



T.C.

SELÇUK ÜNİVERSİTESİ

FEN BİLİMLERİ ENSTİTÜSÜ

HİBRİT SANSÜRLEME ALTINDA BAZI DAĞILIMLAR İÇİN ARALIK VE NOKTA TAHMİNİ

Egemen ÖZKAN

YÜKSEK LİSANS

İstatistik

Haziran-2019
KONYA
Her Hakkı Saklıdır

ÖZET

YÜKSEK LİSANS

HİBRİT SANSÜRLEME ALTINDA BAZI DAĞILIMLAR İÇİN NOKTA VE ARALIK TAHMİNİ

Egemen ÖZKAN

Selçuk Üniversitesi Fen Bilimleri Enstitüsü
İstatistik Anabilim Dalı

Danışman: Prof. Dr Buğra SARAÇOĞLU
2019, 47 Sayfa

Jüri
Prof. Dr. Buğra SARAÇOĞLU
Prof. Dr. Coşkun KUŞ
Prof. Dr. Gülhayat GÖLBAŞI ŞİMŞEK

Bu tez çalışmasında hibrit sansürleme altında, son zamanlarda önerilen yeni sürekli dağılımların bilinmeyen parametreleri için nokta ve aralık tahmini konusu incelenmiştir. Tip-I hibrit sansürlü örneklem altında Log-Kumaraswamy dağılımı çalışılırken, ilerleyen tür Tip-I hibrit sansürlü örneklem altında Kumaraswamy-Ters Rayleigh dağılımı çalışılmıştır. Nokta tahminleri için en çok olabilirlik yöntemi kullanılmıştır. En çok olabilirlik tahminlerini elde etmek için (EM) algoritması kullanılmıştır. Hibrit sansürleme şeması altında, bahsi geçen dağılımların parametrelere ilişkin Fisher bilgi matrisi formülize türetilmiş ve en çok olabilirlik tahmin edicilerine dayalı asimptotik güven aralıkları oluşturmada kullanılmıştır. Elde edilen sonuçlar, Monte Karlo simülasyon çalışmaları ve gerçek veri analizleriyle desteklenmiştir.

Anahtar Kelimeler: Asimptotik Güven Aralığı, EM Algoritması, En Çok Olabilirlik Tahmin Edicileri (MLE), İlerleyen Tip I Hibrit Sansürleme, Kumaraswamy- Ters Rayleigh (Kw-TR) Dağılımı, Log-Kumarawamy (LKw) Dağılımı, Tip I Hibrit Sansürleme

ABSTRACT

MS THESIS

POINT AND INTERVAL ESTIMATION FOR SOME DISTRIBUTIONS UNDER HYBRID CENSORING

Egemen ÖZKAN

THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCE OF SELÇUK UNIVERSITY
THE DEGREE OF MASTER OF SCIENCE IN STATISTICS

Advisor: Prof. Dr. Buğra SARAÇOĞLU

2019, 47 Pages

Jury
Prof. Dr. Buğra SARAÇOĞLU
Prof. Dr. Coşkun KUŞ
Prof. Dr. Gülhayat GÖLBAŞI ŞİMŞEK

In this thesis study, under hybrid censoring schemes, point and interval estimation problems for unknown parameters of new continuous distributions obtained in recent years are examined. While the Log-Kumaraswamy distribution is studied under the Type I hybrid censoring sample, Kumaraswamy-Reverse Rayleigh distribution is considered under the progressive type I hybrid censoring sample.Maximum likelihood method for point estimates is used. EM algorithm is used to obtain maximum likelihood estimators. Under hybrid censoring scheme, fisher information matrix is derived and it has been used to construct the asymptotic confidence intervals for maximum likelihood estimator. The results have been supported by Monte Carlo simulation studies and real data analysis.

Keywords: Asymptotic Confidence Interval, Expectation Maximization Algorithm, Maximum Likelihood Estimation, Progressive Type I Hybrid Censoring, Kumaraswamy-Ters Rayleigh Distribution, Log-Kumaraswamy Distribution, Type I Hybrid Censoring

ÖNSÖZ

Bu çalışmanın gerçekleşmesinde hiçbir zaman bilgisini esirgemeyen, değerli danışman hocam Prof. Dr. Buğra SARAÇOĞLU' na, bilgi birikimiyle çalışmama destek olan hocam Prof. Dr. Coşkun KUŞ'a, Dr. Öğr. Üyesi Yunus AKDOĞAN' a, Arş. Gör. Caner TANIŞ' a, Arş. Gör. Kadir KARAKAYA' ya, bu süreçte hiçbir zaman desteklerini esirgemeyen canım aileme ve arkadaşlarıma teşekkürlerimi sunarım.

Egemen ÖZKAN
KONYA-2019

İÇİNDEKİLER

ÖZET ..... **iv**

ABSTRACT ..... **v**

ÖNSÖZ ..... **vi**

İÇİNDEKİLER ..... **vii**

SİMGELER VE KISALTMALAR ..... **ix**

ÇİZELGELER VE ŞEKİLLER ..... **x**

1. GİRİŞ VE KAYNAK ARAŞTIRMASI **1**

2. TEMEL KAVRAMLAR **6**

2.1. Sansürleme 6
2.1.1. Tip I Sağdan Sansürlü Örneklem 6
2.1.2 Tip II Sağdan Sansürlü Örneklem 7
2.1.3. İlerleyen Tür Tip II Sağdan Sansürlü Örneklem 7
2.1.4. Tip I Hibrit Sansürlü Örneklem 8
2.1.5. İlerleyen Tür Tip I Hibrit Sansürlü Örneklem 8
2.1.6. İlerleyen Tür Tip I Hibrit Sansürlü Örneklemden Sayı Üretme 9
2.2 Tahmin 11
2.2.1. Aralık Tahmini 11
2.2.2. Nokta Tahmini 11
2.2.3. En Çok Olabilirlik Yöntemi 12
2.2.4. Fisher Bilgisi 12
2.2.5. Düzgünlük Koşulları 12
2.2.6. Delta Methodu 13
2.2.7. EM Algoritması 14
2.3.Uyum İyiliği Kriterleri 15
2.3.1 Akaike Bilgi Kriteri (AIC) 15
2.3.2. Bayes Bilgi Kriteri (BIC) 15
2.3.3. Kolmogorov Smirnov Test İstatistiği 16
2.4. Bazı Sürekli Dağılımların Olasılık Yoğunluk ve Dağılım Fonksiyonları 16

3. LOG-KUMARASWAMY (LKW) DAĞILIMININ TİP I HİBRİT SANSÜRLÜ ÖRNEKLEM ALTINDA NOKTA VE ARALIK TAHMİNİ **18**

3.1. Log-Kumaraswamy Dağılımı 18
3.2. Log-Kumaraswamy Dağılımı İçin Nokta Tahmini 18
3.3. Log-Kumaraswamy Dağılımı İçin Aralık Tahmini 22
3.4. Simülasyon Çalışması 23
3.5. Gerçek Veri Analizi 25

4. KUMARASWAMY-TERS RAYLEIGH DAĞILIMININ İLERLEYEN TÜR TİP I HİBRİT SANSÜRLEME ALTINDA NOKTA VE ARALIK TAHMİNİ **28**

4.1.Kumaraswamy-Ters Rayleigh Dağılımı 28
4.2. Kumaraswamy Ters Rayleigh Dağılımı İçin Nokta Tahmini 29
4.3. Kumaraswamy Ters Rayleigh Dağılımı İçin Aralık Tahmini 35
4.4. Simülasyon Çalışması 37
4.5. Gerçek Veri Analizi 38

5. SONUÇLAR VE ÖNERİLER 42

5.1 Sonuçlar 42
5.2 Öneriler 43

KAYNAKLAR 44

ÖZGEÇMİŞ.....47

SİMGELER VE KISALTMALAR

Simgeler

$h(x)$: Hazard fonksiyonu

$L_s(a,b)$: Pseudo log-olabilirlik fonksiyonu

$L_c(\mathbf{W};a,b)$: Tam veriler için olabilirlik fonksiyonu

ε : Hata terimi

$F(x)$: Dağılım fonksiyonu

$\ell(\theta)$: Log-olabilirlik fonksiyonu

$f(x)$: Olasılık yoğunluk fonksiyonu

Θ : Parametre uzayı

Kısaltmalar

AIC : Akaike Bilgi Kriteri

BIC : Bayes Bilgi Kriteri

DTR : Dönüştürülmüş Ters Rayleigh

DW : Dönüştürülmüş Weibull

EM : Beklenti Maksimizasyonu

GR : Genelleştirilmiş Rayleigh

HKO : Hata Kareler Ortalaması

KS : Kolmogorov Smirnov

Kw-TR : Kumaraswamy-Ters Rayleigh

LKw : Log-Kumaraswamy

LL :Log-Lojistik

LN : Log-Normal

MLE : En Çok Olabilirlik Tahmin Edicisi

ÜÜ : Üstelleştirilmiş Üstel

W : Weibull

ÇİZELGELER VE ŞEKİLLER

Çizelgeler

Çizelge3.1**:** $T=10$ ve farklı n ve r değerleri için LKw dağılımı için nokta ve aralık tahmin sonuçları

Çizelge 3.2: $T=2$ ve farklı n ve r değerleri için LKw dağılımı için nokta ve aralık tahmin sonuçları

Çizelge 3.3: Yağış verisi İçin Uyum İyiliği Sonuçları

Çizelge 3.4: Farklı r ve T değerleri için ML tahminleri (HKO) ve asimptotik güven aralıkları

Çizelge 3.5: Karbon fiber verisi için uyum iyiliği sonuçları

Çizelge 3.6: Farklı r ve T değerleri için ML tahminleri (HKO) ve asimptotik güven aralıkları

Çizelge 4.1: $T=10$ ve farklı n ve r değerleri için Kw-TR dağılımı için nokta ve aralık tahmin sonuçları

Çizelge 4.2: Hastaların rahatlama zamanları ile ilgili uyum iyiliği sonuçları

Çizelge 4.3: Sansürleme şemaları ve üretilen veriler

Çizelge 4.4: Farklı sansürleme şemaları n, m Ve T değerleri için MLE tahminleri (HKO) ve asimptotik güven aralığı

Çizelge 4.5: Mekanik parçaların bozulma süreleri ile ilgili uyum iyiliği sonuçları

Çizelge 4.6: Sansürleme şemaları ve Üretilen veriler

Çizelge 4.7: Farklı sansürleme şemaları n, m ve T değerleri için MLE tahminleri (HKO) ve asimptotik güven aralığı

Şekiller

Şekil 3.1: Yağış verisi için uydurulan dağılım fonksiyonu grafikleri

Şekil 3.2: Karbon fiber verisi için uydurulan dağılım fonksiyonu grafikleri

Şekil 4.1: Hastaların rahatlama zamanları verisi için uydurulan dağılım fonksiyonu grafikleri

Şekil 4.2: Mekanik parçaların bozulma süreleri verisi için uydurulan dağılım fonksiyonu grafikleri

1. GİRİŞ VE KAYNAK ARAŞTIRMASI

Yaşam testi deneyleri bazı olasılık modelleri tarafından yürütülen güvenilirlik verileri çalışmaları ile ilgilidir. Yaşam testleri; biyoloji, klinik deneyler, finansal, tarım ve diğer çalışma alanları gibi yaşam analizlerini içeren bir çok uygulamadan oluşmaktadır. Genel olarak yaşam testi deneyleri, zaman ve maliyet durumları gibi çeşitli kısıtlar altında yürütülür. Maliyet ve zaman gibi kısıtlamalar dikkate alındığında genellikle test edilen tüm bileşenlerin bozulma zamanları kaydedilmez. Bu tür deneyler sansürlü verilere neden olmaktadır. (Sultana ve ark., 2018)

En yaygın iki sansürleme şeması geleneksel Tip I ve geleneksel Tip II sansürleme şemalarıdır. Kısaca bu sansürleme şemalarını açıklayalım. n birimin yer aldığı bir yaşam testi düşünelim. Geleneksel Tip I sansürleme şemasında yaşam testi önceden belirlenen T zamanına ulaştığında durdurulurken, geleneksel Tip II sansürleme şemasında önceden belirlenen $r(r<n)$ sayıda bozulma oluştuğunda yaşam testi durdurulur. (Kundu ve Pradhan, 2009)

Tip I sansürleme şemasında önceden belirlenen T zamanına kadar çok az bozulma hatta hiç bozulma gerçekleşmeyebilir. Tip II sansürleme de ise $r.$ bozulmanın zamanı belli olmadığı için deney süresi çok uzun olabilir. Bu tür nedenlerden dolayı alternatif bir sansürleme şeması olarak hibrit sansürleme şeması tanıtılmıştır.

Hibrit sansürleme şeması geleneksel Tip I ve Tip II sansürleme şemasının karışımı olarak bilinir. n bileşenin yer aldığı, bu bileşenlerin yaşam zamanlarının birbirinden bağımsız ve aynı dağılıma sahip olduğunu düşünelim ve bileşenlerin sıralı yaşam zamanları $X_{1:n}, X_{2:n}, ..., X_{n:n}$ olsun. Yaşam testi, n bileşenden ilk r tanesi veya önceden belirlenen T zamanına ulaştığında sona erer. Bir başka ifade ile $T_1^* = \min(X_{r:n}, T)$ zamanında yaşam testi sonlanır. (Epstein, 1954; Balakrishnan ve Kundu, 2013)

Geleneksel Tip I ve Tip II sansürleme şemalarında, test birimleri sonlandırma zamanı hariç herhangi bir zaman noktasında testten çıkarılmaz. Bu nedenden dolayı son zamanlarda, ilerleyen sansürleme şemaları altında literatürde oldukça çok çalışma yer almaktadır.

Daha genel bir sansürleme şeması olarak bilinen, İlerleyen Tip II sansürleme daha esnek bir yapıya sahiptir ve sonlandırma zamanından önce test birimleri belirli zamanlarda yaşam testinden çıkarılır. İlerleyen tip II sansürleme aşağıdaki gibi ifade

edilir; n test birimi 0 başlangıç zamanında teste tabi tutulur. İlk bozulma zamanında rasgele olarak r_1 tane yaşayan parça testten çıkarılır. İkinci bozulma zamanında geriye kalan $n-r_1-1$ kalan parçadan rastgele olarak r_2 tanesi testten çıkarılır. Devam eden bu süreçte, m.bozulma zamanında geriye kalan parçaların tamamı $r_m = n-m-r_1-\ldots-r_m$ testten çıkarılır. (Balakrishnan ve Aggarwala, 2000)

İlerleyen tip II sansürlemenin en büyük dezavantajı, *m.* bozulma zamanının süresinin çok zaman almasıdır. Bu nedenden dolayı ilerleyen Tip II sansürleme ve hibrit sansürlemelerinin karışımı olan ilerleyen Tip I hibrit sansürleme şeması tanıtılmıştır. Bu sansürleme şemasında yaşam testi, önceden belirlenen zamana ulaştığında veya önceden belirlenen bozulmaların sayısı gözlendiğinde yaşam testi durdurulur. (Kundu ve Joarder, 2006)

n birimin yer aldığı, test birimlerinin yaşam zamanlarının birbirinden bağımsız ve aynı dağılıma sahip olduğunu varsayalım. $r=(r_1,r_2,\ldots,r_m)$ önceden belirlenen sansürleme şemaları olsun. Burada $r_1+r_2+\ldots+r_m+m=n$ şeklindedir.. Ayrıca T zamanı önceden belirlenmiştir. İlk bozulma $X_{1:m:n}$ oluştuğunda yaşayan test birimlerinden rasgele olarak r_1 tanesi testten çıkarılır. Benzer şekilde ikinci bozulma zamanında $X_{2:m:n}$ oluştuğunda yaşayan test birimlerinden rasgele olarak r_2 tanesi yaşam testinden çıkarılır. Yaşam testi $T_* = min(X_{m:m:n}, T)$ zamanında son bulur. Başka bir anlatımla, şayet *m.* bozulma zamanı T zamanından önce gerçekleşirse yaşam testi $X_{m:m:n}$ anında son bulur. Aksi taktirde deney $X_{j:m:n} < T < X_{j+1:m:n}$ 'e karşılık gelen T zamanında son bulur ve geriye kalan tüm $r_j^* = n-r_1-r_2-\ldots-r_j-j$ adet test birimi testten çıkarılır.

Son yıllarda, zaman ve maliyetten kazanç sağladığı için hibrit sansürlü örneklemler altında çalışmalar oldukça önem kazanmıştır. Bu tez çalışmasında hibrit sansürlü örneklem altında, yeni yaşam dağılımlarının bilinmeyen parametreleri için nokta ve aralık tahminleri elde edilmiştir. Hibrit sansürleme ve EM algoritmasıyla ilgili bazı çalışmalar aşağıda sırasıyla verilmiştir.

(Epstein, 1954), tarafından yapılan çalışmada deney birimlerinin üstel yaşam zamanı varsayımı altında verilerin analizini yapmıştır. Ayrıca bu çalışmada, ortalama yaşam zamanı için iki yanlı güven aralığı oluşturulmuştur.

(Dempster ve ark., 1977) tarafından yapılan çalışmada Beklenti Maksimizasyonu (EM) algoritması tanıtılmıştır.

(Balakrishnan ve Aggarwala, 2000) tarafından yazılan kitapta, sansürlü örneklemlerden özellikle ilerleyen tip II sansürlü örneklemlere ilişkin temel kavramlar tanıtılmış ve ilerleyen tür sansürleme altında bazı sürekli dağılımlara ilişkin karakterizasyonlara yer verilmiştir. Ayrıca bazı sürekli dağılımlar kullanılarak geliştirilen simülasyon algoritmaları tanıtılmıştır. İlerleyen tür tip II sansürlü örneklemler altında bazı sürekli dağılımlara ilişkin istatistiksel sonuç çıkarım konuları da bu kitapta yer almaktadır.

(Kundu ve Joarder, 2006), çalışmalarında ilerleyen tür tip II Hibrit sansürleme şemasını tanıtmışlar ve bu şema altında Üstel dağılımın bilinmeyen parametreleri için en çok olabilirlik ve bayes tahmin edicilerini elde etmişlerdir. Ayrıca bu çalışmada, olabilirlik oran testine dayalı asimptotik güven aralıkları ve bootstrap güven aralıkları oluşturulmuştur.

(Kundu ve Pradhan, 2009), çalışmalarında test bileşenlerinin yaşam dağılımının iki parametreli genelleştirilmiş üstel dağılım varsayımı ile Tip I hibrit sansürlü örneklem altında nokta ve aralık tahminleri elde etmişlerdir. Bu çalışmada, en çok olabilirlik tahmin edicileri kapalı formda elde edilemediği için, bu tahmin ediciler EM algoritması yardımıyla elde edilmiştir. Ayrıca kayıp gözlem prensibi kullanılarak Fisher Bilgisi matrisinin elemanları kullanılarak asimptotik güven aralıkları oluşturulmuştur.

(Morabbi ve Razmkhah, 2009) tarafından yapılan çalışmada hibrit sansürleme şemaları altında, entropinin ayrışması konusu incelenmiştir. Tip I hibrit sansürleme şemasının entropisi formülize edilerek elde edilen bilgiler tip II hibrit sansürleme şemasının entropisini belirlemek için kullanılmıştır. Sonuçlar bazı yaşam zamanı dağılımlarına uygulanmış ve son olarak hibrit sansürleme şemalarının maksimum entropisi konusu ele alınmıştır.

