



**YAŞAM ANALİZİ İÇİN YENİ DAĞILIM AİLELERİ
VE İSTATİSTİKSEL ÖZELLİKLERİ**

Doktora Tezi

İbrahim ARIK

Eskişehir 2018

**YAŞAM ANALİZİ İÇİN YENİ DAĞILIM AİLELERİ
VE İSTATİSTİKSEL ÖZELLİKLERİ**

İbrahim ARIK

DOKTORA TEZİ

İstatistik Anabilim Dalı

Danışman: Prof. Dr. Yeliz MERT KANTAR

Eskişehir

Anadolu Üniversitesi

Fen Bilimleri Enstitüsü

Kasım 2018

JÜRİ VE ENSTİTÜ ONAYI

.....'nın “.....
.....” başlıklı
tezi .../.../20.. tarihinde aşağıdaki jüri tarafından değerlendirilerek “Anadolu Üniversitesi
Lisansüstü Eğitim-Öğretim ve Sınav Yönetmeliği”nin ilgili maddeleri uyarınca,
..... Anabilim dalında Doktora tezi olarak kabul edilmiştir.

Jüri Üyeleri

Unvanı Adı Soyadı

İmza

Üye (Tez Danışmanı)

:

Üye

:

Üye

:

Üye

:

Üye

:

Prof.Dr. Ersin YÜCEL
Fen Bilimleri Enstitüsü Müdürü

ÖZET

YAŞAM ANALİZİ İÇİN YENİ DAĞILIM AİLELERİ VE İSTATİSTİKSEL ÖZELLİKLERİ

İbrahim ARIK

İstatistik Anabilim Dalı

Anadolu Üniversitesi, Fen Bilimleri Enstitüsü, Kasım 2018

Danışman: Prof. Dr. Yeliz MERT KANTAR

Birçok disiplinde önemli bir yere sahip olan yaşam analizi, belirli bir olayın gerçekleştiği zamana kadar geçen sürenin (başarısızlık süresi veya hayatta kalma süresi) analizi konularını kapsamaktadır. Dolayısıyla, rassal değişken olarak bu sürenin iyi modellenmesi, yaşam süresi ve tehlike oranı tahminlerinin doğru yapılması açısından önemlidir. Bu tez çalışmasında, yaşam analizinin temel kavramları verilerek, $T-X$ metoduna dayalı, yaşam ve güvenilirlik alanında kullanılabilecek yeni genelleştirilmiş dağılımlar önerilmektedir. Bu dağılımların yoğunluk, yaşam, tehlike oranı, kantil ve moment fonksiyonları gibi istatistiksel özellikleri ve en çok olabilirlik tahmincileri de çalışılmıştır. Yeni dağılımların performansını göstermek için çeşitli uygulamalar yapılmış, önerilen dağılımların, mevcut temel dağılımlardan ve alternatif dağılımlardan, uyum iyiliği kriterlerine göre yaşam verisini modellemede daha iyi performans gösterdiği sonucuna ulaşılmıştır.

Anahtar Sözcükler: Yaşam analizi, Odd Burr-Rayleigh dağılımı, Odd Burr-Pareto dağılımı, Odd Burr-lojistik dağılımı, Karakterizasyon.

ABSTRACT

NEW DISTRIBUTION FAMILIES AND THEIR STATISTICAL PROPERTIES FOR SURVIVAL ANALYSIS

İbrahim ARIK

Department of Statistics

Anadolu University, Graduate School of Sciences, November 2018

Supervisor: Prof. Dr. Yeliz MERT KANTAR

Survival analysis, which has an important place in many disciplines, involves the analysis issues of elapsed time period until a particular event occurs (failure time, or survival time). Therefore, properly describing the distribution of a ‘survival time’ as random variable is important to get the accurate estimates on survival and hazard rate. In this thesis, by giving the basic concepts of survival analysis, we introduce new generalized distributions, which can be used in survival and reliability analysis, based on the $T-X$ method. We have studied the statistical properties of these distributions such as the density, hazard rate, survival, quantile, moment functions and maximum likelihood method for parameter estimation. Various real data applications are conducted to show the performance of the new distributions. It is concluded that according to goodness of fit tests, the newly introduced distributions outperform existing base distribution and alternative distributions in modelling survival data.

Keywords: Survival analysis, Odd Burr-Rayleigh distribution, Odd Burr-Pareto distribution, Odd Burr-logistic distribution, Characterization.

09/11/2018

ETİK İLKE VE KURALLARA UYGUNLUK BEYANNAMESİ

Bu tezin bana ait, özgün bir çalışma olduğunu; çalışmamın hazırlık, veri toplama, analiz ve bilgilerin sunumu olmak üzere tüm aşamalarında bilimsel etik ilke ve kurallara uygun davrandığımı; bu çalışma kapsamında elde edilen tüm veri ve bilgiler için kaynak gösterdiğimi ve bu kaynaklara kaynakçada yer verdiğimi; bu çalışmanın Anadolu Üniversitesi tarafından kullanılan “bilimsel intihal tespit programı”yla tarandığını ve hiçbir şekilde “intihal içermediğini” beyan ederim. Herhangi bir zamanda, çalışmamla ilgili yaptığım bu beyana aykırı bir durumun saptanması durumunda, ortaya çıkacak tüm ahlaki ve hukuki sonuçları kabul ettiğimi bildiririm.

İbrahim ARIK

TEŞEKKÜR

Lisansüstü eğitimi ve tezin hazırlanması sürecinde değerli bilgilerini ve tecrübesini benimle paylaşan, yol gösteren ve bu süreçte beni cesaretlendiren, danışman hocam Sayın Prof. Dr. Yeliz Mert KANTAR 'a,

Tez çalışması sürecinde görüş ve önerileriyle katkı sağlayan tez izleme kurulu üyeleri Prof. Dr. Serkan Ali DÜZCE ve Doç. Dr. İlhan USTA'ya,

Çalışma süresince destekleriyle yanımda olan, huzurlu bir çalışma ortamı sağlayan Prof. Dr. Berna YAZICI, Dr. Öğr. Üyesi Çiğdem KALATHİLPARMBİL ve Dr. Öğr. Üyesi Şükrü ACITAŞ'a,

Doğduğum günden bu yana bana yardım ve desteğini esirgemeyen aileme,

Sabırları ve özverileri ile her zaman yanımda olduklarını hissettiğim değerli eşim, meslektaşım Özlem ARIK ve oğlum Ahmet Erdem ARIK'a,

En içten teşekkürü bir borç bilirim.

İbrahim ARIK

Kasım 2018

İÇİNDEKİLER

Sayfa

BAŞLIK SAYFASI	i
JÜRİ VE ENSTİTÜ ONAYI.....	ii
ÖZET	iii
ABSTRACT.....	iv
ETİK İLKE VE KURALLARA UYGUNLUK BEYANNAMESİ.....	v
TEŞEKKÜR	vi
İÇİNDEKİLER.....	vii
TABLolar DİZİNİ.....	x
ŞEKİLLER DİZİNİ.....	xi
SİMGELER VE KISALTMALAR DİZİNİ.....	xii
1. GİRİŞ	1
2. YAŞAM ANALİZİNDE TEMEL KAVRAMLAR.....	5
2.1. Yaşam Verilerinin Özellikleri	5
2.2. Yaşam Fonksiyonu ve Tehlike Fonksiyonu	6
2.3. Yaşam Fonksiyonunun Tahmininde Kullanılan Parametrik Olmayan Yöntemler.....	7
2.3.1. Yaşam fonksiyonunun yaşam tablosu yöntemi ile tahmini	8
2.3.2. Yaşam fonksiyonunun Kaplan-Meier tahmini	9
2.4. Klasik Cox Regresyon.....	10
2.5. Parametrik Yaşam Modelleri (Hızlandırılmış Başarısızlık Zamanı Modeli).....	10
2.5.1. Hızlandırılmış başarısızlık zamanı modelinin log-lineer formda yazılması.....	12
2.5.2. Hızlandırılmış başarısızlık zamanı modelinin tahmini.....	13
2.5.3. Weibull hızlandırılmış başarısızlık süresi modeli.....	14
2.6. Parametrik Model İçin Artıklar	16
2.6.1. Standartlaştırılmış artıklar	16
2.6.2. Cox-Snell artıkları.....	17
2.6.3. Martingale artıkları	17
2.6.4. Deviance artıkları.....	17
2.6.5. Skor artıkları	18
2.6.5. Artıkların analizi	19
2.7. Model Seçim Kriterleri	19

3. DAĞILIM GENELLEŞTİRME YÖNTEMLERİ	21
3.1. Diferansiyel Denklemler Yöntemi	21
3.2. Dönüştürme Yöntemi.....	21
3.3. Kantil Fonksiyonu Metodu.....	22
3.4. Çarpık Dağılım Elde Etme Yöntemi	23
3.5. Beta'dan Üretme Yöntemi.....	24
3.6. Parametre Ekleyerek Dağılım Üretme Yöntemi	24
3.6.1 Üstelleştirme yöntemi (Exponentiated method)	24
3.6.2 Marshall-Olkin yöntemi.....	25
3.7. Composite Yöntemi.....	25
3.8. Transformed-Transformer Yöntemi (T-X Ailesi).....	26
4. YAŞAM ANALİZİ İÇİN YENİ DAĞILIM AİLELERİ.....	27
4.1. <i>OBu-R</i> Dağılımı ve İstatistiksel Özellikleri	29
4.1.1. <i>OBu-R</i> dağılım ve olasılık yoğunluk fonksiyonları.....	31
4.1.2. <i>OBu-R</i> yaşam ve tehlike oranı fonksiyonları	33
4.1.3. <i>OBu-R</i> 'nin asimptotik özellikleri	34
4.1.4. <i>OBu-R</i> 'nin kantil fonksiyonu	36
4.1.5. <i>OBu-R</i> 'nin alternatif matematiksel gösterimi	36
4.1.6. <i>OBu-R</i> 'nin momentleri.....	39
4.1.7. <i>OBu-R</i> 'nin koşullu momentleri	40
4.1.8. <i>OBu-R</i> 'nin sıra istatistiği	42
4.1.9. <i>OBu-R</i> 'nin en çok olabilirlik tahmincisi.....	43
4.1.10. <i>OBu-R</i> 'nin en çok olabilirlik tahmincisi için benzetim çalışması	46
4.1.11. <i>OBu-R</i> 'nin uygulamaları	48
4.2. <i>OBu-P</i> Dağılımı ve İstatistiksel Özellikleri	51
4.2.1. <i>OBu-P</i> dağılım ve olasılık yoğunluk fonksiyonları.....	52
4.2.2. <i>OBu-P</i> yaşam ve tehlike oranı fonksiyonları	54
4.2.3. <i>OBu-P</i> 'nin asimptotik özellikleri	56
4.2.4. <i>OBu-P</i> 'nin kantil fonksiyonu.....	57
4.2.5. <i>OBu-P</i> 'nin alternatif matematiksel gösterimi.....	58
4.2.6. <i>OBu-P</i> 'nin momentleri.....	58
4.2.7. <i>OBu-P</i> 'nin tamamlanmamış momentleri	61
4.2.8. <i>OBu-P</i> 'nin moment üreten fonksiyonu.....	61
4.2.9. <i>OBu-P</i> 'nin ortalamadan sapması.....	62

4.2.10. <i>OBu-P</i> 'nin sıra istatistiği.....	63
4.2.11. <i>OBu-P</i> 'nin en çok olabilirlik tahmincisi.....	64
4.2.12. <i>OBu-P</i> 'nin en çok olabilirlik tahmincisi için benzetim çalışması.....	67
4.2.13. <i>OBu-P</i> 'nin uygulamaları.....	69
4.3. <i>OBu-LL</i> Dağılımı ve İstatistiksel Özellikleri.....	73
4.3.1. <i>OBu-LL</i> dağılım ve olasılık yoğunluk fonksiyonları	74
4.3.2. <i>OBu-LL</i> yaşam ve tehlike oranı fonksiyonları	76
4.3.3. <i>OBu-LL</i> 'nin kantil fonksiyonu	77
4.3.4. <i>OBu-LL</i> 'nin alternatif matematiksel gösterimi	77
4.3.5. <i>OBu-LL</i> 'nin momentleri	78
4.3.6. <i>OBu-LL</i> 'nin en çok olabilirlik tahmincisi	79
4.3.7. <i>OBu-LL</i> regresyon modeli.....	80
4.3.8. <i>OBu-LL</i> 'nin uygulamaları	83
5. SONUÇ	88
KAYNAKÇA.....	91
ÖZGEÇMİŞ	

TABLÖLAR DİZİNİ

	<u>Sayfa</u>
Table 4.1. <i>OBu-R</i> dağılımı için çarpıklık ve basıklık değerleri	41
Tablo 4.2. <i>OBu-R</i> için MLE'nin Bias ve MSE değerleri	47
Tablo 4.3. Cam elyaf verisi	48
Tablo 4.4. Cam elyaf verisi için RD, GD, ERD, EWD ve <i>OBu-R</i> dağılımlarının parametre tahminleri ve kriterler sonuçları.....	48
Tablo 4.5. Rüzgar hızı verisi için RD, GD, ERD, EWD ve <i>OBu-R</i> dağılımlarının parametre tahminleri ve kriterleri sonuçları.....	50
Tablo 4.6. <i>OBu-P</i> dağılımı için ortalama ve varyans değerleri.....	60
Tablo 4.7. <i>OBu-P</i> dağılımı için S ve K değerleri	61
Tablo 4.8. <i>OBu-P</i> için MLE'nin Bias ve MSE değerleri	68
Tablo 4.9. Mesane Kanseri Verisi.....	69
Tablo 4.10. Mesane Kanseri verisi için PD, WPD, BPD, BEPD ve <i>OBu-P</i> dağılımının parametre tahminleri ve kriterleri sonuçları.....	70
Tablo 4.11. Taşkın verisi için PD, EPD, BPD, KwPD ve <i>OBu-P</i> dağılımlarının parametre tahminleri ve kriterler sonuçları.....	71
Tablo 4.12. Göğüs Kanseri Verisi.....	84
Tablo 4.13. Göğüs kanseri verisi için KLL, BLL ve <i>OBu-LL</i> dağılımlarının parametre tahmin ve karşılaştırma kriterleri sonuçları	84
Tablo 4.14. Voltaj Verisi.....	85
Tablo 4.15. Voltaj Verisi verisi için <i>OBu-L</i> , Lojistik ve Log-Weibull dağılımının parametre tahminleri ve kriter sonuçları	85
Tablo 4.16. Olabilirlik oran testi sonuçları (Voltaj verisi).....	86

ŞEKİLLER DİZİNİ

Sayfa

Şekil 4.1. Farklı parametre değerleri için Rayleigh dağılımının oyf grafiği	30
Şekil 4.2. Farklı parametre değerleri için <i>OBu-R</i> dağılımının oyf grafiği.....	32
Şekil 4.3. Farklı parametre değerleri için <i>OBu-R</i> dağılımının tehlike oranı fonksiyonu grafiği.....	34
Şekil 4.4. Olasılık yoğunluk fonksiyonları ve histogram grafiği (Cam elyaf verisi)	49
Şekil 4.5. Olasılık yoğunluk fonksiyonları ve histogram grafiği (Rüzgar hızı verisi) ...	50
Şekil 4.6. Farklı parametre değerleri için <i>OBu-P</i> dağılımının oyf grafiği.....	53
Şekil 4.7. Farklı parametre değerleri için <i>OBu-P</i> dağılımının tehlike oranı fonksiyonu grafiği.....	56
Şekil 4.8. Olasılık yoğunluk fonksiyonları ve histogram grafiği (Mesane kanseri).....	70
Şekil 4.9. Olasılık yoğunluk fonksiyonları ve histogram grafiği (Taşkın verileri)	72
Şekil 4.10. Farklı parametre değerleri için <i>OBu-LL</i> dağılımının oyf grafiği.....	75
Şekil 4.11. Farklı parametre değerleri için <i>OBu-LL</i> dağılımının tehlike oranı fonksiyonu grafiği.....	76
Şekil 4.12. Deviance tipi artıkların Q-Q plot grafiği.....	87

SİMGELER VE KISALTMALAR DİZİNİ

α	: Alfa
β	: Beta
δ	: Delta
ε	: Epsilon
η	: Eta
γ	: Gamma
ξ	: Ksi
λ	: Lamda
μ	: Mu
ϕ	: Phi
σ	: Sigma
θ	: Teta
$f(.)$	: Olasılık yoğunluk fonksiyonu
$F(.)$	: Dağılım fonksiyonu
$h(.)$	: Tehlike oranı fonksiyonu
$H(.)$	: Kümülatif tehlike fonksiyonu
$S(.)$	: Yaşam fonksiyonu
$\Gamma(.,.)$	: Gamma fonksiyonu
AIC	: Akaike bilgi kriteri
AFT	: Hızlandırılmış başarısızlık zamanı (Accelated failure time)
BIC	: Bayesci bilgi kriteri
Bias	: Yanlılık
CAIC	: Düzeltilmiş Akaike bilgi kriteri
cdf	: Dağılım fonksiyonu (cumulative distribution function)
EÇO	: En çok olabilirlik
HQIC	: Hannan Quinn bilgi kriteri
K	: Çarpıklık (Kurtosis)
hrf	: Tehlike oranı fonksiyonu (hazard rate function)

MSE	: Hata kareler ortalaması
$OBu-R$	: Odd Burr Rayleigh
$OBu-P$	: Odd Burr Pareto
$OBu-LL$	: Odd Burr log-lojistik
$OBu-L$	: Odd Burr lojistik
oyf	: Olasılık yoğunluk fonksiyonu
S	: Basıklık (Skewness)



1. GİRİŞ

Yaşam çözümlemesi literatürde yaygın olarak çalışılan alanların başında gelmektedir. Yaşam çözümlemesinde ilgilenilen asıl olgu, ele alınan olayın içerisindeki birimlerin ya da bireylerin sağ kalım süreleridir. Bu sebeple olayların farklı olması nedeniyle, farklı isimler altında tıp, mühendislik, ekonomi ve sosyal bilimler gibi alanlarda yaşam çözümlemesi yaygın uygulama alanına sahiptir. Örneğin sosyal bilimlerde doğum, evlilik, iş yerlerinin çalışma süreleri gibi birçok farklı olay türleri ile ilgilenilirken bu alanda yaşam analizi, olay geçmişi analizi olarak adlandırılmaktadır. Öte yandan mühendislikte bir makinanın ya da bir arabanın herhangi bir parçasının arıza yapma süresi olay olarak ele alınmakta, böylece yaşam çözümlemesi, güvenilirlik analizi adı ile ifade edilmektedir (Balay, 2014; Kleinbaum ve Klein, 2010).

Yaşam süresinin çözümlemesine ilişkin ilk çalışmalar yaşam tablosu yöntemi ile başlamıştır. Daha sonra yaşam fonksiyonunun tahmini için Kaplan Meier yöntemi (Kaplan ve Meier, 1958) geliştirilmiştir. İki yaşam fonksiyonunun karşılaştırılması için log-rank testinin önerilmesi bu alandaki diğer gelişmelerdendir. Bu döneme kadar yapılan çalışmalarda sadece farklı gruptaki birey ya da birimlerin yaşam fonksiyonları karşılaştırılabiliyordu, bireye ya da birime etki edebilecek diğer faktörlerin yaşam süresine olan etkileri dikkate alınamıyordu. Cox'un 1972 yılında önerdiği klasik Cox regresyon modeli ile yaşam süresi üzerine etkisi olan farklı etkenlerinde dikkate alınabilmesi sağlanmıştır.

Yaşam çözümlemesi literatüründe büyük öneme sahip bir diğer konu da yaşam süresinin modellenmesi konusudur. Weibull, log-lojistik, gamma log-normal gibi dağılımlar yaşam süresinin modellenmesinde kullanılan klasik istatistiksel dağılımlardandır. Fakat farklı alanlarda çok farklı özelliklere sahip yaşam verilerinin bulunması klasik istatistiksel dağılımların yetersiz kaldığı kanaatini oluşturmuştur. Son yıllarda özellikle bilgisayarlar gibi hesaplama araçlarının ve programlarının gelişmesinin de etkisiyle modelleme başarısı daha yüksek olabilecek başka bir ifade ile daha esnek dağılımlar elde etme yönünde çalışmalar oldukça ilgi görmektedir. Bu amaçla farklı türdeki dağılım genişletme yöntemlerinden yararlanılarak daha fazla parametreye sahip esnek dağılımlar yaşam süresinin modellenmesinde kullanılmak üzere önerilmiştir. Bu alanda literatürde yapılan çalışmaların bir kısmı aşağıda özetlenmiştir:

Lee ve ark. (2007), çalışmalarında sansürlü yaşam süresini modellemek için yeni beta-Weibull dağılımını çalışmışlar, bu dağılımın modelleme başarısını göstermek için sansürlü gözlemler içeren otobüs motorlarının bozulma süreleri ve çene kanseri verilerini kullanmışlardır.

Senhuesa ve ark. (2008), çalışmalarında yaşam süresini modellemek için Birnbaum-Saunders dağılımını genelleştirmişlerdir.

Cordeiro ve ark. (2010), Kumaraswamy-Weibull dağılımını önermişler ve önerilen yeni dağılımın başarısını bozulma sürelerini modelledikleri gerçek veri seti uygulamasında göstermişlerdir.

Nadarajah ve ark. (2011), çalışmalarında genelleştirilmiş Lindley dağılımını önermişler ve yeni önerdikleri dağılımın yaşam süresini modellemedeki başarısını göstermek amacıyla gerçek veri seti üzerine bir uygulama yapmışlar ve bu uygulamada önerdikleri dağılımın bu literatürde yaygın olarak kullanılan klasik istatistiksel dağılımlardan olan Weibull, gamma ve log-normal dağılımlarından daha başarılı olduğu sonucuna varmışlardır.

De Pascoa ve ark. (2011), çalışmalarında yeni Kumaraswamy-genelleştirilmiş gamma dağılımını sunmuşlardır.

Lemonte ve Cordeiro (2013), yaşam süresini modellemede kullanılabilecek beş parametrelili McDonald-Lomax dağılımını önermişlerdir.

Bourguignon ve ark. (2014), dağılım genişletme yöntemi olarak Weibull-G ailesini önerdikleri çalışmalarında, Weibull-üstel ve Weibull-Burr XII dağılımlarını önermişler ve bu dağılımların ömür sürelerini modellemedeki başarılarını birer gerçek veri seti üzerinde göstermişlerdir.

Cordeiro ve ark. (2015), çalışmalarında gamma-Lomax dağılımını önermişlerdir ve elde edilen yeni dağılımın modelleme başarısını kanser verileri üzerinde göstermişlerdir.

