat model seçimi ve karşılaştırması için uygun olmayacağını belirtmişlerdir. Ayrıca sonsal model olasılıklarının hesaplanmasında uygun önsel bilginin kullanımının gerekliliğini vurgularken, açıklayıcı olmayan önsel bilgiyle yürütülen çalışmalardan elde edilen parametre tahminlerinin gelecekte yapılacak çalışmalar için kullanılmayacağını ifade etmişlerdir (İbrahim ve ark., 2001). Gelfand ve Mallick (1995) ise açıklayıcı olmayan önsel bilgi kullanıldığı büyük veri setlerinden elde edilen parametre tahminlerinin, klasik en çok olabilirlik yönteminden elde edilen parametre tahminlerine yakın olacağını ifade etmişlerdir.

Tüm istatistiksel dağılımlarda olduğu gibi sağkalım süresine ilişkin dağılımları tanımlarken de dağılımın olasılık yoğunluk fonksiyonu, birikimli dağılım fonksiyonu, sağkalım fonksiyonu ve risk (hazard) fonksiyonundan faydalanılır. Bu çalışmada kullanılan meme kanseri verilerinin yukarıda sayılan fonksiyon özellikleri incelenmiş ve dağılım şeklinin lognormal olduğu görülmüştür. Lognormal dağılım, olaya ilişkin verilerin dağılımın sol kuyrukta yoğunlaştığı ve yapılan logaritmik dönüşüm sonucunda sağkalım süresinin dağılımının normal dağılıma yaklaştığı görülmüştür. Lognormal dağılımının dezavantajı ise izlem süresi arttıkça Hazard fonksiyonun değerinin sifıra yaklaşmasıdır.

Bu çalışmada, meme kanserine yakalanmış hastaların hastalığa yakalandıktan ölene kadar geçirdikleri sürenin dağılımı incelenmiş, bu süreye etki eden prognostik faktörleri belirlemek için ise Klasik ve Bayesçi parametrik sağkalım analizi uygulanmıştır. Verilerin dağılımının şeklinin belirlenmesinin ardından; KPSA ve açıklayıcı olan ve olmayan önsel bilgiyi kullanan BSA uygulanmıştır.

Önsel bilginin elde edilmesi için, meme kanserli hastaların sağkalım sürelerine ilişkin önceki araştırmalar incelenmiştir. Yapılan inceleme sonucunda ilk olarak Raman ve çalışma arkadaşlarının 2012 yılının Şubat ayında yaptıkları çalışma baz alınmıştır. 522 meme kanserli hastanın sağkalım süresi üzerinde evre, düşük sayısı, çocuk sayısı, kemoterapi, radyoterapi, ailedeki kanser hikayesi ve menepoz durumu değişkenlerinin etkileri incelenmiş ve bu amaçla üç farklı dağılım (Weibull, Üstel ve Lognormal) için KPSA uygulanmıştır. İncelenen prognostik faktörlerden; evre ve

kemoterapi deęişkenlerin etkisinin önemli olduęu saptanmıştır. Baz alınan bir dięer çalışma ise Chapman ve arkadaşlarının 2008 yılında 5187 meme kanserli hastanın saękalım süresi üzerine yaptıkları çalışmadır. Bu çalışmadaki prognostik faktörler; yaş, menepoz durumu, tamoksifen tedavi, radyasyon terapisi, kemik kırığı, kemik erimesi, kalp-damar hastalığı, hormon tedavisi durumu, lenf nodu durumu ve alternatif tedavi durumu olarak belirlenmiştir. Deęişkenlerinin saękalım süresi üzerinde etkisini incelemek amacıyla Lognormal model uygulanmış, sadece yaş ve lenf nodu durumu deęişkenlerinin önemli olduğunu belirlemişlerdir. Son olarak Sun ve arkadaşlarının 2002 yılında yaptığı çalışmadan önsel bilgi edinilmeye çalışılmış ve 678 meme kanserli hastanın saękalım süresi üzerinde etkili olan yaş, tümör büyüklüęü, lenf nodu durumu, östrojen reseptör durumu, progesteron reseptör durumu, alternatif tedavi alma durumu, kemoterapi, radyoterapi ve alternatif hormon terapisi deęişkenleri incelenmiştir. Saękalım süresi üzerinde etkisini incelemek amacıyla Lognormal model uygulanması sonucunda tümör büyüklüęü, lenf nodu durumu ve östrojen reseptör durumu deęişkenlerinin önemli olduęu belirlenmiştir.