(Dube ve ark., 2011), çalışmalarında hibrit sansürlü verilere dayalı log-normal dağılımının parametrelerinin tahmini konusunu incelemişlerdir. Bu dağılımın bilinmeyen parametreleri için en çok olabilirlik tahmin edicileri, EM algoritması kullanılarak elde edilmiştir. Fisher Bilgi Matrisi elde edilerek bu matris asimptotik güven aralıklarının oluşturulmasında kullanılmıştır.

(Park ve Lee, 2012), çalışmalarında tip I sağdan sansürlü örneklem altında parametre tahmini yapmışlardır. En çok olabilirlik tahminleri EM ve Monte Carlo EM (MCEM) tekniği kullanılarak elde edilmiştir. Bu algoritmalar Normal, Laplace ve Rayleigh dağılımlarının en çok olabilirlik tahmin edicilerini bulmada kullanılmıştır.

(Rastogi ve Tripathi, 2013), tarafından yapılan çalışmada, hibrit sansürlü örneklem altında, iki parametreli Burr XII dağılımının parametrelerinin istatistiksel çıkarımları ele alınmıştır. EM algoritması kullanılarak bilinmeyen parametreler için en çok olabilirlik tahmin edicileri elde edilmiştir. Fisher bilgi matrisi kullanılarak bilinmeyen parametreler için asimptotik güven aralıkları oluşturulmuştur. Bayes tahmin edicileri ve olağan güven aralıkları elde edilmiştir. Bilinmeyen parametrelerin Bayes tahminlerini hesaplayabilmek için Lindey yaklaşım metodu ve Markov Zincirleri Monte Karlo tekniği kullanılarak önerilen tahmin edicilerin performansları hata kareler ortalaması bakımından karşılaştırılmıştır. Son olarak iki adet gerçek veri analizi yapılmıştır.

(Balakrishnan ve Kundu, 2013), tarafından yapılan çalışmada Tip I ve Tip II Hibrit sansürleme konularını incelenmiş daha sonra genelleştirilmiş hibrit sansürleme yapısı ile bileşik hibrit sansürleme yapısı ile ilerleyen tür hibrit sansürleme yapıları hakkında geniş bilgilere yer verilerek bu şemalara dayalı sonuç çıkarım metotları incelenmiştir.

(Mao ve ark., 2014), tarafından yapılan makale çalışmasında, genelleştirilmiş tip I hibrit sansürlü üstel veriler için rekabetçi risk modeline ilişkin çıkarımsal sonuçlar ve parametre tahminleri yapılmıştır. Bilinmeyen parametreler için en çok olabilirlik tahmini kullanılarak, koşullu moment üreten fonksiyon ile parametrelerin koşullu dağılımları ve parametreler için tam güven aralığı elde edilmiştir. Boostrap metodu ve asimptotik dağılımlar kullanılarak asimptotik güven aralıkları oluşturulmuştur. Sonuçlar Monte Carlo simülasyonları ve gerçek veri analizi ile desteklenmiştir.

(Kohansal ve ark., 2015), tarafından yapılan çalışmada, tip II hibrit sansürlü örneklem durumunda, ağırlıklandırılmış üstel dağılım için parametre tahmini problemi incelenmiştir. EM algoritması kullanılarak parametreler için en çok olabilirlik tahminleri elde edilmiştir. Fisher bilgi matrisi kullanılarak asimptotik güven aralıkları oluşturulmuştur. Markov Zinciri Monte Carlo (MCMC) simülasyon teknikleri kullanılarak parametrelere ilişkin bayescil güven aralıkları ve bayes tahminleri elde edilmiştir. Son olarak sonuçların açıklayıcı olması için bir adet gerçek veri analizi yapılmıştır.

(Zhu ve Balakrishnan, 2015), tarafından yapılan çalışmada, Birnbaum–Saunders dağılımının parametrelerinin tip I, tip II ve hibrit sansürleme şemaları altında en çok olabilirlik tahmin edicilerinin varlığı ve tekliği problemi tartışılmıştır. Parametrelerin en çok olabilirlik tahmin edicilerini hesaplamak için nümerik yöntemler kullanılmış ve son olarak sonuçların açıklayıcı olması için gerçek veri analizi yapılmıştır.

(Ateya, 2017), çalışmasında Balakrishnan'ın bileşik hibrit sansürleme şeması altında, Ters Weibull dağılımının parametreleri için nokta ve aralık tahmin konuları incelenmiştir. Nokta tahminleri için, en çok olabilirlik ve bayes metodu kullanılmıştır. Monte Carlo simülasyonları kullanılarak karesel hata kayıp fonksiyonu ve linex kayıp fonksiyonuna dayanan bayes tahminleri hesaplanmıştır. Bayes tahminleri, bütün tahmin edicilerin hata kareler ortalaması hesaplanarak karşılık gelen en çok olabilirlik tahminleri ile kıyaslanmıştır. Son olarak bir adet gerçek veri analizi yapılmıştır.

(Kayal ve ark., 2017) tarafından yapılan çalışmada, ilerleyen tür tip I hibrit sansürleme şeması altında, Burr XII dağılımının bilinmeyen parametreleri EM algoritması kullanılarak en çok olabilirlik tahmin edicileri ve en çok olabilirlik tahmin edicilerine dayalı asimptotik güven aralıkları elde edilmiştir. Son olarak iki adet gerçek veri analizi yapılmıştır.

(Sultana ve ark., 2018),tarafından yapılan çalışmada hibrit sansürlü örneklem ile iki parametreli Kumaraswamy dağılımının bilinmeyen parametrelerinin tahminleri elde edilmeye çalışılmıştır. EM algoritması kullanılarak en çok olabilirlik tahmin edicileri türetilmiştir. Tahmin edicilere ilişkin asimptotik ve bootstrap güven aralıkları elde edilmiştir. Farklı yaklaşım metotları kullanılarak karesel kayıp fonksiyonu altında bayes tahminleri elde edilmiştir. Bir adet simülasyon çalışması yapılarak farklı yaklaşım metotları kullanılarak hesaplanan bayes tahmin edicileri ile en çok olabilirlik tahmin edicileri kıyaslanmıştır.

2. TEMEL KAVRAMLAR

Bu bölümde, tez içerisinde kullanılan tanım ve kavramlar tanıtılmıştır.

2.1. Sansürleme

Yaşam testi deneyleri; biyoloji, klinik, finans tarım gibi bir çok alanda yaygın bir şekilde kullanılmaktadır. Mekanik veya elektronik bileşenlerin saat, gün, hafta ,ay veya yıl olarak ölçülen çalışma süreleri yaşam veya başarısızlık zamanları olarak adlandırılır. Klinik çalışmalarda bu süreler, deneğin tedavi altına girdiği andan itibaren deneğin yaşam süresini temsil eder. Klinik çalışmalarda yaşam testine tabi tutulan deneklerden bazıları ilgi eksikliğinden veya çeşitli durumlardan dolayı tedaviden vazgeçebilir. Bu durumda tedaviden vazgeçen deneklerin geriye kalan yaşam zamanları hakkında bilgi sahibi olunmadığından bu türdeki veriler "sansürlü veriler" olarak adlandırılır. Genel olarak ifade edilirse yaşam testi deneyleri, zaman ve maliyet durumları gibi çeşitli kısıtlar altında yürütüldüğünde genellikle test edilen tüm bileşenlerin bozulma zamanları hakkında bilgi sahibi olunmaz. Bu tür deneyler sansürlü verilere neden olmaktadır. (Sultana ve ark., 2018;Balakrishnan ve Rao, 1998a).

2.1.1. Tip I Sağdan Sansürlü Örneklem

n bileşenin yer aldığı bir sistemdeki bileşenlerin yaşam zamanlarının birbirinden bağımsız ve aynı dağılımlı rasgele değişkenler olduğu varsayılsın. Yaşam testi önceden belirlenen bir $T\left(T>0\right)$ zamanına geldiğinde sonlandırılır. Ayrıca bozulan bileşenlerin değiştirilmediği varsayılmaktadır. Bu tür sansürleme yapısında sadece örneklemin belirli bir bölümüyle ilgilenildiği için bazı bilgiler kayıptır. Tip I sansürleme yapısında testin süresi belirlidir ancak bozulmaların sayısı bir rasgele değişkendir. Bu tip sansürlemede önceden belirlenen T zamanına kadar hiç bozulma gerçekleşmeyebilir veya çok az bozulma gerçekleşebilir. Bu tip I sansürlü örneklemin dezavantajıdır. (Balakrishnan ve Kundu, 2013). T anına kadar bu n tane bileşenden $k\left(k\le n\right)$ tanesinin yaşam zamanı gözlensin. Bu durumda $X_{1:n},X_{2:n},\ldots X_{k:n}$ şeklinde elde edilen örneklem tip I sağdan sansürlü örneklem olarak adlandırılır. Burada $X_{i:n}$, i. sıra

istatistiğini göstermektedir. Tip I sağdan sansürlü örneklem altında ortak olasılık yoğunluk fonksiyonu aşağıdaki gibidir.

$$f(x_{1:n}, x_{2:n}, \ldots x_{k:n} \mid \theta) = \frac{n!}{(n-k)!} \prod_{i=1}^{k} f(x_{i:n}) \left[1 - F(T)\right]^{n-k} \qquad (2.1)$$

(Balakrishnan ve Rao, 1998b).

2.1.2 Tip II Sağdan Sansürlü Örneklem

n tane bileşenden oluşan bir sistemdeki bileşenlerin yaşam zamanları birbirinden bağımsız ve aynı dağılımlı rasgele değişkenler olsun. Tip II sansürleme yapısında deney önceden belirlenen bozulmaların sayısı olan $r\left(r \le n\right)$ sayısına ulaştığında test sonlandırılır. Tip II sansürleme yapısında bozulmaların sayısı önceden belirlidir ancak deney süresi bir rasgele değişkendir. $r.$ bozulmanın tam zamanı belirli olmadığı için deney süresi çok uzun olabilir. Bu istenmeyen durum Tip II sağdan sansürleme için bir dezavantajdır. (Balakrishnan ve Rao, 1998b). $X_{1:n}, X_{2:n}, \ldots X_{k:n}$ ile gösterilen Tip II sansürleme altında ortak olasılık yoğunluk fonksiyonu aşağıdaki gibi ifade edilir.

$$f(x_{1:n}, x_{2:n}, \ldots, x_{r:n} \mid \theta) = \frac{n!}{(n-r)!} \prod_{i=1}^{r} f(x_{i:n}) [1 - F(x_{r:n})]^{n-r} \qquad (2.2)$$

(Balakrishnan ve Rao, 1998b) (Kuş, 2004).

2.1.3. İlerleyen Tür Tip II Sağdan Sansürlü Örneklem

Birbirinden bağımsız ve aynı dağılıma sahip n bileşenin yer aldığı bir yaşam testi deneyi düşünülsün. Deneyci n parçadan m $\left(m \le n\right)$ tanesinin bozulma zamanını gözlemektedir. İlk bozulma zamanında kalan $n-1$ yaşayan parçadan rasgele seçilen r_1 tanesi testten atılır. İkinci bozulma zamanında kalan $n - r_1 - 1$ yaşayan parçadan rasgele seçilen r_2 tanesi testten atılır. Bu şekilde devam eden süreçte son olarak $m.$ bozulma zamanında yaşayan parçaların tamamı olan $r_m = n - m - r_1 - \ldots - r_{m-1}$ adet parça yaşam testinden atılır. r_i ile gösterilen sansürlemelerin sayısı ve m ile gösterilen bozulmaların sayısı, deney başlamadan önce belirlenmiştir. Burada $n = \sum_{i=1}^{m} r_i + m$ şeklindedir.

$X_{1:m:n}, X_{2:m:n}..., X_{m:m:n}$'ler sıralanmış m tane bozulma zamanını gösterir ve ilerleyen tür tip II sansürlenmiş sıra istatistikleri olarak adlandırılır. İlerleyen tip II sansürleme altında ortak olasılık yoğunluk fonksiyonu aşağıdaki gibi ifade edilir.

$$f(x_{1:m:n},...x_{m:m:n}/\theta) = C\prod_{i=1}^{m} f(x_i)[1-F(x_i)]^{r_i},\ 0<x_1<x_2.....<x_m<\infty \qquad (2.3)$$

Burada $C = n(n-r_1-1)\times...\times(n-r_1-r_2-...-r_{m-1}-m+1)$'dir. (Balakrishnan ve Saleh, 2012); (Balakrishnan ve Aggarwala, 2000; Kundu ve Joarder, 2006).

2.1.4. Tip I Hibrit Sansürlü Örneklem

Hibrit sansürleme şeması geleneksel Tip I ve Tip II sansürleme şemasının karışımı olarak bilinir. n bileşenin yer aldığı bir sistemdeki bileşenlerin yaşam zamanlarının birbirinden bağımsız ve aynı dağılıma sahip olduğu düşünülsün. Bu bileşenlerin sıralı yaşam zamanları $X_{1:n}, X_{2:n},..., X_{n:n}$ olsun. Yaşam testi, n bileşenden r adet bozulmaya ulaşıldığında veya önceden belirlenen T zamanına ulaştığında sona erer. Bir başka ifade ile $T_1^* = \min(X_{r:n}, T)$ zamanında yaşam testi sonlanır. Gözlenen örneklemler aşağıdaki iki durumdan biri olabilir.

Durum1: $X_{1:n}, X_{2:n},..., X_{r:n}$ eğer $X_{r:n} \le T$

Durum2 : $X_{1:n}, X_{2:n},..., X_{m:n}$ eğer $X_{m+1:n} > T$, $m < r$

Bu örneklemler altında ortak olasılık yoğunluk fonksiyonları;

$$f(x_{1:n}, x_{2:n},..., x_{r:n}|\boldsymbol{\theta}) = \left(\frac{n!}{(n-r)!}\right)\prod_{i=1}^{r} f(x_{i:n})[1-F(x_{r:n})]^{(n-r)}$$
$$f(x_{1:n}, x_{2:n},..., x_{m:n}|\boldsymbol{\theta}) = \prod_{i=1}^{m} f(x_{i:n})[1-F(T)]^{(n-m)} \qquad (2.4)$$

şeklinde olacaktır (Rastogi ve Tripathi, 2013).

2.1.5. İlerleyen Tür Tip I Hibrit Sansürlü Örneklem

$X_1, X_2,..., X_n$ birbirinden bağımsız aynı dağılıma sahip yaşam zamanları olsun. Bu n tane bileşen yaşam testine tabi tutulduğunda, m bozulma sayısını, $r = (r_1, r_2, \ldots, r_m)$ testten çıkarılan parçalardan oluşan ve $r_1 + r_2 + \ldots + r_m + m = n$ özelliğine sahip sansür şemasını, T önceden belirlenen zamanı ve $X_{i:m:n}$'de i. parçanın

bozulma zamanını göstersin. Yaşam testinde ilk bozulma zamanında yaşayan test birimlerinden rasgele olarak r_1 tanesi, ikinci bozulma zamanında yaşayan test birimlerinden rasgele olarak r_2 tanesi çıkarılarak devam eden süreç $T_* = min\left(X_{m:m:n}, T\right)$ zamanında son bulur. Eğer m. bozulma zamanı T zamanından önce gerçekleşirse yaşam testi $X_{m:m:n}$ 'de son bulur. Aksi taktirde deney $X_{j:m:n} < T < X_{j+1:m:n}$ 'e karşılık gelen T zamanında son bulur ve hayatta kalan $r_j^* = n - r_1 - r_2 - \ldots - r_j - j$ adet test birimi testten çıkarılır. Burada $X_{j:m:n}$, T zamanına kadar oluşan bozulmaların sayısını gösterir. Bu koşullar altında ilerleyen Tip I hibrit sansürleme yapısında gözlenen örneklemler aşağıdaki iki durumdan biri olabilir.

$$\begin{cases} I : \{X_{1:m:n}, X_{2:m:n}, ..., X_{m:m:n}\}, eğer\ X_{m:m:n} < T \\ II : \{X_{1:m:n}, X_{2:m:n}, ..., X_{j:m:n}\}, eğer\ X_{j:m:n} < T < X_{j+1:m:n} \end{cases} \tag{2.5}$$

Bu örneklemler altında ortak olasılık yoğunluk fonksiyonları ise aşağıdaki gibi ifade edilir.

$$\begin{cases} I : f(x_{1:m:n}, x_{2:m:n}, ..., x_{m:m:n} \,|\boldsymbol{\theta}) = C\prod_{i=1}^{m} f(x_{i:m:n})[1-F(x_{i:m:n})]^{r_i} \\ II : f(x_{1:m:n}, x_{2:m:n}, ..., x_{j:m:n} \,|\boldsymbol{\theta}) = D\prod_{i=1}^{j} f(x_{i:m:n})[1-F(x_{i:m:n})]^{r_i}[1-F(T)]^{r_j^*} \end{cases} \tag{2.6}$$

Buradaki C ve D sabitleri;

$C = n(n-r_1-1)...(n-r_1-r_2-...r_{m-1}-m+1)$

$D = n(n-r_1-1)...(n-r_1-r_2-...r_{m-1}-j+1)$

biçimindedir. (Kayal ve ark., 2017).

2.1.6. İlerleyen Tür Tip I Hibrit Sansürlü Örneklemden Sayı Üretme

Balakrishnan ve Sandhu (1995) nun algoritmasını kullanarak ilerleyen tür tip II sağdan sansürlü örneklemden sayı üretmek için aşağıdaki yöntem kullanılır.

$U_{1:m:n}, U_{2:m:n}, ..., U_{m:m:n}$, $R_1, R_2, .., R_m$ sansür şemalı ve Düzgün$(0,1)$ dağılımdan alınmış ilerleyen tür tip II sağdan sansürlü sıra istatistikleri olmak üzere;

$$V_1 = \frac{1-U_{m:m:n}}{1-U_{m-1:m:n}}$$
$$V_2 = \frac{1-U_{m-1:m:n}}{1-U_{m-2:m:n}}$$
$$\vdots$$
$$V_{m-1} = \frac{1-U_{2:m:n}}{1-U_{1:m:n}}$$
$$V_m = 1-1-U_{1:m:n}$$

dönüşümleri uygulanarak birbirinden bağımsız ve her biri $Beta\left(i+\sum_{j=m-i+1}^{m} R_j, 1\right)$ dağılımına sahip V_i rasgele değişkenleri elde edilir. Bu durumda $W_i = V_i^{\,i+\sum_{j=m-i+1}^{m} R_j}$ $(i=1,...,m)$ dönüşümü uygulanarak birbirinden bağımsız ve Düzgün$(0,1)$ dağılıma sahip W_i rasgele değişkenleri elde edilir. Bu bilgiler ışığında ilerleyen tür tip II sağdan sansürlü örneklemden sayılar aşağıdaki gibi üretilir.