Saboor ve Pagony (2016), Marshall-Olkin gamma Weibull dağılımını önermişler ve bu dağılımın modelleme başarısını yaşam verileri içeren gerçek veri seti uygulamalarıyla göstermişlerdir.

Tahir ve ark. (2016), çalışmalarında Weibull-Pareto dağılımını ele almışlar ve bu dağılımı yaşam süresini modellemek için kullanmışlardır.

Alizadeh ve ark. (2017), çalışmalarında yaşam süresi dağılımı olarak dört parametrelili üstelleştirilmiş power Lindley geometrik dağılımını önermişler ve bu dağılımın modelleme başarısını gerçek veri seti uygulamasıyla göstermişlerdir.

Altun ve ark. (2017), yaşam süresi verilerini modellemede kullanılabilecek, Odd Burr Lindley dağılımını önermiş ve dağılımın özelliklerini incelemişlerdir.

Yaşam sürelerinin üzerine farklı açıklayıcı değişkenlerin etkilerini incelemeye olanak sağlayan klasik Cox regresyon modelinin (Cox, 1972) geliştirilmesinin ardından, parametrik regresyon modelleri de yaşam çözümlemesi alanında son yıllarda yoğun olarak çalışılan konuların başında gelmektedir. Bu amaçla Weibull, log-normal, log-lojistik gibi istatistiksel dağılımlar, regresyon modelinde başarısızlık süresini modellemede kullanılan yaygın modellerdir. Bu amaçla farklı yapıdaki yaşam sürelerini dahi iyi modelleyebilmek için yine dağılım genişletme yöntemleri aracılığıyla yeni dağılımlar önerilmiş ve bu dağılımlar üzerinden yeni parametrik regresyon modelleri elde edilmiş olup bu modellerin klasik dağılımlardan daha başarılı oldukları üzerine literatürde birçok çalışma yapılmıştır. Bu alanda literatürde yapılan çalışmaların bir kısmı aşağıdaki gibi özetlenebilir:

Silva ve ark. (2008), sansürlü yaşam sürelerinin de bulunduğu verilerin üzerinde ilgili açıklayıcı değişkenlerin etkilerini inceleyebilmek için log-Burr XII regresyon modelini önermişlerdir.

Ortega ve ark. (2011), beta-Weibull dağılımı için log-lineer regresyon modelini yaşam verileri için uygulamışlardır.

Ortega ve ark. (2012), beta-Birnbaum-Saunders dağılımı için log-lineer regresyon modelini ele almışlardır.

Cordeiro ve ark. (2013), beta üstelleştirilmiş Weibull dağılımını önermiş oldukları çalışmalarında bu dağılımı temel alan regresyon modelini de bazı parametrik regresyon modelleri ile karşılaştırmışlardır.

Lemonte (2013), yaşam süresini modellemek amacıyla genişletilmiş Birnbaum-Saunders regresyon modelini önermiştir.

Cordeiro ve ark. (2014), dağılım genelleştirme yöntemi olarak Lomax-G ailesini önerdikleri çalışmalarında bu aileye dayalı log-Lomax-Weibull regresyon modelini kurmuşlardır.

Gomes ve ark. (2014), çalışmalarında Kumaraswamy genelleştirilmiş Rayleigh dağılımını önererek bu dağılımın yaşam süresi için regresyon modelini elde etmişlerdir.

Pescim ve ark. (2017), çalışmalarında yaşam süresi verilerini modellemek için odd log-lojistik genelleştirilmiş yarı-normal dağılımını temel alan regresyon modelini önermişlerdir.

Bu tez çalışmasında yukarıda da ifade edildiği gibi yaşam analizinde sıklıkla ele alınan konulardan olan yaşam süresinin modellenmesine ve log-yaşam süresi regresyon modeline katkı sağlaması amacıyla dağılım genelleştirme yöntemlerinden olan T-X dağılımlar ailesi (Alzaatreh ve ark., 2013) yöntemini temel alan Odd-Burr-G (Alizadeh ve ark., 2016) metodu kullanılarak üç farklı dağılım genelleştirilmiştir.

Öncelikle Rayleigh dağılımının temel dağılım olarak kullanılmış olup, Odd-Burr Rayleigh dağılımı elde edilmiştir. Dağılımın bir dizi istatistiksel özellikleri ele alınmış ve yaşam süresini modellemedeki başarısını ortaya koyabilmek amacıyla gerçek veri seti üzerindeki modelleme performansı incelenmiştir. Ayrıca Rayleigh dağılımı rüzgâr hızı analizi literatüründe de önemli bir yere sahiptir. Temel alınan dağılımın literatürde önem arz ettiği alandaki başarısını da gösterebilmek amacıyla elde edilen yeni dağılımın performansı rüzgâr hızı verileri üzerinde de incelenmiştir.

Daha sonra daha çok taşkın verileri gibi ekstrem verileri modellemede kullanılan Pareto dağılımı temel dağılım olarak alınmış ve Odd Burr Pareto dağılımı elde edilmiştir. Bu dağılımın istatistiksel özelliklerinin incelenmesinin yanı sıra bazı dağılımlar ile olan ilişkileri de ele alınmıştır. Yine yaşam süresi verisini modelleme performansını incelemek amacıyla yaşam verilerini içeren gerçek veri seti üzerinde bir uygulamanın yanı sıra elde edilen yeni dağılımın, temel alınan dağılımın sıklıkla kullanıldığı alanlardaki başarısını da inceleyebilmek için nehir taşkın verilerini ele alarak bir uygulama yapılmıştır.

Son olarak yaşam analizinde de önem arz eden klasik istatistiksel dağılımlardan olan log-lojistik dağılımı temel alınarak Odd Burr Log-Lojistik dağılımı önerilmiştir. Elde edilen bu yeni dağılımın yaşam süresini modellemedeki performansının incelenmesinin yanı sıra dağılımı temel alan log-doğrusal regresyon modeli de elde edilmiştir. Elde edilen bu yeni regresyon modelinin performansını incelemek amacıyla sansürlü gözlemlerin de olduğu gerçek veri seti üzerinde uygulama yapılmıştır.

Yukarıda ifade edilen konular dikkate alınarak tezin bölümleri beş ana başlıkta verilmiştir. İkinci bölümde, yaşam çözümlemesi konusuna ait temel kavramlar özetle ele alınmıştır. Üçüncü bölümde, literatürde kullanılan dağılım genelleştirme yöntemleri anlatılmıştır. Dördüncü bölümde, bu tez çalışması kapsamında genelleştirilen yaşam dağılımları, istatistiksel özellikleri ve gerçek veri uygulamaları ele alınmıştır. Sonuç ve önerilerin bulunduğu beşinci bölüm ile tez çalışması tamamlanmıştır.

2. YAŞAM ANALİZİNDE TEMEL KAVRAMLAR

Yaşam analizi, ilgilenilen belirli bir olayın, belirli bir başlangıç zamanından itibaren gerçekleşmesine kadar geçen süreyi veri olarak kullanıp analizini yapan teknikleri kapsamaktadır. Bu bölümde yaşam analizi ile ilgili temel kavramlar ele alınacaktır.

2.1. Yaşam Verilerinin Özellikleri

Yaşam verilerinin, standart verilerin çözümlenmesinde kullanılan istatistiksel tekniklerle analiz edilememesinin sebebi iki başlıkta ele alınabilir. Bu sebeplerden ilki; yaşam verileri genellikle simetrik dağılmaz. Aynı özellikteki bireylere ait yaşam verilerinin histogramı çizilip incelendiğinde sağa (pozitif) çarpık bir yapı göze çarpar. Bu da dağılımın sağda daha uzun kuyruklu olduğunun göstergesidir. Bu dezavantaj veride dönüşümler yaparak (logaritmik) verinin daha simetrik hale getirilmesi ile giderilebilir. Fakat daha uygun olan çözüm, orijinal veriyi alternatif daha uygun dağılımla modelleyerek çözümlenmesidir. Diğer bir sebep ise yaşam verilerinin sansürlenmiş veriler içermesidir. Bir birey için yaşam süresinin bitiş noktası, ilgilenilen olgu için bilinmiyorsa, bu yaşam süresi sansürlenmiş olarak adlandırılır. Bu durum ilgilenilen olgunun gözlem süresi boyunca gerçekleşmemesi ya da bireyin çalışmadan ayrılmasından kaynaklanabilir. Bir birey için gözlemlenen yaşam süresi, şayet bireyin ölümü gözlemlenen olgu ile ilgili değilse de sansürlenmiş veri olarak değerlendirilir. Örneğin akciğer kanserinden tedavi gören hastanın ölümünün kalp krizi sebebiyle olması, o hastanın akciğer kanseri gözlemi için sansürlenmiş veri olarak değerlendirilmesini gerektirir (Collett, 2003).

t_0 zamanında tedavisine başlanan bir hastanın tam olarak bilinmeyen $t_0 + t$ anında öldüğü varsayalım. Eğer hastanın en son $t_0 + c$ zamanında yaşıyor olduğu biliniyorsa, c süresi sansürlü yaşam süresi olarak adlandırılır. Buradaki sansürleme bireyin çalışmaya dâhil olmasından sonra gerçekleşir. Yani bilinen son yaşam süresinin sağında olan bir zamanla ilgilidir. Dolayısıyla bu veri tipi sağdan sansürlenmiş olarak adlandırılır. Sağdan sansürlenmiş yaşam süresi bilinmeyen gerçek yaşam süresinden daha azdır (Collett, 2003; Kleinbaum ve Klein, 2010).

Sansürlemenin bir diğer türü de soldan sansürlemedir ki gerçek yaşam süresi bilinen yaşam süresinden daha azdır. Örneğin tümörleri alınan bir hastanın üç ay sonraki

kontrolünde tekrar tümör gözlenmiş ise bu hasta için kanserin nüksetmesi üç aydan daha azdır. Yani soldan sansürlenmiş yaşam süresidir (Collett, 2003).

Bir diğer sansürleme türü olan aralıklı sansürleme, olgunun gerçekleştiği yaşam aralığının bilindiği durumdur. Yani kanserden tedavi edilen bir hastanın altı ay sonraki kontrolünde tümöre rastlanmamış, on iki ay sonraki kontrolünde ise tekrar tümör bulunduğu tespit edilmiştir. Yani kanserin nüksetme zamanı 6 ay – 12 ay aralığında olmuştur. Bu gözlemlenen nüksetme zamanı aralıklı sansürlenmiştir (Collett, 2003; Kleinbaum ve Klein, 2010).

Sansürlenmiş verilerin analizinde yapılan önemli bir varsayım; bireyin gerçek yaşam süresi olan t 'nin, bireyin c zamanında ($c < t$) sansürlenmiş veri olmasına neden olan herhangi bir sansürleme mekanizmasından bağımsız olmasıdır (non-informative censoring) (Collett, 2003).

2.2. Yaşam Fonksiyonu ve Tehlike Fonksiyonu

T negatif değerler almayan, yaşam süresi için tanımlı bir rassal değişken olmak üzere, $f(t)$, T 'nin olasılık yoğunluk fonksiyonu olsun. T 'nin dağılım fonksiyonu,

$$F(t) = P(T < t) = \int_0^t f(u) du \quad (2.1)$$

şeklindedir. Bu fonksiyon yaşam süresinin t 'den daha az olması olasılığını ifade eder.

$S(t)$ ile gösterilen yaşam fonksiyonu ise yaşam süresinin t 'den daha fazla olması olasılığını verir ve (2.2)'deki gibi ifade edilir:

$$S(t) = P(T \geq t) = 1 - F(t) \quad (2.2)$$

$h(t)$ ile gösterilen tehlike fonksiyonu t anındaki ölüm riskini verir ve (2.3) denklemiyle ifade edilir:

$$h(t) = \lim_{\varepsilon \rightarrow 0} \left\{ \frac{P(t \leq T < t + \varepsilon | T \geq t)}{\varepsilon} \right\} \quad (2.3)$$

Tehlike fonksiyonu, olasılık yoğunluk ve yaşam fonksiyonu yardımıyla (2.4) ve (2.5) eşitlikleri ile ifade edilebilir:

$$h(t) = \frac{f(t)}{S(t)} \quad (2.4)$$

$$h(t) = \frac{f(t)}{S(t)} = \frac{f(t)}{[1-F(t)]} = -\frac{d}{dt} \{\log [1-F(t)]\} = -\frac{d}{dt} \{\log S(t)\}. \quad (2.5)$$

(2.5) eşitliğinden yararlanılarak yaşam fonksiyonu,

$$S(t) = \exp\{-H(t)\} \quad (2.6)$$

olarak yazılabilir. Burada $H(t)$,

$$H(t) = \int_0^t h(u) du \quad (2.7)$$

olarak ifade edilir ve kümülatif tehlike fonksiyonu olarak adlandırılır. (2.6) eşitliğinden yararlanarak kümülatif tehlike fonksiyonu,

$$H(t) = -\log S(t) \quad (2.8)$$

şeklinde de ifade edilebilir (Collett, 2003).

2.3. Yaşam Fonksiyonunun Tahmininde Kullanılan Parametrik Olmayan Yöntemler

Empirik yaşam fonksiyonu $\hat{S}(t)$,

$$\hat{S}(t) = \frac{\text{Yaşam süresi } \geq t \text{ olan birey sayısı}}{\text{Veri setindeki tüm bireylerin sayısı}} \quad (2.9)$$

yardımla hesaplanabilir.

Yaşam fonksiyonunun değeri henüz ölümün gerçekleşmediği zamanda 1'e eşit olup, son ölüm zamanından sonra ise 0'a eşittir. $\hat{S}(t)$ iki ölüm zamanı arasında sabit olup $\hat{S}(t)$ - t grafiği basamak fonksiyonu şeklindedir (Collett, 2003; Kleinbaum ve Klein, 2010).

(2.9) eşitliği ile yapılan yaşam fonksiyonu tahmini veri setinde durdurulmuş gözlem değeri olduğunda kullanılamaz. Durdurulmuş verilerin bulunduğu durumlarda $\hat{S}(t)$ 'nin hesabında kullanılabilecek parametrik olmayan yöntemler izleyen bölümlerde açıklanmıştır (Collett, 2003).

2.3.1. Yaşam fonksiyonunun yaşam tablosu yöntemi ile tahmini

Bu yöntem ile öncelikle gözlemin yapıldığı zaman periyodu aralıklara ayrılır. Bu aralıkların uzunluklarının eşit olması zorunluluğunun olmamasının yanında genellikle eşit olacak şekilde ayarlanır. Aralık sayısı çalışmada yer alan birey sayısına bağlı olarak belirlenir ve genellikle 5-15 olarak alınır (Collett, 2003).

Gözlem süresi m adet aralığa ayrılınsın ve (t'_j, t'_{j+1}) , $j. (j=1,2,...,m)$ aralık olmak üzere, d_j ve c_j sırasıyla bu aralıkta gerçekleşen ölüm sayısını ve durdurulmuş zaman sayısını ifade etsin. n_j 'de $j.$ aralığın başında yaşayan birey sayısı, yani risk grubunda olan bireylerin sayısı olsun. Durdurma işleminin $j.$ aralık boyunca durdurulmuş yaşam sürelerinin düzgün bir şekilde dağılacığı şekilde yapıldığı varsayımı altında, bu aralıktaki riskte olan bireylerin ortalama sayısı,

$$n'_j = n_j - c_j / 2 \quad (2.10)$$

olarak hesaplanır. Bu varsayım aktüeryal varsayım olarak ifade edilir (Collett, 2003). j 'inci aralıktaki ölüm olasılığı d_j / n'_j olarak tahmin edilebilir. Dolayısıyla bu aralıktaki yaşam olasılığı ise $(n'_j - d_j) / n'_j$ şeklindedir. Şimdi herhangi bir t'_k ($k=1,2,...,m$) zamanından sonraki, bir an için bireyin yaşam olasılığı ele alınsın. Burada t'_k herhangi bir zaman aralığının başlangıç anıdır. Bu olasılığın yaşam tablosu yöntemi tahmini,

$$S^*(t) = \prod_{j=1}^k \left(\frac{n'_j - d_j}{n'_j} \right), \quad t'_k \leq t \leq t'_{k+1}, \quad k = 1, 2, \dots, m \quad (2.11)$$

yardımıyla hesaplanır. Yaşam fonksiyonunun tahmini ilk zaman aralığının başlangıç anı olan t'_1 'den önce bire eşit olup t'_{m+1} 'den sonra ise sıfıra eşittir. Bu yöntemle tahmin edilen yaşam fonksiyonunun grafiği de basamak fonksiyonu şeklinde olup, her bir zaman aralığında sabit değerler alır (Collett, 2003).

Yaşam tablosu yöntemi tahmin aşamasında oluşturulan aralıkların seçiminden etkilenir. Ayrıca bu yöntemin, ölüm zamanlarının tam olarak bilinmediği, sadece ardışık zaman aralıklarında meydana gelen ölüm sayısının ve durdurulmuş zaman sayısının bilindiği durumlarda kullanımı daha uygundur. Gerçek yaşam zamanlarının bilindiği

durumlarda bu yöntemin kullanılması, yaşam sürelerinin gruplanması sebebiyle bilgi kaybına neden olur. Bu durum özellikle birim sayısının 30'dan küçük olduğu örneklerde ortaya çıkar (Collett, 2003).

2.3.2. Yaşam fonksiyonunun Kaplan-Meier tahmini

Yaşam fonksiyonunun Kaplan-Meier tahminini elde etmek için yaşam-tablosu yönteminde olduğu gibi zaman aralıkları oluşturulmaktadır. Yalnız burada her bir zaman aralığı bir tek ölüm zamanı içermekte olup aralıkların başlangıç anları bu ölüm zamanlarıdır.

$t_1, t_2, \dots, t_n$ yaşam sürelerinden oluşan n büyüklüğündeki veri seti olsun. Bu verilerden bazılarının sağdan durdurulmuş olabileceği ve aynı süreye sahip birden çok birey olabileceği varsayalım. Bireyler içerisinde $r \leq n$ olmak üzere, r tane ölüm sayısı olsun ve bu ölüm süreleri $t_{(1)} < t_{(2)} < \dots < t_{(r)}$ şeklinde sıralansın. $t_{(j)}$ j 'inci ölüm zamanı olmak üzere, n_j bu zamandan hemen önce yaşayan bireylerin sayısı ($t_{(j)}$ zamanında ölen bireyler dâhil) ve d_j de $t_{(j)}$ zamanında ölen bireylerin sayısı olsun. Ölüm zamanları ve durdurulmuş zamanların çakıştığı örneklemelerle karşılaşılabilir. Bu durumda durdurulmuş zamanın ölüm zamanından hemen sonra gerçekleşmiş olduğu kabul edilir ve n_j hesap edilir (Collett, 2003; Kleinbaum ve Klein, 2010).

Örneklemde bulunan bireylerin ölüm zamanlarının bağımsız olduğu varsayımı altında t zamanı için yaşam fonksiyonunun Kaplan-Meier tahmini,

$$\hat{S}(t) = \prod_{j=1}^k \left(\frac{n_j - d_j}{n_j} \right), \quad t_{(k)} \leq t \leq t_{(k+1)}, \quad k = 1, 2, \dots, r \quad (2.12)$$

formülü yardımıyla hesaplanır. Burada $t_{(r+1)}$ sonsuz olarak alınır. Eğer örneklemdeki en büyük değer durdurulmuş yaşam süresi (t^*) ise, $t > t^*$ için $\hat{S}(t)$ tanımsızdır. Öte yandan en büyük gözlem değeri ölüm zamanı ise bu gözlemde daha büyük bir zaman için $\hat{S}(t)$ sıfıra eşittir. Kaplan-Meier tahmincisi de yaşam-tablosu tahmincisi gibi tahmin edilmiş olasılıkların çarpımlarından oluşur. Aslında Kaplan-Meier tahmincisi yaşam-tablosu tahmincisinin, zaman aralıkları sayısının sonsuza gittiği ve zaman aralıkları uzunluğunun sıfıra gittiği limit durumudur (Collett, 2003; Kleinbaum ve Klein, 2010).

2.4. Klasik Cox Regresyon

Önceki bölümlerde ele alınan parametrik olmayan metotlar tek bir yaşam verisi örnekleminin analizinde ya da iki veya daha fazla grubun yaşam sürelerinin karşılaştırılmasında kullanılabilir. Fakat yaşam sürelerini etkileyen farklı etkenler dikkate alınamaz. Yaşam süresini etkileyen açıklayıcı değişkenlerin belirlenip modele dâhil edildiği orantılı tehlikeler modeli olarak adlandırılan model Cox (1972) tarafından önerilmiştir. Cox tarafından önerildiği için klasik Cox regresyon modeli olarak da adlandırılmaktadır. Model orantılı tehlikeler varsayımına dayanmasına rağmen, yaşam süreleri için herhangi bir dağılımsal varsayım gerektirmemektedir. Dolayısıyla klasik Cox regresyon modeli yarı parametrik model olarak adlandırılmaktadır (Cox, 1972; Collett, 2003; Ata, 2010).

$X_1, X_2, \dots, X_p$, tehlike fonksiyonunu etkileyen p tane açıklayıcı değişken olsun. $h_0(t)$, tüm açıklayıcı değişken değerlerinin sıfır olduğu bireyin temel tehlike fonksiyonu olmak üzere i . birey için klasik Cox regresyon modeli,

$$h_i(t) = \exp(\beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_p x_{pi}) h_0(t) \quad (2.13)$$

şeklinde ifade edilir (Cox, 1972; Collet, 2003; Ata, 2010).

2.5. Parametrik Yaşam Modelleri (Hızlandırılmış Başarısızlık Zamanı Modeli)

Klasik Cox regresyon modeli kullanıldığında yaşam verilerinin herhangi bir olasılık dağılımına uygunluğu hakkında bir varsayıma ihtiyaç bulunmamaktadır. Dolayısıyla temel tehlike fonksiyonu herhangi bir dağılıma göre hesaplanmamaktadır. Bu sebeple klasik Cox regresyon modeli, parametrik regresyon modellerine göre daha esnek ve yaygın kullanıma sahiptir. Diğer yandan, eğer yaşam verileri herhangi bir dağılıma uygunluk gösteriyor ise bu dağılıma dayalı modelden yapılan çıkarımlar daha kesin olmaktadır. Özellikle göreceli tehlikeler ve yaşam süresinin medyanı değerlerinin tahminleri herhangi bir parametrik varsayımın olmadığı duruma göre daha küçük standart hata değerine sahip olur (Collett, 2003).