Bu çalışmada meme kanserli hastaların saękalım süresi üzerinde yaş, aile hikayesi, menepoz durumu, tümör büyüklüęü, TNM evresi IIA, TNM evresi IIB, histolojik tip, östrojen reseptör durumu, progesteron reseptör durumu, tedavi (radyoterapi ve hormon terapisi) ve aksiller lenf nodu tutulumu deęişkenlerinin etkileri incelenmiştir. Bu deęişkenler içinden tümör büyüklüęü, TNM evresi IIA, TNM evresi IIB, östrojen reseptör durumu ve aksiller lenf nodu tutulumu deęişkenlerinin saękalım süresi üzerindeki etkilerinin önemli olduęuna KPSA, BSA-I ve BSA-II analizleri yardımıyla karar verilmiştir. Lognormal dağılıma sahip saękalım süresi verilerine ilk olarak KPSA ardından önsel bilgi yardımıyla açıklayıcı önsel bilgiye dayalı BSA-I uygulanmıştır. Bir sonraki adımda önsel bilginin elde edilemedięi varsayılarak açıklayıcı olmayan önsel bilgiye dayalı BSA-II uygulanarak üç yöntem için karşılaştırma yapılmıştır. Sapma ve AIC kriterlerine göre yapılan karşılaştırma sonucunda, açıklayıcı önsel bilgiye dayalı Bayesci saękalım analizinin dięerlerine nazaran daha küçük sapma ve AIC deęerine sahip olduęu ve dolayısı ile tercih edilmesi gerektięine karar verilmiştir. Ayrıca BSA-I yöntemi ile elde edilen parametre tahminlerine ilişkin standart hataların, BSA-II ve KPSA için elde edilen standart hatalardan daha küçük olduęu gözlenmiştir.

Yapılan deęerlendirme sonucunda, daha nce yapılan alıřmalarda da olduęu gibi BSA ile etkin bir sonu elde etmek iin uygun ve gereęi yansıtın nsel bilgi ile alıřmak gerektięi sylenebilir. Ayrıca aıklayıcı olmayan nsel bilgi kullanıldıęında, BSA ile KPSA'nın benzer sonular verdięi de gzden kaırılmamalıdır.

lmcl hastalıęa sahip bireylerin bilgileri doęrultusunda saękalım modelleri kullanarak hastalıęa neden olan prognostik faktrler incelenmesinde ileriye ynelik alıřmalar, varolan bilgiler birleřtirilerek daha kesin sonular elde edilmesi saęlanabilir. lkemizde kanser trleri iin gerekli veri tabanlarının mevcut olmaması, bu tr uygulamaların yapılmasını zorlařtırmaktadır. Dolayısı ile kanser gibi nemli bir hastalıęa neden olan prognostik faktrler belirlenememektedir. Yapılan tez alıřmasının nemi bu noktada ortaya ıkmakta ve nsel bilgiye ihtiya duyan dięer alıřmalara baz oluřturması mit edilmektedir.

Bayesci saękalım analizinde son yıllarda olduka gncel bir konu olan saękalım sresine etki eden prognostik faktrler dıřında, llemeyen veya bilinmeyen etkenler zerinde ileriye ynelik alıřma yapılması planlanmaktadır.

KAYNAKLAR

- Abu-Talep, A.A., Smadi, M.M., Alawneh, A.J.**, 2007, Bayes Estimation of the Lifetime Parameters for the Exponential Distribution, *Journal of Mathematics and Statistics*, **3(3)**, 106-108.
- Ahmed, A.M., Ibrahim, N.A.**, 2011, Bayesian Survival Estimator for Weibull Distribution with Censored Data, *Journal of Applied Sciences*, **11(2)**, 393-396.
- Arjas, E.**, 1988, A Graphical Method for Assessing Goodness of Fit in Cox's Proportional Hazards Model, *Journal of the American Statistical Association*, **83**, 204-212.
- Beamonte, E. and J. D. Bermúdez, J. D.**, 2003, A Bayesian Semiparametric Analysis for Additive Hazards Models with Censored Observations, **2**, 101–117.
- Bekiroğlu N.**, 1998, Açıklamalı Biyoistatistik Terimleri Sözlüğü. 1. baskı. Nobel Tıp Kitabevleri, İstanbul.
- Bekiroğlu N.**, 1999, İki veya İkiden Fazla Sağkalım Olasılıkları Karşılaştırılmasının Bir Uygulamayla İrdelenmesi. 4. Ulusal Biyoistatistik Kongresi Bildirileri Kitabı, Ankara Üniversitesi. Ankara. 23-34.
- Bolstad, W.M.**, 2007, Introduction to Bayesian Statistics, John Wiley&Sons Inc, New York.
- Box, G.E.P. and Tiao, C.G.**, 1973, Bayesian Inference in Statistical Analysis, Addison-Wesley, London.
- Brooks, S. P.**, 1998, Markov Chain Monte Carlo Method and its Application, *The Statistician*, **47**, 69-100.
- Brooks, S. P. and Gelman, A.**, 1997, General Methods for Monitoring Convergence of Iterative Simulations, *Journal of Computational and Graphical Statistics*, **7**, 434–455.
- Brooks, S. P. and Roberts, G. O.**, 1998, Assessing Convergence of Markov Chain Monte Carlo Algorithms, *Statistics and Computing*, **8**, 319–335.