1. Düzgün$(0,1)$ dağılımdan m adet $W_1, W_2, ..., W_m$ sayıları üretilir.

2. $W_1, W_2, ..., W_m$ kullanılarak $V_i = W_i^{1/\left(i+\sum_{j=m-i+1}^{m} R_j\right)}$, $i=1,2,...,m$ sayıları elde edilir.

3. $U_{i:m:n} = 1-V_m V_{m-1}...V_{m-i+1}, i=(1,2,...,m)$ dönüşümü uygulanarak elde edilen $U_{1:m:n}, U_{2:m:n}, ..., U_{m:m:n}$' ler Düzgün$(0,1)$ dağılımdan üretilen m birimlik ilerleyen tür tip II sağdan sansürlü örneklemler elde edilir.

Son olarak $X_{i:m:n} = F^{-1}(U_{i:m:n})$, $i=1,2,...,m$ rasgele değişkenleri elde edilir. burada $F^{-1}(.)$, ilgilenilen yaşam zamanı dağılımının dağılım fonksiyonunun tersi olarak ifade edilir. Böylece elde edilen $X_{1:m:n}, X_{2:m:n}, ..., X_{m:m:n}$'ler $F(\cdot)$ dağılımından üretilen m birimlik ilerleyen tür tip II sağdan sansürlü örneklemlerdir. Eğer $X_{1:m:n}, X_{2:m:n}, ..., X_{m:m:n}$ örnekleminde $X_{m:m:n} < T$ ise eşitlik (2.5)'te yer alan Durum I örnekleminden aksi halde yani $X_{m:m:n} > T$ ise eşitlik (2.6)'da yer alan Durum II örnekleminden sayı üretilerek ilerleyen tür tip I hibrit sansürlü örneklemler elde edilir (Balakrishnan ve Sandhu, 1995; Balakrishnan ve Aggarwala, 2000; Kundu ve Joarder, 2006).

2.2 Tahmin

Bir konu hakkında araştırma yapılmak istendiğinde kitleye ilişkin karakteristik özelliklerin incelenmesi ve kitlenin bilinmeyenleri hakkında istatistiksel çıkarımların yapılması araştırma sonuçlarının yorumlanabilmesi açısından oldukça önemlidir. Bu istatistiksel çıkarım konularından biri kitle parametrelerinin tahmini problemidir. Tahmin problemi matematiksel olarak aşağıdaki biçimde tanımlanabilir. $X_1, X_2, ..., X_n$, $f(.;\underline{\theta})$, $\underline{\theta} \in \Theta \subset \mathbb{R}^r$ dağılımından alınmış n birimlik örneklem olsun. Örneklemin herhangi bir fonksiyonu olarak tanımlanan bir istatistik kitle parametresini tahmin etmek amacıyla kullanıldığında tahmin edici adını alır. Tahmin edicinin aldığı değere ise tahmin denir. Tahmin nokta ve aralık tahmini olmak üzere ikiye ayrılır.

2.2.1. Aralık Tahmini

$X_1, X_2, ..., X_n$, $f(\cdot;\theta)$ $\theta \in \Theta \subset \mathbb{R}$ dağılımına sahip bağımsız ve aynı dağılımlı rasgele değişkenler olsun. Burada θ gerçel değerli bir parametredir. Rasgele aralık, en az bir sınır noktasında rasgele değişken olan sonlu veya sonsuz aralık olarak tanımlanır. $L(X_1,...,X_n)$ ve $U(X_1,...,X_n)$, $L(X_1,...,X_n) \le U(X_1,...,X_n)$ olacak şekilde iki farklı istatistik olmak üzere $P_\theta\left[L(X_1,...X_n) \le \theta \le U(X_1,...X_n)\right] \ge 1-\alpha, \forall \theta \in \Theta$ koşulunu sağlayan $\left[L(X_1,...,X_n), U(X_1,...,X_n)\right]$ rasgele aralığına θ parametresi için $1-\alpha(0<\alpha<1)$ güven seviyeli güven aralığı denir. (Roussas, 1997).

2.2.2. Nokta Tahmini

Bilinmeyen kitle parametreleri için nokta tahmin edicilerinin elde edilebilmesi için literatürde bir çok yöntem kullanılmaktadır. Bu yöntemlerden en sık kullanılanları en çok olabilirlik, momentler, en küçük kareler ve bayes tahmin edicilerini elde etme yöntemleridir. Bu tez çalışmasında en çok olabilirlik yöntemine ağırlık verilmiştir.

2.2.3. En Çok Olabilirlik Yöntemi

$X_1, X_2, ..., X_n$ $f(\cdot;\theta), \theta \in \Theta \subseteq \mathbb{R}^p$ dağılımına sahip n birimlik örneklem ve bu örnekleme ilişkin ortak olasılık (yoğunluk) fonksiyonu $f(x_1, x_2, \ldots, x_n | \theta)$ olsun. Bu durumda, $L(\theta | \underline{x}) = f(\theta | x_1, x_2, ..., x_n) = \prod_{i=1}^{n} f(\theta | x_i)$ biçiminde tanımlanan, θ'nın bir fonksiyonu olan $L(\theta | \underline{x})$ fonksiyonuna olabilirlik fonksiyonu adı verilir. Olabilirlik fonksiyonunun logaritması yani log-olabilirlik fonksiyonu ise;

$$\ell(\theta) = Log\left(L(\theta | \mathbf{x})\right) = \sum_{i=1}^{n} \log\left(f(\theta | x_i)\right) \tag{2.7}$$

biçiminde tanımlanır. Olabilirlik fonksiyonunu veya log-olabilirlik fonksiyonunu maksimum yapan θ değeri θ'nın en çok olabilirlik (ML) tahmini olarak adlandırılır. Logaritma fonksiyonu monoton artan bir fonksiyon olduğu için $L(\theta | x)$ ve $LogL(\theta | x)$ aynı θ değerinde maksimize olacaktır. (Roussas, 1997).

2.2.4. Fisher Bilgisi

$X_1, X_2, ..., X_n$ olasılık (yoğunluk) fonksiyonu $f(x | \theta)$, $\theta \in \Theta \subset R^p$ olan n birimlik örneklem ve bu örnekleme ilişkin log-olabilirlik fonksiyonu $\log\left(L(\boldsymbol{\theta} | x)\right), \boldsymbol{\theta} \in \Theta \subset R^p$ olmak üzere bu örnekleme dayalı bazı düzgünlük koşullarının sağlanması durumunda fisher bilgisi aşağıdaki gibi ifade edilir (Roussas, 1997).

$$I(\boldsymbol{\theta}) = -E_\theta\left[\frac{\partial^2}{\partial \boldsymbol{\theta}^2} \log\left(L(\boldsymbol{\theta} | x)\right)\right] \tag{2.8}$$

2.2.5. Düzgünlük Koşulları

$X_1, X_2, ..., X_n$ birbirinden bağımsız $f(x | \theta)$ dağılımına sahip n birimlik örneklem olmak üzere düzgünlük koşulları aşağıdaki gibidir.

1- Parametre uzayı açık bir aralıktır.

2- Olasılık yoğunluk fonksiyonu $f(x | \theta)$, $\theta \in \Theta$ noktasında türevlenebilirdir.

3- X_i 'ler aynı destek kümesine sahiptir ve $A=\{x: f(x|\theta)>0\}$ ile gösterilen destek kümesi θ 'dan bağımsızdır.

4- $\theta \neq \theta'$ ise $f(x|\theta) \neq f(x|\theta')$ biçimindedir.

5- $\forall \underline{x} \in \chi$ için, $f(x|\theta)$ yoğunluk fonksiyonu θ ' ya göre üç kez türevlenebilirdir. Burada χ örneklem uzayını ifade etmektedir.

6- $\forall \theta_0 \in \Theta$ için θ_0 'a bağlı pozitif bir c sayısı ve bir $M(x)$ fonksiyonu mevcut olsun. $\forall \underline{x} \in \chi$ için $E_{\theta_0}[M(X)]<\infty$ olması durumunda;

$$\left|\frac{\partial^3}{\partial\theta^3}\log f(x|\theta)\right| \leq M(x), \quad \theta_0 - c < \theta < \theta_0 + c.$$

şeklindedir. $X_1, X_2, ..., X_n$' ler $f(x|\theta)$ dağılımına sahip n birimlik örneklem ve olabilirlik fonksiyonu $L(\theta|\underline{x}) = \prod_{i=1}^{n} f(x_i|\theta)$ olmak üzere, θ' nın ML tahmini $\hat{\theta}$ olsun. En çok olabilirlik tahmin edicisinin (MLE) değişmezlik özelliğinden $\tau(\hat{\theta})$ 'da θ'nın sürekli bir fonksiyonu olan $\tau(\theta)$ için MLE olur. Düzgünlük koşullarının ilk dört koşulu altında $\varepsilon > 0$ ve $\theta \in \Theta$ için, $\lim_{n\to\infty} P_\theta\left(\left|\tau\left(\hat{\theta}\right)-\tau(\theta)\right| \geq \varepsilon\right)=0$ olması durumunda $\tau(\hat{\theta}), \tau(\theta)$ için tutarlıdır. Düzgünlük koşullarının beş ve altı durumlarının sağlanması durumunda ise $\tau(\hat{\theta})$, $\tau(\theta)$ için asimptotik etkin tahmin edici olur ve ayrıca bu tahmin edici asimptotik normal dağılıma sahiptir. Yani $\sqrt{n}\left(\tau\left(\hat{\boldsymbol{\theta}}\right)-\tau(\boldsymbol{\theta})\right) \xrightarrow{d} N\left(0, I^{-1}(\boldsymbol{\theta})\right)$ 'dır.(Casella ve Berger, 2001; Lehmann ve Casella, 2006)

2.2.6. Delta Methodu

$\left(Y_n\right)_{n\geq 1}$, dağılım fonksiyonu $F_n(y)$ olan rasgele değişkenlerin bir dizisi ve Y, $F(y)$ dağılım fonksiyonuna sahip bir rasgele değişken olmak üzere F 'in sürekli olduğu noktalarda $\lim_{n\to\infty} F_n(y) = F(y)$ oluyorsa $\left(Y_n\right)_{n\geq 1}$ rasgele değişkenler dizisi Y rasgele değişkenine dağılımda yakınsar denir ve $Y_n \xrightarrow{d} Y$ şeklinde gösterilir.

$(Y_n)_{n\geq 1}$ rasgele değişkenlerin bir dizisi olmak üzere $\sqrt{n}(Y_n-\theta)\xrightarrow{d}N(0,\sigma^2)$ 'dir. Verilen g fonksiyonu ve bir θ değeri için $g'(\theta)$ mevcut ve sıfırdan farklı olsun. Bu durumda;

$\sqrt{n}\left[g(Y_n)-g(\theta)\right]\xrightarrow{d}g'(\theta)\sqrt{n}(Y_n-\theta)$'dir (Casella ve Berger, 2001).

2.2.7. EM Algoritması

Beklenti maksimizasyonu; veriler tam olarak gözlenmediği durumlarda ve sansürlü veri problemlerinde parametrik modellerin en çok olabilirlik tahminlerini bulmak için kullanılan bir tekniktir. Olabilirlik fonksiyonlarının maksimize edilmesi zor olduğundan dolayı, (Dempster ve ark., 1977) EM algoritmasını tanıtmışlardır. Beklenti maksimizasyonu (EM) algoritması tam olmayan veri problemini, tam veri haline getirerek çözmektir. EM algoritması iki adımdan oluşur. Her iterasyonun E adımı (beklenti adımı) koşullu tam verinin dağılımı üzerinden beklenen değerini almayı içerir. Her iterasyonun M adımı (maksimizasyon adımı) ise basit kapalı formda olan tam verinin en çok olabilirlik tahminini gerektirir. Tam olmayan veri problemleri için EM algoritması diğer optimizasyon tekniklerine göre en önemli iki özelliği basitliği ve durağan olmasıdır. Tam olmayan verinin karmaşık log-olabilirlik fonksiyonunu maksimize etmek yerine eksik veriler verildikten sonra tam verilere dayalı log-olabilirlik fonksiyonu tekrarlı olarak maksimize edilir. Ayrıca EM algoritmasının ardışık iterasyonları log-olabilirlik fonksiyonunun asla azalmamasını sağlar. Bu genellikle Newton-Raphson gibi gradient yöntemler için sağlanmayabilir. Bundan dolayı EM algoritması, herhangi bir başlangıç noktası aldığında küresel maksimum noktasına yakınsar. Sansürlü veriler kayıp gözlem durumlarının özel bir durumu olduğu için, sansürlü veriler için de EM algoriması uygulanabilir (Park ve Lee, 2012).

EM'nin matematiksel olarak anlatımı aşağıdaki gibidir. θ parametresinin mevcut tahmini $\theta^{(t)}$ olsun. EM'nin E-adımında beklenen tam verinin log-olabilirlik fonksiyonu

$$Q(\theta\,|\,\theta^{(t)})=\int \ell(\theta\,|\,y)\,f(Y_{kayıp}\,|\,Y_{gözlenen},\theta=\theta^{(t)})dY_{kayıp}$$

şeklinde elde edilir. EM'nin M-adımı, bütün θ değerleri için beklenen tam verinin log-olabilirlik fonksiyonunu maksimize ederek $\theta^{(t+1)}$'i belirler.(Little ve Rubin, 2014)

burada t iterasyonu belirtir ve $Q(\theta^{(t+1)}\,|\,\theta^{(t)})\geq Q(\theta\,|\,\theta^{(t)}))$ biçimindedir.

2.3.Uyum İyiliği Kriterleri

2.3.1 Akaike Bilgi Kriteri (AIC)

(Kullback ve Leibler, 1951) tahmin değeri gerçek değere yakınsadığında kaybolan bilgiyi elde etmek için bir ölçüm kriteri geliştirdi. Kulback ve Leibler ölçümü bilgi kaybını minimize eden iyi bir model kriteridir. Japon istatistikçi (Akaike, 1998) Kulback-Leibler ölçüsü ile en çok olabilirlik tahmin yöntemi arasında bir ilişki kurdu ve yeni bir model seçim kriteri olarak Akaike bilgi kriterini (AIC) tanıttı. Bu yöntem farklı modellerin kıyaslanması için önerilen bir yöntemdir. AIC aşağıdaki gibi formülize edilir.

$$AIC = -2\ell(\boldsymbol{\theta}) + 2k \tag{2.9}$$

Burada $\ell(\theta)$ log-olabilirlik fonksiyonunu, k ise parametre sayısını göstermektedir. Modeller kıyaslanırken en küçük AIC değeri tercih edilir (Hurvich ve Tsai, 1989; Cavanaugh, 1997; UCAL, 2006).

2.3.2. Bayes Bilgi Kriteri (BIC)

Bayes bilgi kriteri (BIC) veya Schwarz bilgi kriteri sonlu bir model kümesi arasında, model seçimi için kullanılan bir kriterdir. Bu kısmen olabilirlik fonksiyonuna dayanır ve AIC ile ilişkilidir. Uyum iyiliği ölçülen modelin parametre sayısının fazla olması BIC değerinin hesaplanmasında model için bir dezavantajdır. BIC değeri uydurulan modelin parametre tahminleri ile hesaplanan log-olabilirlik fonksiyonunun değeri ile parametre sayısına bağlıdır. BIC aşağıdaki gibi ifade edilir.

$$BIC = -2\ell\left(\hat{\theta}\right) + k\log(n) \tag{2.10}$$

Burada $\ell\left(\hat{\theta}\right)$ log-olabilirlik fonksiyonunu, k parametre sayısını, n örneklem büyüklüğünü göstermektedir. Modeller karşılaştırılırken en küçük BIC değeri tercih edilir (Hurvich ve Tsai, 1989; Cavanaugh, 1997).

2.3.3. Kolmogorov Smirnov Test İstatistiği

Uyum iyiliği testleri, yokluk hipotezini test etmek için bazı bilinmeyen dağılımlardan rasgele olarak seçilen bir örneklemin gerçekte bilinen bir dağılımından gelip gelmediğinin incelenmesini içerir. En popüler parametrik olmayan uygunluk testi Kolmogorov-Smirnov testidir. F dağılım fonksiyonu ile $X_1, X_2, \ldots, X_n$ örneklemi göz önüne alındığında, $H_0 : F = F_0$ hipotezinin $H_1 : F = F_0$'ın hipotezine göre test edilme problemini düşünelim. Burada F_0, H_0 yokluk hipotezinde verilen bazı özel dağılım fonksiyonunu ve $F_n(x)$ ampirik dağılım fonksiyonunu göstermektedir. Buna göre Kolmogorov- Smirnov test istatistiği (D),

$$D = \sup_x \left| F_n(x) - F_0(x) \right| \tag{2.11}$$

şeklinde ifade edilir (Zamar ve ark., 1994). Burada $F_n(x)$ örneklemin ampirik dağılım fonksiyonu ifade etmektedir. Bu istatistiğin küçük değerleri verilerin dağılıma uygunluk gösterdiğini ifade eder. Ayrıca, tek örneklem için D istatistiği, Kolmogorov-Smirnov testine karşılık gelen tablodan n ve $1-\alpha$ değerine göre bulunan D_k ve örneklemden hesaplanan değer D_h olmak üzere $D_h \geq D_k$ ise H_0 reddedilir (Yildirim ve Gökpinar, 2012) .