Varsayalım ki aynı hastalık sebebiyle tedavi gören hastalar rasgele iki gruba ayrılsın. Gruplardan birine standart tedavi yöntemi diğerine ise yeni geliştirilen tedavi yöntemi uygulansın. Standart tedavi uygulanan grubun yaşam fonksiyonu $S_s(t)$ ve yeni

yöntemin uygulandığı grubun yaşam fonksiyonu $S_N(t)$ olsun. Hızlandırılmış başarısızlık zamanı modeli,

$$S_N(t) = S_S(t/\phi)$$

olarak ifade edilir. Burada ϕ bilinmeyen bir sabit olmak üzere ϕ^{-1} hızlandırma faktörü olarak adlandırılır. Tehlike fonksiyonu ile yaşam fonksiyonu arasındaki ilişkiden yararlanarak yukarıda verilen eşitlik,

$$h_N(t) = \phi^{-1} h_S(t/\phi)$$

biçiminde yazılabilir. X_i ,

$$x_i = \begin{cases} 1, & \text{yeni tedavi (N)} \\ 0, & \text{standart tedavi (S)} \end{cases}$$

değerlerini alan gösterge değişkeni olmak üzere, i . birey için tehlike fonksiyonu,

$$h_i(t) = \phi^{-x_i} h_0(t/\phi^{x_i}) \quad (2.14)$$

şeklinde ifade edilebilir. $\phi > 0$ olduğundan $\phi = e^\alpha$ olarak alınabilir. Dolayısıyla (2.14) eşitliğindeki hızlandırılmış başarısızlık zamanı modeli,

$$h_i(t) = e^{-\alpha x_i} h_0(t/e^{\alpha x_i}) \quad (2.15)$$

olarak elde edilir. Bu eşitlik p tane açıklayıcı değişkenin bulunduğu durum için genelleştirilebilir:

$$h_i(t) = e^{-\eta_i} h_0(t/e^{\eta_i}) \quad (2.16)$$

Buradaki η_i ,

$$\eta_i = \alpha_1 x_{1i} + \alpha_2 x_{2i} + \dots + \alpha_p x_{pi}$$

olup modelin lineer bileşenidir. Bu modelde i . bireyin yaşam fonksiyonu,

$$S_i(t) = S_0\{t/\exp(\eta_i)\}$$

şeklinde ifade edilebilir (Collett, 2003; Kleinbaum ve Klein, 2010).

2.5.1. Hızlandırılmış başarısızlık zamanı modelinin log-lineer formda yazılması

i. bireyin yaşam süresine ilişkin T rassal değişkeni aşağıdaki gibi lineer formda ifade edilsin:

$$\log T_i = \mu + \alpha_1 x_{1i} + \alpha_2 x_{2i} + \dots + \alpha_p x_{pi} + \sigma \varepsilon_i \quad (2.17)$$

Bu modeldeki $\alpha_1, \alpha_2, \dots, \alpha_p$ açıklayıcı değişkenlerin bilinmeyen katsayıları, modelin hata dağılımının μ ve σ sırasıyla konum ve ölçek parametrelerini ve ε_i hataları ifade etmektedir. Bu modelde hataların belirli bir dağılımdan gelen rassal değişken olduğu varsayılır. $\alpha_1, \alpha_2, \dots, \alpha_p$ parametreleri açıklayıcı değişkenlerin yaşam süresine olan etkilerini yansıtır. Örneğin eğer belirli bir açıklayıcı değişkene ait katsayı pozitif ise o açıklayıcı değişken değerinin artmasının yaşam süresini arttırdığı manasına gelir.

(2.16) eşitliği ile (2.17)'in ilişkisini göstermek için *i.* bireye ait t zamanındaki yaşam fonksiyonu (2.17)'ten yararlanarak,

$$\begin{aligned} S_i(t) &= P(T_i \geq t) = P\left\{\exp(\mu + \alpha_1 x_{1i} + \dots + \alpha_p x_{pi} + \sigma \varepsilon_i) \geq t\right\} \\ S_i(t) &= P\left\{\exp(\mu + \sigma \varepsilon_i) \geq t / \exp(\alpha_1 x_{1i} + \dots + \alpha_p x_{pi})\right\} \end{aligned}$$

şeklinde ifade edilebilir. Tüm açıklayıcı değişkenlerin sıfır değerini aldığı temel yaşam fonksiyonu $S_0(t)$,

$$S_0(t) = P\left\{\exp(\mu + \sigma \varepsilon_i) \geq t\right\}$$

şeklindedir. Bu eşitlikten yararlanarak *i.* bireye ait yaşam fonksiyonu,

$$S_i(t) = S_0\left\{t / \exp(\alpha_1 x_{1i} + \dots + \alpha_p x_{pi})\right\} \quad (2.18)$$

olarak ifade edilebilir. Bu eşitlik hızlandırılmış başarısızlık zamanı modelindeki *i.* bireyin yaşam fonksiyonunun genel formudur. Burada *i.* birey için hızlandırma faktörü $\exp(-\alpha' x_i)$ 'dir. (2.18) eşitliğinde her iki tarafın logaritması alınıp, -1 ile çarpılıp t 'ye göre türevi alınırsa,

$$h_i(t) = \exp(-\alpha' x_i) h_0\left\{t / \exp(\alpha' x_i)\right\}$$

eşitliği elde edilir ve bu da (2.16) ile verilen modele karşılık gelir.

i. birey için yaşam fonksiyonu,

$$S_i(t) = P(T_i \geq t) = P(\log T_i \geq \log t)$$

şeklinde ifade edilebilir. (2.17)'ten yararlanarak,

$$\begin{aligned} S_i(t) &= P(\mu + \alpha_1 x_{1i} + \alpha_2 x_{2i} + \dots + \alpha_p x_{pi} + \sigma \varepsilon_i \geq \log t) \\ &= P\left(\varepsilon_i \geq \frac{\log t - \mu - \alpha_1 x_{1i} - \alpha_2 x_{2i} - \dots - \alpha_p x_{pi}}{\sigma}\right) \end{aligned} \quad (2.19)$$

olarak yazılabilir. $S_{\varepsilon_i}(t)$ i . hata rassal değişkeninin yaşam fonksiyonu olmak üzere, i . bireyin yaşam fonksiyonu,

$$S_i(t) = S_{\varepsilon_i}\left(\frac{\log t - \mu - \alpha_1 x_{1i} - \alpha_2 x_{2i} - \dots - \alpha_p x_{pi}}{\sigma}\right) \quad (2.20)$$

biçiminde ifade edilebilir. Bu sonuç T_i rassal değişkenine ait yaşam fonksiyonunun, ε_i hata terimine ait yaşam fonksiyonundan bulunabileceğini gösterir (Collett, 2003).

2.5.2. Hızlandırılmış başarısızlık zamanı modelinin tahmini

Hızlandırılmış başarısızlık zamanı modelinin parametreleri en çok olabilirlik yöntemi kullanılarak tahmin edilir. Bu amaçla kullanılacak olabilirlik fonksiyonu modelin lineer formundan elde edilir. $t_1, t_2, \dots, t_n$ gözlemlerine ait olabilirlik fonksiyonu,

$$L = \prod_{i=1}^n \{f_i(t_i)\}^{\delta_i} \{S_i(t_i)\}^{1-\delta_i}$$

şeklindedir. Burada $f_i(t_i)$ ve $S_i(t_i)$ i . bireyin t_i anındaki olasılık ve yaşam fonksiyonları, δ_i ise,

$$\delta_i = \begin{cases} 1, & \text{başarısız} \\ 0, & \text{sansürlenmiş} \end{cases}$$

değerlerini alan gösterge değişkenidir. (2.20)'den yararlanarak,

$$S_i(t) = S_{\varepsilon_i}(z_i)$$

olarak yazılabilir. Burada $z_i = (\log t - \mu - \alpha_1 x_{1i} - \alpha_2 x_{2i} - \dots - \alpha_p x_{pi}) / \sigma$ şeklindedir. Her iki tarafın t 'ye göre türevi alındığında,

$$f_i(t_i) = \frac{1}{\sigma t_i} f_{\varepsilon_i}(z_i)$$

biçiminde elde edilir. Bu sonuç yardımıyla olabilirlik fonksiyonu,

$$L = \prod_{i=1}^n (\sigma t_i)^{-\delta_i} \{f_{\varepsilon_i}(z_i)\}^{\delta_i} \{S_{\varepsilon_i}(z_i)\}^{1-\delta_i}$$

şeklinde olur. Sonuç olarak modelin olabilirlik fonksiyonunun logaritması,

$$\log L = \sum_{i=1}^n \left\{ -\delta_i \log(\sigma t_i) + \delta_i \log f_{\varepsilon_i}(z_i) + (1-\delta_i) S_{\varepsilon_i}(z_i) \right\} \quad (2.21)$$

olarak elde edilir (Collett, 2003; Lee ve Wang, 2003). Modelde bulunan $(p+2)$ adet bilinmeyen parametrenin tahminlerine, $\log L$ 'in maksimize edilmesiyle ulaşılır. Bu amaçla Newton-Raphson veya Nelder-Mead yöntemi kullanılabilir.

2.5.3. Weibull hızlandırılmış başarısızlık süresi modeli

Yaşam sürelerinin, ölçek parametresi λ ve şekil parametresi γ olan Weibull $(W(\lambda, \gamma))$ dağılımına sahip olduğu varsayalım. Bu durumda temel tehlike fonksiyonu,

$$h_0(t) = \lambda \gamma t^{\gamma-1} \quad (2.22)$$

biçiminde olur. i . bireyin tehlike fonksiyonu,

$$h_i(t) = e^{-\eta_i} \lambda \gamma (e^{-\eta_i} t)^{\gamma-1} = (e^{-\eta_i})^\gamma \lambda \gamma t^{\gamma-1} \quad (2.23)$$

şeklinde ifade edilir. Weibull dağılımı için genel hızlandırılmış başarısızlık süresi modeli,

$$\begin{aligned} h_i(t) &= e^{-\eta_i} h_0(t / e^{\eta_i}) \\ h_i(t) &= e^{-\eta_i} \lambda \gamma (e^{-\eta_i} t)^{\gamma-1} = (e^{-\eta_i})^\gamma \lambda \gamma t^{\gamma-1} \end{aligned} \quad (2.24)$$

olarak elde edilir. Dolayısıyla i . bireyin yaşam süresi $W(\lambda e^{-\gamma \eta_i}, \gamma)$ dağılımına sahip olur. Sonuç olarak Weibull dağılımı hızlandırılmış başarısızlık süresi özelliğini sağlar (Collett, 2003).

Modelin log-lineer gösterimi ele alınsın. T_i Weibull dağılımına sahip olduğunda ε_i Gumbel dağılımına sahip olur. Bu dağılım asimetrik bir dağılımdır ve yaşam fonksiyonu aşağıda ifade edildiği gibidir:

$$S_{\varepsilon_i}(t) = \exp(-e^t), \quad -\infty < t < \infty \quad (2.25)$$

Gumbel dağılımının kümülatif tehlike ve tehlike fonksiyonları sırasıyla aşağıdaki gibidir:

$$\begin{aligned} H_{\varepsilon_i}(t) &= e^t \\ h_{\varepsilon_i}(t) &= e^t \end{aligned} \quad (2.26)$$

$T_i = \exp(\mu + \alpha_1 x_{1i} + \dots + \alpha_p x_{pi} + \sigma \varepsilon_i)$ rassal değişkeninin Weibull dağıldığını göstermek için (2.20) eşitliğinden yararlanarak aşağıdaki gibi ifade edilebilir:

$$S_i(t) = \exp \left\{ -\exp \left(\frac{\log t - \mu - \alpha_1 x_{1i} - \alpha_2 x_{2i} - \dots - \alpha_p x_{pi}}{\sigma} \right) \right\} \quad (2.27)$$

Yukarıda verilen eşitlik,

$$S_i(t) = \exp(-\lambda_i t^{1/\sigma})$$

biçiminde de ifade edilebilir. Buradaki λ_i ,

$$\lambda_i = \exp \left\{ -(\mu + \alpha_1 x_{1i} + \dots + \alpha_p x_{pi}) / \sigma \right\}$$

şeklinde. Dolayısıyla yukarıda ifade edilen yaşam fonksiyonu ölçek parametresi λ şekil parametresi σ^{-1} olan Weibull dağılımına ait yaşam fonksiyonudur. Sonuç olarak (2.27) eşitliği Weibull yaşam fonksiyonunun hızlandırılmış başarısızlık süresi formunda gösterimdir.

Weibull hızlandırılmış başarısızlık zamanı modelinin kümülatif tehlike fonksiyonu yaşam fonksiyonu ile olan ilişki yardımıyla aşağıdaki gibi ifade edilebilir:

$$H_i(t) = -\log S_i(t) = \exp \left(\frac{\log t - \mu - \alpha_1 x_{1i} - \alpha_2 x_{2i} - \dots - \alpha_p x_{pi}}{\sigma} \right) \quad (2.28)$$

Bu eşitlik $H_i(t) = \lambda_i t^{1/\sigma}$ biçiminde de ifade edilebilir. Tehlike fonksiyonu ise,

$$h_i(t) = \frac{1}{\sigma t} \exp \left(\frac{\log t - \mu - \alpha_1 x_{1i} - \alpha_2 x_{2i} - \dots - \alpha_p x_{pi}}{\sigma} \right) \quad (2.29)$$

biçimindedir. Bu eşitlik $h_i(t) = \lambda_i \sigma^{-1} t^{\frac{1}{\sigma}-1}$ şeklinde de yazılabilir (Collett, 2003).

Hızlandırılmış başarısızlık zamanı modeli Weibull dağılımının yanında üstel, log-lojistik, log-normal, gamma gibi dağılımlar için de elde edilebilir.

2.6. Parametrik Model İçin Artıklar

T_i , $i = 1, 2, \dots, n$, i 'inci bireyin yaşam süresini ifade eden rassal değişken ve i 'inci bireye ait açıklayıcı değişken değerleri olmak üzere hızlandırılmış başarısızlık zamanı modeli log-lineer formda aşağıdaki gibi ifade edilir:

$$\log T_i = \mu + \alpha_1 x_{1i} + \alpha_2 x_{2i} + \dots + \alpha_p x_{pi} + \sigma \varepsilon_i.$$

Burada ε_i rassal değişken olup dağılımı T_i 'nin dağılımı ile ilişkilidir. Bilindiği üzere μ, σ ve α_j , $j = 1, 2, \dots, p$ bilinmeyen parametrelerdir. Eğer i 'inci bireye ait yaşam süresi sansürlü ise o bireyin artık değeri de sansürlüdür. Parametrik modelin analizinde kullanılan artık türleri başlıklar halinde izleyen bölümlerde açıklanmıştır.

2.6.1. Standartlaştırılmış artıklar

Hızlandırılmış başarısızlık zamanı modeli için elde edilen standartlaştırılmış artıklar,

$$r_{s_i} = \{ \log t_i - \hat{\mu} - \hat{\alpha}_1 x_{1i} + \hat{\alpha}_2 x_{2i} + \dots + \hat{\alpha}_p x_{pi} \} / \hat{\sigma} \quad (2.30)$$

şeklinde tanımlanır. Burada t_i , i 'inci bireyin gözlemlenmiş yaşam süresi, $\hat{\mu}$, $\hat{\sigma}$ ve $\hat{\alpha}_j$ tahmin edilmiş parametre değerleridir. Eğer model doğru kurulmuş ise standartlaştırılmış artık değerlerinin, hızlandırılmış başarısızlık zamanı modelindeki ε_i 'ler ile aynı dağılıma sahip olması beklenir. Örneğin; T_i , Weibull dağılımına sahip olsun, eğer kurulan model doğru ise standartlaştırılmış artıkların Gumbel dağılımına sahip olmaları beklenir. Artıkların tahmin edilen yaşam fonksiyonu ile ε_i 'nin yaşam fonksiyonu olan $S_{\varepsilon_i}(\varepsilon)$ 'nin benzer olması beklenir. Bilindiği üzere $-\log(S_{\varepsilon_i}(\varepsilon))$ $\lambda = 1$ ile üstel dağılıma sahiptir. Dolayısıyla eğer model doğru kurulmuşsa $-\log(S_{\varepsilon_i}(r_{s_i}))$ yaklaşık olarak 1 parametrelili üstel dağılıma sahip olmalıdır. Bu husus model uygunluğunun kontrol edilmesinde kullanılan grafiklerin temelini oluşturur (Collett, 2003).

2.6.2. Cox-Snell artıkları

Cox-Snell artığı olarak ifade edilen artıklar esas olarak ilgili bireyin tahmin edilmiş olan kümülatif tehlike fonksiyonuna eşittir. Hızlandırılmış başarısızlık süresi modelinde i 'inci bireye ait yaşam fonksiyonu,

$$S_i(t) = S_{\varepsilon_i} \left(\frac{\log t_i - \mu - \alpha_1 x_{1i} - \alpha_2 x_{2i} - \dots - \alpha_p x_{pi}}{\hat{\sigma}} \right)$$

eşitliği yardımıyla elde edilir. Dolayısıyla parametrik model için Cox-Snell artığı,

$$r_{C_i} = \hat{H}_i(t_i) = -\log \hat{S}(t_i) \quad (2.31)$$

şeklinde ifade edilir. Bilindiği üzere $\hat{H}_i(t_i)$, i 'inci bireye ait t_i zamanı için tahmin edilen kümülatif tehlike fonksiyonudur (Collett, 2003; Kleinbaum ve Klein, 2010).

$S(T)$ 'nin yapısına bağlı kalmaksızın $Y = -\log S(T)$ rassal değişkeni 1 parametrelili üstel dağılıma sahip olduğu için eğer kurulan model doğru ise Cox-Snell artıkları 1 parametrelili üstel dağılıma sahip olmalıdır.

2.6.3. Martingale artıkları

Martingale artıkları $(0, t_i)$ aralığında gözlemlenmiş ölüm sayısı ile model aracılığıyla tahmin edilmiş ölüm sayısı arasındaki farkları vermektedir ve aşağıdaki gibi tanımlanır:

$$r_{M_i} = \delta_i - r_{C_i} = \delta_i + \log \hat{S}_i(t_i) \quad (2.32)$$

Bilindiği üzere eşitlikteki δ_i değeri eğer i 'inci birey gözlemlenmiş ise 1 sansürlenmiş ise 0 değerini alır (Collett, 2003).

2.6.4. Deviance artıkları

Martingale artıkları çarpık bir dağılıma sahip oldukları için Therneau (1990) bu artıkları simetrik hale getirmek için deviance artıklarını aşağıdaki gibi tanımlamıştır:

$$r_{D_i} = \text{sign}(r_{M_i}) \left[-2 \left\{ r_{M_i} + \delta_i \log(\delta_i - r_{M_i}) \right\} \right]^{1/2}, i = 1, \dots, n. \quad (2.33)$$

2.6.5. Skor artıkları

Skor artıkları, log-olabilirlik fonksiyonunun bilinmeyen parametrelere göre kısmi türevlerini oluşturan bileşenlere karşılık gelir. N birimlik bir gözlem için hızlandırılmış başarısızlık modelinin olabilirlik fonksiyonu:

$$\log L = \sum_{i=1}^n \left\{ -\delta_i \log(\sigma t_i) + \delta_i \log f_{\varepsilon_i}(z_i) + (1 - \delta_i) S_{\varepsilon_i}(z_i) \right\}$$

şeklinde. Burada $z_i = (\log t - \mu - \alpha_1 x_{1i} - \alpha_2 x_{2i} - \dots - \alpha_p x_{pi}) / \sigma$ olarak ifade edilir. Log olabilirlik fonksiyonunun parametrelere göre kısmi türevleri,

$$\begin{aligned} \frac{\partial \log L}{\partial \mu} &= \sigma^{-1} \sum_{i=1}^n g(z_i), \\ \frac{\partial \log L}{\partial \sigma} &= \sigma^{-1} \sum_{i=1}^n \{ z_i g(z_i) \delta_i \}, \\ \frac{\partial \log L}{\partial \alpha_j} &= \sigma^{-1} \sum_{i=1}^n \{ x_{ji} g(z_i) \}, \end{aligned}$$

şeklinde olup burada $g(z_i)$,

$$g(z_i) = \frac{(1 - \delta_i) f_{\varepsilon_i}(z_i)}{S_{\varepsilon_i}(z_i)} - \frac{\delta_i f'_{\varepsilon_i}(z_i)}{f_{\varepsilon_i}(z_i)}$$

olarak ifade edilir. En çok olabilirlik tahmini için hesaplanan bu kısmi türev değerlerinin her bir i 'inci bileşeni ilgili terime karşılık gelen skor artığını ifade eder. Sonuç olarak standartlaştırılmış artıkların tanımından da yararlanarak μ için i 'inci skor artığı;

$$\hat{\sigma}^{-1} g(r_{s_i}) \quad (2.34)$$

ölçek parametresi σ için i 'inci skor artığı;

$$\hat{\sigma}^{-1} \{ r_{s_i} g(r_{s_i}) - \delta_i \} \quad (2.35)$$

son olarak j 'inci açıklayıcı değişken olan x_j için i 'inci skor artığı;

$$\hat{\sigma}^{-1} x_{ji} g(r_{s_i}) \quad (2.36)$$

şeklinde ifade edilir (Collett, 2003; Lawless, 2003).

2.6.5. Artıkların analizi

Modelin uygunluğu için yapılan artık analizinde en sık kullanılan grafik Cox-Snell artıklarının grafiğidir. Bu amaçla $-\log \hat{S}(r_{C_i}) - r_{C_i}$ grafiği oluşturulur. Eğer 1 eğimli ve 0'dan geçen doğru şeklinde grafik elde edilmişse kurulan modelin uygun olduğu sonucuna varılır.

Martingale ya da deviance artıkları ile elde edilen grafikler aykırı değerlerin tespitinde kullanılır. Martingale ya da deviance artıklarının sıralı yaşam sürelerine ya da açıklayıcı değişkenlere karşı çizilen grafikler yardımıyla modele uygun olmayan yaşam sürelerinin varlığı ya da açıklayıcı değişkenlerin varlığı kontrol edilebilir. Aynı şekilde açıklayıcı değişkenlere ait skor artıklarının, sıralı yaşam sürelerine karşı elde edilen grafikler yardımıyla veri setinde aykırı değerlerin olup olmadığının tespiti mümkün olabilir (Collett, 2003).

2.7. Model Seçim Kriterleri

Yaşam verisinin modellenmesinde kullanılabilecek en uygun modele karar verebilmek için çeşitli kriterler kullanılmaktadır. Bu amaçla Akaike bilgi kriteri (*AIC*) ve Bayesci bilgi kriteri (*BIC*) sıklıkla kullanılan iki kriterdir.