Calle, M.L., Hough, G., Curia, A., Gómez, G., 2006, Bayesian Survival Analysis Modeling Applied to Sensory Shelf Life of Foods, *Food Quality And Preference*, **17**, 3-4.

Carlin, B. P., Louis, T. A., 2000, Bayes and Empirical Bayes Methods for Data Analysis, Chapman & Hall, New York.

Chapman, J.W., Meng, D., Shepherd, L., Parulekar, W., Ingle, J.N., Muss, H.B., Palmer, M., Yu, C., and Goss, P.E., 2008, Competing Causes Of Death From A Randomized Trial Of Extended Adjuvant Endocrine Therapy For Breast Cancer, *J Natl Cancer Inst*, **100**, 252-260.

Christensen, D., Johnson, W., Branscum, A. and Hanson, T. E., 2011, Bayesian Ideas and Data Analysis, Chapman & Hall, New York.

Cohen, J., Nagin, D., Wallstorm, G. and Wasserman, L., 1998, Hierarchical Bayesian Analysis of Arrest Rates, *Journal of the American Statistical Association*, **93**, 120-1270.

Collet, D., 2003, Modelling Survival Data in Medical Research, Chapman & Hall, New York.

Congdon, P. D., 2010, Applied Bayesian Hierarchical Methods, Chapman & Hall, New York.

Congdon, P., 2003, Applied Bayesian Modelling, John Wiley & Sons, England.

Congdon, P., 2006, Bayesian Statistical Modelling, John Wiley & Sons, England.

Cox, D.R., 1972, Regression Models and Life Tables (With Discussion), *Journal Of The Royal Statistical Society, Series B*, **34**, 187-220.

Cox, D.R., 1975, Partial Likelihood. *Biometrika*, **62**, 269–276.

De Finetti, B., 1974, Theory of Probability: A Critical Introductory Treatment, Wiley, London, New York.

Demirhan, H., 2004, Logaritmik Doğrusal Modellerde Parametrelerin ve Beklenen Göze Sıklıklarının Bayesci Kestirimi, *Yüksek Lisans Tezi*, Hacettepe Üniversitesi Fen Bilimleri Enstitüsü, Ankara.

Dişçi, R., 2008, Temel ve Klinik Biyoistatistik, İstanbul Medikal Yayıncılık Ltd. Şti., İstanbul.

Evans, M., Hastings, N., Peacock, B., 2000, Statistical Distributions. Eds: Cressie NAC, Fisher NI, Johnstone IM, Kadane JB, Scott DW, Silverman BW, Smith AFM, Teugels JL, Barnett V, Bradley RA, Hunter JS, Kendall DG., 3rd ed, John Wiley&Sons Inc, New York.