2.4. Bazı Sürekli Dağılımların Olasılık Yoğunluk ve Dağılım Fonksiyonları

Üstelleştirilmiş Üstel (ÜÜ) dağılımının, dağılım fonksiyonu ve olasılık yoğunluk fonksiyonları aşağıda sırasıyla verilmiştir (Gupta ve Kundu, 2001).

$$f(x) = ab\exp(-ax)\left(1-\exp(-ax)\right)^{b-1} \tag{2.12}$$

$$F(x) = \left(1-e^{-ax}\right)^b, x,\ a,b>0 \tag{2.13}$$

Weibull (W) dağılımının, dağılım fonksiyonu ve olasılık yoğunluk fonksiyonları aşağıda sırasıyla verilmiştir (Weibull, 1951).

$$f(x) = \frac{b}{a}\left(\frac{x}{a}\right)^{b-1} e^{-\left(\frac{x}{a}\right)^b} \tag{2.14}$$

$$F(x) = 1-\exp\left(-\left(\frac{x}{a}\right)\right)^b,\ a,b,x>0 \tag{2.15}$$

Dönüştürülmüş Weibull (DW) dağılımının, dağılım fonksiyonu ve olasılık yoğunluk fonksiyonları aşağıda sırasıyla verilmiştir (Aryal ve Tsokos, 2011).

$$F(x)=\left(1-e^{\left(-\left(\frac{x}{b}\right)^{a}\right)}\right)\left(1+\lambda e^{\left(-\left(\frac{x}{b}\right)^{a}\right)}\right),\ a,b,x>0,-1<\lambda<1 \tag{2.16}$$

$$f(x)=\frac{b}{a}\left(\frac{x}{b}\right)^{a-1}e^{-\left(\frac{x}{b}\right)^{a}}\left(1-\lambda+2\lambda e^{-\left(\frac{x}{b}\right)^{a}}\right) \tag{2.17}$$

Log-Normal (LN) dağılımın yoğunluk fonksiyonu aşağıda verilmiştir (Aitchison ve Brown, 1957).

$$f(x)=\frac{1}{b\sqrt{2\pi}x}e^{-\frac{(\log x-a)^2}{2\sigma^2}},\ x,\sigma>0,a\in\mathbb{R} \tag{2.18}$$

Log-lojistik (LL) dağılımının, dağılım fonksiyonu ve olasılık yoğunluk fonksiyonları aşağıda sırasıyla verilmiştir (Kantam ve ark., 2001).

$$f(x)=\frac{b}{a}\frac{(x/a)^{b-1}}{\left(1+(x/a)^{b}\right)^{2}} \tag{2.19}$$

$$F(x)=\frac{1}{1+(x/a)^{b-1}},\ x,\ a,b>0 \tag{2.20}$$

Genelleştirilmiş Rayleigh (GR) dağılımının, dağılım fonksiyonu ve olasılık yoğunluk fonksiyonları aşağıda sırasıyla verilmiştir (Kundu ve Raqab, 2005).

$$f(x)=2a^{2}bxe^{-(ax)^{2}}\left(1-e^{-(ax)^{2}}\right)^{b-1} \tag{2.21}$$

$$F(x)=\left(1-e^{-(ax)^{2}}\right)^{b},\ x,\ a,b>0 \tag{2.22}$$

Dönüştürülmüş Ters Rayleigh (DTR) dağılımının, dağılım fonksiyonu ve olasılık yoğunluk fonksiyonları aşağıda sırasıyla verilmiştir (Ahmad ve ark., 2014).

$$f(x)=\frac{2b}{x^{3}}e^{-\frac{b}{x^{2}}}\left(1+a-ae^{-\frac{b}{x^{2}}}\right) \tag{2.23}$$

$$F(x)=e^{-\frac{b}{x^{2}}}\left(1+a-ae^{-\frac{b}{x^{2}}}\right),\ b,x>0,-1<a<1 \tag{2.24}$$

3. LOG-KUMARASWAMY (LKW) DAĞILIMININ TİP I HİBRİT SANSÜRLÜ ÖRNEKLEM ALTINDA NOKTA VE ARALIK TAHMİNİ

Bu bölümde Tip-I hibrit sansürlü örneklem altında LKw dağılımının bilinmeyen parametreleri için nokta ve aralık tahmin edicileri elde edilmiştir.

3.1. Log-Kumaraswamy Dağılımı

Kumaraswamy dağılımı ilk olarak 1980 yılında Kumaraswamy tarafından ortaya atılmıştır. Olasılık yoğunluk fonksiyonu ve dağılım fonksiyonu aşağıda verilmiştir.

$$f_X(x,a,b) = abx^{a-1}\left(1-x^a\right)^{(b-1)} \tag{3.1}$$

$$F_X(x;a,b) = 1-\left(1-x^a\right)^b \tag{3.2}$$

Burada $0 \le x \le 1, a > 0$ ve $b > 0$ 'dır (Kumaraswamy, 1980). Log-Kumaraswamy (LKw) dağılımı Lemonte ve ark. (2013) tarafından tanıtıldı ve yaşam analizi testlerinde kullanıldı. X rasgele değişkeni a ve b parametreli Kumaraswamy dağılımına sahip olsun. $Y = -\log(1-X)$ dönüşümü kullanılarak LKw dağılımının olasılık yoğunluk fonksiyonu ve dağılım fonksiyonu aşağıda verilmiştir (Lemonte ve ark., 2013).

$$f(y;a,b) = abe^{-y}\left(1-e^{-y}\right)^{(a-1)}\left(1-\left(1-e^{-y}\right)^a\right)^{b-1}, y > 0 \tag{3.3}$$

$$F(y;a,b) = 1-\left(1-\left(1-e^{-y}\right)^a\right)^b, y > 0 \tag{3.4}$$

3.2. Log-Kumaraswamy Dağılımı İçin Nokta Tahmini

$Y_{1:n}, Y_{2:n}, ..., Y_{n:n}$ a ve b parametreli LKw dağılımından alınmış sıralı yaşam zamanları olsun. Tip I hibrit sansürleme durumunda yaşam testi $T_1 = \min(Y_{r:n}, T)$ zamanında durdurulur ve gözlenen örneklemler aşağıdaki iki durumdan biri olabilir (Rastogi ve Tripathi, 2013).

Durum1: $I : \{Y_{1:n}, Y_{2:n}, ..., Y_{r:n}\}$ eğer $Y_{r:n} \le T$

Durum2 : $\{Y_{1:n}, Y_{2:n}, ..., Y_{m:n}\}$ eğer $m < r,\ Y_{r:n} > T$

Bu örneklemler altında olabilirlik fonksiyonu

$$L(a,b|\mathbf{y}) = \left(\frac{n!}{(n-r)!}\right)\prod_{i=1}^{r} f(y_{i:n})[1-F(y_{r:n})]^{(n-r)}$$
$$L(a,b|\mathbf{y}) = \left(\frac{n!}{(n-m)!}\right)\prod_{i=1}^{m} f(y_{i:n})[1-F(T)]^{(n-m)} \quad (3.5)$$

biçimindedir. Eşitlik (3.5) kullanılarak LKw dağılımı için olabilirlik fonksiyonu;

$$L(a,b) = \left(\frac{n!}{(n-d)!}\right)\prod_{i=1}^{d} abe^{-y}[1-e^{-y}]^{(a-1)}[1-(1-e^{-y})^{a}]^{(b-1)}[1-(1-e^{(-c)})^{a}]^{b(n-m)} \quad (3.6)$$

Burada c ve d ;

$$c = \begin{cases} y_{r:n} & \text{durum 1 için} \\ T & \text{durum 2 için} \end{cases} \quad (3.7)$$

$$d = \begin{cases} r & \text{durum 1 için} \\ m & \text{durum 2 için} \end{cases} \quad (3.8)$$

şeklindedir. Log- olabilirlik fonksiyonu ise;

$$\log L(a,b|\mathbf{y}) = \log\left(\frac{n!}{(n-d)!}\right) + d\log(a) + d\log(b) - \sum_{i=1}^{d} y_{i:n} + (a-1)\sum_{i=1}^{d}\log(1-e^{-y_{i:n}})$$
$$+(b-1)\sum_{i=1}^{d}\log(1-(1-e^{-y_{i:n}})^{a}) + b(n-d)\log(1-(1-e^{-c})^{a}) \quad (3.9)$$

olarak elde edilir. (3.5) eşitliğinin a ve b parametrelerine göre türevleri aşağıdaki gibi elde edilir.

$$\frac{\partial \log L}{\partial a} = \frac{d}{a} + \sum_{i=1}^{d}\log(1-e^{-y_{i:n}}) + (b-1)\sum_{i=1}^{d}\frac{(1-e^{-y})^{a}\log(1-e^{-y})}{1-(1-e^{-y})^{a}}$$
$$-b(n-d)\frac{(1-e^{-c})\log(1-e^{-c})}{1-(1-e^{-c})} = 0 \quad (3.10)$$

$$\frac{\partial \log L}{\partial b} = \frac{d}{b} + \sum_{i=1}^{d}\log(1-(1-e^{-y_{i:n}})^{a}) + (n-d)\log(1-(1-e^{-y_{i:n}})^{a}) = 0 \quad (3.11)$$

a ve b parametrelerinin tahminleri kapalı formda elde edilemediği için bu tahminleri hesaplamak için bazı nümerik tekniklere ihtiyaç vardır. En çok olabilirlik tahminlerini bulmak için beklenti maksimizasyonu (EM) algoritması kullanılır.

$\mathbf{Y} = (Y_{1:n}, Y_{2:n}, ..., Y_{d:n})$ ve $\mathbf{Z} = (Z_1, Z_2, ..., Z_{n-d})$ sırasıyla gözlenen ve gözlenemeyen verileri göstersin. Burada Z kayıp gözlem verisi ve $\mathbf{W} = (\mathbf{Y}, \mathbf{Z})$ tam veri setini göstermektedir (Dempster ve ark., 1977). $\mathbf{W}$ için olabilirlik fonksiyonu aşağıdaki gibi ifade edilir.

$$L_c(a,b|\mathbf{W}) = \prod_{i=1}^{d} f\left(y_{i:n}\right) \prod_{i=1}^{n-d} f\left(z_i\right)$$
$$= \prod_{i=1}^{d} abe^{-y_{i:n}} \left(1-e^{-y_{i:n}}\right)^{(a-1)} \left(1-(1-e^{-y_{i:n}})^a\right)^{(b-1)} \times \tag{3.12}$$
$$\prod_{i=1}^{n-d} abe^{-z_i} \left(1-e^{-z_i}\right)^{(a-1)} \left(1-(1-e^{-z_i})^a\right)^{(b-1)}$$

Tam örneklem için log-olabilirlik fonksiyonu ise;

$$\log L_c(a,b|\mathbf{W}) = n\log a + n\log b - \left(\sum_{i=1}^{d} y_{i:n} + \sum_{i=1}^{n-d} z_i\right)$$
$$+(a-1)\left(\sum_{i=1}^{d} \log(1-e^{-y_{i:n}}) + \sum_{i=1}^{n-d} \log(1-e^{-z_i})\right) \tag{3.13}$$
$$+(b-1)\left(\sum_{i=1}^{d} \log\left(1-(1-e^{-y_{i:n}})^a\right) + \sum_{i=1}^{n-d} \log\left(1-(1-e^{-z_i})^a\right)\right)$$

biçiminde elde edilir. Beklenti Maksimizasyonunun (EM) E-adımı, $L_s(a,b)$ ile gösterilen pseudo Log olabilirlik fonksiyonuna eşit olan $E[LogL_c(\mathbf{W};a,b)/\mathbf{Y}=\mathbf{y}]$ koşullu beklenen değerinin hesaplanmasını gerektirir. Pseudo Log-olabilirlik fonksiyonu aşağıdaki gibi yazılır.

$$L_s(a,b) = n\log a + n\log b - \sum_{i=1}^{d} y_{i:n} + (a-1)\sum_{i=1}^{d} \log(1-e^{y_{i:n}})$$
$$+(b-1)\sum_{i=1}^{d} \log\left(1-(1-e^{-y_{i:n}})^a\right) - \sum_{i=1}^{n-d} E\left(Z_i \mid z_i > c\right)$$
$$+(a-1)\sum_{i=1}^{n-d} E\left(\log(1-e^{-Z_i}) \mid z_i > c\right) \tag{3.14}$$
$$+(b-1)\sum_{i=1}^{n-d} E\left(\log\left(1-(1-e^{-Z_i})^a\right) \mid z_i > c\right)$$

(3.14) eşitliğindeki koşullu beklenen değerleri bulmak için aşağıdaki teoremden yararlanılır.

Teorem 3.2. Verilen $Y_1 = y_1, ..., Y_d = y_d < c$, Z_i'nin koşullu olasılık yoğunluk fonksiyonu;

$$f_{\mathbf{Z}|\mathbf{Y}}(z_i \mid Y_1 = y_1, ..., Y_d = y_d < c) = \frac{f_{LKw}(z_i;a,b)}{1-F_{LKw}(c;a,b)}; \qquad z_i > c \tag{3.15}$$

olarak elde edilir (Ng ve ark., 2002; Kundu ve Pradhan, 2009). Bu teorem kullanılarak koşullu beklenen değerler aşağıdaki şekilde hesaplanabilir.

$$E\left(Z_i \mid Z_i > c\right) = \frac{ab}{1-F_{LKw}(c;a,b)} \int_c^{\infty} y_{i:n} e^{-y_{i:n}} (1-e^{-y_{i:n}})^{a-1} (1-(1-e^{-y_{i:n}})^a)^{b-1} dy_{i:n} \\ = A(c,a,b) \tag{3.16}$$

$$E\left(\log(1-e^{-z_i}) \mid z_i > c\right) = \frac{ab}{1-F_y(c;a,b)} \\ \times \int_c^{\infty} \log(1-e^{-y_{i:n}}) e^{-y} (1-e^{-y_{i:n}})^{(a-1)} \left(1-(1-e^{-y_{i:n}})^a\right)^{(b-1)} dy_{i:n} \\ = B(c,a,b) \tag{3.17}$$

$$E\left(\log(1-(1-e^{-z_i})^a) \mid z_i > c\right) = \frac{ab}{1-F_{LKw}(c;a,b)} \\ \times \int_c^{\infty} \log(1-(1-e^{-y_{i:n}})^a) e^{-y_{i:n}} (1-e^{-y_{i:n}})^{(a-1)} \\ \times \left(1-(1-e^{-y_{i:n}})^a\right)^{b-1} dy \\ = C(c,a,b) \tag{3.18}$$

(3.16), (3.17) ve (3.18) eşitlikleriyle birlikte (3.14) eşitliğindeki pseudo Log-olabilirlik fonksiyonu

$$L_s(a,b) = n\log a + n\log b - \sum_{i=1}^{d} y_{i:n} + (a-1)\sum_{i=1}^{d} \log(1-e^{-y_{i:n}}) \\ +(b-1)\sum_{i=1}^{d} \log\left(1-(1-e^{-y_{i:n}})^a\right) - (n-d)A(c,a^{(k)},b^{(k)}) \\ +(a-1)(n-d)B(c,a^{(k)},b^{(k)}) + (b-1)(n-d)C(c,a^{(k)},b^{(k)}) \tag{3.19}$$

biçiminde yazılabilir. M adımında düzenlenen Eşitlik (3.19)' te yer alan pseudo log-olabilirlik fonksiyonunun aşağıdaki gibi a ve b'ye göre türevleri alınarak maksimize edilir.

$$\frac{\partial L_s(a,b)}{\partial a} = \frac{n}{a} + \sum_{i=1}^{d} \log(1-e^{-y_{i:n}}) - (b-1)\sum_{i=1}^{d} \frac{(1-e^{-y_{i:n}})^a \log(1-e^{-y_{i:n}})}{1-(1-e^{-y_{i:n}})^a} = 0 \\ (n-d)B\left(c,a^{(k)},b^{(k)}\right) \tag{3.20}$$

$$\frac{\partial L_s(a,b)}{\partial b} = \frac{n}{b} + \sum_{i=1}^{d} \log(1-(1-e^{-y_{i:n}})^a) + (n-d)C\left(c,a^{(k)},b^{(k)}\right) \tag{3.21}$$

Yukarıdaki lineer olmayan denklem sistemleri Kundu ve Pradhan (2009) tarafından kullanılan sabit nokta yöntemiyle çözülür. EM algoritması uygulanırken öncelikle başlangıç değerleri a^0 ve b^0 seçilir. $k.$ adımda a ve b parametrelerinin tahmini

$\left(a^k, b^k\right)$ olmak üzere eşitlik (3.18)' in maksimize edilmesiyle $\left(a^{k+1}, b^{k+1}\right)$ iterasyonu elde edilir.

3.3. Log-Kumaraswamy Dağılımı İçin Aralık Tahmini

Bu bölümde Louis (1982) 'in kayıp değer prensibi kullanılarak Fisher bilgi matrisi elde edilmiştir. Gözlenen Fisher bilgi matrisi asimptotik güven aralıkları oluşturmak için kullanılabilir. Kayıp değer prensibi aşağıda gibi açıklanmıştır.

Gözlenen Bilgi=Tam Bilgi-Kayıp Bilgi (3.22)

$\boldsymbol{\theta} = (a, b)$, $\mathbf{Y}$ gözlenen vektör , $\mathbf{W}$ tam veri, $I_{\mathbf{Y}}(\theta)$ gözlenen bilgi ve $I_{W/Y}(\boldsymbol{\theta})$ ise kayıp bilgi (Sansürlü gözlemlerin Fisher Bilgisi) olmak üzere , (3.22) aşağıdaki gibi elde edilir.

$$I_{\mathbf{Y}}(\theta) = I_{\mathbf{W}}(\boldsymbol{\theta}) - I_{\mathbf{W/Y}}(\boldsymbol{\theta}) \tag{3.23}$$

$$I_{\mathbf{W}}(\theta) = -E\left[\frac{\partial^2 (L_c(\mathbf{W}|\boldsymbol{\theta}))}{\partial \boldsymbol{\theta}^2}\right] \tag{3.24}$$

$$I_{\mathbf{W/Y}}(\theta) = -(n-d)E\left[\frac{\partial^2 \log f_z(z/Y, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}^2}\right] \tag{3.25}$$

(Louis, 1982; Dube ve ark., 2011). $\hat{\boldsymbol{\theta}}$ 'nın asimptotik varyans-kovaryans matrisi, $I_{\mathbf{Y}}\left(\hat{\boldsymbol{\theta}}\right)$' nin tersi alınarak elde edilir. Bu durumda $I_{\mathbf{W}}(\boldsymbol{\theta})$ matrisinin elemanları;

$$\alpha_{11} = \frac{n}{a^2} + n(b-1)ab\int_0^{\infty} \ln(1-e^{-y})^2 (1-e^{-y})^{2a-1} (1-(1-e^{-y})^a)^{b-3} e^{-y} dy \tag{3.26}$$

$$\alpha_{12} = \alpha_{21} = nab\int_0^{\infty} (1-e^{-y})^{2a-1} \log(1-e^{-y}) (1-(1-e^{-y})^a)^{b-a-1} e^{-y} dy \tag{3.27}$$

$$a_{22} = \frac{n}{b^2} \tag{3.28}$$

şeklinde hesaplanır. $I_{\mathbf{W|Y}}(\boldsymbol{\theta})$ matrisinin elemanları;

$$I_{\mathbf{W|Y}}(\theta) = (n-d)\begin{bmatrix} \beta_{11}(c;a,b) & \beta_{12}(c;a,b) \\ \beta_{21}(c;a,b) & \beta_{22}(c;a,b) \end{bmatrix} \tag{3.29}$$

$$\beta_{11} = \frac{1}{a^2} + h_1(c;a,b) + \frac{b(1-e^{-c})^a (\log(1-e^{-c}))^2}{(1-(1-e^{-c})^a)^2} \tag{3.30}$$

$$\beta_{22} = \frac{1}{b^2} \tag{3.31}$$

$$\beta_{12} = \beta_{21} = h_2\left(c;a,b\right) + \frac{(1-e^{-c})^a \log(1-e^{-c})}{1-(1-e^{-c})^a} \tag{3.32}$$

olarak hesaplanır. Burada $h_1(c;a,b)$ ve $h_2(c;a,b)$ aşağıdaki gibi hesaplanır.

$$h_1(c;a,b) = \frac{ab}{\left(1-(1-e^{-c})\right)^b} \int_{1-e^{-c}}^{1} u^{2a-1}\left(\log(u)\right)^2\left(1-u^a\right)^{b-a-2} du \tag{3.33}$$

$$h_2(c;a,b) = \int_{1-e^{-c}}^{\infty} u^{2a-1}\log(u)\left(1-u^a\right)^{b-2} du \tag{3.34}$$

Bilinmeyen parametreler için güven aralıkları her zaman kapalı formda yazılamadığı için asimptotik güven aralıkları elde edilir. Asimptotik güven aralıklarını elde etmek için büyük örneklem yaklaşımı kullanılır. Kısaca $\sqrt{n}\left(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}\right)$ ' nın asimptotik dağılımı aşağıdaki gibidir.

$$\sqrt{n}\left(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}\right) \xrightarrow{d} N\left(0, I_Y^{-1}(\boldsymbol{\theta})\right) \tag{3.35}$$

Burada $I_Y^{-1}(\boldsymbol{\theta})$ matrisi $I_Y(\boldsymbol{\theta})$ matrisinin tersidir. Eşitlik (3.35) kullanılarak a ve b parametreleri için asimptotik güven aralıkları oluşturulabilir. Bu yaklaşımla, a ve b parametreleri için $1-\eta$ seviyeli güven aralıkları eşitlik (3.36) ve (3.37) ' da ifade edilmiştir.

$$P\left(\hat{a} - z_{\eta/2}\sqrt{Var(\hat{a})} < a < \hat{a} + z_{\eta/2}\sqrt{Var(\hat{a})}\right) = 1-\eta \tag{3.35}$$

$$P\left(\hat{b} - z_{\eta/2}\sqrt{Var(\hat{b})} < b < \hat{b} + z_{\eta/2}\sqrt{Var(\hat{b})}\right) = 1-\eta \tag{3.36}$$

3.4. Simülasyon Çalışması

Bu bölümde, Tip I hibrit sansürlü örneklem altında LKw dağılımının bilinmeyen a ve b parametrelerinin EM algoritması kullanılarak elde edilen en çok olabilirlik tahmin edicilerinin performanslarına ve bu tahmin edicilere dayalı asimptotik güven aralıklarının elde edilmesine ilişkin 10000 tekrara dayalı bir Monte Karlo simülasyon çalışması yapılmıştır. Bu simülasyon çalışmasında, a ve b parametrelerinin MLE 'lerinin yan ve hata kareler ortalamaları (HKO) ile MLE 'lerine dayalı asimptotik güven aralıklarının genişliği ve kapsama olasılıkları hesaplanmıştır. Burada $a = 0.5$ ve $b = 0.8$ olarak alınmıştır. Simülasyon sonuçları farklı n, T, r değerleri için Çizelge (3.1) ve Çizelge (3.2) 'de verilmiştir.