Akaike bilgi kriteri,

$$AIC = -2\log L + 2q \quad (2.37)$$

şeklinde olup bu eşitlikteki q modeldeki bilinmeyen parametre sayısını ifade eder. Bu istatistik değerinin daha düşük olduğu model daha iyi model anlamına gelir (Collett, 2003; Ata, 2010).

Bayesci bilgi kriteri,

$$BIC = -2\log L + q \log n \quad (2.38)$$

şeklinde olup bu eşitlikteki n toplam gözlem sayısını ifade etmektedir (Collett, 2003; Ata, 2010). Aynı şekilde *BIC* kriterinin değerinin düşük olması da modelin daha iyi olduğunu ifade eder.

Literatürde, *AIC* ve *BIC*'nin yanı sıra tutarlı Akaike bilgi kriteri (*CAIC*) ve Hannan Quinn bilgi kriteri (*HQIC*) kullanılmakta olup formülleri sırasıyla ifade edilmiştir (Tahir ve ark, 2016; Hannan ve Quinn, 1979).

$$CAIC = -2\log L + \frac{2qn}{n-q-1} \quad (2.39)$$

$$HQIC = -2\log L + 2q \ln(\ln(n)) \quad (2.40)$$

AIC ve *BIC*'de olduğu gibi *CAIC* ve *HQIC* değerlerinin düşük olması da modelin daha iyi olduğunu ifade eder.



3. DAĞILIM GENELLEŞTİRME YÖNTEMLERİ

Bu bölümde literatürde geliştirilmiş olan dağılım üretme yöntemlerine kısaca değinilecektir.

3.1. Diferansiyel Denklem Yöntemi

Pearson (1895) tarafından önerilen yöntem ile istatistiksel dağılımlar elde etmek için diferansiyel denklem kullanılır. Bu sistem, simetrik olmayan verileri modellemek amacıyla geliştirilmiştir. (3.1) eşitliği ile verilen denklem sistemini sağlayan fonksiyonlar Pearson sistemi tarafından üretilen yoğunluk fonksiyonlarını ifade eder:

$$\frac{1}{f(x)} \frac{df(x)}{dx} = \frac{a+x}{b_0+b_1x+b_2x^2} \quad (3.1)$$

Burada a, b_0, b_1, b_2 parametreler olup, $f(.)$ fonksiyonunun şekli bu parametrelere bağlıdır. Farklı tipteki dağılımlar yukarıdaki eşitliğin farklı formdaki çözümlerine karşılık gelir (Pearson, 1895; Lee ve ark, 2013).

Diferansiyel denklem yöntemine dayalı farklı bir metot, Burr (1942) tarafından önerilmiş olup (3.2) eşitliğindeki diferansiyel denklemi sağlayan dağılım sistemlerinden oluşmaktadır:

$$dF = F(1-F)g(x)dx \quad (3.2)$$

Burada $0 \leq F \leq 1$ ve $g(x)$, x üzerinde tanımlı negatif olmayan fonksiyonlardır. Verilen eşitliğin 12 adet çözümü, dağılım fonksiyonu formunda Burr tarafından elde edilmiş olup Burr I-XII dağılımları olarak adlandırılmaktadır (Burr, 1942; Lee ve ark., 2013).

3.2. Dönüştürme Yöntemi

Johnson (1949) tarafından dağılım geliştirmek amacıyla önerilen bu yöntemde (3.3) eşitliği ile verilen normalleştirme dönüşümü kullanılmaktadır:

$$Z = \gamma + \delta f\left(\frac{X - \xi}{\lambda}\right) \quad (3.3)$$

Burada $f(\cdot)$ dönüşüm fonksiyonu, Z standartlaştırılmış normal rassal değişken, γ, δ, λ ve ξ parametrelerdir. Bu yöntem, dönüştürme yöntemi olarak bilinir. δ ve λ parametrelerinin pozitif olduğu varsayılır.

Johnson (1949) üç adet dönüştürme fonksiyonu önermiştir:

i) Log-Normal aile (S_L):

$$Z = \gamma + \delta \ln(X - \xi), \quad X \geq \xi \quad (3.4)$$

ii) Sınırlı dağılımlar ailesi (S_B):

$$Z = \gamma + \delta \ln\left(\frac{X - \xi}{\xi + \lambda - X}\right), \quad \xi \leq X \leq \xi + \lambda \quad (3.5)$$

iii) Sınırsız dağılımlar ailesi (S_U):

$$\begin{aligned} Z &= \gamma + \delta \ln\left\{\left(\frac{X - \xi}{\lambda}\right) + \left[\left(\frac{X - \xi}{\lambda}\right)^2 + 1\right]^{1/2}\right\}, \quad -\infty \leq X \leq \infty \\ &= \gamma + \delta \sinh^{-1}\left(\frac{X - \xi}{\lambda}\right). \end{aligned} \quad (3.6)$$

İfade edilen bu dağılım aileleri normal, log-normal, gamma, beta, üstel dağılım gibi literatürde sıklıkla kullanılan dağılımları kapsamaktadır.

Literatürde sıklıkla kullanılan Birnbaum-Saunders dağılımı, Johnson ailesinin bir üyesidir. Birnbaum ve Saunders (1969) tarafından geliştirilen bu dağılım normalleştirme dönüşümü denkleminde $\gamma = \xi = 0$, $f(y) = \sqrt{y} - \sqrt{1/y}$ ve $X = Y / \lambda$ alınarak türetilmiştir. Dolayısıyla Birnbaum-Saunders rassal değişkeni X ile dönüştürülen Z arasındaki ilişki $Z = \delta(\sqrt{X / \lambda} - \sqrt{\lambda / X})$ şeklindedir (Lee ve ark., 2013; Johnson, 1949).

3.3. Kantil Fonksiyonu Metodu

Bu yöntem ile üretilen dağılımlar ailesi kantil fonksiyonuna dayalı olarak (3.7)'deki şekilde tanımlanır:

$$Q(y) = Q(y; \lambda_1, \lambda_2, \lambda_3, \lambda_4) = \lambda_1 + \frac{y^{\lambda_3} - (1-y)^{\lambda_4}}{\lambda_2}, \quad 0 \leq y \leq 1 \quad (3.7)$$

λ_1 ve λ_2 parametreleri sırasıyla konum ve ölçek parametreleridir. λ_3 çarpıklığı, λ_4 ise basıklığı kontrol eder. Verilen eşitliğe karşılık gelen olasılık yoğunluk fonksiyonu (3.8)'deki gibidir:

$$f(x) = \frac{\lambda_2}{\lambda_3 y^{\lambda_3-1} + \lambda_4 (1-y)^{\lambda_4-1}}, \quad x = Q(y) \quad (3.8)$$

Bu yöntem üzerine yapılan ilk çalışmalar lamda dağılımını elde etmeye yöneliktir. Daha sonraki yıllarda bu dağılım, Ramberg ve Schmeiser (1974) ve Ramberg ve ark. (1979) tarafından geliştirilmiş lamda dağılımları olarak adlandırılmıştır (Lee ve ark., 2013).

3.4. Çarpık Dağılım Elde Etme Yöntemi

İki simetrik dağılımın bileşiminden çarpık dağılım elde etme yöntemi ilk olarak Azzalini (1985) tarafından önerilmiştir. Çarpık normal dağılım ailesini de ilk tanıtan Azzalini'dir. Çarpık normal aile şu şekilde açıklanır: X rassal değişkeni 0 noktasında simetrik olasılık yoğunluk fonksiyonuna sahip olsun. Y ise $G'(\cdot)$ simetrik olan, mutlak sürekli $G(\cdot)$ dağılım fonksiyonuna sahip olsun. Sonuç olarak herhangi bir $\lambda \in \mathbb{R}$ için,

$$0,5 = P(Y - \lambda X < 0) = E_x \left[P(Y < \lambda X | X = x) \right] = \int_{-\infty}^{\infty} f_0(x) G(\lambda x) dx \quad (3.9)$$

sağlanır. Bu nedenle,

$$2f_0(x)G(\lambda x), \quad -\infty < x < \infty \quad (3.10)$$

geçerli bir olasılık yoğunluk fonksiyonunu ifade eder.

Z rassal değişkeni standart normal dağılıma sahip olsun ve Z_λ rassal değişkeninin olasılık yoğunluk fonksiyonu (3.11)'deki gibi tanımlansın:

$$f_{Z_\lambda}(z, \lambda) = 2\varphi(z)\Phi(\lambda z), \quad z \in \mathbb{R}, \lambda \in \mathbb{R} \quad (3.11)$$

Burada $\varphi(z)$ ve $\Phi(z)$ standart normal dağılımın olasılık yoğunluk ve dağılım fonksiyonlarıdır. Dolayısıyla Z_λ rassal değişkeni çarpık normal rassal değişkenidir ve $Z \sim SN(\lambda)$ ile gösterilir. Burada λ çarpıklık parametresidir (Azzalini, 1985; Lee ve ark., 2013).

3.5. Beta'dan Üretim Yöntemi

Eugene ve ark. (2002) tarafından önerilen bu yöntemde beta dağılımının olasılık yoğunluk fonksiyonu üretici fonksiyon olarak kullanılmaktadır. Beta sınıfına dâhil olan X rassal değişkeninin dağılım fonksiyonu (3.12)'deki gibi tanımlanır:

$$G(x) = \int_0^{F(x)} b(t) dt \quad (3.12)$$

Burada $b(t)$ beta dağılımının olasılık yoğunluk fonksiyonu ve $F(x)$ ise herhangi bir rassal değişkenin dağılım fonksiyonudur. Yukarıda ifade edilen bu eşitlik yardımıyla üretilen beta sınıfına dâhil rassal değişkenin olasılık yoğunluk fonksiyonu (3.13)'teki şekilde elde edilir:

$$g(x) = \frac{1}{B(\alpha, \beta)} f(x) F^{\alpha-1}(x) (1-F(x))^{\beta-1} \quad (3.13)$$

Bu yöntem yeni dağılım üretmek için kolay bir yol sunmaktadır. Literatürde bu yöntem kullanılarak üretilen beta-normal, beta-Gumbel, beta-üstel, beta-Weibull, beta-Pareto gibi birçok dağılım bulunmaktadır (Eugene, 2002; Lee ve ark., 2013).

3.6. Parametre Ekleyerek Dağılım Üretim Yöntemi

Hali hazırdaki bir dağılıma parametre ekleyerek “genelleştirilmiş” dağılım elde etmek sıklıkla kullanılan yöntemlerdendir. Weibull, Pareto, gamma, lojistik gibi birçok dağılımın bu yöntem kullanılarak elde edilen genelleştirilmiş tipleri bulunmaktadır. Parametre ekleyerek dağılım elde etmek için kullanılan çeşitli metotlar olmasına rağmen en çok bilinenleri aşağıda kısaca verilen üstelleştirme ve Marshall-Olkin yöntemidir.

3.6.1 Üstelleştirme yöntemi (Exponentiated method)

Mudholkar ve Srivastava (1993) bu yöntemi kullanarak üstelleştirilmiş Weibull dağılımını (3.14)'teki gibi ifade etmişlerdir:

$$F(x) = \{1 - \exp[-(\alpha x)^\beta]\}^\gamma, x > 0, \alpha, \beta, \gamma > 0. \quad (3.14)$$

Açıkça görülmektedir ki $\gamma=1$ olduğunda Weibull dağılımı elde edilmektedir. Genel olarak üstelleştirme yöntemi F ve F_E sırasıyla X rassal değişkeni ve üstelleştirilmiş X rassal değişkeninin dağılım fonksiyonları olmak üzere,

$$F_E = (F)^\gamma, \gamma > 0 \quad (3.15)$$

eşitliği yardımıyla elde edilir (Mudholkar ve Srivastava, 1993; Lee ve ark., 2013).

3.6.2 Marshall-Olkin yöntemi

Mevcut bir dağılıma parametre ekleyerek yeni bir dağılım elde etmek amacıyla, Marshall-Olkin (1997) tarafından yaşam fonksiyonuna dayalı olarak önerilen yöntem (3.16)'daki gibidir:

$$\bar{G}(x; \alpha) = \frac{\alpha S(x)}{1 - \bar{\alpha} S(x)} = \frac{\alpha S(x)}{F(x) + \alpha S(x)}, \quad -\infty < x < \infty, \quad 0 < \alpha < \infty \quad (3.16)$$

Burada $S(x)$ herhangi bir dağılımın yaşam fonksiyonu, α dağılıma eklenen yeni parametre ve $\bar{\alpha} = 1 - \alpha$ 'dır (Marshall ve Olkin, 1997; Lee ve ark., 2013).

3.7. Composite Yöntemi

Cooray ve Ananda (2005) tarafından önerilen bu yöntem ile iki dağılımın (3.17)'de ifade edildiği biçimde birleştirilmesiyle yeni bir dağılım elde edilir:

$$f(x) = \begin{cases} cf_1(x), & x \in (0, \theta] \\ cf_2(x), & x \in [\theta, \infty) \end{cases} \quad (3.17)$$

Burada c normalleştirme sabiti olup, f_1 ve f_2 pozitif tanımlı olasılık yoğunluk fonksiyonlarıdır. Bilinmeyen θ sabiti öyle seçilmelidir ki elde edilen $f(x)$ yoğunluk fonksiyonu o noktada sürekli ve türevlenebilir olsun. (3.18) ile ifade edilen kısıt dikkate alınarak belirtilen noktada süreklilik ve türevlenebilirlik sağlanır:

$$f_1(\theta) = f_2(\theta), \quad f_1'(\theta) = f_2'(\theta) \quad (3.18)$$

f_1 ile ifade edilen dağılım n parametrelidir, f_2 ile ifade edilen dağılım m parametrelidir olmak üzere yeni elde edilen f dağılımının en fazla $n+m-1$ adet parametresi olabilir (Cooray ve Ananda, 2005; Lee ve ark., 2013).

3.8. Transformed-Transformer Yöntemi (T-X Ailesi)

Beta'dan üretme yönteminde kullanılan üretici dağılımların $(0,1)$ aralığında tanımlı olması, diğer sürekli dağılımların üretici dağılım olarak kullanılabiliyor kullanılamayacağı sorusunu akla getirmiştir. Alzaatreh (2013) bu amaçla, betadan üretim yöntemini tanım aralığı sadece $(0,1)$ olmayan sürekli dağılımlar için de geliştirmiştir (Lee ve ark. 2013).

$T \in [a,b]$, $-\infty < a < b < \infty$, olmak üzere $r(t)$, T rassal değişkeninin olasılık yoğunluk fonksiyonu olsun. $W(F(x))$ herhangi bir X rassal değişkeninin dağılım fonksiyonu $F(x)$ 'in bir fonksiyonu olmak üzere aşağıdaki özellikleri sağlamaktadır:

- i. $W(F(x)) \in [a,b]$,
- ii. $W(F(x))$ türevlenebilirdir ve monoton azalmayandır,
- iii. $x \rightarrow -\infty$, $W(F(x)) \rightarrow a$ ve $x \rightarrow \infty$, $W(F(x)) \rightarrow b$

Yeni elde edilen dağılım ailesinin dağılım fonksiyonu (3.19)'daki gibi tanımlanır:

$$G(x) = \int_a^{W(F(x))} r(t) dt = R(W(F(X))). \quad (3.19)$$

Verilen bu eşitlikte $R(t)$, T rassal değişkeninin dağılım fonksiyonudur. Bu eşitliğe karşılık gelen olasılık yoğunluk fonksiyonu:

$$g(x) = \left\{ \frac{d}{dx} W(F(x)) \right\} r\{W(F(x))\} \quad (3.20)$$

Bu dağılımlar ailesi “dönüştürülen-dönüştürücü (transformed-transformer)” aile olarak adlandırılır. Burada $r(t)$ olasılık yoğunluk fonksiyonu, $W(F(x))$ dönüştürücü fonksiyonu aracılığıyla, $G(x)$ dağılım fonksiyonuna dönüştürülmüştür. X rassal değişkeni kesikli olabilir, bu durumda $G(x)$, kesikli dağılımlar ailesinin dağılım fonksiyonu olur (Alzaatreh ve ark., 2013).

$W(F(x))$ fonksiyonunun tanımı T rassal değişkeninin tanım aralığına bağlıdır. Dikkate alınan farklı $W(F(x))$ fonksiyonları, farklı dağılım ailelerini oluşturur. T 'nin tanım aralığına bağlı olarak $W(F(x))$ için bazı örnekler aşağıdaki gibi verilebilir:

a. Eğer T 'nin tanım aralığı sınırlı ise, örneğin T 'nin tanım aralığı $[0,1]$ olsun. Bu tip dağılımlar düzgün $(0,1)$, beta, Kumaraswamy ve diğer genelleştirilmiş beta dağılımları olabilir. Bu durumda $W(F(x))$, $F(x)$ ya da $F^\alpha(x)$ olarak tanımlanabilir. Böylelikle son dönemde üzerinde sıklıkla çalışılan betadan üretilmiş dağılımlar ailesi elde edilir.

b. $T \in [a, \infty)$, $a \geq 0$ olsun. $a=0$ alındığında $W(F(x))$;

$$\text{i. } -\log[1-F(x)]$$

$$\text{ii. } \frac{F(x)}{1-F(x)}$$

$$\text{iii. } -\log[1-F^\alpha(x)], \alpha > 0$$

$$\text{iv. } \frac{F^\alpha(x)}{1-F^\alpha(x)}, \alpha > 0$$

olarak tanımlanabilir.

c. $T \in (-\infty, \infty)$ olsun. Bu durumda $W(F(x))$:

$$\text{i. } \log\{-\log[1-F(x)]\}$$

$$\text{ii. } \log\left\{\frac{F(x)}{1-F(x)}\right\}$$

$$\text{iii. } \log\{-\log[1-F^\alpha(x)]\}, \alpha > 0$$

$$\text{iv. } \log\left\{\frac{F^\alpha(x)}{1-F^\alpha(x)}\right\}, \alpha > 0$$

şeklinde ifade edilebilir (Alzaatreh ve ark., 2013).

4. YAŞAM ANALİZİ İÇİN YENİ DAĞILIM AİLELERİ

İstatistiksel dağılımlar, istatistiksel olguları tanımlamak, modellemek ve tahmin etmek için çok yararlıdır. Bu sebeple, mühendislik, ekonomi, sosyal ve yaşam bilimleri gibi çeşitli alanlarda gerçekleşmiş olayları modellemek ve tanımlamak için istatistiksel dağılımlar yaygın bir şekilde kullanılmaktadır (Johnson ve ark., 1994). Literatürde, bilinen istatistiksel dağılımlardan daha esnek olan çok sayıda yeni dağılım önerilmiş ve farklı veri türlerini modellemede kullanılmıştır. Bölüm 3'te bahsedildiği gibi, son zamanlarda yeni dağılım ailelerinin tanımlanma süreci genellikle temel dağılımların genişletilmesi ile yapılmaktadır. Bu amaçla, literatürde var olan temel dağılıma bir veya daha fazla parametre ekleyerek yeni ve esnek dağılımlar üretmek için çeşitli yaklaşımlar ve yöntemler önerilmiştir. Bunlardan sıklıkla kullanılanlar aşağıdaki gibi listelenebilir:

- i. Marshall ve Olkin (1997), bir dağılım ailesine yeni bir parametre eklemek için yeni yöntem geliştirmişlerdir.
- ii. Cordeiro ve de Castro (2011), dağılım fonksiyonu G olan her hangi bir sürekli dağılımı temel dağılım olarak alan ve iki yeni şekil parametresi eklenmesine olanak sağlayan dağılım genelleştirme yöntemi önermişlerdir (Kumaraswamy-G).
- iii. Eugene ve ark. (2002), beta dağılımının üretici dağılım olarak kullanıldığı yeni dağılım üretme yöntemi önermişlerdir.
- iv. Alexander ve ark. (2012), Eugene ve ark. (2002) tarafından kullanılan yönteme benzer olarak, genelleştirilmiş beta dağılımının üretici dağılım olarak kullanıldığı dağılım üretme yöntemini önermişlerdir.
- v. Alzaatreh ve ark. (2013), üretici dağılım olarak her hangi bir dağılımın kullanılabileceği daha geniş kapsamlı dağılım üretme yöntemini geliştirmişlerdir (*T-X ailesi*).

Yukarıda listelenen bu yöntemlere ilaveten farklı dağılım üretme yöntemleri de literatürde mevcuttur (Bourguignon ve ark, 2014; Azzalini, 1985; Pearson, 1895, Jones, 2004; Potdar ve Shirke, 2013).

Bu bölümde Alzaatreh ve ark. (2013) tarafından önerilen yönteme dayalı olarak, yaşam analizinde kullanılabilecek esnek dağılımlar önerilerek, dağılımların istatistiksel özellikleri araştırılacaktır. Dağılımların performansı, gerçek verileri modelleme bakımından incelenecek ve ayrıca önerilen yeni dağılıma dayalı regresyon modeli olarak kendi temel dağılımlarına ve makul alternatiflerine karşı test edilecektir.