- Feigl, P. and Zelen, M.**, 1965, Estimation of Exponential Survival Probabilities with Concomitant Information, *Biometrics*, **21**,826–838.
- Feller, W.**, 1968, An Introduction to Probability Theory and Its Applications, John Wiley & Sons, New York.
- Fisher L.D., Belle Gv.**, 1993, Biostatistics, A Methodology for the Health Sciences. John Wiley&Sons Inc, New York.
- Gelfand, A. E., Smith, A.F.M.**, 1990, Sampling Based Approaches to Calculating Marginal Densities, *Journal of the American Statistical Association*, **85**, 398-409.
- Gelfand, A. E., Hills, S. E., Racine-Poon, A., and Smith, A. F. M.**, 1990, Illustration of Bayesian Inference in Normal Data Models Using Gibbs Sampling, *Journal of the American Statistical Association*, **85**, 972–985.
- Gelfand, A.E. and Mallick, B.K.**, 1995, Bayesian Analysis of Proportional Hazards Models Built from Monotone Functions, *Biometrics*, **51**(3).
- Gelman A., Carlin J.B., Stern H.S., Rubin D.B.**, 2004, Bayesian Data Analysis. Chapman-Hall, Florida.
- Gelman, A.**, 2006, Prior Distributions for Variance Parameters in Hierarchical Models, *International Society for Bayesian Analysis*, **3**, 515-533.
- Gelman, A., Carlin, J.B., Stern, H.S. and Rubin, D.B.**, 2004, Bayesian Data Analysis, Second Edition, Chapman & Hall, New York.
- Geman, S., Geman, D.**, 1984, Stochastic Relaxation, Gibbs Distribution, and the Bayesian Restoration of Images, *IEEE Transaction on Pattern Analysis and Machine Intelligence*, **6**, 721–741.
- Gilks, W. R. and Wild, P.**, 1992, Adaptive Rejection Sampling for Gibbs Sampling, *Applied Statistics*, **41**, 337–348.
- Gilks, W.R., Richardson, S., Spiegelhalter D.J.**, 1996, Markov Chain Monte Carlo in Practice. Chapman-Hall, London.
- Gill, J.**, 2008, Bayesian Methods: A Social and Behavioral Sciences Approach, Second Edition, Chapman & Hall, New York.
- Grambsch, P.M., Therneau, T.M.**, 1994, Proportional Hazards Tests and Diagnostics Based on Weighted Residuals, *Biometrika*, **81**, 515-526.
- Hosmer, D.W., Lemeshow, S.**, 1999, Applied Survival Analysis: Regression Modelling of Time to Event Data, Wiley&Sons, New York.
- Ibrahim, J.G., Chen, M.H., Sinha, D.**, 2001, Bayesian Survival Analysis, Springer-Verlag, New York.

- Jaynes, E. T.**, 1985, Some Random Observations, *Synthese* , **63**, 115-138.
- John, W., Gamel, M.D., Vogel, R.L., Valagussa, P. and Bonadonna, G.**, 1994, Parametric Survival Analysis of Adjuvant Therapy for Stage II Breast Cancer, *Cancer*, **74**, 2483-2490.
- Kalbfleisch, J.D., Prentice, R.L.**, 1980, The Statistical Analysis of Failure Time Data, Wiley, New York.
- Kaplan, E.L. and Meier, P.**, 1958 Non-Parametric Estimation from Incomplete Observations. *Journal of The American Statistical Association*, **53**, 457-481.
- Klein, J.P., Moeschberger, M.L.**, 1997, Survival Analysis Techniques for Censored and Truncated Data, Springer, New York.
- Kleinbaum, D.G. and Klein, M.**, 1996, Survival Analysis: A Self-Learning Text, Springer, USA.
- Kumru, Ö.**, 2003, Markov Zinciri Monte Carlo Yöntemi, *Yüksek Lisans Tezi*, Hacettepe Üniversitesi Fen Bilimleri Enstitüsü, Ankara.
- Kurt, İ.**, 2008, Bayesgil Yaşam Analizi ve Cox Regresyon Yaşam Analizi'nin Türetilmiş ve Gerçek Veri Setlerinde Uygulanması, Doktora Tezi, Eskişehir Osmangazi Üniversitesi Sağlık Bilimleri Enstitüsü.
- Kurt, İ., Türe, M., Özdamar, K.**, 2011, Comparison of Bayesian and Classical Analysis of Weibull Regression Model: A Simulation Study, *Türkiye Klinikleri J Biostat*, **3**(1):8-13.
- Lecture 17, BIOST 515, 2004, Cox Proportional Hazard Models.
- Lee, E.T. and Wang, J.W.**, 2003, Statistical Methods for Survival Data Analysis, John Wiley & Sons, Canada.
- Menteşe, T.**, 2011, Use of Bayesian Approach in Chemistry, *Hacettepe Journal of Biology & Chemistry*, **39**, 19-22.
- Mostafa, A. A., Ghorbal, A.B.**, 2011, Using WinBUGS to Cox Model with Changing from the Baseline Hazard Function, *Applied Mathematical Sciences*, **5**, 2217-2240.
- Ntzoufras, I.**, 2009, Bayesian Modeling Using Winbugs, John Wiley & Sons, Canada.
- O'Neill, P.D.**, 2002, A Tutorial Introduction to Bayesian Inference For Stochastic Epidemic Models Using Markov Chain Monte Carlo Methods. *Mathematical Biosciences* **180**.

O" Hagan, A., 1986, Probability: Methods and Measurement, Chapman & Hall, New York.

O" Hagan, A., 1994, Kendall's Advanced Theory of Statistics: Volume 2B, Bayesian Inference, London: Arnold.

Persson, I., 2002, Essays on the Assumption of Proportional Hazards in CoxRegression, <http://www.diva-portal.org>.