Çizelge 3.1: $T=10$ ve farklı n ve r değerleri için LKw dağılımı için nokta ve aralık tahmin sonuçları

(n,r)	Parametreler	MLE	Yan	HKO	Aralık Uzunluğu	Kapsama Olasılığı
(30,15)	a	0.5913	0.0913	0.0241	0.7216	0.9994
	b	0.9778	0.1778	0.0667	1.6726	0.9998
(30,20)	a	0.5868	0.0865	0.0241	0.6446	0.9960
	b	0.9458	0.1458	0.0650	1.2502	0.9850
(30,25)	a	0.5722	0.0722	0.0238	0.5822	0.9832
	b	0.9117	0.1117	0.0619	0.9872	0.9937
(50,25)	a	0.5802	0.0806	0.0154	0.5974	0.9997
	b	0.9498	0.1498	0.0428	1.2365	0.9999
(50,30)	a	0.5773	0.0713	0.0150	0.5103	0.9955
	b	0.9290	0.1290	0.0399	1.0243	0.9996
(50,40)	a	0.5600	0.0600	0.0136	0.4512	0.9822
	b	0.8875	0.0875	0.0331	0.7612	0.9953
(50,45)	a	0.5456	0.0461	0.0129	0.4261	0.9710
	b	0.8659	0.0659	0.0297	0.6703	0.9852
(100,50)	a	0.5725	0.0725	0.0094	0.3820	0.9920
	b	0.9267	0.1267	0.0261	0.8412	0.9998
(100,70)	a	0.5589	0.0589	0.0079	0.3326	0.9790
	b	0.8809	0.0809	0.0181	0.5933	0.9983
(100,90)	a	0.5306	0.0306	0.0058	0.2947	0.9685
	b	0.8390	0.0390	0.0133	0.4550	0.9747
(200,100)	a	0.5673	0.0673	0.0066	0.2675	0.9656
	b	0.9156	0.1156	0.0183	0.5835	0.9999
(200,150)	a	0.5477	0.0477	0.0044	0.2261	0.9512
	b	0.8607	0.0607	0.0092	0.3829	0.9875
(200,160)	a	0.5309	0.0409	0.0039	0.2194	0.9548
	b	0.8505	0.0503	0.0080	0.3591	0.9829

Çizelge (3.1)'e bakıldığı zaman örneklem hacmi arttıkça yan ve hata kareler ortalamasının (HKO) azaldığı görülmektedir. Ayrıca örneklem hacmi artıkça kapsama olasılıklarının 0.95' e yakınsadığı ve aralık uzunluklarının daraldığı görülmektedir

Çizelge 3.2 : T=2 ve farklı n ve r değerleri için LKw dağılımı için nokta ve aralık tahmin sonuçları

(n,r)	Parametreler	MLE	Yan	HKO	Aralık Uzunluğu	Kapsama Olasılığı
(30,15)	a	0.5921	0.0921	0.0243	0.7224	0.9992
	b	0.9756	0.1756	0.0658	1.6668	0.9998
(30,20)	a	0.5868	0.0869	0.0242	0.6447	0.9963
	b	0.9458	0.1459	0.0649	1.2505	0.9888
(30,25)	a	0.5812	0.0812	0.0243	0.5934	0.9679
	b	0.9123	0.1123	0.0648	0.9906	0.9813
(50,25)	a	0.5805	0.0805	0.0153	0.5487	0.9983
	b	0.9497	0.1497	0.0428	1.2360	0.9999
(50,30)	a	0.5776	0.0776	0.0152	0.5106	0.9949
	b	0.9267	0.1267	0.0388	1.0246	0.9997
(50,40)	a	0.5617	0.0617	0.0146	0.4512	0.9822
	b	0.8843	0.0843	0.0321	0.7612	0.9953
(50,45)	a	0.5456	0.0461	0.0129	0.4526	0.9770
	b	0.8659	0.0659	0.0297	0.7579	0.9912
(100,50)	a	0.5711	0.0711	0.0090	0.3810	0.9927
	b	0.9252	0.1252	0.0257	0.8392	0.9998

(100,70)	*a*	0.5775	0.0578	0.0077	0.3320	0.9814
	b	0.8811	0.0811	0.0177	0.5935	0.9983
(100,90)	*a*	0.5552	0.0552	0.0082	0.3058	0.8937
	b	0.8525	0.0525	0.0176	0.4705	0.9350
(200,100)	*a*	0.5677	0.0677	0.0066	0.2676	0.9647
	b	0.9136	0.1136	0.0180	0.5814	0.9998
(200,150)	*a*	0.5575	0.0575	0.0054	0.2403	0.9523
	b	0.8800	0.0800	0.0118	0.4445	0.9958
(200,160)	*a*	0.5405	0.0405	0.0039	0.2192	0.9558
	b	0.8453	0.0483	0.0080	0.3579	0.9820

Çizelge (3.2)'ye bakıldığı zaman örneklem hacmi arttıkça yan ve HKO'nın azaldığı görülmektedir. Ayrıca örneklem hacmi arttıkça kapsama olasılıklarının 0.95' e yakınsadığı ve aralık uzunluklarının daraldığı görülmektedir.

3.5. Gerçek Veri Analizi

Bu bölümde iki adet gerçek veri analiz çalışması yapılmıştır. Birinci veri grubu, Minneapolis / St Paul'da Mart ayı yağışının 30 gözleminden oluşmaktadır. Birinci veri grubu Usman ve ark.(2017) tarafından Kumaraswamy Half-Logistic ve Ul Haq (2016) tarafından Dönüştürülmüş Üstelleştirilmiş Ters Rayleigh dağılımının modellenmesinde kullanılmıştır (ul Haq, 2016; Usman ve ark., 2017). İkinci veri grubunda, 50 gözlemden oluşan 50 mm uzunluğunda karbon fiberlerin çekme gerilimine dayanması göz önünde bulundurulmuştur. İkinci veri grubu Nichols ve Padgett (2006) tarafından Weibull dağılımının modellenmesinde kullanılmıştır (Nichols ve Padgett, 2006). Veri grupları aşağıda sırasıyla verilmiştir.

Veri grubu 1

0,77 1,74 0,81 1,20 1,95 1,20 0,47 1,43 3,37 2,20 3,00 3,09 1,51 2,10 0,52 1,62 1,31 0,32 0,59 0,81 2,81 1,87 1,18 1,35 4,75 2,48 0,96 1,89 0,90 2,05

Veri Grubu 2

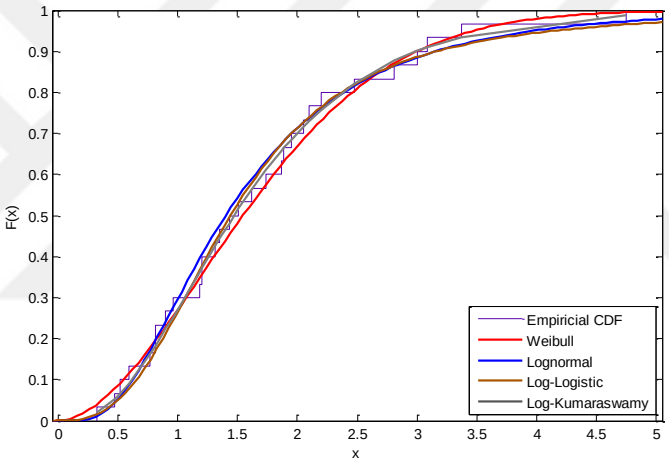
1,41 3,68 2,97 1,36 0,98 2,76 4,91 3,68 1,84 1,59 3,19 1,57 0,81 5,56 1,73 1,59 2 1,22 1,12 1,71 2,17 1,17 5,08 2,48 1,18 3,51 2,17 1,69 1,25 4,38 1,84 0,39 3,68 2,48 0,85 1,61 2,79 4,70 2,03 1,80 1,57 1,08 2,03 1,61 2,12 1,89 2,88 2,82 2,05 3,65

Üzerinde çalışılan veri gruplarının Log-Kumaraswamy dağılımından gelip gelmediğini test etmek için uyum iyiliği testleri uygulanmıştır.

Birinci veri grubu W, LL, LN ve LKw dağılımlarını karşılaştırmak için kullanılan AIC, BIC, -2logL, Kolmogorov-Smirnov (KS) uyum iyiliği test sonuçları aşağıdaki Çizelge (3.3)'te verilmiştir. Ayrıca veri grubu I için bütün dağılımların, dağılım fonksiyonu grafikleri şekil (3.1) verilmiştir.

Çizelge 3.3 Yağış verisi İçin Uyum İyiliği Sonuçları

Model	$\hat{a}$	$\hat{b}$	AIC	BIC	-2logL	K.S	p değeri
LKw(a,b)	**3.2148**	**1.2232**	**80.1981**	**82.9914**	**76.1891**	**0.0616**	**0.9998**
LL(a,b)	1.4407	2.7879	80.9509	84.6639	77.8615	0.0741	0.9965
W(a,b)	1.8923	1.8088	81.2865	84.0889	77.2865	0.0689	0.9988
LN(a,b)	0.3373	0.6226	80.9509	83.7533	76.9509	0.0912	0.9640.



Şekil 3.1. Yağış verisi için uydurulan dağılım fonksiyonu grafikleri

Çizelge (3.3) tüm kriterler bakımından incelendiğinde veri gurubunu en iyi modelleyen dağılımın LKw olduğu görülmektedir. Gerçek veri gurubunun LKw dağılımına uygunluğu sağlandıktan sonra, rasgele olarak seçilen farklı r ve T değerleri için en çok olabilirlik tahminleri (MLE), standart hatalar ve asimptotik güven aralıkları Çizelge (3.4)' te yer almaktadır. Bu tabloda standart hatalar parantez içinde belirtilmiştir.

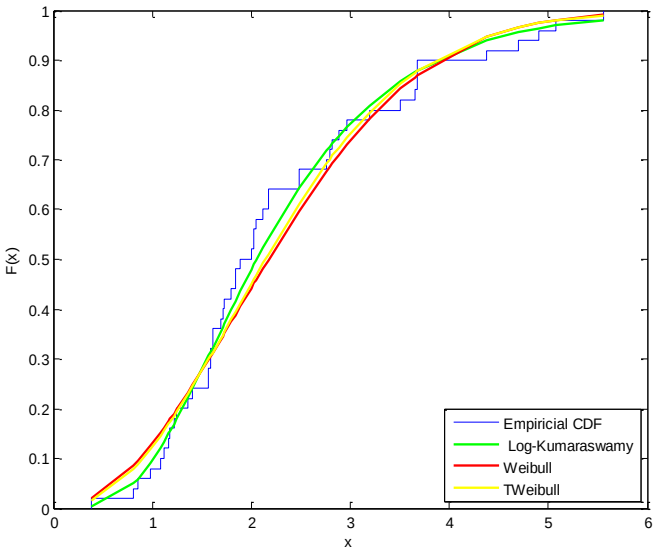
Çizelge 3.4: Farklı r ve T değerleri için ML tahminleri ve asimptotik güven aralıkları

		T=1.5		T=2.0		T=2.5	
r	Parametre	MLE	Asimptotik Güven Arağı	MLE	Asimptotik Güven Arağı	MLE	Asimptotik Güven Arağı
15	a	2.8851 (0.8906)	(1.1394,4.6307)	2.8851 (0.8906)	(1.1394,4.6307)	2.8851 (0.8906)	(1.1394,4.6307)
	b	1.1158 (0.5051)	(0.1258 2.1058)	1.1158 (0.5051)	(0.1258 2.1058)	1.1158 (0.5051)	(0.1258 2.1058)
20	a	2.9582 (0.9042)	(1.1852,4.7310)	2.9509 (0.8059)	(1.3714,4.5304)	2.9509 (0.8059))	(1.3714,4.5304)
	b	1.0844 (0.4802)	(0.1425,2.0264)	1.1152 (0.3902)	(0.3504,1.8801)	1.1152 (0.3902)	(0.3504,1.8801)
25	a	2.9582 (0.9042)	(1.1852,4.7310)	3.2141 (0.8384)	(1.5709,4.8573)	(3.0787) (0.7600)	(1.5891,4.5684)
	b	1.0844 (0.4802)	(0.1425,2.0264)	1.2117 (0.4098)	(0.4086,2.0148)	1.1946 (0.3492)	(0.5101,1.8790)

İkinci veri grubu LKw,W ve TW dağılımlarını karşılaştırmak için kullanılan AIC, BIC,- 2logL, Kolmogorov-Smirnov uyum iyiliği test sonuçları aşağıdaki Çizelge (3.5)'te verilmiştir. Ayrıca veri grubu II için bütün dağılımların, uydurulan dağılım fonksiyonu grafikleri şekil (3.2) 'de verilmiştir.

Çizelge 3.5: Karbon Fiber verisi İçin Uyum İyiliği Sonuçları

Model	$\hat{a}$	$\hat{b}$	$\hat{\lambda}$	AIC	BIC	-2logL	K-S	p value
LKw(a,b)	**5.1047**	**1.0075**	-	**151.0413**	**154.8654**	**147.0413**	**0.0987**	**0.7139**
W(a,b)	2.0559	2.0559	-	154.7975	158.6215	150.7975	0.1417	0.2676
TW(a,b,λ)	2.2145	2.9868	0.5062	155.6293	161.3653	149.6293	0.1305	0.3616



Şekil 3.2. Karbon fiber verisi için uydurulan dağılım fonksiyonu grafikleri

Çizelge (3.5) tüm kriterler bakımından incelendiğinde veri gurubunu en iyi modelleyen dağılımın LKw olduğu görülmektedir. Gerçek veri grubunun LKw dağılımına uygunluğu sağlandıktan sonra, rasgele olarak seçilen farklı r ve T değerleri için en çok olabilirlik tahminleri (MLE), standart hatalar ve asimptotik güven aralıkları Çizelge (3.6)' da yer almaktadır.

Çizelge 3.6: Farklı r ve T değerleri için ML tahminleri ve asimptotik güven aralıkları

		T=2.0		T=3.0		T=4.0	
R	Parametre	MLE	Asimptotik Güven Arağı	MLE	Asimptotik Güven Arağı	MLE	Asimptotik Güven Arağı
25	a	5.1068 (1.2444)	(2.6677,7.5458)	5.1068 (1.2444)	(2.6677,7.5458)	5.1068 (1.2444)	(2.6677,7.5458)
	b	1.1555 (0.4189)	(0.3444,1.9765)	1.1555 (0.4189)	(0.3444,1.9765)	1.1555 (0.4189)	(0.3444,1.9765)
35	a	5.3413 (1.2789)	(2.8346,7.8479)	5.0890 (1.0713)	(2.9893,7.1887)	5.0890 (1.0713)	(2.9893,7.1887)
	b	1.1180 (0.3930)	(0.3478,1.8882)	1.0091 (0.2546)	(0.5102,1.5081)	1.1152 (0.2546)	(0.5102,1.5081)
45	a	5.3413 (1.2789)	(2.8346,7.8479)	5.7809 (1.1198)	(3.5861,7.9756)	5.1923 (0.9867)	(3.2584,7.1262)
	b	1.1180 (0.3930)	(0.3478,1.8882)	1.1638 (0.2726)	(0.6296,1.6980)	1.1946 (0.2136)	(0.6270,1.4645)

4. KUMARASWAMY-TERS RAYLEIGH DAĞILIMININ İLERLEYEN TÜR TİP I HİBRİT SANSÜRLEME ALTINDA NOKTA VE ARALIK TAHMİNİ

4.1.Kumaraswamy-Ters Rayleigh Dağılımı

Kumaraswamy dağılımı ilk olarak 1980 yılında Kumaraswamy tarafından ortaya atılmıştır ve yaşam verilerini modellemede kullanılmıştır. Cordeiro ve Castro (2011) eşitlik (3.1) ve (3.2) 'i kullanarak olasılık dağılımlarının bir genellemesini önerdi. Corderio ve Castro tarından genelleştirilen sınıfın olasılık yoğunluk ve dağılım fonksiyonu aşağıdaki şekilde verilmiştir (Cordeiro ve de Castro, 2011).

$$f(x) = abg(x)\left(G(x)\right)^{a-1}\left[1-\left(G(x)\right)^{\alpha}\right]^{b-1} \tag{4.1}$$

$$F(x) = 1-\left[1-\left(G(x)\right)^{\alpha}\right]^{b} \tag{4.2}$$

(4.1) ve (4.2) eşitlikleri kullanılarak Kumaraswamy dağılımına dayanan diğer dağılımlar ortaya atılmıştır. Shahbaz ve ark. (2012) tarafından önerilen Kumaraswamy-Ters Weibull (Kw-TW) dağılımı bu yöntem ile elde edilen dağılımlardan biridir. Kw-

TW dağılımının olasılık yoğunluk fonksiyonu ve dağılım fonksiyonu sırasıyla aşağıda verilmiştir.

$$f_{Kw-IW}(x)=\frac{ab\alpha\beta}{x^{\beta+1}}e^{-\frac{a\alpha}{x^{\beta}}}\left(1-e^{-\frac{a\alpha}{x^{\beta}}}\right)^{b-1};(x,a,b,\alpha,\beta)>0 \tag{4.3}$$

$$F_{Kw-IW}(x)=1-\left(1-e^{-\frac{a\alpha}{x^{\beta}}}\right)^{b} \tag{4.4}$$

Kw-TW dağılımının dağılım fonksiyonunda $\alpha=1$ ve $\beta=2$ alınırsa Kumaraswamy-Ters Rayleigh (Kw-TR) dağılımı elde edilmiş olur. Bu dağılıma ilişkin olasılık yoğunluk fonksiyonu ve dağılım fonksiyonu

$$f_{Kw-IR}(x)=2abx^{-3}e^{-ax^{-2}}\left(1-e^{-ax^{-2}}\right)^{b-1};(x,a,b)>0 \tag{4.5}$$

$$F_{Kw-IR}(x)=1-\left(1-e^{-ax^{-2}}\right)^{b} \tag{4.6}$$

şeklindedir. Burada a şekil ve b ölçek parametresidir.