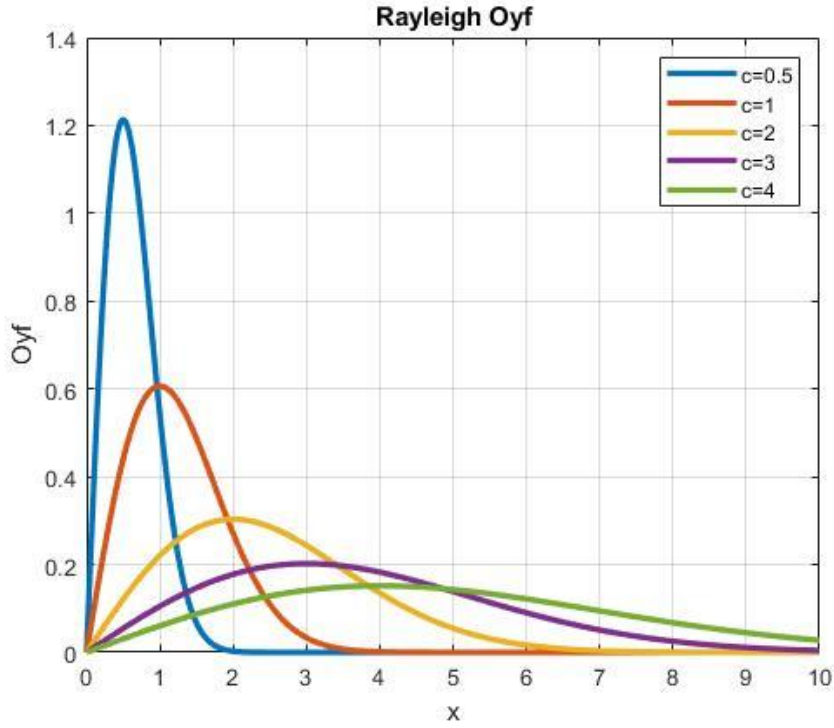
4.1. OBU-R Dağılımı ve İstatistiksel Özellikleri

Rayleigh dağılımı yaşam verilerinin analizinde sıklıkla kullanılmasının (Bhattacharya ve ark., 1990) yanı sıra, rüzgar enerjisi alanında belirli bir bölgenin rüzgar gücünün tahmini için rüzgar hızı verilerinin modellenmesi (Ouahabi ve ark., 2017) ve hidroloji alanında dalga verilerinin modellenmesi (Shariff ve ark., 2012) gibi birçok alanda da kullanılmaktadır. Rayleigh dağılımına sahip X rassal değişkeninin dağılım ve olasılık yoğunluk fonksiyonları sırasıyla (4.1) ve (4.2) eşitlikleri ile ifade edilir:

$$G(x) = 1 - \exp\left(\frac{-x^2}{2c^2}\right), x > 0 \quad (4.1)$$

$$g(x) = \frac{x}{c^2} \exp\left(\frac{-x^2}{2c^2}\right), x > 0 \quad (4.2)$$

Denklemlerde görülen c dağılımın ölçek parametresini ifade eder ve dağılımın şekil parametresi yoktur. Rayleigh dağılımının farklı ölçek parametre değerleri için elde edilen oyf grafiği Şekil 4.1'de verilmiştir. Bununla birlikte, Rayleigh dağılımının yaygın kullanımına rağmen, yalnızca bir parametresi olduğundan dağılım yeterince esnek değildir. Bu sebeple, literatürde Rayleigh dağılımının çeşitli genelleştirmelerine ilişkin çalışmalar bulunmaktadır. Örneğin genelleştirilmiş Gompertz-Rayleigh (Bradley ve ark., 1984), Kumaraswamy genelleştirilmiş Rayleigh (Gomes ve ark., 2014), Kumaraswamy üstelleştirilmiş Rayleigh (Rashwan, 2016), odd genelleştirilmiş üstelleştirilmiş Rayleigh (Luguterah, 2014), Marshall-Olkin genelleştirilmiş Rayleigh (MirMostafae ve ark., 2017) ve Pareto-Rayleigh (Godase ve ark., 2017) dağılımları yaşam, güvenilirlik ve ilgili farklı alanlarda kullanılabilecek alternatif dağılımlar olarak önerilmiştir.



Şekil 4.1. Farklı parametre değerleri için Rayleigh dağılımının oyf grafiği

Alizadeh ve ark. (2017) tarafından $T-X$ ailesi yönteminden yararlanılarak elde edilen Odd Burr-G ailesi kullanılarak, temel dağılımın Rayleigh olduğu Odd Burr-Rayleigh ($OBu-R$) dağılımı elde edilecektir. Böylelikle iki şekil parametresi ilave edilerek elde edilen üç parametrelili yeni dağılımın, tek bir ölçek parametresine sahip olan Rayleigh dağılımına göre daha esnek olması hedeflenmektedir.

Alizadeh ve ark. (2017) tarafından tanımlanan dağılım ailesinin dağılım fonksiyonu (4.3)'teki gibidir:

$$F(x) = \int_0^{G(x)} \frac{abt^{a-1}}{(1+t^a)^{b+1}} dt = 1 - \frac{[1-G(x)]^{ab}}{\{G(x)^a + [1-G(x)]^a\}^b}, \quad x > 0 \quad (4.3)$$

(4.3) dağılım fonksiyonunun türevinin alınması ile elde edilen olasılık yoğunluk fonksiyonu;

$$f(x) = \frac{abg(x)G(x)^{a-1}[1-G(x)]^{ab-1}}{\{G(x)^a + [1-G(x)]^a\}^{b+1}}, \quad x > 0 \quad (4.4)$$

şeklindedir. a ve b elde edilen yeni dağılımın şekil parametreleridir. İlerleyen bölümlerde yukarıda ifade edilen yöntem kullanılarak elde edilen $OBu-R$ dağılımının özellikleri açıklanmıştır.

4.1.1. $OBu-R$ dağılım ve olasılık yoğunluk fonksiyonları

(4.3) denkleminde $G(x)$ temel dağılımı yerine (4.1) ile ifade edilen Rayleigh dağılım fonksiyonu kullanılırsa $OBu-R$ dağılımının dağılım fonksiyonu;

$$F(x) = 1 - \frac{\left[\exp\left(\frac{-x^2}{2c^2}\right) \right]^{ab}}{\left\{ \left[1 - \exp\left(\frac{-x^2}{2c^2}\right) \right]^a + \left[\exp\left(\frac{-x^2}{2c^2}\right) \right]^a \right\}^b}, \quad a, b, c, x > 0 \quad (4.5)$$

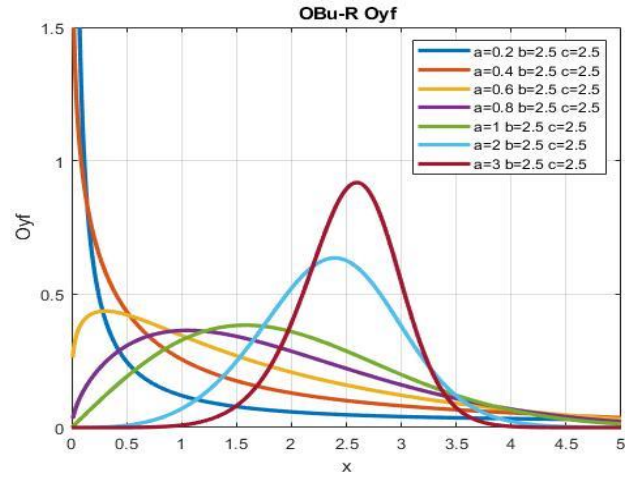
(4.5) dağılım fonksiyonuna karşılık gelen olasılık yoğunluk fonksiyonu ise;

$$f(x) = \frac{abx \left[\exp\left(\frac{-x^2}{2c^2}\right) \right]^{ab} \left[1 - \exp\left(\frac{-x^2}{2c^2}\right) \right]^{a-1}}{c^2 \left\{ \left[1 - \exp\left(\frac{-x^2}{2c^2}\right) \right]^a + \left[\exp\left(\frac{-x^2}{2c^2}\right) \right]^a \right\}^{b+1}}, \quad (4.6)$$

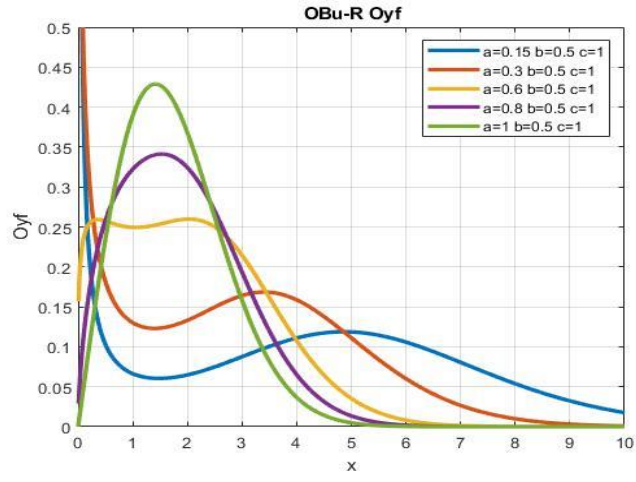
şeklinde ifade edilir. Burada c ölçek parametresi, a ve b ise şekil parametreleridir. (4.6)

oyf'na sahip olan X rassal değişkeni $X \sim OBu - R(a, b, c)$ şeklinde ifade edilir.

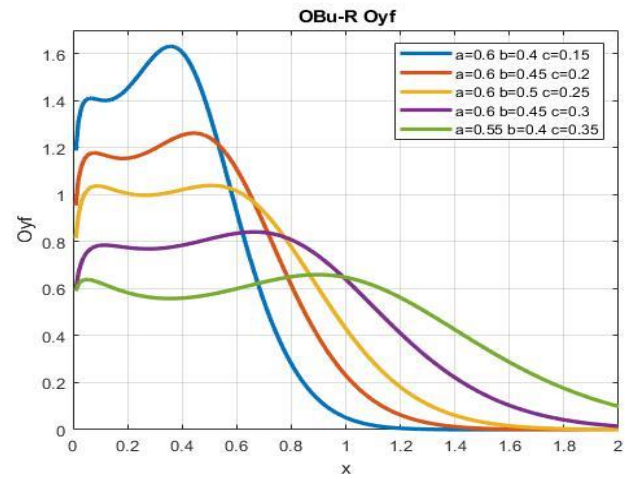
Farklı parametre değerleri için $OBu-R$ dağılımın oyf'nu grafikleri Şekil 4.2'de verilmiştir. Şekillerden de görüldüğü üzere elde edilen yeni dağılım, Şekil 4.1'de verilen Rayleigh dağılımına göre oldukça esnektir ve ayrıca $OBu-R$ dağılımı çift modlu da olabilmektedir.



(a)



(b)



(c)

Şekil 4.2. Farklı parametre değerleri için OBU-R dağılımının oyf grafiği

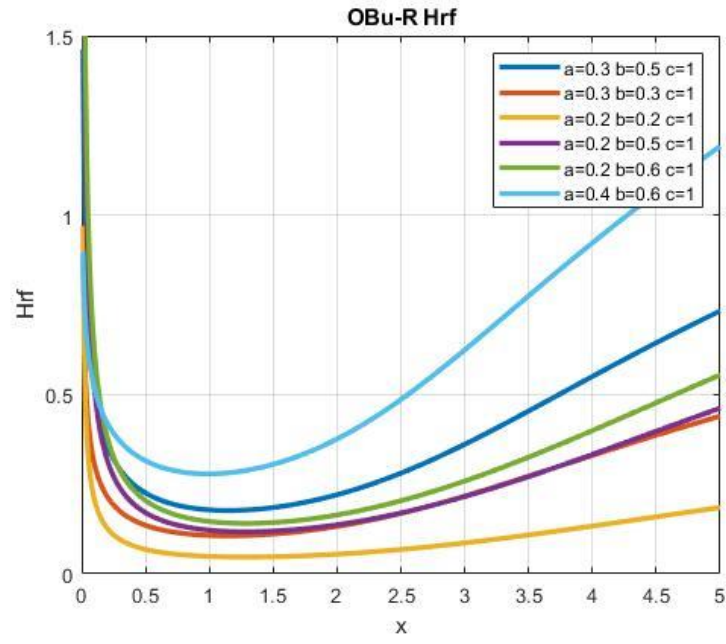
4.1.2. *OBu-R* yaşam ve tehlike oranı fonksiyonları

OBu-R dağılımının yaşam ve tehlike oranı fonksiyonları sırasıyla (4.7) ve (4.8) ile ifade edilmiştir:

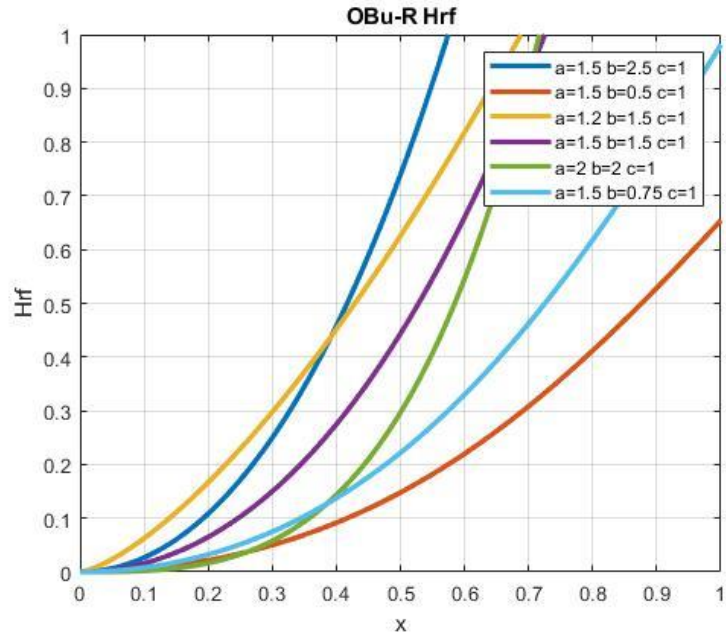
$$S(x) = \frac{\left[\exp\left(\frac{-x^2}{2c^2}\right) \right]^{ab}}{\left\{ \left[1 - \exp\left(\frac{-x^2}{2c^2}\right) \right]^a + \left[\exp\left(\frac{-x^2}{2c^2}\right) \right]^a \right\}^b} \quad (4.7)$$

$$h(x) = \frac{abx \left[1 - \exp\left(\frac{-x^2}{2c^2}\right) \right]^{a-1}}{c^2 \left\{ \left[1 - \exp\left(\frac{-x^2}{2c^2}\right) \right]^a + \left[\exp\left(\frac{-x^2}{2c^2}\right) \right]^a \right\}} \quad (4.8)$$

Farklı parametre değerleri için tehlike oranı fonksiyonu (hrf) grafikleri Şekil 4.3'te verilmiştir. Grafiklerden de görüldüğü üzere parametre değerlerine bağlı olarak *OBu-R* dağılımı monoton artan ve azalan-artan (bathtub eğrisi) şeklinde grafiklere sahip olup elde edilen yeni dağılımın tehlike fonksiyonu Rayleigh dağılımına nazaran daha esnek bir yapıya sahiptir. *OBu-R* rasgele değişkeninin tehlike oranı fonksiyonunun bu cazip esnekliği, yaşam verileri analizi için kullanılabilecek bir dağılım olduğunun göstergesidir.



(a)



(b)

Şekil 4.3. Farklı parametre değerleri için OBU-R dağılımının tehlike oranı fonksiyonu grafiği

4.1.3. OBU-R'nin asimptotik özellikleri

Dağılım, olasılık yoğunluk ve tehlike oranı fonksiyonlarının asimptotik durumları aşağıdaki teoremlerde ifade edilmiştir (Alizadeh ve ark., 2017).

Teorem 4.1: X rassal değişkeninin negatif olmayan değerleri ve $x \rightarrow 0$ iken dağılım, olasılık yoğunluk ve tehlike oranı fonksiyonlarının asimptotik durumları (4.9)'daki gibidir:

$$\begin{aligned} F(x) &\sim bG(x)^a \\ f(x) &\sim abg(x)G(x)^{a-1} \\ h(x) &\sim abg(x)G(x)^{a-1} \end{aligned} \quad (4.9)$$

X rassal değişkeninin negatif olmayan değerleri ve $x \rightarrow \infty$ iken dağılım, olasılık yoğunluk ve tehlike oranı fonksiyonlarının asimptotik durumları (4.10)'daki gibidir:

$$\begin{aligned} 1 - F(x) &\sim [a\bar{G}(x)]^b \\ f(x) &\sim ba^b g(x)\bar{G}(x)^{b-1} \\ h(x) &\sim \frac{bg(x)}{\bar{G}(x)} \end{aligned} \quad (4.10)$$

Kanıt 4.1: Olasılık yoğunluk ve tehlike oranı fonksiyonları dağılım fonksiyonunun türevi olduğu için kanıt yalnızca dağılım fonksiyonu için ifade edilecektir. $x \rightarrow 0$ iken (4.11)'de ifade edilen oran dikkate alınsın:

$$\frac{F(x)}{G(x)^a} = \frac{1 - \left[\exp\left(\frac{-x^2}{2c^2}\right) \right]^{ab}}{\left\{ \left[1 - \exp\left(\frac{-x^2}{2c^2}\right) \right]^a + \left[\exp\left(\frac{-x^2}{2c^2}\right) \right]^a \right\}^b} \quad (4.11)$$

Açıkça görülmektedir ki (4.11) eşitliğinde $x \rightarrow 0$ iken 0/0 belirsizliği bulunmaktadır. Dolayısıyla L'hospital kuralı uygulandığında (4.12)'deki denklem elde edilir:

$$b \left[\left(\frac{1}{\exp\left(\frac{-x^2}{2c^2}\right)} - 1 \right)^a + 1 \right]^{-b-1} \left[\exp\left(\frac{-x^2}{2c^2}\right) \right]^{-2} \quad (4.12)$$

(4.12) denkleminde görülmektedir ki $x \rightarrow 0$ iken sonuç b olur. Sonuç olarak kanıt tamamlanmış olur ve $x \rightarrow 0$ iken $F(x) \sim bG(x)^a$ elde edilir.

4.1.4. *OBu-R*'nin kantil fonksiyonu

(4.5) denkleminde ifade edilen dağılım fonksiyonuna dayalı *OBu-R* dağılımının kantil fonksiyonu (4.13)'teki gibi elde edilir:

$$\begin{aligned}
 F(x) &= 1 - \frac{\left[\exp\left(\frac{-x^2}{2c^2}\right) \right]^{ab}}{\left\{ \left[1 - \exp\left(\frac{-x^2}{2c^2}\right) \right]^a + \left[\exp\left(\frac{-x^2}{2c^2}\right) \right]^a \right\}^b} \\
 [1 - F(x)]^{1/b} &= \frac{\left[\exp\left(\frac{-x^2}{2c^2}\right) \right]^a}{\left[1 - \exp\left(\frac{-x^2}{2c^2}\right) \right]^a + \left[\exp\left(\frac{-x^2}{2c^2}\right) \right]^a} \\
 \left\{ \left[[1 - F(x)]^{-1/b} - 1 \right]^{1/a} + 1 \right\}^{-1} &= \exp\left(\frac{-x^2}{2c^2}\right) \\
 X &= \sqrt{2c^2 \log \left\{ \left[[1 - F(x)]^{-1/b} - 1 \right]^{1/a} + 1 \right\}} \quad (4.13)
 \end{aligned}$$

OBu-R rassal değişkeni $F(x) = u \sim U(0,1)$ alınarak (4.13) kantil fonksiyonu yardımıyla kolaylıkla üretilebilir. Burada U düzgün dağılım rassal değişkenidir.

4.1.5. *OBu-R*'nin alternatif matematiksel gösterimi

Bu bölümde *OBu-R* dağılım ve olasılık yoğunluk fonksiyonlarının kuvvet serilerine dayalı alternatif gösterimleri elde edilecektir. İlk olarak, *OBu-R* dağılım fonksiyonu ele alınsın (Altun ve ark., 2017):

$$F(x) = 1 - \frac{[1 - G(x)]^{ab}}{\left\{ G(x)^a + [1 - G(x)]^a \right\}^b} = 1 - \left\{ 1 - \frac{G(x)^a}{G(x)^a + \bar{G}(x)^a} \right\}^b$$

burada $\bar{G}(x) = 1 - G(x)$ şeklindedir. Denklemin ikinci terimi için binom açılımı uygulandığında dağılım fonksiyonu (4.14)'deki formda elde edilebilir:

$$F(x) = 1 - \sum_{i=0}^{\infty} (-1)^i \binom{b}{i} \frac{G(x)^{ai}}{[G(x)^a + \bar{G}(x)^a]^i} \quad (4.14)$$

Daha sonra $G(x)^{ai}$ ifadesi için binom açılımı yardımıyla elde edilen alternatif gösterim (4.15)'deki gibi elde edilmiştir:

$$\begin{aligned} G(x)^{ai} &= \{1 - [1 - G(x)]\}^{ai} = \sum_{j=0}^{\infty} (-1)^j \binom{ai}{j} [1 - G(x)]^j \\ [1 - G(x)]^j &= \sum_{k=0}^{\infty} (-1)^k \binom{j}{k} G(x)^k \\ G(x)^{ai} &= \sum_{k=0}^{\infty} \alpha_k G(x)^k \end{aligned} \quad (4.15)$$

burada $\alpha_k = \sum_{j=k}^{\infty} (-1)^{k+j} \binom{ai}{j} \binom{j}{k}$ şeklindedir.

(4.14) denkleminin son teriminin paydasındaki ifade binom açılımı yardımıyla aşağıdaki gibi ifade edilebilir:

$$\begin{aligned} G(x)^a &= \{1 - [1 - G(x)]\}^a = \sum_{j=0}^{\infty} (-1)^j \binom{a}{j} [1 - G(x)]^j \\ [1 - G(x)]^j &= \sum_{k=0}^{\infty} (-1)^k \binom{j}{k} G(x)^k \\ G(x)^a &= \sum_{k=0}^{\infty} \eta_k G(x)^k \\ G(x)^a + \bar{G}(x)^a &= \sum_{k=0}^{\infty} \eta_k G(x)^k + \sum_{k=0}^{\infty} (-1)^k \binom{a}{k} G(x)^k \\ &= \sum_{k=0}^{\infty} \left\{ \eta_k + (-1)^k \binom{a}{k} \right\} G(x)^k \\ &= \sum_{k=0}^{\infty} t_k G(x)^k, \end{aligned}$$

burada $\eta_k = \sum_{j=k}^{\infty} (-1)^{k+j} \binom{a}{j} \binom{j}{k}$ ve $t_k = \eta_k + (-1)^k \binom{a}{k}$ şeklindedir. (Gradshteyn ve Ryzhik, 2002)'de herhangi bir kuvvet serisinin, pozitif tam sayı olan i 'inci kuvveti (4.16)'daki gibi ifade edilmiştir:

$$\left(\sum_{k=0}^{\infty} t_k x^k \right)^i = \sum_{k=0}^{\infty} c_{i,k} x^k \quad (4.16)$$

burada $c_{i,k} = (kt_0)^{-1} \sum_{m=1}^k [m(i+1) - k] t_m c_{i,k-m}$ ve $c_{i,0} = t_0^i$ şeklindedir. (4.16) eşitliği yardımıyla,

$$\begin{aligned} \left[G(x)^a + \bar{G}(x)^a \right]^i &= \left[\sum_{k=0}^{\infty} t_k G(x)^k \right]^i \\ &= \sum_{k=0}^{\infty} c_{i,k} G(x)^k \end{aligned} \quad (4.17)$$

şeklinde elde edilir. Sonuç olarak,

$$\frac{G(x)^{ai}}{\left[G(x)^a + \bar{G}(x)^a \right]^i} = \frac{\sum_{k=0}^{\infty} \alpha_k G(x)^k}{\sum_{k=0}^{\infty} \beta_k G(x)^k}$$

şeklinde iki kuvvet serisinin oranı olarak elde edilir. Bu oran alternatif olarak (4.18)'deki gibi ifade edilebilir (Gradshteyn ve Ryzhik, 2002):

$$\frac{G(x)^{ai}}{\left[G(x)^a + \bar{G}(x)^a \right]^i} = \sum_{k=0}^{\infty} \lambda_k G(x)^k \quad (4.18)$$

burada $k > 0$ için $\lambda_k = \beta_0^{-1} \left[\alpha_k - \beta_0^{-1} \sum_{r=1}^k \beta_r \lambda_{k-r} \right]$ ve $\lambda_0 = \frac{\alpha_0}{\beta_0}$ şeklindedir. Sonuç olarak

OBu-R dağılım fonksiyonunun alternatif gösterimi (4.19)'daki gibi elde edilir:

$$\begin{aligned}
F(x) &= 1 - \sum_{k,i=0}^{\infty} (-1)^i \binom{b}{i} \lambda_k G(x)^k \\
&= 1 - \sum_{k=0}^{\infty} m_k G(x)^k \\
&= \sum_{k=0}^{\infty} n_k G(x)^k = \sum_{k=0}^{\infty} n_k H_k(x)
\end{aligned} \tag{4.19}$$

ve olasılık yoğunluk fonksiyonunun alternatif gösterimi,

$$f(x) = \sum_{k=0}^{\infty} n_{k+1} h_{k+1}(x) \tag{4.20}$$

şeklindedir. Burada $H_k(x)$ ölçek parametresi c ve şekil parametresi k olan üstelleştirilmiş Rayleigh dağılımı olup, $m_k = \sum_{i=0}^{\infty} (-1)^i \binom{b}{i} \lambda_k$, $n_0 = 1 - m_0$ ve $n_k = -m_k$ şeklindedir. (4.19) eşitliğine göre yeni elde edilen dağılım fonksiyonu, temel alınan dağılım fonksiyonunun ağırlıklandırılmış kuvvet serisi toplamı olarak ifade edilmiştir. Bu sonuç *OBu-R* dağılımının moment, koşullu moment ve sıra istatistikleri gibi bazı istatistiksel özelliklerinin elde edilmesinde kullanılacaktır.