Pourhoseingholi, M.A., Moghimi-Dehkordi, B., Safae, A., Hajizadeh, E., Solhpour, A. and Zali, M.R., 2009. Prognostic Factors in Gastric Cancer Using Log-normal Censored Regression Model, *Indian J Med Res*, **129**, 262-267.

Press, J.S., 1989, Bayesian Statistics: Principles, Models and Applications, Wiley, New York.

Richard, A., Becker, J. M., Chambers, A. R., Wilks, W. S., Cleveland, 1998, S-PLUS 4 Guide to Statistics, MathSoft, Washington.

Royston, P., 2001, The Lognormal Distribution as a Model for Survival Time in Cancer, with an Emphasis on Prognostic Factors, *Statistica Neerlandica*, **55**, 89-104. SAS Institute, 2006, Preliminary Capabilities for Bayesian Analysis in SAS/STATR Software, SAS Institute Inc., Cary, NC, USA.

Schoenfeld, D., 1982, Partial Residuals for the Proportional Hazards Model, *Biometrika*, **69**, 551-555.

Spiegelhalter, D., Thomas, A., N. Best, N. and Gilks, W., 1996, BUGS 0.5: Examples Volume 1, MRC Biostatistics Unit, Institute of Public Health, Cambridge.

Sun, J., Chapman, J.W., Gordon, R., Sivaramakrishna, M.A., Fish, E.B., 2002 Survival from Primary Breast Cancer After Routine Clinical Use of Mammography. *The Breast Journal* , **8**.

Tai, P., Yu, E., Vinh-Hung, V., Cserni, G. and Vlastos, G., 2004, Survival of Patients with Metastatic Breast Cancer: Twenty-Year Data from Two SEER Registries, *Biomedcentral*, **60**, 1471-2407.

Therneau, T.M., Grambsch, P.M., 2000, Modelling Survival Data: Extending the Cox Model, Springer, New York.

Türkan, A.H., 2007, Güvenilirlik Analizinde Kullanılan İstatistiksel Dağılım Modelleri, *Yüksek Lisans Tezi*, Çukurova Üniv. Fen Bilimleri Enstitüsü, Adana.

Valsecchi, M., Silvestri, D., Sasieni, P., 1996, Evaluation of Long-Term Survival: Use of Diagnostics and Robust Estimators with Cox's Proportional Hazards Model, *Statistics in Medicine*, **15**, 2763-2780.

Van Belle, G., Fisher, L.D., Heagerty, P.J., Lumley, T.S., 2004, Biostatistics: A Methodology for the Health Sciences, John Wiley & Sons, Inc., New York.

Venkatesan, P., Raman, T.T., 2012, Accelerated Failure Time Frailty Model in Survival Analysis, *International Journal of Science and Technology*, **2**, 65-69.

Wong, M.C.M., Lam, K.F., Lo, E.C.M., 2005, Bayesian Analysis of Clustered Interval-Censored Data, *J Dent Res* , **84(9)**, 817-821.

Yardımcı, A., 1992, Çoklubağlantılı Çoklu Doğrusal Regresyonda Bayes Yaklaşımı, *Basılmamış Bilim Uzmanlığı Tezi*, Hacettepe Üniversitesi, Fen Bilimleri Enstitüsü, Ankara.

Yin, G. and Ibrahim, J.G., 2006, Bayesian Transformation Hazard Model, IMS Lecture Notes–Monograph Series, *2nd Lehmann Symposium-Optimality*, **49**, 170-182.

ÖZGEÇMİŞ

Adı-Soyadı : Esin AVCI
Doğum tarihi ve yeri : 1979 Erbil
Uyruğu : T.C.

Eğitim Durumu

Lisans : Ondokuz Mayıs Üniversitesi, Fen Edebiyat Fakültesi,
İstatistik Bölümü (2001) (Bölüm Birinciliği)
Yüksek Lisans : Ondokuz Mayıs Üniversitesi, Fen Bilimleri Enstitüsü,
İstatistik A.B.D.(2005)
Doktora : Mimar Sinan Güzel Sanatlar Üniversitesi, Fen
Bilimleri Enstitüsü, İstatistik A.B.D.

Yabancı Dil : İngilizce

Mesleki Deneyim

<i>Görev Unvanı</i>	<i>Görev Yeri</i>	<i>Yıl</i>
Araştırma Görevlisi	Sinop Üniversitesi Fen Edebiyat Fakültesi İstatistik Bölümü	2004-2008
İstatistikçi	Türkiye İstatistik Kurumu Başkanlığı İstanbul Bölge Müdürlüğü	2008-