4.2. Kumaraswamy Ters Rayleigh Dağılımı İçin Nokta Tahmini

$X_1,X_2,...,X_n$ Kw-TW dağılımından alınmış n birimlik örneklem olsun. İlerleyen tür tip I hibrit sansürleme durumunda yaşam testi $T_1=\min(X_{m:m:n},T)$ zamanında durdurulur. Bu durumda gözlenen örneklemler aşağıdaki iki durumdan biri olabilir.

DurumI: $\mathrm{X}_{1:m:n},X_{2:m:n},...,X_{m:m:n}$ eğer $\mathrm{X}_{m:m:n}<T$

Durum II: $\mathrm{X}_{1:m:n},X_{2:m:n},...,X_{j:m:n}$ eğer $\mathrm{X}_{j:m:n}<T<X_{j+1:m:n}$

Bu örneklem durumları altında olabilirlik fonksiyonları,

$$\begin{aligned} I&: L(a,b|\mathbf{x})=C\prod_{i=1}^{m}f\left(x_{i:m:n}\right)\left(1-F\left(x_{i:m:n}\right)\right)^{r_i} \\ II&: L(a,b|\mathbf{x})=D\prod_{i=1}^{j}f\left(x_{i:m:n}\right)\left(1-F\left(x_{i:m:n}\right)\right)^{r_i}\left(1-F\left(T\right)\right)^{r_j^*} \end{aligned} \tag{4.7}$$

şeklindedir. Burada $\mathbf{x}=\left(x_{1:n},x_{2:n},...,x_{r:n}\right)$, $C=n(n-r_1-1)...(n-r_1-r_2-...r_{m-1}-m+1)$ ve $D=n(n-r_1-1)...(n-r_1-r_2-...r_{m-1}-j+1)$ biçimindedir. Eşitlik (4.7) kullanılarak $Kw-TR(a,b)$ dağılımı için olabilirlik fonksiyonu aşağıdaki gibi ifade edilebilir.

$$L(a,b\,|\,\mathbf{x}) = 2^d a^d b^d \prod_{i=1}^{d} x^{-3} e^{-a x_{i:m:n}^{-2}} e^{W(a,b)} \tag{4.8}$$

Burada d ve $W(a,b)$

$$d = \begin{cases} m, \text{ durum I için} \\ j \text{ , durum II için} \end{cases} \tag{4.9}$$

$$W(a,b) = \begin{cases} \sum_{i=1}^{m} \left(b(r_i+1)-1\right) \log\left(1-e^{-a x_{i:m:n}^{-2}}\right) & \text{, durum I için} \\ \sum_{i=1}^{m} \left(b(r_i+1)-1\right) \log\left(1-e^{-a x_{i:m:n}^{-2}}\right) + b r_j^* \log\left(1-e^{-aT^{-2}}\right) & \text{, durum II için} \end{cases} \tag{4.10}$$

şeklindedir. Log-olabilirlik fonksiyonu ise aşağıdaki gibi yazılır.

$$\begin{aligned} \log L(a,b\,|\,x) &= d\log(2) + d\log(a) + d\log(b) - 3\sum_{i=1}^{d} \log(x_{i:m:n}) \\ &\quad - a\sum_{i=1}^{d} x_{i:m:n}^{-2} + W(a,b) \end{aligned} \tag{4.11}$$

Eşitlik (4.11) 'in bilinmeyen a ve b parametrelerine göre türevi alınıp sıfıra eşitlenmesi durumunda aşağıdaki olabilirlik eşitlikleri elde edilir.

$$\frac{\partial \log L(a,b\,|\,x)}{\partial a} = \frac{d}{a} - \sum_{i=1}^{d} x_{i:m:n}^{-2} + W(a) = 0 \tag{4.12}$$

$$\frac{\partial \log L(a,b\,|\,x)}{\partial b} = \frac{d}{b} + W(b) = 0 \tag{4.13}$$

Burada;

$$W(a) = \begin{cases} \sum_{i=1}^{m} \left(b(r_i+1)-1\right) \dfrac{x_i^{-2} e^{-a x_{i:m:n}^{-2}}}{1-e^{-a x_{i:m:n}^{-2}}} & \text{, durum I} \\ \sum_{i=1}^{j} \left(b(r_i+1)-1\right) \dfrac{x_i^{-2} e^{-a x_{i:m:n}^{-2}}}{1-e^{-a x_{i:m:n}^{-2}}} + b r_j^* \dfrac{T^{-2} e^{-aT^{-2}}}{1-e^{-aT^{-2}}} & \text{, durum II} \end{cases} \tag{4.14}$$

$$W(b) = \begin{cases} \sum_{i=1}^{m} (r_i+1) \log\left(1-e^{-a x_{i:m:n}^{-2}}\right) & \text{, durum I} \\ \sum_{i=1}^{j} (r_i+1) \log\left(1-e^{-a x_{i:m:n}^{-2}}\right) + r_j^* \left(1-e^{-aT^{-2}}\right) & \text{, durum II} \end{cases} \tag{4.15}$$

şeklindedir. a ve b 'nin MLE 'leri açık bir şekilde elde edilemediği için beklenti maksimizasyonu (EM) algoritması kullanılır. Durum I için gözlenen örneklemler $X = (X_1, X_2, ..., X_m)$, gözlenemeyen örneklemler $Z = (Z_1, Z_2, ..., Z_m)$ ve $Z_g = \left(Z_{g1}, Z_{g2}, ..., Z_{gr_g}\right)_{1\times r_g}$ $g = (1,2,...,m)$ olsun. Bu durumda tam veri seti için

$W=\left(X,Z\right)$ şeklinde olacaktır. Durum II için gözlenen örneklemler $X=\left(X_1,X_2,...,X_j\right)$, gözlenemeyen örneklemler $Z=\left(Z_1,Z_2,...,Z_j\right)$ olmak üzere burada $Z_s=\left(Z_{s1},Z_{s2},...,Z_{sr_s}\right)_{1\times r_s}$ $s=\left(1,2,...,j\right)$ ve $Z^{'}=\left(Z_1^{'},Z_2^{'},...,Z_{r_j^*}^{'}\right)$ olsun. Bu durumda tam veri seti $W=\left(X,Z,Z^{'}\right)$ olacaktır. Bu iki durum altında olabilirlik fonksiyonları;

$$L(a,b\mid x)=\begin{cases}P & ,\text{durum I için}\\ P+Q,&\text{durum II için}\end{cases} \tag{4.16}$$

biçimindedir. Burada

$$\begin{aligned}P&=\prod_{i=1}^{d}f\left(x_{i:m:n}\right)\prod_{i=1}^{d}\prod_{k=1}^{r_i}f\left(z_{ik}\right)\\&=\prod_{i=}^{d}\left(2abx_{i:m:n}^{-3}e^{-ax_{i:m:n}^{-2}}\left(1-e^{-ax_{i:m:n}^{-2}}\right)^{b-1}\right)\times\prod_{i=1}^{d}\prod_{k=1}^{r_i}f\left(2abz_{ik}^{-3}e^{-az_{ik}^{-2}}\left(1-e^{-az_{ik}^{-2}}\right)^{b-1}\right)\end{aligned} \tag{4.17}$$

ve

$$\begin{aligned}Q&=\prod_{p=1}^{r_j^*}f\left(z_p^{'}\right)\\&=r_j^*\left(\log(2)+\log(a)+\log(b)\right)+\sum_{p=1}^{r_j^*}z_p^{'-3}+\sum_{i=1}^{r_j^*}e^{-az_p^{'-2}}\\&\quad+(b-1)\sum_{p=1}^{r_j^*}\log\left(1-e^{-az_p^{'-2}}\right)\end{aligned} \tag{4.18}$$

olarak yazılır. P ve Q' ya ilişkin log-olabilirlik fonksiyonları aşağıdaki şekilde yazılır.

$$\log L(a,b\mid x)=\begin{cases}P^* & \text{durum I için}\\ P^*+Q^* & \text{durum II için}\end{cases} \tag{4.19}$$

Burada

$$\begin{aligned}P^*&=n\log\left(2\right)+n\log\left(a\right)+n\log\left(b\right)-3\sum_{i=1}^{d}\log\left(x_{i:m:n}\right)-a\sum_{i=1}^{d}x_{i:m:n}^{-2}\\&\quad+\left(b-1\right)\sum_{i=1}^{d}\log\left(1-e^{-ax_{i:m:n}^{-2}}\right)-3\sum_{i=1}^{d}\sum_{k=1}^{r_i}\log\left(z_{ik}\right)-a\sum_{i=1}^{d}\sum_{k=1}^{r_i}z_{ik}^{-2}\\&\quad+(b-1)\sum_{i=1}^{d}\sum_{k=1}^{r_i}\log\left(1-e^{-az_{ik}^{-2}}\right)\end{aligned} \tag{4.20}$$

ve

$$Q^* = r_j^* \left(\log(2)+\log(a)+\log(b)\right) - 3\sum_{p=1}^{r_j^*}\log\left(z_p^{'}\right) - a\sum_{p=1}^{r_j^*} z_p^{'\,-2} + (b+1)\sum_{p}^{r_j^*}\log\left(1-e^{-az_p^{'\,-2}}\right) \tag{4.21}$$

şeklindedir. EM algoritmasının E-adımı P^* ve Q^*'a göre beklenen değer almayı içerir. Buna göre;

$$P^* = n\log(2)+n\log(a)+n\log(b) - 3\sum_{i=1}^{d}\log\left(x_{i:m:n}\right) - a\sum_{i=1}^{d} x_{i:m:n}^{-2} + (b-1)\sum_{i=1}^{d}\log\left(1-e^{-ax_{i:m:n}^{-2}}\right) - 3\sum_{i=1}^{d}\sum_{k=1}^{r_i} E\left[\log(z_{ik}) \mid z_{ik} > x_{i:m:n}\right] - a\sum_{i=1}^{d}\sum_{k=1}^{r_i} E\left[z_{ik}^{-2} \mid z_{ik} > x_{i:m:n}\right] + (b-1)\sum_{i=1}^{d}\sum_{k=1}^{r_i} E\left[\log\left(1-e^{-az_{ik}^{-2}}\right) \mid z_{ik} > x_{i:m:n}\right] \tag{4.22}$$

ve

$$Q^* = r_j^* \left(\log(2)+\log(a)+\log(b)\right) - 3\sum_{p=1}^{r_j^*} E\left(\log z_p^{'} \mid Z' > T\right) - a\sum_{p=1}^{r_j^*} E\left[z_p^{'\,-2} \mid Z' > T\right] + (b+1)\sum_{i=1}^{r_j^*} E\left[\log\left(1-e^{-az_p^{'\,-2}}\right) \mid Z' > T\right] \tag{4.23}$$

şeklinde elde edilir. Buradaki koşullu beklenen değerler teorem 4.1 ve teorem 4.2 yardımıyla hesaplanır.

Teorem 4.1

$X_{1:m:n} = x_{1:m:n}, X_2 = x_{2:m:n}, ..., X_{i:m:n} = x_{i:m:n}$ verildiğinde Z_{ik}'nin $k = 1, ..., r_i, i = 1, 2, ..., m$ koşullu dağılımı ;

$$f_{Z_{ik}|X}\left(z_{ik} \mid x_{1:m:n}, x_{2:m:n}, ..., x_{i:m:n}\right) = \frac{f\left(z_{ik}\right)}{\left(1-F\left(x_{i:m:n}\right)\right)}, z_{ik} > x_{i:m:n} \tag{4.24}$$

şeklinde yazılır. Z_{ik} ve Z_{jl} $k \neq l$, $X_{i:m:n} = x_{i:m:n}$ verildiğinde koşullu bağımsızdır.

Teorem 4.2

$X_{1:m:n} = x_{1:m:n}, X_2 = x_{2:m:n}, ..., X_{i:m:n} = x_{i:m:n}$ ve $x_{m+1:n}, x_{m+2:n}, ..., x_{i:n}$ verildiğinde Z_{ik} *ve* $Z_P^{'}$'nın koşullu dağılımı;

$$f_{Z,Z^{'}|X,T}\left(z_{ik},z_{p}^{'}\mid x_{1:m:n},x_{2:m:n},...,x_{j:m:n},x_{m+1:n},x_{m+1:n},x_{m+2:n},...,x_{i:n},T\right)=\frac{f\left(z_{ik}\right)f\left(z_{p}^{'}\right)}{\left(1-F\left(x_{i:m:n}\right)\right)\left(1-F\left(T\right)\right)} \quad (4.25)$$

şeklinde elde edilir. Teorem 4.1 ve teorem 4.2 yardımıyla koşullu beklenen değerler aşağıdaki gibi hesaplanır.

$$\begin{aligned}E\left[\log(Z_{ik})\mid Z_{ik}>x_{i}\right]&=\frac{2ab}{\left(1-F\left(x_{i:m:n},a,b\right)\right)}\int_{x_{i}}^{\infty}\log\left(x\right)x^{-3}e^{-ax^{-2}}\left(1-e^{-ax^{-2}}\right)\times\\&\left(1-e^{-ax^{-2}}\right)^{b-1}.dx\\&=A\left(x_{i:m:n},a,b\right)\end{aligned} \quad (4.26)$$

$$\begin{aligned}E\left[Z_{ik}^{-2}\mid Z_{ik}>x_{i}\right]&=\frac{2ab}{\left(1-F\left(x_{i:m:n},a,b\right)\right)}\int_{x_{i}}^{\infty}x^{-2}x^{-3}e^{-ax^{-2}}\left(1-e^{-ax^{-2}}\right)\times\\&\left(1-e^{-ax^{-2}}\right)^{b-1}.dx\\&=B\left(x_{i:m:n},a,b\right)\end{aligned} \quad (4.27)$$

$$\begin{aligned}E\left[\log\left(1-e^{-aZ_{ik}^{-2}}\right)\mid Z_{ik}>x_{i:m:n}\right]&=\frac{2ab}{\left(1-F\left(x_{i:m:n},a,b\right)\right)}\int_{x_{i:m:n}}^{\infty}l\log\left(1-e^{-ax^{-2}}\right)x^{-3}e^{-ax^{-2}}\times\\&\left(1-e^{-ax^{-2}}\right)\left(1-e^{-ax^{-2}}\right)^{b-1}.dx\\&=C\left(x_{i:m:n},a,b\right)\end{aligned} \quad (4.28)$$

$$\begin{aligned}E\left[\log(Z_{p}^{'})\mid Z_{p}^{'}>T\right]&=\frac{2ab}{\left(1-F\left(T,a,b\right)\right)}\int_{T}^{\infty}\log\left(x\right)x^{-3}e^{-ax^{-2}}\left(1-e^{-ax^{-2}}\right)\times\\&\left(1-e^{-ax^{-2}}\right)^{b-1}.dx\\&=D\left(T,a,b\right)\end{aligned} \quad (4.29)$$

$$\begin{aligned}E\left[Z_{p}^{'-2}\mid Z_{p}^{'}>T\right]&=\frac{2ab}{\left(1-F\left(T,a,b\right)\right)}\int_{T}^{\infty}x^{-2}x^{-3}e^{-ax^{-2}}\left(1-e^{-ax^{-2}}\right)\times\\&\left(1-e^{-ax^{-2}}\right)^{b-1}.dx\\&=E\left(T,a,b\right)\end{aligned} \quad (4.30)$$

$$E\left[\log\left(1-e^{-aZ_p^{'-2}}\right)\mid Z_p^{'}>T\right]=\frac{2ab}{\left(1-F\left(T,a,b\right)\right)}\int_T^{\infty} l\log\left(1-e^{-ax^{-2}}\right)x^{-3}e^{-ax^{-2}}\times$$
$$\left(1-e^{-ax^{-2}}\right)\left(1-e^{-ax^{-2}}\right)^{b-1}.dx \qquad (4.31)$$
$$=F\left(T,a,b\right)$$

Buna göre (4.26)-(4.31) arasındaki eşitlikler yardımıyla log-olabilirlik fonksiyonları P^* ve Q^* tekrar düzenlenirse;

$$P^*=n\log(2)+n\log(a)+n\log(b)-3\sum_{i=1}^{d}\log\left(x_{i:m:n}\right)-a\sum_{i=1}^{d}x_{i:m:n}^{-2}$$
$$+(b-1)\sum_{i=1}^{d}\log\left(1-e^{-ax_{i:m:n}^{-2}}\right)-3\sum_{i=1}^{d}A\left(x_{i:m:n},a,b\right)-a\sum_{i=1}^{d}B\left(x_{i:m:n},a,b\right) \qquad (4.32)$$
$$+(b-1)\sum_{i=1}^{d}C\left(x_{i:m:n},a,b\right)$$

$$Q^*=r_j^*\left(\log(2)+\log(a)+\log(b)-3D\left(T,a,b\right)-aE\left(T,a,b\right)+\left(b+1\right)F\left(T,a,b\right)\right) \qquad (4.33)$$

şeklinde elde edilir. EM algoritmasının M adımı a ve b parametreleri için eşitlik (4.32) ve (4.33)' de düzenlenen fonksiyonlarının maksimizasyon adımını içerir. Buna göre durum I için P^* fonksiyonunun a ve b 'ye göre türevleri alınıp sıfıra eşitlenirse olabilirlik denklemleri aşağıdaki gibi yazılır.

$$\frac{\partial P^*}{\partial a}=\frac{n}{a}-\sum_{i=1}^{d}x_{i:m:n}^{-2}+(b-1)\sum_{i=1}^{d}\frac{e^{-ax_{i:m:n}^{-2}}x_{i:m:n}^{-2}}{1-e^{-ax_{i:m:n}^{-2}}}-\sum_{i=1}^{d}B\left(x_{i:m:n},a^k,b^k\right)=0 \qquad (4.34)$$

$$\frac{\partial P^*}{\partial b}=\frac{n}{b}+\sum_{i=1}^{d}\log\left(1-e^{-ax_{i:m:n}^{-2}}\right)+C\left(x_{i:m:n},a^k,b^k\right)=0 \qquad (4.35)$$

şeklinde elde edilir. Durum II için P^*+Q^* fonksiyonunun a ve b 'ye göre türevleri alınıp sıfıra eşitlenirse olabilirlik denklemleri aşağıdaki gibi yazılır.

$$\frac{\partial\left(P^*+Q^*\right)}{\partial a}=\frac{n}{a}-\sum_{i=1}^{d}x_i+(b-1)\sum_{i=1}^{d}\frac{x_i^{-2}e^{-ax_i^{-2}}}{1-e^{-ax_i^{-2}}}-\sum_{i=1}^{d}B\left(x_i,a^k,b^k\right)+$$
$$r_j^*\left(\frac{1}{a}+E\left(T,a^k,b^k\right)\right)=0 \qquad (4.36)$$

$$\frac{\partial\left(P^*+Q^*\right)}{\partial b}=\frac{n}{b}+\sum_{i=1}^{d}\log\left(1-e^{-ax_i^{-2}}\right)+C\left(x_i,a^k,b^k\right)+r_j^*\left(\frac{1}{b}+F\left(T,a^k,b^k\right)\right)=0 \qquad (4.37)$$

EM algoritması uygulanırken öncelikle başlangıç değerleri a^0 ve b^0 seçilir. E-adım ve M- adım iterasyonları yakınsama elde edilene kadar tekrar edilir.