4.1.6. *OBu-R*'nin momentleri

Bu bölümde *OBu-R* dağılımının moment fonksiyonu (4.20) eşitliği yardımıyla elde edilecektir. Bu amaçla üstelleştirilmiş Rayleigh dağılımının r 'inci momenti (4.21)'deki gibi ifade edilir.

$$\mu_{ER} = E_{ER}(X^r) = \frac{\alpha}{c^2} \int_0^{\infty} x^{r+1} e^{-\frac{x^2}{2c^2}} \left(1 - e^{-\frac{x^2}{2c^2}}\right)^{\alpha-1} dx. \tag{4.21}$$

Burada α ve c sırasıyla şekil ve ölçek parametreleridir. Binom açılımı yardımıyla üstelleştirilmiş Rayleigh dağılımının r 'inci momenti (4.22) eşitliği gibi de ifade edilebilir (Mahmoud ve Ghazal, 2017):

$$\mu_{ER} = E_{ER}(X^r) = 2^{r/2} c^r \alpha \sum_{j=0}^{\alpha-1} \frac{(-1)^j \binom{\alpha-1}{j}}{(j-1)^{\frac{r}{2}+1}} \Gamma\left(\frac{r}{2}+1\right) \tag{4.22}$$

Sonuç olarak (4.20) eşitliği ile ifade edilen ilişki yardımıyla $OBu-R$ dağılımının r 'inci momenti (4.23)'teki gibi ifade edilebilir:

$$\mu_{OBu-R} = E_{OBu-R}(X^r) = 2^{r/2} c^r \Gamma\left(\frac{r}{2} + 1\right) \sum_{k=0}^{\infty} n_{k+1} (k+1) \left[\sum_{j=0}^k \frac{(-1)^j \binom{k}{j}}{(j-1)^{\frac{r}{2}+1}} \right] \quad (4.23)$$

4.1.7. $OBu-R$ 'nin koşullu momentleri

Yaşam süresi modellerinde çıkarım için koşullu momenlerin elde edilmesi önemlidir. Bu amaçla öncelikle üstelleştirilmiş Rayleigh dağılımının r 'inci koşullu momenti (4.24)'teki gibi elde edilmiştir:

$$E(X^r | X > t) = 2^{r/2} c^r \alpha \sum_{j=0}^{\alpha-1} \frac{(-1)^j \binom{\alpha-1}{j}}{(j-1)^{\frac{r}{2}+1}} \Gamma\left(\frac{r}{2} + 1, t\right) \quad (4.24)$$

Burada $\Gamma(a, t) = \int_t^{\infty} x^{a-1} e^{-x} dx$ ifadesi tamamlanmamış gamma fonksiyonudur. Sonuç olarak (4.24) yardımıyla $OBu-R$ dağılımının r 'inci koşullu momenti,

$$E(X^r | X > t) = 2^{r/2} c^r \Gamma\left(\frac{r}{2} + 1, t\right) \sum_{k=0}^{\infty} n_{k+1} (k+1) \left[\sum_{j=0}^k \frac{(-1)^j \binom{k}{j}}{(j-1)^{\frac{r}{2}+1}} \right] \quad (4.25)$$

şeklinde ifade edilir.

İstatistiksel dağılımların esnekliği genel olarak çarpıklık (S) ve basıklık (K) katsayılarının incelenmesiyle karakterize edilir (Caudill., 2012). $OBu-R$ dağılımının ilk dört momenti ile S ve K değerleri Tablo 4.1'de verilmiştir. Tablodan da görüldüğü üzere $OBu-R$ dağılımı oldukça geniş bir aralıkta çarpıklık ve basıklık değerlerine sahiptir.

Dolayısıyla sabit çarpıklık $\left(\frac{2\sqrt{\pi(\pi-3)}}{(4-\pi)^{3/2}} \approx 0.631 \right)$ ve basıklık

$\left(-\frac{6\pi^2 - 24\pi + 16}{(4 - \pi)^2} \approx 3.245 \right)$ değerlerine sahip Rayleigh dağılımına göre oldukça esnek bir dağılımdır.

Öte yandan, dağılım ailelerinin esnekliğini karşılaştırmak için, S ve K arasındaki (4.26)'daki teorik ilişki de kullanılabilir (Caudill, 2012):

$$K \geq S^2 + 1 \quad (4.26)$$

Tablo 4.1 incelendiğinde, $OBu-R$ dağılımının olası tüm S ve K kombinasyonlarının (4.26) ile verilen eşitsizliği sağladığı görülmektedir.

Table 4.1. $OBu-R$ dağılımı için çarpıklık ve basıklık değerleri

a	b	c	S	K
0,5	0,5	1	0,4910	2,7035
0,5	1	1	0,7509	2,9005
0,5	2	1	1,2094	4,0631
0,5	5	1	1,9281	7,7019
0,5	10	1	2,2332	10,4098
1,5	0,5	1	0,7347	3,6447
1,5	1	1	0,5309	3,4895
1,5	2	1	0,3006	3,0952
1,5	5	1	0,1399	2,8024
1,5	10	1	0,1057	2,7233
2	0,5	1	0,8088	4,1178
2	1	1	0,4526	3,6610
2	2	1	0,0969	3,1498
2	5	1	-0,1289	2,8946
2	10	1	-0,1795	2,8295
5	0,5	1	0,9647	5,2909
5	1	1	0,2263	4,0477
5	2	1	-0,3259	3,7070
5	5	1	-0,6485	3,8647
5	10	1	-0,7391	3,9510
10	0,5	1	0,9690	5,6440
10	1	1	0,1190	4,1560
10	2	1	-0,4638	4,0334
10	5	1	-0,8071	4,4064
10	10	1	-0,9116	4,5880

Kantillere dayalı olarak hesaplanan robust çarpıklık (S_R) (Galton, 1983) ve basıklık (K_R) (Moors, 1988) ölçüleri de literatürde kullanılmakta olup aşağıdaki formüller ile ifade edilen bu ölçüler yardımıyla herhangi bir dağılımın momenti olmasa da bu ölçüler hesaplanabilir.

$$S_R = \frac{Q(6/8) - 2Q(4/8) + Q(2/8)}{Q(6/8) - Q(2/8)}$$

$$K_R = \frac{Q(7/8) - Q(5/8) + Q(3/8) - Q(1/8)}{Q(6/8) - Q(2/8)}$$

4.1.8. *OBu-R*'nin sıra istatistiği

$X_1, X_2, \dots, X_n$ *OBu-R* dağılımından n birimlik örneklem olmak üzere $X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(j)} \leq \dots \leq X_{(n)}$, bu örneklemde sıra istatistikleri olsun. $X_{(j)}$ sıra istatistiğinin oyf,

$$\begin{aligned} f_{x_{(j)}}(x) &= \frac{n!}{(n-j)!(j-1)!} f(x) [F(x)]^{j-1} [1-F(x)]^{n-j} \\ &= \frac{n!}{(n-j)!(j-1)!} \sum_{i=0}^{n-j} (-1)^i \binom{n-j}{i} [F(x)]^{j+i-1} f(x). \end{aligned} \quad (4.27)$$

(4.27) fonksiyonu (4.19) ve (4.20) eşitliklerinden yararlanarak (4.28)'deki formda da yazılabilir.

$$\begin{aligned} f_{x_{(j)}}(x) &= \frac{n!}{(n-j)!(j-1)!} \\ &\sum_{i=0}^{n-j} (-1)^i \binom{n-j}{i} \left[\sum_{l=0}^{\infty} (l+1) b_{l+1} g(x) G(x)^l \right] \left[\sum_{k=0}^{\infty} b_k G(x)^k \right]^{j+i-1} \end{aligned} \quad (4.28)$$

Kuvvet serinin kuvveti ile ilgili eşitlik yardımıyla (Gradshteyn ve Ryzhik, 2002) $X_{(j)}$ sıra istatistiğinin oyf,

$$\begin{aligned} f_{x_{(j)}}(x) &= \frac{n!}{(n-j)!(j-1)!} \\ &\sum_{i=0}^{n-j} (-1)^i \binom{n-j}{i} \left[\sum_{l=0}^{\infty} (l+1) b_{l+1} g(x) G(x)^l \right] \left[\sum_{k=0}^{\infty} a_k G(x)^k \right] \end{aligned} \quad (4.29)$$

olarak yazılabilir. Bu eşitlikteki $a_0 = b_0^{j+i-1}$ ve $k \geq 1$ için

$a_m = (mb_0)^{-1} \sum_{k=1}^m [k(j+i) - m] b_k a_{m-k}$ şeklindedir. (4.29) denklemini bir dizi cebirsel işlemin ardından,

$$\begin{aligned} f_{x_{(j)}}(x) &= \sum_{i=0}^{n-j} \sum_{k,l=0}^{\infty} \frac{n!}{(n-j)!(j-1)!} (-1)^i \binom{n-j}{i} (l+1) b_{l+1} a_k g(x) G(x)^{l+k} \\ &= \sum_{i=0}^{n-j} \sum_{k,l=0}^{\infty} \frac{n!}{(n-j)!(j-1)!} (-1)^i \binom{n-j}{i} \frac{(l+1) b_{l+1} a_k}{(l+k+1)} h(x)^{l+k+1}. \end{aligned} \quad (4.30)$$

şeklinde elde edilebilir.

$$d_{i,k,l} = \frac{n!}{(n-j)!(j-1)!} (-1)^i \binom{n-j}{i} \frac{(l+1) b_{l+1} a_k}{(l+k+1)}$$

olarak alınsın. Sonuç olarak $X_{(j)}$ sıra istatistiğinin oyf,

$$f_{x_{(j)}}(x) = \sum_{i=0}^{n-j} \sum_{k,l=0}^{\infty} d_{i,k,l} h_{(l+k+1)}(x) \quad (4.31)$$

şeklinde elde edilir. Bu eşitlikteki $h(x)$ şekil ve ölçek parametreleri sırasıyla $(l+k+1)$ ve c olan üstelleştirilmiş Rayleigh dağılımının oyf'dur. Sonuç olarak $OBu-R$ dağılımının sıra istatistiğinin oyf $h(x)$ 'in lineer kombinasyonu olarak ifade edilebilir.

4.1.9. $OBu-R$ 'nin en çok olabilirlik tahmincisi

Önerilen $OBu-R$ dağılımının parametrelerinin tahmini için en çok olabilirlik (EÇO) yöntemi kullanılacaktır. Bu amaçla $X_1, X_2, \dots, X_n$, parametreleri a, b ve c olan $OBu-R$ dağılımından örneklem olmak üzere, log-olabilirlik fonksiyonu,

$$\begin{aligned} \log L(a, b, c) &= n \log(ab) + \sum_{i=1}^n \log\left(\frac{x_i}{c^2}\right) - \frac{ab}{c^2} \sum_{i=1}^n x_i^2 \\ &\quad + (a-1) \sum_{i=1}^n \log\left\{1 - \exp\left(\frac{-x_i^2}{2c^2}\right)\right\} \\ &\quad - (b+1) \sum_{i=1}^n \log\left\{\left[1 - \exp\left(\frac{-x_i^2}{2c^2}\right)\right]^a + \left[\exp\left(\frac{-x_i^2}{2c^2}\right)\right]^a\right\}. \end{aligned} \quad (4.32)$$

şeklinde ifade edilir. Bilindiği üzere parametrelerin olabilirlik tahmincileri log-olabilirlik fonksiyonunun parametrelere göre kısmi türevlerinin sıfıra eşitlenmesi ile elde edilir. Bu amaçla (4.32) eşitliğinin parametrelere göre kısmi türevleri (4.33), (4.34) ve (4.35) eşitlikleri şeklinde elde edilmiştir.

$$\begin{aligned} \frac{\partial \log L}{\partial a} &= \frac{n}{a} - \frac{b}{2c^2} \sum_{i=1}^n x_i^2 - \sum_{i=1}^n \log \left\{ 1 - \exp \left(\frac{-x_i^2}{2c^2} \right) \right\} \\ &- (b+1) \sum_{i=1}^n \frac{\left[1 - \exp \left(\frac{-x_i^2}{2c^2} \right) \right]^a \log \left[1 - \exp \left(\frac{-x_i^2}{2c^2} \right) \right] + \left[\exp \left(\frac{-x_i^2}{2c^2} \right) \right]^a \log \left[\exp \left(\frac{-x_i^2}{2c^2} \right) \right]}{\left[1 - \exp \left(\frac{-x_i^2}{2c^2} \right) \right]^a + \left[\exp \left(\frac{-x_i^2}{2c^2} \right) \right]^a} = 0 \end{aligned} \quad (4.33)$$

$$\frac{\partial \log L}{\partial b} = \frac{n}{b} - \frac{a}{2c^2} \sum_{i=1}^n x_i^2 - \sum_{i=1}^n \log \left\{ \left[1 - \exp \left(\frac{-x_i^2}{2c^2} \right) \right]^a + \left[\exp \left(\frac{-x_i^2}{2c^2} \right) \right]^a \right\} = 0 \quad (4.34)$$

$$\begin{aligned} \frac{\partial \log L}{\partial c} &= -\frac{2n}{c} + \frac{ab}{c^3} \sum_{i=1}^n x_i^2 - (a-1) \sum_{i=1}^n \frac{x^2}{c^3 \left[\exp \left(\frac{x^2}{2c^2} \right) - 1 \right]} \\ &- (b+1) \sum_{i=1}^n \frac{\frac{ax^2 \left[\exp \left(-\frac{x^2}{2c^2} \right) \right]^a}{c^3} - \frac{ax^2 \exp \left(-\frac{x^2}{2c^2} \right) \left[1 - \exp \left(-\frac{x^2}{2c^2} \right) \right]^{a-1}}{c^3}}{\left[1 - \exp \left(-\frac{x^2}{2c^2} \right) \right]^a + \left[\exp \left(-\frac{x^2}{2c^2} \right) \right]^a} = 0 \end{aligned} \quad (4.35)$$

Parametrelerin en çok olabilirlik tahminleri, yukarıda verilen denklemlerin eş zamanlı çözümü ile elde edilebilir. Bu amaçla denklemlerin açık çözümü mümkün olmadığından Newton-Raphson gibi nümerik yöntemlerin kullanılması gerekir.

Hipotez testi ve aralık tahmini için gözlemlenmiş bilgi matrisinin tersi alınarak elde edilen asimptotik varyans-kovaryans matrisi kullanılır. Gözlemlenmiş bilgi matrisinin tersi,

$$I(\hat{\theta}) = \begin{bmatrix} -\frac{\partial^2 \log L}{\partial a^2} & -\frac{\partial^2 \log L}{\partial a \partial b} & -\frac{\partial^2 \log L}{\partial a \partial c} \\ -\frac{\partial^2 \log L}{\partial b \partial a} & -\frac{\partial^2 \log L}{\partial b^2} & -\frac{\partial^2 \log L}{\partial b \partial c} \\ -\frac{\partial^2 \log L}{\partial a \partial c} & -\frac{\partial^2 \log L}{\partial b \partial c} & -\frac{\partial^2 \log L}{\partial c^2} \end{bmatrix}^{-1} \quad (4.36)$$

burada $\hat{\theta} = (\hat{a}, \hat{b}, \hat{c})$, tahmin edilen parametre vektörüdür. (4.36) matrisinin elemanları ifade edildiği gibidir:

$$\frac{\partial^2 \log L}{\partial a^2} = -\frac{n}{a^2} - (b+1) \sum_{i=1}^n \frac{(1-u_i)^a u_i^a [\log(1-u_i) - \log(u_i)]^2}{[u_i^a + (1-u_i)^a]^2},$$

$$\frac{\partial^2 \log L}{\partial b^2} = -\frac{n}{b^2},$$

$$\begin{aligned} \frac{\partial^2 \log L}{\partial c^2} &= \frac{2n}{c^2} - \frac{3ab}{c^4} \sum_{i=1}^n x_i^2 + (a-1) \sum_{i=1}^n \frac{x_i^2 [3c^2 u_i - (1-u_i) x_i^2]}{c^6 u_i^2} \\ &\quad - (b+1) \sum_{i=1}^n \frac{ax_i^2 \left[(3c^2 - x_i^2) u_i^{2a} e^{\frac{x_i^2}{2c^2}} - 3c^2 u_i^{2a} \right] e^{\frac{ax_i^2}{c^2}}}{c^6 \left(e^{\frac{x_i^2}{2c^2}} - 1 \right)^2 \left(u_i^a e^{\frac{ax_i^2}{2c^2}} + 1 \right)^2} \\ &\quad - (b+1) \sum_{i=1}^n \frac{ax_i^2 \left[(ax_i^2 - 3c^2) u_i^a e^{\frac{x_i^2}{c^2}} + (9c^2 - x_i^2) u_i^a e^{\frac{x_i^2}{2c^2}} - 6c^2 u_i^a \right] e^{\frac{ax_i^2}{2c^2}}}{c^6 \left(e^{\frac{x_i^2}{2c^2}} - 1 \right)^2 \left(u_i^a e^{\frac{ax_i^2}{2c^2}} + 1 \right)^2} \\ &\quad + (b+1) \sum_{i=1}^n \frac{ax_i^2 \left(3c^2 e^{\frac{x_i^2}{c^2}} + 6c^2 e^{\frac{x_i^2}{2c^2}} + 3c^2 \right)}{c^6 \left(e^{\frac{x_i^2}{2c^2}} - 1 \right)^2 \left(u_i^a e^{\frac{ax_i^2}{2c^2}} + 1 \right)^2}, \end{aligned}$$

$$\frac{\partial^2 \log L}{\partial a \partial b} = -\frac{1}{2c^2} \sum_{i=1}^n x_i^2 - \sum_{i=1}^n \frac{u_i^a \log(u_i) + (1-u_i)^a \log(1-u_i)}{u_i^a + (1-u_i)^a},$$

$$\frac{\partial^2 \log L}{\partial b \partial c} = \frac{a}{c^3} \sum_{i=1}^n x_i^2 - \sum_{i=1}^n \frac{ax_i^2 (1-u_i)^a - ax_i^2 (1-u_i) u_i^{a-1}}{c^3 [u_i^a + (1-u_i)^a]},$$

$$\begin{aligned}
\frac{\partial^2 \log L}{\partial a \partial c} &= \frac{b}{c^3} \sum_{i=1}^n x_i^2 + \sum_{i=1}^n \frac{x_i^2}{c^3 u_i} \\
&- (b+1) \sum_{i=1}^n \frac{x_i^2 \log(1-u_i)(1-u_i)^a a - x_i^2 (1-u_i) u_i^{a-1} \log(u_i) a + x_i^2 (1-u_i) a - x_i^2 (1-u_i) u_i^{a-1}}{c^3 [u_i^a + (1-u_i)^a]} \\
&+ (b+1) \sum_{i=1}^n \frac{\left[\log(1-u_i)(1-u_i)^a + u_i^a \log(u_i) \right] \left[\frac{x_i^2 (1-u_i)^a a - x_i^2 (1-u_i) u_i^{a-1} a}{c^3} \right]}{[u_i^a + (1-u_i)^a]^2}
\end{aligned}$$

Yukarıdaki eşitliklerde ifade edilen u_i ,

$$u_i = 1 - \exp\left(-\frac{x_i^2}{2c^2}\right)$$

şeklindedir.

4.1.10. *OBu-R*'nin en çok olabilirlik tahmincisi için benzetim çalışması

Bu bölümde, *OBu-R* dağılımı için EÇO yönteminin performansını değerlendirmek amacıyla bir benzetim çalışması yapılmıştır. *OBu-R* dağılımından rassal değişkenler üretmek için dağılımın kantil fonksiyonu kullanılmıştır. Her bir parametre değeri için örneklem büyüklükleri, $n=50, 100, 500$ ve 1000 olarak ele alınmıştır. Her bir durum için 1000 tekrar yapılmıştır. Elde edilen sonuçların değerlendirilmesi için, ortalama (*Mean*) hata kareler ortalaması (*MSE*) ve yanlılık (*Bias*) kriterlerinden yararlanılmıştır.

$$MSE(\hat{\theta}) = \frac{1}{1000} \sum (\hat{\theta} - \theta)^2$$

$$Bias(\hat{\theta}) = \left(\frac{1}{1000} \sum \hat{\theta} \right) - \theta$$

Benzetim sonuçları Tablo 4.2'de verilmiştir. Tablodaki sonuçlardan görüldüğü üzere örneklem büyüklüğü n arttıkça *MSE* ve *Bias* değerleri azalmakta olup yine n arttıkça tahmin edilen parametrelerin ortalamaları gerçek parametre değerlerine oldukça yaklaşmaktadır.