4.3. Kumaraswamy Ters Rayleigh Dağılımı İçin Aralık Tahmini

Bu bölümde varyans-kovaryans matrisini elde etmek için Louis'in (1982) kayıp gözlem prensibi kullanılarak Fisher bilgi matrisi elde edilmiştir. Fisher bilgi matrisi bazı düzgünlük şartlarını sağlaması durumunda ilgilenilen dağılımın bilinmeyen parametreleri için asimptotik güven aralıklarını elde etmede kullanılır. Kayıp değer prensibi aşağıdaki gibi açıklanır.

$$I_X\left(\boldsymbol{\theta}\right) = I_W\left(\boldsymbol{\theta}\right) - I_{W/X}\left(\boldsymbol{\theta}\right) \tag{4.38}$$

burada $\boldsymbol{\theta} = \left(a, b\right)$,

$$I_{\mathbf{W|X}}\left(\boldsymbol{\theta}\right) = \begin{cases} \sum_{i=1}^{m} r_i I_{W|X}^{(i)}\left(\boldsymbol{\theta}\right) & \text{durum I için} \\ \sum_{i=1}^{j} r_i I_{W|X}^{(i)}\left(\boldsymbol{\theta}\right) + r_j^{*} I_{W|T}\left(\boldsymbol{\theta}\right) & \text{durum II için} \end{cases} \tag{4.39}$$

ve

$$I_W\left(\theta\right) = -E\left[\frac{\partial^2\left(L_c\left(\mathbf{W};\boldsymbol{\theta}\right)\right)}{\partial\boldsymbol{\theta}^2}\right] \tag{4.40}$$

şeklindedir. Sansürlü gözlemler altında fisher bilgisi;

$$I_{W|X}^{(i)}\left(\boldsymbol{\theta}\right) = -E\left[\frac{\partial^2 \log f_{Z_i}\left(z_i \mid x_i, \boldsymbol{\theta}\right)}{\partial\boldsymbol{\theta}}\right] \tag{4.41}$$

ve

$$I_{W|T}^{(i)}\left(\boldsymbol{\theta}\right) = -E\left[\frac{\partial^2 \log f_{Z'}\left(z' \mid T, \boldsymbol{\theta}\right)}{\partial\boldsymbol{\theta}}\right] \tag{4.42}$$

şeklinde hesaplanır. $\hat{\boldsymbol{\theta}}$ ' nın asimptotik varyans-kovaryans matrisi, $I_X\left(\hat{\boldsymbol{\theta}}\right)$ ' nin tersi alınarak elde edilir Buna göre Kw-TR dağılımı için $I_W\left(\boldsymbol{\theta}\right)$ 'nın elemanları;

$$\alpha_{11} = \begin{cases} \dfrac{n}{a^2} + n\left(b-1\right)2ab\displaystyle\int_0^{\infty} x^{-7} e^{-2ax^{-2}}\left(1-e^{-ax^{-2}}\right)^{b-3} dx & \text{durum I} \\ \left(\dfrac{n+r_j^{*}}{\alpha^2}\right) + \left(n+r_j^{*}\right)\left(b-1\right)2ab\displaystyle\int_0^{\infty} x^{-7} e^{-2ax^{-2}}\left(1-e^{-ax^{-2}}\right)^{b-3} dx & \text{durum II} \end{cases} \tag{4.43}$$

$$\alpha_{12}=\alpha_{21}=\begin{cases}\dfrac{n}{a^2}-n2ab\displaystyle\int_0^{\infty}x^{-5}e^{-2ax^{-2}}\left(1-e^{-ax^{-2}}\right)^{b-2}dx & \text{durum I}\\ \left(\dfrac{n+r_j^*}{a^2}\right)+\left(\dfrac{n+r_j^*}{a^2}\right)2ab\displaystyle\int_0^{\infty}x^{-5}e^{-2ax^{-2}}\left(1-e^{-ax^{-2}}\right)^{b-2}dx & \text{durum II}\end{cases} \tag{4.44}$$

$$\alpha_{22}=\begin{cases}\dfrac{n}{b^2} & \text{durum I}\\ \left(\dfrac{n+r_j^*}{b^2}\right) & \text{durum II}\end{cases} \tag{4.45}$$

şeklinde hesaplanır. Sansürlü gözlemler altında $I_{W|X}^{(i)}(\boldsymbol{\theta})$ matrisinin elemanları ise;

$$I_{W|X}^{(i)}(\boldsymbol{\theta})=\begin{bmatrix}\beta_{11}(x_i;a,b) & \beta_{12}(x_i;a,b)\\ \beta_{21}(x_i;a,b) & \beta_{22}(x_i;a,b)\end{bmatrix} \tag{4.46}$$

burada;

$$\beta_{11}(x_i;a,b)=\frac{1}{a^2}+(b-1)h_1(x_i;a,b)-k_1 \tag{4.47}$$

$$\beta_{12}(x_i;a,b)=\beta_{21}(x_i;a,b)=-h_2(x_i;a,b)+k_2 \tag{4.48}$$

$$\beta_{22}(x_i;a,b)=\frac{1}{b^2} \tag{4.49}$$

olarak elde edilir. $h_1(x_i;a,b)$, $h_2(x_i;a,b)$, k_1 ve k_2 ise aşağıdaki gibidir.

$$h_1(x_i;a,b)=\frac{2ab}{\left(1-\left(1-\left(1-e^{-ax_i^{-2}}\right)^b\right)\right)}\int_{x_i}^{\infty}x^{-7}e^{-2ax^{-2}}\left(1-e^{-ax^{-2}}\right)^{b-3}dx \tag{4.50}$$

$$k_1=b\frac{e^{-ax_i^{-2}}}{x_i^4\left(1-e^{-ax_i^{-2}}\right)} \tag{4.51}$$

$$h_2(x_i;a,b)=\frac{2ab}{\left(1-\left(1-\left(1-e^{-ax_i^{-2}}\right)^b\right)\right)}\int_{x_i}^{\infty}x^{-5}e^{-2ax^{-2}}\left(1-e^{-ax^{-2}}\right)^{b-2}dx \tag{4.52}$$

$$k_2=\frac{e^{-ax^{-2}}}{x^2\left(1-e^{-ax^{-2}}\right)} \tag{4.53}$$

$I_{W|T}(\boldsymbol{\theta})$ matrisinin elemanları ise;

$$I_{W|T}(\boldsymbol{\theta})=\begin{bmatrix} c_{11}(x_i;a,b) & c_{12}(x_i;a,b) \\ c_{21}(x_i;a,b) & c_{22}(x_i;a,b) \end{bmatrix} \tag{4.54}$$

şeklindedir. Burada;

$$c_{11}(T;a,b)=\frac{1}{a^2}+(b-1)g_1(T;a,b)-l_1 \tag{4.55}$$

$$c_{12}(T;a,b)=c_{21}(T;a,b)=-g_2(T;a,b)+l_2 \tag{4.56}$$

$$c_{22}(T;a,b)=\frac{1}{b^2} \tag{4.57}$$

olarak hesaplanır. $g_1(T;a,b)$, $g_2(T;a,b)$, l_1 ve l_2 elde edilir.

$$g_1(T;a,b)=\frac{2ab}{\left(1-\left(1-\left(1-e^{-aT^{-2}}\right)^b\right)\right)}\int_T^{\infty} x^{-7}e^{-2ax^{-2}}\left(1-e^{-ax^{-2}}\right)^{b-3}dx \tag{4.58}$$

$$l_1=b\frac{e^{-aT^{-2}}}{T^4\left(1-e^{-aT^{-2}}\right)} \tag{4.59}$$

$$g_2(T;a,b)=\frac{2ab}{\left(1-\left(1-\left(1-e^{-aT^{-2}}\right)^b\right)\right)}\int_T^{\infty} x^{-5}e^{-2ax^{-2}}\left(1-e^{-ax^{-2}}\right)^{b-2}dx \tag{4.60}$$

$$l_2=\frac{e^{-aT^{-2}}}{T^2\left(1-e^{-aT^{-2}}\right)} \tag{4.61}$$

a ve b parametreleri için $1-\eta$ güven seviyeli asimptotik güven aralığı eşitlik (4.62) ve (4.63) ’de ifade edilmiştir.

$$P\left(\hat{a}-z_{\eta/2}\sqrt{Var(\hat{a})}<a<\hat{a}+z_{\eta/2}\sqrt{Var(\hat{a})}\right)=1-\eta \tag{4.62}$$

$$P\left(\hat{b}-z_{\eta/2}\sqrt{Var(\hat{b})}<b<\hat{b}+z_{\eta/2}\sqrt{Var(\hat{b})}\right)=1-\eta \tag{4.63}$$

4.4. Simülasyon Çalışması

Bu bölümde, ilerleyen tip I hibrit sansürlü örneklem altında Kw-TR dağılımının bilinmeyen a ve b parametrelerinin en çok olabilirlik tahmin edicilerinin

performansları farklı sansürleme şemaları için yan, hata kareler ortalaması (HKO) ve aralık uzunluğu bakımından incelenmiştir. Yanlar, HKO ve aralık uzunluğu farklı sansürleme şemaları durumunda $a = 0.9$ ve $b = 1.2$ için denemenin 10.000 kez tekrarlanması durumunda elde edilmiştir ve bu durumlara ilişkin sonuçlar Çizelge (4.4) 'de verilmiştir.

Çizelge 4.1: : T=10 ve farklı n ve r değerleri için Kw-TR dağılımı için nokta ve aralık tahmin sonuçları

(n,m)	Sansür Şeması	Yan		HKO		Aralık Uzunluğu		Kapsama Olasılığı	
		$\hat{a}$	$\hat{b}$	$\hat{a}$	$\hat{b}$	$\hat{a}$	$\hat{b}$	$\hat{a}$	$\hat{b}$
(20,10)	$(9*0,10)$	-0.0372	-0.0253	0.0383	0.0509	1.2920	2.6696	0.9919	0.9985
(20,15)	$(14*0,5)$	0.0232	0.0855	0.0513	0.0907	1.1441	2.8013	0.9872	0.9878
(30,15)	$(14*0,15)$	-0.0500	-0.0556	0.0251	0.0376	1.0374	2.0887	0.9938	0.9982
(30,20)	$(20*0,10)$	-0.0263	-0.0001	0.0279	0.0564	0.9372	1.7182	0.9862	0.9904
(40,15)	$(14*0,25)$	-0.2679	-0.7968	0.6475	0.0800	0.8880	2.5160	0.9034	0.9408
(40,30)	$(29*0,10)$	-0.0180	0.0057	0.0215	0.0489	0.7765	1.3375	0.9834	0.9874
(50,25)	$(29*0,10)$	-.0.0601	-0.0804	0.0169	0.0264	0.7928	1.5603	0.9920	0.9983
(50,35)	$(39*0,10)$	-0.0377	0.0075	0.0166	0.0257	0.6775	1.1266	0.9808	0.9832
(60,35)	$(34*0,25)$	-.0.0572	-0.0699	0.0148	0.0252	0.6790	1.2557	0.9850	0.9938
(60,45)	$(44*0,15)$	-0.0302	-0.0168	0.0144	0.0300	0.6264	1.0626	0.9834	0.9828
(100,60)	$(59*0,40)$	-0.0621	-0.0804	0.0104	0.0189	0.5167	0.9340	0.9804	0.9942
(100,80)	$(79*0,20)$	-0.0334	-0.0264	0.0094	0.0199	0.4702	0.7654	0.9721	0.9790
(150,90)	$(89*0,60)$	-0.0654	-0.0880	0.0086	0.0160	0.4201	0.7547	0.9717	0.9914
(150,120)	$(119*0,30)$	-0.0366	-0.0352	0.0070	0.0148	0.3828	0.6181	0.9642	0.9775
(150,140)	$(139*0,10)$	-0.0052	0.0040	0.0076	0.0170	0.3685	0.5576	0.9628	0.9686

Çizelge 4.4 incelendiğinde n değeri sabit tutulup m değeri arttığında HKO' da ve yan değerlerinde azalma aralık uzunluğunun ise daraldığı görülmektedir. Ayrıca örneklem hacmi n değeri arttığında HKO azalmakta ve kapsama olasılıklarının %95'e yaklaştığı gözlenmektedir.

4.5. Gerçek Veri Analizi

Bu bölümde iki adet gerçek veri analiz çalışması yapılmıştır. Birinci veri grubu, ağrı kesici alan 20 hastanın rahatlama zamanlarından oluşmaktadır. Birinci veri grubu Afify ve ark. (2017) tarafından dönüştürülmüş beta dağılımının modellenmesinde kullanılmıştır(Afify ve ark., 2017). İkinci veri grubunda, 20 mekanik parçanın bozulma

süresinden oluşan veri grubundan oluşmaktadır. İkinci veri grubu Genç ve ark .(2014) tarafından Beta moyal dağılımının modellenmesinde kullanılmıştır (Genc ve ark., 2014). Veri grupları sırasıyla aşağıda verilmiştir.

Veri Grubu I

1.1 1.2000 1.3000 1.4000 1.4000 1.5000 1.6000 1.6000 1.7000 1.7000 1.7000 1.8000 1.8000 1.9000 2.0000 2.2000 2.3000 2.7000 3.0000 4.1000

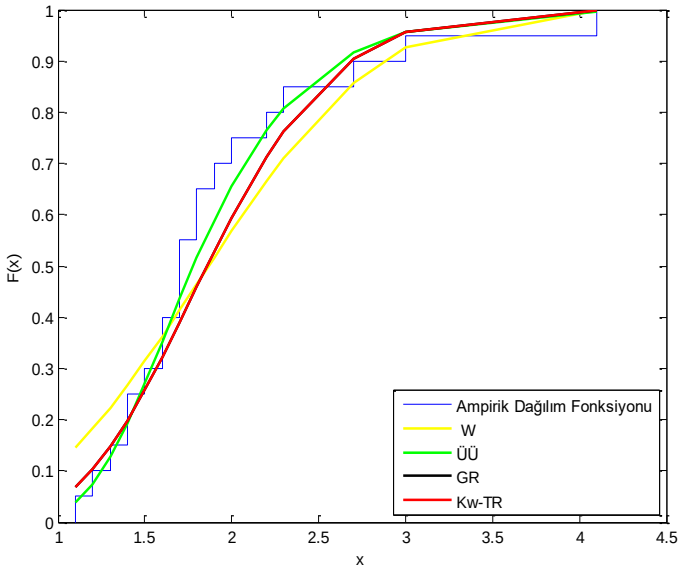
Veri Grubu II

0,067 0,068 0,076 0,081 0,084 0,085 0,085 0,086 0,089 0,098 0,098 0,114 0,114 0,115 0,121 0,125 0,131 0,149 0,160 0,485

Üzerinde çalışılan veri gruplarının Kw-TR dağılımından gelip gelmediğini test etmek için uyum iyiliği kriterleri uygulanmıştır. Birinci veri grubu W, Ü.Ü, GR, Kw-TR dağılımlarını karşılaştırmak için kullanılan AIC,BIC,-2logL,Kolmogorov-Smirnov (KS) uyum iyiliği test sonuçları aşağıdaki Çizelge (4.2)'de verilmiştir. Ayrıca veri grubu 1 için bütün dağılımların, dağılım fonksiyonu grafikleri şekil (4.1) 'de verilmiştir.

Çizelge 4.2: Hastaların rahatlama zamanları ile ilgili uyum iyiliği sonuçları

	$\hat{a}$	$\hat{b}$	AIC	BIC	-2logL	K.S	p değeri
Kw-TR(a,b)	**5.4553**	**3.6098**	**35.7359**	**37.7273**	**31.7359**	**0.1264**	**0.9065**
GR(a,b)	0.6901	3.2459	40.8045	42.7960	36.8045	0.1901	0.4650
W(a,b)	2.1299	2.7870	45.1728	47.1642	41.1728	0.1849	0.5005
Ü.Ü(a,b)	2.23	36.6831	36.5212	38.5126	32.5212	0.1343	0.8633



Şekil 4.1: Hastaların rahatlama zamanları verisi için uydurulan dağılım fonksiyonu grafikleri

Çizelge (4.5) tüm kriterler bakımından incelendiğinde veri gurubunu en iyi modelleyen dağılımın Kw-TR olduğu görülmektedir. 20 hastanın rahatlama zamanları ile ilgili veri grubundan $m=10$ ve $m=15$ için rasgele olarak üretilen ilerleyen tür tip II hybrit sansürlü veriler ve sansürleme şemaları Çizelge (4.3)'da verilmiştir. Çizelge (4.4)'de farklı sansürleme şemaları ve farklı n, m ve T değerleri için (MLE) , (HKO) ve güven aralıkları ile ilgili sonuçlar yer almaktadır.