Tablo 4.2. OBU-R için EÇO tahmincisinin Bias ve MSE değerleri

α	b	c	n	Ortalama			Bias			MSE		
				$\hat{a}$	$\hat{b}$	$\hat{c}$	$\hat{a}$	$\hat{b}$	$\hat{c}$	$\hat{a}$	$\hat{b}$	$\hat{c}$
0,5	0,5	0,5	50	0,5089	0,6977	0,5722	-0,0089	-0,1977	-0,0722	0,0131	3,1674	0,2941
0,5	0,5	0,5	100	0,5065	0,5586	0,5260	-0,0065	-0,0586	-0,0260	0,0066	0,0676	0,0238
0,5	0,5	0,5	250	0,5032	0,5193	0,5089	-0,0032	-0,0193	-0,0089	0,0027	0,0181	0,0074
0,5	0,5	0,5	500	0,5021	0,5074	0,5037	-0,0021	-0,0074	-0,0037	0,0012	0,0072	0,0031
0,5	0,5	0,5	1000	0,5001	0,5058	0,5025	-0,0001	-0,0058	-0,0025	0,0006	0,0039	0,0016
α	b	c	n	$\hat{a}$	$\hat{b}$	$\hat{c}$	$\hat{a}$	$\hat{b}$	$\hat{c}$	$\hat{a}$	$\hat{b}$	$\hat{c}$
3	2	0,5	50	3,2711	8,8595	0,5343	-0,2711	-6,8595	-0,0343	0,7062	429,1074	0,0169
3	2	0,5	100	3,1099	4,3295	0,5165	-0,1099	-2,3295	-0,0165	0,1458	122,3603	0,0064
3	2	0,5	250	3,0469	2,6727	0,5093	-0,0469	-0,6727	-0,0093	0,0508	12,6095	0,0024
3	2	0,5	500	3,0121	2,2670	0,5066	-0,0121	-0,2670	-0,0066	0,0235	0,9923	0,0009
3	2	0,5	1000	3,0002	2,1181	0,5035	-0,0002	-0,1181	-0,0035	0,0103	0,2325	0,0003
α	b	c	n	$\hat{a}$	$\hat{b}$	$\hat{c}$	$\hat{a}$	$\hat{b}$	$\hat{c}$	$\hat{a}$	$\hat{b}$	$\hat{c}$
2,5	1,5	0,5	50	2,7687	6,5412	0,5417	-0,2687	-5,0412	-0,0417	0,4381	297,6375	0,0280
2,5	1,5	0,5	100	2,5823	3,2603	0,5258	-0,0823	-1,7603	-0,0258	0,1092	76,0992	0,0108
2,5	1,5	0,5	250	2,5352	1,7602	0,5086	-0,0352	-0,2602	-0,0086	0,0383	0,9966	0,0023
2,5	1,5	0,5	500	2,5129	1,5999	0,5035	-0,0129	-0,0999	-0,0035	0,0161	0,2439	0,0009
2,5	1,5	0,5	1000	2,5036	1,5541	0,5025	-0,0036	-0,0541	-0,0025	0,0089	0,0871	0,0004
α	b	c	n	$\hat{a}$	$\hat{b}$	$\hat{c}$	$\hat{a}$	$\hat{b}$	$\hat{c}$	$\hat{a}$	$\hat{b}$	$\hat{c}$
2	2	1	50	2,2131	9,5806	1,1544	-0,2131	-7,5806	-0,1544	0,2699	358,8525	0,2380
2	2	1	100	2,0998	5,6680	1,0922	-0,0998	-3,6680	-0,0922	0,0668	139,1724	0,1256
2	2	1	250	2,0339	2,8383	1,0342	-0,0339	-0,8383	-0,0342	0,0212	15,5435	0,0313
2	2	1	500	2,0159	2,2784	1,0175	-0,0159	-0,2784	-0,0175	0,0095	1,3022	0,0108
2	2	1	1000	2,0039	2,1587	1,0122	-0,0039	-0,1587	-0,0122	0,0045	0,4871	0,0053

4.1.11. *OBu-R*'nin uygulamaları

Bu bölümde, *OBu-R* dağılımının performansını incelemek amacıyla iki ayrı gerçek veri seti üzerinde uygulama yapılarak literatürde bu veri setleri için kullanılan dağılımlar ile karşılaştırması yapılmıştır.

Uygulama I

İlk veri seti İngiltere Ulusal Fizik laboratuvarın gerçekleştirilen 63 gözlemden oluşan 1,5 cm'lik cam elyafın mukavemetinin ölçülmesi deneyinden alınmıştır (Smith ve Naylor, 1987). Veri seti Tablo 4.3.'te verilmiştir.

Tablo 4.3. *Cam elyaf verisi*

0,55	0,74	0,77	0,81	0,84	0,93	1,04	1,11	1,13	1,24
1,25	1,27	1,28	1,29	1,30	1,36	1,39	1,42	1,48	1,48
1,49	1,49	1,50	1,50	1,51	1,52	1,53	1,54	1,55	1,55
1,58	1,59	1,60	1,61	1,61	1,61	1,61	1,62	1,62	1,63
1,64	1,66	1,66	1,66	1,67	1,68	1,68	1,69	1,70	1,70
1,73	1,76	1,76	1,77	1,78	1,81	1,82	1,84	1,84	1,89
2,00	2,01	2,24							

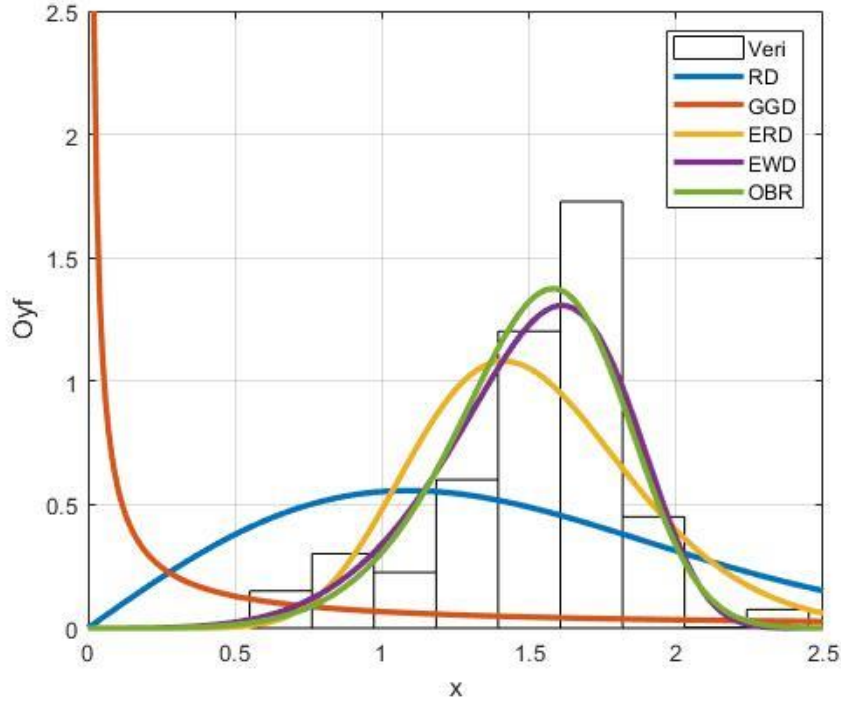
OBu-R dağılımı, literatürde sıklıkla kullanılan dağılımlardan olan Rayleigh (RD), Gamma (GD), üstelleştirilmiş Rayleigh (ERD) ve üstelleştirilmiş Weibull (EWD) dağılımları ile model seçmede yaygın olarak kullanılan Akaike bilgi kriteri (AIC), Bayesci bilgi kriteri (BIC) ve negatif log-olabilirlik (-logL) kriteri dikkate alınarak karşılaştırılmıştır. Elde edilen sonuçlar Tablo 4.4'te verilmiştir.

Tablo 4.4. *Cam elyaf verisi için RD, GD, ERD, EWD ve OBU-R dağılımlarının parametre tahminleri ve kriterler sonuçları*

	RD	GD	ERD	EWD	<i>OBu-R</i>
<i>c</i>	1,5407	0,08640	0,7164	0,5820	1,8414
<i>a</i>	-	17,4395	5,4860	0,6712	2,6500
<i>b</i>	-	-	-	7,2845	7,5276
-logL	49,7909	23,9515	23,9288	14,6755	14,6362
AIC	101,5818	51,9031	51,8575	35,3510	35,2723
BIC	103,7249	56,1893	56,1438	41,7804	41,7017

AIC, BIC and -logL'in küçük değerleri ele alınan modeller arasından en iyi olanı gösterdiği bilinmektedir. Tablo 4.4.'ten de görüldüğü üzere model seçim kriterlerine göre

OBu-R dağılımı, ele alınan diğer dağılımlara göre veri setini daha iyi modellemiştir. Ayrıca önerilen dağılımın veri setini modellemede görsel olarak değerlendirilmesi için *oyf*'ları Şekil 4.4'de verilmiştir. Grafikten de anlaşılacağı üzere *OBu-R* dağılımı veriyi modellemede başarılıdır.



Şekil 4.4. Olasılık yoğunluk fonksiyonları ve histogram grafiği (Cam elyaf verisi)

Uygulama II

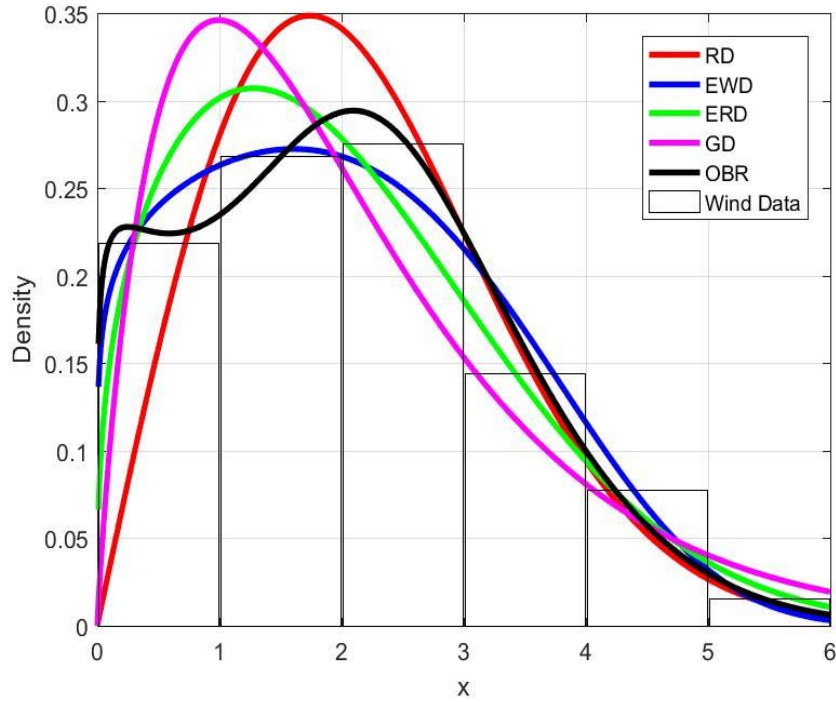
İkinci uygulamada, rüzgâr enerjisi tahmini açısından yenilenebilir enerji alanında çok önemli olan (Kantar ve Usta, 2015; Kantar ve ark., 2018), aylık rüzgâr hızı verilerinin modellenmesi ele alınmıştır. Yenilenebilir Enerji Genel Müdürlüğü'nden alınan Deniz seviyesinden 10 m yükseklikte ölçülen Göztepe bölgesine ait olan rüzgâr hızı verileri kullanılmıştır.

İlk uygulamada ele alınan dağılımlar burada da kullanılmış olup bu dağılımlara ilişkin parametre tahmin değerlerine ilaveten değerlendirme kriterleri AIC, BIC, negatif log-olabilirlik, belirlilik katsayısı R^2 ve hata ortalamalarının karekökü (RMSE) sonuçları Tablo 4.5.'te verilmiştir.

Tablo 4.5. Rüzgar hızı verisi için RD, GD, ERD, EWD ve OBU-R dağılımlarının parametre tahminleri ve kriterleri sonuçları

	RD	GD	ERD	EWD	OBu-R
c	2,4611	1,1162	1,9827	0,2786	0,7719
a	-	1,8905	0,6760	0,3310	0,5812
b	-	-	-	3,4571	0,3259
$-\log L$	1168,783	1167,353	1128,750	1116,653	1114,044
AIC	2339,565	2339,070	2261,501	2239,305	2234,089
BIC	2344,129	2348,197	2270,629	2252,997	2247,780
R^2	0,8921	0,8509	0,9454	0,9717	0,9901
RMSE	0,0393	0,0402	0,0234	0,0163	0,0099

Tablo 4.5.'ten de anlaşıldığı üzere OBU-R dağılımı, diğer dağılımlardan tüm kriterlere göre daha iyi sonuçlar elde edilmiştir. Öte yandan önerilen dağılımın görsel olarak değerlendirilmesi için Şekil 4.5.'te dağılımları oyf grafiği elde edilmiştir. Grafikten de anlaşılabacağı üzere OBU-R dağılımı ele alınan diğer dağılımlardan daha iyi modelleme başarısı göstermiştir.



Şekil 4.5. Olasılık yoğunluk fonksiyonları ve histogram grafiği (Rüzgar hızı verisi)

4.2. OBU-P Dağılımı ve İstatistiksel Özellikleri

Vilfredo Pareto'nun (1896) servet ve geliri modellemesi için önerdiği Pareto dağılımı, ağır kuyruk özellikleri nedeniyle aşırı yağış, rüzgâr ve aşırı kayıplar da dahil olmak üzere sağkalım verilerini, aktüeryal verileri ve aşırı değerlere sahip verileri modellemede yaygın olarak kullanılmaktadır. Ayrıca ekonomi, mühendislik, kuyruk teorisi gibi farklı alanlarda geniş uygulamalara sahiptir (Arnold, 2008). Pareto dağılımına sahip X rassal değişkeninin dağılım ve olasılık yoğunluk fonksiyonları sırasıyla (4.37) ve (4.38)'deki gibi ifade edilir.

$$G(x) = 1 - \left(\frac{x}{\theta} \right)^{-\alpha}, \quad x \geq \theta > 0, \alpha > 0 \quad (4.37)$$

$$g(x) = \frac{\alpha \theta^\alpha}{x^{\alpha+1}}, \quad (4.38)$$

Denklemlerde görülen θ ve α parametreleri sırasıyla ölçek ve şekil parametreleridir. Pareto dağılımının yaygın kullanımına rağmen, iki parametrelili Pareto dağılımı büyük bir esnekliğe sahip değildir. Bu sebeple, literatürde Pareto dağılımının çeşitli genelleştirmelerine ilişkin çalışmalar bulunmaktadır. Örneğin, Nadarajah (2005), üstelleştirilmiş Pareto dağılımını tanımlamış ve bazı istatistiksel özelliklerini incelemiştir. Akinsete ve ark. (2008), Pareto dağılımının genelleştirilmesi üzerine çalışmışlar ve beta-Pareto dağılımını önermişlerdir. Yine beta dağılımının temel dağılım olarak alındığı, beta-Pareto ve üstelleştirilmiş genelleştirilmiş Pareto gibi dağılımları kapsayan beta genelleştirilmiş Pareto dağılımı Mahmoudi (2011) tarafından önerilmiştir. Benzer şekilde beta üstelleştirilmiş Pareto dağılımı Zea ve ark. (2012) tarafından önerilmiş olup mesane kanseri verilerine uygulanmıştır. Öte yandan gamma-Pareto dağılımı T - X dağılım aileleri yöntemi kullanılarak Alzaatreh ve ark. (2013) tarafından tanımlanmıştır. Kumaraswamy-Pareto dağılımı ise Bourguignon ve ark. (2013) tarafından yaşam süresi verilerinin analizinde kullanılmak üzere önerilmiştir. Literatürde önerilmiş olan iki farklı Weibull-Pareto dağılımı bulunmaktadır. Bunlardan ilki Aljarrah ve ark. (2015) tarafından T - X dağılımlar ailesi yöntemine dayalı olarak dönüştürücü fonksiyonun log-lojistik dağılımının kantil fonksiyonu olarak seçilmesiyle tanımlanmış olup bir diğer Weibull-Pareto dağılımı ise Tahir ve ark. (2016) tarafından Bourguignon ve ark. (2014)'nin önerdiği yöntemine dayalı olarak tanımlanmıştır. Yine T - X dağılımlar ailesi yöntemine

dayalı olarak Godase ve ark. (2017) Pareto-Rayleigh dağılımı tanımlanmışlar ve dağılımın bazı istatistiksel özelliklerini incelenmiştir.

Bu bölümde, T-X dağılımlar ailesi yöntemine (Alizadeh ve ark., 2017) dayalı olarak Odd-Burr Pareto (*OBu-P*) dağılımı tanımlanmıştır. Tanımlanan yeni dağılım iki yeni şekil parametresine sahiptir. Elde edilen yeni dağılımın esnekliği, basıklık-çarpıklık değerlerinin yanı sıra olasılık yoğunluk ve tehlike oranı grafikleri yardımıyla ve gerçek veri seti üzerinde yapılan uygulamalar ile gösterilmiştir.

4.2.1. *OBu-P* dağılım ve olasılık yoğunluk fonksiyonları

(4.3) denkleminde $G(x)$ temel dağılımı yerine (4.37) ile ifade edilen Pareto dağılım fonksiyonu kullanılırsa *OBu-P* dağılımı (4.39)'daki şekilde elde edilir.

$$F(x) = 1 - \frac{\left(\frac{x}{\theta}\right)^{-\alpha ab}}{\left\{ \left[1 - \left(\frac{x}{\theta}\right)^{-\alpha}\right]^a + \left(\frac{x}{\theta}\right)^{-\alpha a} \right\}^b}, \quad a, b, \alpha > 0, x \geq \theta > 0 \quad (4.39)$$

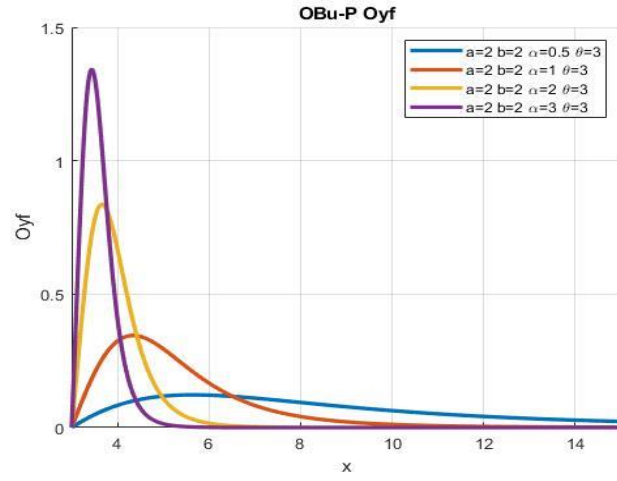
(4.39) dağılım fonksiyonuna karşılık gelen olasılık yoğunluk fonksiyonu ise;

$$f(x) = \frac{ab\alpha\theta^{\alpha ab} x^{-(\alpha ab+1)} \left[1 - \left(\frac{x}{\theta}\right)^{-\alpha}\right]^{a-1}}{\left\{ \left[1 - \left(\frac{x}{\theta}\right)^{-\alpha}\right]^a + \left(\frac{x}{\theta}\right)^{-\alpha a} \right\}^{b+1}}, \quad (4.40)$$

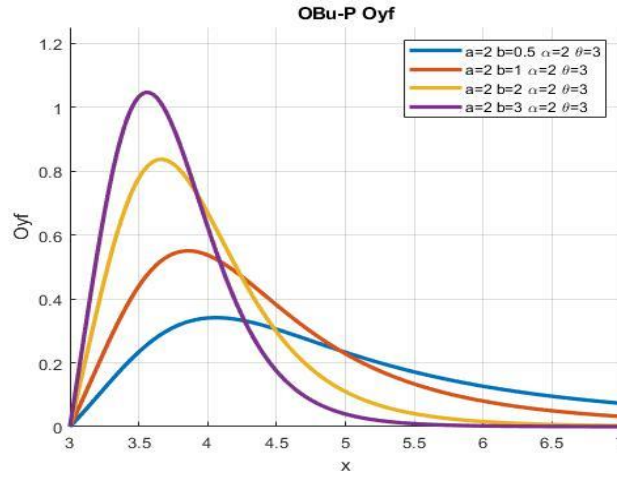
şeklinde ifade edilir. Burada θ ölçek parametresi, a , b ve α şekil parametreleridir. (4.40)

OBu-P oyf'na sahip olan X rassal değişkeni $X \sim OBu-P(a, b, \alpha, \theta)$ şeklinde ifade edilir.

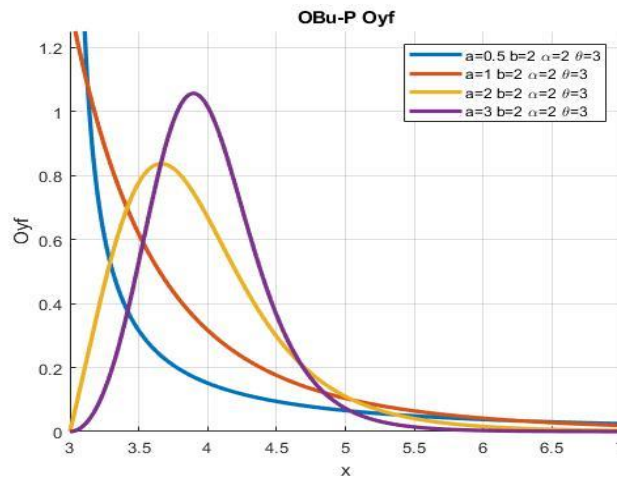
Farklı parametre değerleri için *OBu-P* dağılımının oyf'nu grafikleri Şekil 4.6'da verilmiştir. Grafiklerden görülebileceği gibi, önerilen *OBu-P* oyf, şekil parametrelerinin değerlerine bağlı olarak, farklı kuyruk davranışları ve çarpıklık ve/veya basıklık sunmaktadır.