Çizelge 4.3: Sansürleme şemaları ve Üretilen veriler

Sansürleme Şeması	İlerleyen Tür Tip II Sansürlemeden Üretilen Veriler
$s_{Ş1}=(0^{*2},3,4,1,0^{*3},1)$	1.1 1.20001.3000 1.6000 1.800 1.900 2.2000 2.3000 2.7000 3.0000
$s_{Ş2}=(0^{*2},2,7,0^{*5},1)$	1.1 1.2000 1.300 1.500 1.9000 2.000 2.2000 2.3000 2.7000 3000
$s_{Ş3}=(0^{*2},0,1,2^{*3},0^{*3},1)$	1.1 1.200 1.3000 1.4000 1.5000 1.6000 1.7000 1.8000 1.9000 2.2000 2.3000 2.7000 3.0000 4.1000
$s_{Ş4}=(0^{*3},1,0,1,2,1^{*2},0^{*5})$	1.1 1.2000 1.3000 1.4000 1.5000 1.6000 1.7000 1.8000 1.9000 2.2000 2.3000 2.7000 3.0000 4.1000

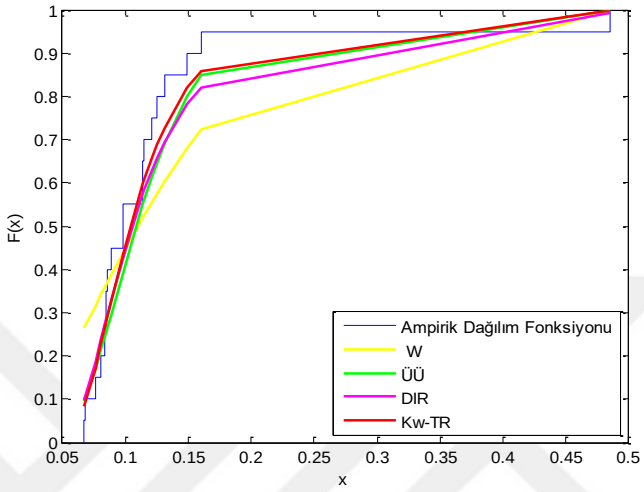
Çizelge 4.4: Farklı sansürleme şemaları n , m ve T değerleri için MLE tahminleri ve asimptotik güven aralığı

			T=2.0		T=4.0	
m	Sansürleme Şemaları	Parametre	MLE	Asimptotik Güven Arağı	MLE	Simptotik Güven Arağı
10	$s_{Ş1}$	a	5.0460 (1.7861)	(1.5452,8.5468)	5.6193 (1.7114)	(2.2649,8.9737)
		b	1.6896 (1.0943)	(0.0000, 3.8335)	2.1816 (1.1992)	(0.0000, 4.5320)
10	$s_{Ş2}$	a	5.1398 (1.7467)	(1.7154,8.5623)	5.6253 (1.6548)	(2.3819,8.8688)
		b	1.8666 (1.1945)	(0.0000,4.2079)	2.3291 (1.2527)	(0.0000,5.6468)
14	$s_{Ş3}$	a	5.1347 (1.5741)	(2.0494,8.2201)	4.9491 (1.3772)	(2.2497,7.6484)
		b	2.2629 (1.2770)	(0.0000,4.7728)	2.0935 (0.9570)	(0.2178,3.9693)
14	$s_{Ş4}$	a	5.1069 (1.4736)	(2.2188,7.9951)	5.2982 (1.3819)	(2.5897,8.0068)
		b	2.5660 (1.3734)	(0.0000,5.2579)	2.7916 (1.2941)	(0.2552,5.3281)

İkinci veri grubu W, Ü.Ü, DIR, Kw-TR dağılımlarını karşılaştırmak için kullanılan AIC,BIC,-2logL,Kolmogorov-Smirnov (KS) uyum iyiliği test sonuçları aşağıdaki Çizelge (4.5)'de verilmiştir. Ayrıca veri grubu II için bütün dağılımların, dağılım fonksiyonu grafikleri şekil (4.2) verilmiştir.

Çizelge 4.5: Mekanik parçaların bozulma süreleri ile ilgili uyum iyiliği sonuçları

	$\hat{a}$	$\hat{b}$	AIC	BIC	-2logL	K.S	P Değeri
Kw-TR(a,b)	**0.0149**	**2.3897**	**-70.1140**	**-68.1225**	**-74.1140**	**0.1261**	**0.9081**
DIR(a,b)	0.9362	0.0133	-69.9714	-67.9800	-73.9714	0.1588	0.6936
W(a,b)	0.9362	0.1337	-48.8456	-46.8542	-52,8456	0.2641	0.1227
Ü.Ü(a,b)	27.7525	13.8250	-61.9529	-59.9614	-65.9529	0.1602	0.6830



Şekil 4.2: Mekanik parçaların bozulma süreleri verisi için uydurulan dağılım fonksiyonu grafikleri

Çizelge (4.5) tüm kriterler bakımından incelendiğinde veri gurubunu en iyi modelleyen dağılımın Kw-TR olduğu görülmektedir. 20 mekanik parçanın bozulma zamanları ile ilgili veri grubundan $m=10$ ve $m=14$ için rasgele olarak üretilen ilerleyen tür tip II sansürlü üretilen veriler ve sansürleme şemaları Çizelge (4.6)'da verilmiştir. Farklı n, m ve T değerleri için hesaplanan (MLE) , (HKO) ve güven aralıkları Çizelge (4.7)' de yer almaktadır.

Çizelge 4.6: Sansürleme şemaları ve Üretilen veriler

Sansürleme Şeması	İlerleyen Tür Tip II Sansürlemeden Üretilen Veriler
$s_{Ş1}$=(3,2,0,2,1,0^{*4},2)	0.0670;0.0840;0.0860;0.890;0.1140;0.1150;0.1210;0.1250;0.1310;0.1490
$s_{Ş2}$=(0^{*4},3,0,1^{*2},0,5)	0.0670;0.0680;0.0760;0.0810;0.0840;0.0890;0.0980;0.1140;0.1150;0.1210
$s_{Ş3}$=(0^{*4},2,0,4,0^{*7})	0.0670;0.0680;0.0760;0.0810;0.0840;0.0860;0.0890;0.1150;0.1210;0.1250;0.1310; 0.1490;0.1600;0.4850

Çizelge 4.7: Farklı sansürleme şemaları n , m ve T değerleri için MLE tahminleri ve asimptotik güven aralığı

			T=0.10		T=0.14	
m	Sansürleme Şemaları	Parametre	MLE	Asimptotik Güven Arağı	MLE	Asimptotik Güven Arağı
10	ss_1	a	0.0170 (0.0100)	(0.0000, 0.0367)	0.0234 (0.0830)	(0.0071,0.0396)
		b	1.0381 (1.3763)	(0.0000, 3.7356)	2.4130 (1.7270)	(0.0000, 5.7980)
10	ss_2	a	0.0190 (0.0072)	(0.0049,0.0332)	0.0185 (0.0058)	(0.0070,0.0299)
		b	3.3134 (2.9142)	(0.0000,8.8252)	2.7026 (1.7413)	(0.0000,6.1155)
14	ss_3	a	0.0195 (0.0073)	(0.0053,0.0338)	0.0179 (0.0052)	(0.0077,0.0281)
		b	3.6929 (3.5892)	(0.0000,10.7276)	2.7269 (1.5352)	(0.0000,5.7319)

5. SONUÇLAR VE ÖNERİLER

5.1 Sonuçlar

Bu tez çalışmasının 3. bölümünde Tip I hibrit sansürlü örneklem altında, Lkw dağılımının bilinmeyen parametreleri için nokta ve aralık tahmin konuları ele alınmıştır. Nokta tahminlerini elde etmek için, en çok olabilirlik tahmin edicileri metodu düşünülmüştür ve en çok olabilirlik tahmin edicilerini elde etmek için EM algoritması kullanılmıştır. Tip I hibrit sansürleme şeması altında bazı düzgünlük koşullarının sağlanması durumunda LKw dağılımının parametrelerine ilişkin Fisher Bilgi matrisi elde edilmiştir. Elde edilen Fisher Bilgi matrisi asimptotik güven aralıklarını elde etmede kullanılmıştır. Monte Karlo simülasyonları yardımıyla farklı sansürleme şemaları altında en çok olabilirlik tahmin edicilerinin performansları; yan, HKO, aralık uzunluğu ve kapsama olasılıkları bakımından incelenmiştir. Açıkça görüldüğü üzere T zamanı sabit tutulduğunda, n ve r değeri arttığında ortalama yanın ve HKO'nın azaldığı görülmüştür. Ayrıca T zamanı sabit tutulduğunda, n ve r değeri arttığında aralık uzunluğunun daraldığı ve kapsama olasılıklarının 0.95'e yakınsadığı görülmektedir. Bununla birlikte iki gerçek veri uygulaması gerçekleştirilmiştir ve farklı sansürleme şemaları altında standart hata, en çok olabilirlik tahminleri, asimptotik güven aralıkları değerlendirilmiştir. Gerçek veri uygulmasında da görüldüğü üzere n, r ve T değeri değerleri arttıkça standart hatanın azaldığı ve aralık uzunluklarının daraldığı görülmüştür.

Tez çalışmasının 4. Bölümünde ise diğer bir sansürleme şaması olan ilerleyen tür tip I hibrit sansürleme altında, Kw-TR dağılımının bilinmeyen parametreleri için nokta ve aralık tahminleri elde edilmiştir. Nokta tahminlerini elde etmek için en çok olabilirlik tahmin edicileri metodu kullanılmıştır. Ancak en çok olabilirlik tahmin edicileri açıkça elde edilemediği için, en çok olabilirlik tahmin edicilerini bulmak için EM algoritması kullanılmıştır. İlerleyen tür tip I hibrit sansürleme şeması altında, bazı düzgünlük koşullarının sağlaması durumunda Fisher bilgi matrisi elde edilmiştir. Elde edilen Fisher bilgi matrisi asimptotik güven aralıklarını elde etmede kullanılmıştır. Tahmin edicilerin performanslarını kıyaslamak için bir adet Monte Karlo simülasyonu gerçekleştirilmiştir. Simülasyon sonuçlarını bakıldığında n değeri sabit tutulup m değeri arttığında ortalama yanın, HKO'nın azaldığı aralık uzunluğunun ise daraladığı görülmektedir. Ayrıca örneklem hacmi n değeri arttığında ortalama yan, HKO azalmakta ve kapsama olasılıklarının $\%95$'e yaklaştığı görülmüştür. Son olarak sonuçları güncel hayata uyarlamak için iki adet gerçek veri uygulaması yapılmıştır ve farklı ilerleyen tür tip I hibrit sansürleme şemaları altında parametre tahminleri ve parametre tahminlerine ilişkin standart hatalar incelenmiştir. . Gerçek veri analizindeki en iyi sonucun veri gurubu I için $m=14$ ve SS_4 durumunun ve veri grubu II için $m=14$ ve SS_3 durumunun olduğu görülmektedir.

5.2 Öneriler

İlerideki çalışmalarda, hibrit sansürleme şemaları altında bu dağılımlar için stres dayanıklılık güvenilirliğinin tahmini konusunda çalışmalar yapılabileceği planlanmaktadır.

KAYNAKLAR

Afify, A., Yousof, H. ve Nadarajah, S., 2017, The beta transmuted-H family for lifetime data, *Statistics and Its Interface*, 10 (3), 505-520.

Ahmad, A., Ahmad, S. ve Ahmed, A., 2014, Transmuted Inverse Rayleigh distribution: a generalization of the Inverse Rayleigh distribution, *Mathematical Theory and Modeling*, 4 (7), 90-98.

Aitchison, J. ve Brown, J. A., 1957, The lognormal distribution with special reference to its uses in economics.

Akaike, H., 1998, Information theory and an extension of the maximum likelihood principle, In: Selected papers of hirotugu akaike, Eds: Springer, p. 199-213.

Aryal, G. R. ve Tsokos, C. P., 2011, Transmuted Weibull distribution: A generalization of theWeibull probability distribution, *European Journal of Pure and Applied Mathematics*, 4 (2), 89-102.

Ateya, S. F., 2017, Estimation under inverse Weibull distribution based on Balakrishnan's unified hybrid censored scheme, *Communications in Statistics-Simulation and Computation*, 46 (5), 3645-3666.

Balakrishnan, N. ve Sandhu, R., 1995, A simple simulational algorithm for generating progressive Type-II censored samples, *The American Statistician*, 49 (2), 229-230.

Balakrishnan, N. ve Rao, C. R., 1998a, 1 Order statistics: An introduction, *Handbook of statistics*, 16, 3-24.

Balakrishnan, N. ve Rao, C. R., 1998b, Order statistics: theory & methods, Elsevier Amsterdam, p.

Balakrishnan, N. ve Aggarwala, R., 2000, Progressive censoring: theory, methods, and applications, Springer Science & Business Media, p.

Balakrishnan, N. ve Saleh, H., 2012, Relations for moments of progressively Type-II censored order statistics from log-logistic distribution with applications to inference, *Communications in Statistics-Theory and Methods*, 41 (5), 880-906.

Balakrishnan, N. ve Kundu, D., 2013, Hybrid censoring: models, inferential results and applications, *Computational Statistics & Data Analysis*, 57 (1), 166-209.

Casella, G. ve Berger, R. L., 2001, Statistical inference, Duxbury/Thomson Learning, p.

Cavanaugh, J. E., 1997, Unifying the derivations for the Akaike and corrected Akaike information criteria, *Statistics & Probability Letters*, 33 (2), 201-208.

Cordeiro, G. M. ve de Castro, M., 2011, A new family of generalized distributions, *Journal of Statistical Computation and Simulation*, 81 (7), 883-898.

Dempster, A. P., Laird, N. M. ve Rubin, D. B., 1977, Maximum likelihood from incomplete data via the EM algorithm, *Journal of the royal statistical society. Series B (methodological)*, 1-38.

Dube, S., Pradhan, B. ve Kundu, D., 2011, Parameter estimation of the hybrid censored log-normal distribution, *Journal of Statistical Computation and Simulation*, 81 (3), 275-287.

Epstein, B., 1954, Truncated life tests in the exponential case, *The Annals of Mathematical Statistics*, 555-564.

Genc, A. A., Korkmaz, M. C. ve Kus, C., 2014, The Beta Moyal-Slash Distribution, *Journal of Selcuk University Natural and Applied Science*, 3 (4), 88-104.

Gupta, R. D. ve Kundu, D., 2001, Exponentiated exponential family: an alternative to gamma and Weibull distributions, *Biometrical Journal: Journal of Mathematical Methods in Biosciences*, 43 (1), 117-130.

Hurvich, C. M. ve Tsai, C.-L., 1989, Regression and time series model selection in small samples, *Biometrika*, 76 (2), 297-307.

Kantam, R., Rosaiah, K. ve Rao, G. S., 2001, Acceptance sampling based on life tests: Log-logistic model, *Journal of applied statistics*, 28 (1), 121-128.

Kayal, T., Tripathi, Y. M., Rastogi, M. K. ve Asgharzadeh, A., 2017, Inference for Burr XII distribution under Type I progressive hybrid censoring, *Communications in Statistics-Simulation and Computation*, 46 (9), 7447-7465.

Kohansal, A., Rezakhah, S. ve Khorram, E., 2015, Parameter estimation of Type-II hybrid censored weighted exponential distribution, *Communications in Statistics-Simulation and Computation*, 44 (5), 1273-1299.

Kullback, S. ve Leibler, R. A., 1951, On information and sufficiency, *The Annals of Mathematical Statistics*, 22 (1), 79-86.

Kumaraswamy, P., 1980, A generalized probability density function for double-bounded random processes, *Journal of Hydrology*, 46 (1-2), 79-88.

Kundu, D. ve Raqab, M. Z., 2005, Generalized Rayleigh distribution: different methods of estimations, *Computational Statistics & Data Analysis*, 49 (1), 187-200.

Kundu, D. ve Joarder, A., 2006, Analysis of Type-II progressively hybrid censored data, *Computational Statistics & Data Analysis*, 50 (10), 2509-2528.

Kundu, D. ve Pradhan, B., 2009, Estimating the parameters of the generalized exponential distribution in presence of hybrid censoring, *Communications in Statistics—Theory and Methods*, 38 (12), 2030-2041.

Kuş, C., 2004, Bazı Yaşam Zamanı Dağılımlarının Parametrelerinin Tam ve Sansürlü Verilere Dayalı Tahmini, Doktora Tezi, *Selçuk Üniversitesi Fen Bilimleri Enstitüsü*.

Lehmann, E. L. ve Casella, G., 2006, Theory of point estimation, Springer Science & Business Media, p.

Lemonte, A. J., Barreto-Souza, W. ve Cordeiro, G. M., 2013, The exponentiated Kumaraswamy distribution and its log-transform, *Brazilian Journal of Probability and Statistics*, 27 (1), 31-53.

Little, R. J. ve Rubin, D. B., 2014, Statistical analysis with missing data, John Wiley & Sons, p.

Louis, T. A., 1982, Finding the observed information matrix when using the EM algorithm, *Journal of the Royal Statistical Society: Series B (Methodological)*, 44 (2), 226-233.

Mao, S., Shi, Y.-M. ve Sun, Y.-D., 2014, Exact inference for competing risks model with generalized type-I hybrid censored exponential data, *Journal of Statistical Computation and Simulation*, 84 (11), 2506-2521.

Morabbi, H. ve Razmkhah, M., 2009, Entropy of hybrid censoring schemes.

Ng, H., Chan, P. ve Balakrishnan, N., 2002, Estimation of parameters from progressively censored data using EM algorithm, *Computational Statistics & Data Analysis*, 39 (4), 371-386.

Nichols, M. D. ve Padgett, W., 2006, A bootstrap control chart for Weibull percentiles, *Quality and reliability engineering international*, 22 (2), 141-151.

Park, C. ve Lee, S. B., 2012, Parameter estimation from censored samples using the expectation-maximization algorithm, *arXiv preprint arXiv:1203.3880*.

Rastogi, M. K. ve Tripathi, Y. M., 2013, Inference on unknown parameters of a Burr distribution under hybrid censoring, *Statistical Papers*, 54 (3), 619-643.

Roussas, G. G., 1997, A course in mathematical statistics, Elsevier, p.

Sultana, F., Tripathi, Y. M., Rastogi, M. K. ve Wu, S.-J., 2018, Parameter Estimation for the Kumaraswamy Distribution Based on Hybrid Censoring, *American Journal of Mathematical and Management Sciences*, 1-19.

UCAL, M. Ş., 2006, EKONOMETRİK MODEL SEÇİM KRİTERLERİ ÜZERİNE KISA BIR İNCELEME.

ul Haq, M. A., 2016, Transmuted exponentiated inverse Rayleigh distribution.

Usman, R. M., Haq, M. ve Talib, J., 2017, Kumaraswamy half-logistic distribution: properties and applications, *J Stat Appl Probab*, 6, 597-609.

Weibull, W., 1951, A statistical distribution function of wide applicability, *Journal of applied mechanics*, 18 (3), 293-297.

Yildirim, N. ve Gökpinar, F., 2012, Bazı Normallik Testlerinin 1. Tip Hataları ve Güçleri Bakımından Kıyaslanması, *Süleyman Demirel Üniversitesi Fen Bilimleri Enstitüsü Dergisi*, 16 (1).

Zamar, R., Peña Sánchez de Rivera, D. ve Justel, A., 1994, A multivariate Kolmogorov-Smornov test of goodnes of fit, *Universidad Carlos III de Madrid. Departamento de Estadística*.

Zhu, X. ve Balakrishnan, N., 2015, Birnbaum–Saunders distribution based on Laplace kernel and some properties and inferential issues, *Statistics & Probability Letters*, 101, 1-10.

ÖZGEÇMİŞ

KİŞİSEL BİLGİLER

Adı Soyadı : Egemen ÖZKAN
Uyruğu : T.C
Doğum Yeri ve Tarihi : Konya-1993
Telefon : 0546 217 23 09
e-mail : egemen2858@gmail.com

EĞİTİM

Derece		**Adı, İlçe, İl**	**Bitirme Yılı**
Lise	:	Konevi Anadolu Lisesi, Selçuklu, Konya	2011
Üniversite	:	Selçuk Üniversitesi , Selçuklu, Konya	2016
Yüksek Lisans	:	Selçuk Üniversitesi , Selçuklu, Konya	

İŞ DENEYİMLERİ

Yıl	**Kurum**	**Görevi**

UZMANLIK ALANI

İstatistik

YABANCI DİLLER

İngilizce

BELİRTMEK İSTEĞİNİZ DİĞER ÖZELLİKLER

YAYINLAR

Özkan E., Saraçoğlu B., (2019), ‘ Point and Interval Estimation for Log-Kumaraswamy Distrubition based on Type I Hybrid Censoring, International Conference on Applied Analysis and Mathematical Modeling (ICAAMM-2019), (Özet Bildiri /Sözlü Sunum) 10-13 March.