(a)



(b)



(c)

Şekil 4.6. Farklı parametre değerleri için OBU-P dağılımının oyf grafiği

OBu-P rassal değişkeni için özel durumlar

a) $a=b=1$ olduğunda, açıkça görülmektedir ki *OBu-P* dağılımı Pareto dağılımına indirgenmektedir.

b) $a=1$ olduğunda, $OBu-P(a,b,\alpha,\theta)$ dağılımı şekil parametresi $k=b\alpha$ ve ölçek parametresi θ olan oyf aşağıda ifade edilen Pareto dağılımına dönüşmektedir.

$$g(x) = \frac{k\theta^k}{x^{k+1}}, \quad k > 0, \quad x \geq \theta > 0.$$

c) U rassal değişkeni $(0, 1)$ 'de düzgün dağılıma sahip olmak üzere,

$$X = \theta \left\{ \left[(1-U)^{-1/b} - 1 \right]^{1/a} + 1 \right\}^{1/\alpha}$$

rassal değişkeni $OBu-P(a,b,\alpha,\theta)$ dağılımına sahip olur.

d) Y rassal değişkeni $Y \sim BurrXII(a,b)$ dağılımına sahip olmak üzere,

$$X = \theta(Y+1)^{1/\alpha}$$

rassal değişkeni $OBu-P(a,b,\alpha,\theta)$ dağılımına sahip olur.

e) Y rassal değişkeni şekil parametresi b ve ölçek parametresi 1 olan Lomax dağılımına sahip olmak üzere,

$$X = \theta(Y^{1/a} + 1)^{1/\alpha}$$

rassal değişkeni $OBu-P(a,b,\alpha,\theta)$ dağılımına sahip olur.

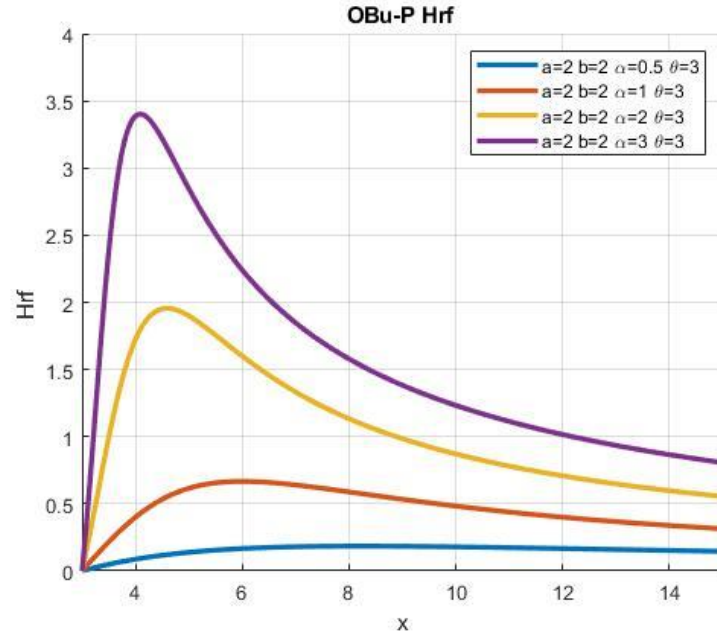
4.2.2. *OBu-P* yaşam ve tehlike oranı fonksiyonları

OBu-P dağılımının yaşam ve tehlike oranı fonksiyonları sırasıyla (4.41) ve (4.42)'de ifade edilmiştir.

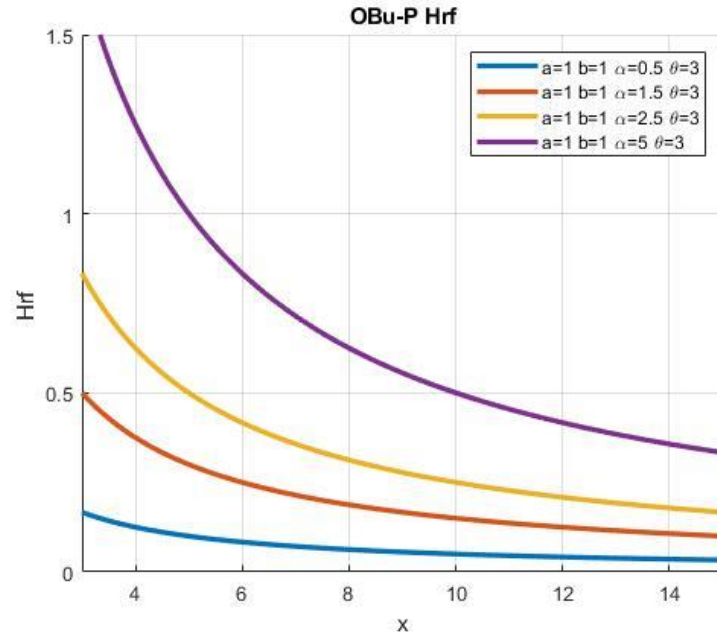
$$S(x) = \frac{\left(\frac{x}{\theta}\right)^{-\alpha ab}}{\left\{ \left[1 - \left(\frac{x}{\theta}\right)^{-\alpha} \right]^a + \left(\frac{x}{\theta}\right)^{-\alpha a} \right\}^b} \quad (4.41)$$

$$h(x) = \frac{ab\alpha \left[1 - \left(\frac{x}{\theta} \right)^{-\alpha} \right]^{a-1}}{x \left\{ \left[1 - \left(\frac{x}{\theta} \right)^{-\alpha} \right]^a + \left(\frac{x}{\theta} \right)^{-\alpha a} \right\}} \quad (4.42)$$

Farklı parametre değerleri için tehlike oranı fonksiyonunun grafikleri Şekil 4.7 'de verilmiştir. Bu grafiklerden görüldüğü gibi, *OBu-P* dağılımının tehlike oranı fonksiyonu, farklı parametre değerlerine bağlı olarak monoton azalan ve artan-azalan (tek modlu) gibi farklı şekillere sahiptir. *OBu-P* dağılımının tehlike fonksiyonunun sahip olduğu bu esneklik, onun yaşam süresi verilerini analiz etmek için uygun olduğunu göstermektedir.



(a)



(b)

Şekil 4.7. Farklı parametre değerleri için OBU-P dağılımının tehlike oranı fonksiyonu grafiği

4.2.3. OBU-P'nin asimptotik özellikleri

Dağılımın, olasılık yoğunluk ve tehlike oranı fonksiyonlarının asimptotik durumları aşağıdaki teoremlerde ifade edilmiştir (Alizadeh ve ark., 2017).

Teorem 4.2: $x \rightarrow \infty$ iken $OBu-P$ dağılımının olasılık yoğunluk fonksiyonu 0 olup $x \rightarrow \theta$ iken asimptotik durumları (4.43)'teki gibidir:

$$\lim_{x \rightarrow \theta} f(x) = \begin{cases} \infty, & 0 < a < 1 \\ \frac{ab\alpha}{\theta}, & a = 1 \\ 0, & a > 1 \end{cases} \quad (4.43)$$

Kanıt 4.2: (4.40) denklemi ile ifade edilen dağılımın oyf'si incelendiğinde yukarıdaki limit durumlarının elde edilebileceği açıkça görülmektedir.

Teorem 4.3: $x \rightarrow \infty$ iken $OBu-P$ dağılımının tehlike oranı fonksiyonu 0 olup $x \rightarrow \theta$ iken asimptotik durumları (4.44)'teki gibidir:

$$\lim_{x \rightarrow \theta} h(x) = \begin{cases} \infty, & 0 < a < 1 \\ \frac{ab\alpha}{\theta}, & a = 1 \\ 0, & a > 1 \end{cases} \quad (4.44)$$

Kanıt 4.3: (4.42) denklemi ile ifade edilen dağılımın tehlike oranı fonksiyonu incelendiğinde yukarıdaki limit durumlarının elde edilebileceği açıkça görülmektedir.

4.2.4. $OBu-P$ 'nin kantil fonksiyonu

(4.39) denkleminde ifade edilen dağılım fonksiyonuna dayalı $OBu-P$ dağılımının kantil fonksiyonu (4.45)'teki gibi elde edilir:

$$F(x) = 1 - \frac{\left(\frac{x}{\theta}\right)^{-\alpha ab}}{\left\{ \left[1 - \left(\frac{x}{\theta}\right)^{-\alpha} \right]^a + \left(\frac{x}{\theta}\right)^{-\alpha a} \right\}^b}$$

$$[1 - F(x)]^{1/b} = \frac{\left(\frac{x}{\theta}\right)^{-\alpha a}}{\left\{ \left[1 - \left(\frac{x}{\theta}\right)^{-\alpha} \right]^a + \left(\frac{x}{\theta}\right)^{-\alpha a} \right\}}$$

$$\left\{ \left[1 - F(x) \right]^{-1/b} - 1 \right\}^{1/a} + 1 = \left(\frac{x}{\theta} \right)^\alpha$$

$$X = \theta \left\{ \left[\left(1 - F(x) \right)^{-1/b} - 1 \right]^{1/a} + 1 \right\}^{1/\alpha} \quad (4.45)$$

OBu-P rassal değişkeni $F(x) = u \sim U(0,1)$ alınarak (4.45) kantil fonksiyonu yardımıyla kolaylıkla üretilebilir.

4.2.5. *OBu-P*'nin alternatif matematiksel gösterimi

Bu bölümde *OBu-R* dağılımında olduğu gibi *OBu-P* dağılımının, dağılım ve olasılık yoğunluk fonksiyonlarının kuvvet serilerine dayalı alternatif gösterimleri elde edilecektir. Bir önceki bölümde izlenilen yöntemlerin bu dağılım için aynı şekilde uygulanması sonucunda *OBu-P* dağılım ve olasılık yoğunluk fonksiyonları için alternatif gösterimler:

$$F(x) = \sum_{k=0}^{\infty} n_k G(x)^k = \sum_{k=0}^{\infty} n_k H_k(x) \quad (4.46)$$

$$f(x) = \sum_{k=0}^{\infty} n_{k+1} h_{k+1}(x) \quad (4.47)$$

şeklinde elde edilir. Burada $H_k(x)$ ölçek parametresi θ ve şekil parametreleri k ve α

olan üstelleştirilmiş Pareto (*EP*) dağılımı olup $m_k = \sum_{i=0}^{\infty} (-1)^i \binom{b}{i} \lambda_k$, $n_0 = 1 - m_0$ ve

$n_k = -m_k$ şeklindedir. *OBu-R* dağılımında olduğu gibi (4.46) eşitliğine göre yeni elde edilen dağılım fonksiyonu, temel alınan dağılım fonksiyonunun ağırlıklandırılmış kuvvet serisi toplamı olarak ifade edilmiştir. Bu sonuç *OBu-P* dağılımının moment, tamamlanmamış moment ve sıra istatistikleri gibi bazı istatistiksel özelliklerinin elde edilmesinde kullanılacaktır.

4.2.6. *OBu-P*'nin momentleri

Bu bölümde *OBu-P* dağılımının moment fonksiyonu (4.47) eşitliği yardımıyla elde edilecektir. Bu amaçla üstelleştirilmiş Pareto dağılımının r 'inci momenti (4.48)'deki gibi ifade edilir:

$$\mu'_{EP} = E_{EP}(Z^r) = \alpha\beta\theta^\alpha \int_{\theta}^{\infty} z^{r-\alpha-1} \left(1 - \left(\frac{z}{\theta}\right)^{-\alpha}\right)^{\beta-1} dz \quad (4.48)$$

burada α ve β şekil parametreleri olup θ ise ölçek parametresidir. Bir dizi cebirsel işlemin ardından üstelleştirilmiş Pareto dağılımının r 'inci momenti,

$$\mu'_{EP} = E_{EP}(Z^r) = \beta\theta^r B\left(1 - \frac{r}{\alpha}, \beta\right), \quad r \leq \alpha \quad (4.49)$$

Burada $B(x, y) = \int_0^1 t^{(x-1)}(1-t)^{(y-1)} dt$ beta fonksiyonunu ifade etmektedir. Sonuç olarak

(4.47) eşitliği ile ifade edilen ilişki yardımıyla $OBu-P$ dağılımının r 'inci momenti (4.50)'deki gibi ifade edilebilir:

$$\begin{aligned} \mu'_{OBu-R} = E_{OBu-R}(X^r) &= \sum_{k=0}^{\infty} n_{k+1} \int_{\theta}^{\infty} x^r h_{(k+1)}(x) dx \\ &= \theta^r \sum_{k=0}^{\infty} n_{k+1} (k+1) B\left(1 - \frac{r}{\alpha}, (k+1)\right), \quad r \leq \alpha \end{aligned} \quad (4.50)$$

$r=1$ olarak alındığında $OBu-P$ dağılımının birinci momenti yani ortalaması aşağıdaki eşitlik yardımıyla elde edilebilir:

$$\mu'_1 = \theta \sum_{k=0}^{\infty} n_{k+1} (k+1) B\left(1 - \frac{1}{\alpha}, (k+1)\right).$$

X 'in j 'inci merkezi momenti (4.50) eşitliği yardımıyla aşağıdaki denklemden kolaylıkla elde edilebilir:

$$\mu_j = E[(X - \mu)^j] = \sum_{k=0}^j \binom{j}{k} \mu'_j (-\mu'_1)^{j-k}.$$

Farklı parametre değerleri için $OBu-P$ dağılımının ortalama ve varyansı ile basıklık (S) ve çarpıklık (K) değerleri sırasıyla Tablo 4.6. ve Tablo 4.7.'de verilmiştir. Çarpıklık ve basıklık değerlerinin her ikisi de a , b ve α parametrelerinin azalan fonksiyonudur. Öte yandan Tablo 4.7. incelendiğinde $OBu-P$ dağılımının olası tüm S ve K kombinasyonlarının bir önceki bölümde ifade edilen (4.26) eşitsizliğini sağladığı görülmektedir.

Tablo 4.6. *OBu-P dağılımı için ortalama ve varyans değerleri*

a	α	θ	$b=0.75$		$b=2$		$b=5$	
			μ	σ^2	μ	σ^2	μ	σ^2
0,7500	5,0000	1,0000	1,5136	0,9862	1,1178	0,0297	1,0321	0,0022
		2,0000	3,0253	3,7077	2,2355	0,1189	2,0642	0,0086
		3,0000	4,5338	7,8742	3,3533	0,2676	3,0962	0,0194
	7,5000	1,0000	1,2876	0,1879	1,0747	0,0107	1,0211	0,0009
		2,0000	2,5753	0,7508	2,1493	0,0428	2,0421	0,0036
		3,0000	3,8629	1,6853	3,2240	0,0964	3,0632	0,0081
	10,0000	1,0000	1,1997	0,0748	1,0547	0,0055	1,0157	0,0005
		2,0000	2,3993	0,2990	2,1093	0,0218	2,0313	0,0020
		3,0000	3,5990	0,6728	3,1640	0,0491	3,0470	0,0044
2,0000	5,0000	1,0000	1,2169	0,0295	1,1148	0,0042	1,0715	0,0013
		2,0000	2,4339	0,1179	2,2297	0,0169	2,1431	0,0053
		3,0000	3,6508	0,2652	3,3445	0,0379	3,2146	0,0120
	7,5000	1,0000	1,1376	0,0104	1,0748	0,0017	1,0470	0,0006
		2,0000	2,2753	0,0414	2,1495	0,0068	2,0940	0,0022
		3,0000	3,4129	0,0932	3,2243	0,0153	3,1410	0,0051
	10,0000	1,0000	1,1008	0,0052	1,0554	0,0009	1,0350	0,0003
		2,0000	2,2016	0,0209	2,1108	0,0037	2,0700	0,0012
		3,0000	3,3024	0,0469	3,1663	0,0082	3,1050	0,0028
5,0000	5,0000	1,0000	1,1666	0,0029	1,1300	0,0009	1,1087	0,0005
		2,0000	2,3333	0,0116	2,2600	0,0036	2,2175	0,0020
		3,0000	3,4999	0,0260	3,3899	0,0080	3,3262	0,0046
	7,5000	1,0000	1,1080	0,0011	1,0848	0,0004	1,0712	0,0002
		2,0000	2,2159	0,0045	2,1696	0,0015	2,1424	0,0009
		3,0000	3,3239	0,0102	3,2544	0,0033	3,2136	0,0019
	10,0000	1,0000	1,0798	0,0006	1,0629	0,0002	1,0529	0,0001
		2,0000	2,1597	0,0024	2,1258	0,0008	2,1058	0,0005
		3,0000	3,2395	0,0054	3,1887	0,0018	3,1587	0,0010

Tablo 4.7. *OBu-P dağılımı için S ve K değerleri*

<i>a</i>	<i>α</i>	<i>b=0,75</i>		<i>b=2,00</i>		<i>b=5,00</i>	
		<i>S</i>	<i>K</i>	<i>S</i>	<i>K</i>	<i>S</i>	<i>K</i>
0,75	5,00	12,6624	477,9953	3,7463	30,3627	3,4758	23,1607
	7,50	5,9603	140,1287	3,1557	19,9182	3,2933	20,4344
	10,00	4,0019	44,3707	2,9285	16,8140	3,2120	19,1609
2,00	5,00	3,1139	25,6883	1,2533	6,3108	0,7112	3,6957
	7,50	2,5483	16,4506	1,1567	5,7384	0,6751	3,5900
	10,00	2,3278	13,7372	1,1109	5,4886	0,6574	3,5400
5,00	5,00	1,5691	9,4037	0,2030	3,6052	-0,2119	3,1115
	7,50	1,4489	8,4355	0,1695	3,5644	-0,2331	3,1253
	10,00	1,3924	8,0184	0,1529	3,5460	-0,2437	3,1328

4.2.7. OBU-P'nin tamamlanmamış momentleri

Momentlerin elde edilışinde olduđu gibi üstelleřtirilmiř Pareto dađılımlarının tamamlanmamıř moment fonksiyonundan yararlanılacaktır. Bu amaçla üstelleřtirilmiř Pareto dađılımlarının *r*'inci tamamlanmamıř momenti,

$$m_{r,z} = \int_{\theta}^w z^r h(z) dz = \beta \theta^r B_w \left(1 - \frac{r}{\alpha}, \beta \right), \quad r \leq \alpha \quad (4.51)$$

burada $B_w(x, y) = \int_0^w t^{(x-1)} (1-t)^{(y-1)} dt$ ifadesi tamamlanmamıř beta fonksiyonudur.

Böylece, OBU-P dađılımlarının *r*'inci tamamlanmamıř momenti (4.52)'deki formda elde edilebilir:

$$m_{r,x}(w) = \theta^r \sum_{k=0}^{\infty} n_{k+1} (k+1) B_w \left(1 - \frac{r}{\alpha}, (k+1) \right), \quad r \leq \alpha. \quad (4.52)$$

4.2.8. OBU-P'nin moment üreten fonksiyonu

EP dađılımlarının moment üreten fonksiyonu ařađıdaki gibi elde edilir (Tahir ve ark., 2016):

$$M_{EP}(t) = \alpha \beta \sum_{i=0}^{\infty} (-1)^i \theta^{(i+1)} \binom{\beta-1}{i} J(\theta, (i+1)\alpha + 1, t)$$

Buradaki $J(.)$ fonksiyonu ařađıdaki gibidir:

$$J(q, p, t) = \int_q^\infty x^{-p} e^{tx} dx$$

$$= (-t)^p q \left[-\frac{\pi \csc(p\pi)}{tq\Gamma(p)} - \frac{p\Gamma(-p)}{qt} + \frac{e^{tq}}{(-t)^{p+1} q^{p+1}} + \frac{p\Gamma(-p, -tq)}{tq} \right]$$

$\Gamma(\alpha, t) = \int_t^\infty x^{\alpha-1} e^{-x} dx$ ifadesi tamamlanmamış gamma fonksiyonu olup *OBu-P* dağılımının moment üreten fonksiyonu (4.47) eşitliği yardımıyla (4.53)'teki gibi ifade edilir:

$$M(t) = \alpha \sum_{k,i=0}^{\infty} n_{k+1} (k+1) \theta^{(i+1)\alpha} \binom{k}{i} J(\theta, (i+1)\alpha + 1, t). \quad (4.53)$$

4.2.9. *OBu-P*'nin ortalamadan sapması

Ortalamadan sapma ve medyandan sapma, bir popülasyonda yayılma ölçüsü için alternatif ölçülerdir (Akinsete ve ark., 2008). X *OBu-P* dağılımına ait $\mu = E(X)$ ortalamalı bir rassal değişken olmak üzere ortalamadan sapma (4.54)'deki gibi hesaplanabilir:

$$\begin{aligned} D(\mu) &= E[|X - \mu|] = \int_{\theta}^{\infty} |x - \mu| f(x) dx = \int_{\theta}^{\mu} (\mu - x) f(x) dx + \int_{\mu}^{\infty} (x - \mu) f(x) dx \\ &= \int_{\theta}^{\mu} \mu f(x) dx - \int_{\theta}^{\mu} xf(x) dx + \int_{\mu}^{\infty} xf(x) dx - \int_{\mu}^{\infty} \mu f(x) dx \\ &= \mu F(x) - \int_{\theta}^{\mu} xf(x) dx + \left[\mu - \int_{\theta}^{\mu} xf(x) dx \right] - \left[1 - \int_{\theta}^{\mu} \mu f(x) dx \right] \\ &= 2\mu F(x) - 2 \int_{\theta}^{\mu} xf(x) dx + \mu - 1 \\ D(\mu) &= 2\mu F(x) - 2m_{1,X}(\mu) + \mu - 1, \end{aligned} \quad (4.54)$$

Burada $m_{1,X}(\mu)$ *OBu-P* dağılımının birinci tamamlanmamış moment fonksiyonu olup (4.52) eşitliği ile ifade edilmiştir.

M , X rassal değişkeninin medyanı olmak üzere, medyandan ortalama sapma (4.55)'deki gibi hesaplanır:

$$\begin{aligned}
D(M) &= E[|X - M|] = \int_{\theta}^{\infty} |x - M| f(x) dx = \int_{\theta}^M (M - x) f(x) dx + \int_M^{\infty} (x - M) f(x) dx \\
&= 2 \int_{\theta}^M (M - x) f(x) dx + \int_{\theta}^{\infty} (x - M) f(x) dx \\
&= 2 \int_{\theta}^M M f(x) dx - 2 \int_{\theta}^M x f(x) dx + E(X - M) \\
&= 2MF(M) - 2m_{1,X}(M) + \mu - M \\
D(M) &= \mu - 2m_{1,X}(M).
\end{aligned}$$

(4.55)

4.2.10. OBU-P'nin sıra istatistiği

$X_1, X_2, \dots, X_n$ OBU-P dağılımından n birimlik örneklem olmak üzere $X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(j)} \leq \dots \leq X_{(n)}$, bu örneklemde sıra istatistikleri olsun. $X_{(j)}$ sıra istatistiğinin oyf,

$$\begin{aligned}
f_{X_{(j)}}(x) &= \frac{n!}{(n-j)!(j-1)!} f(x) [F(x)]^{j-1} [1-F(x)]^{n-j} \\
&= \frac{n!}{(n-j)!(j-1)!} \sum_{i=0}^{n-j} (-1)^i \binom{n-j}{i} [F(x)]^{j+i-1} f(x).
\end{aligned}$$

Yukarıda ifade edilen fonksiyon (4.46) ve (4.47) eşitliklerinden yararlanarak aşağıdaki formda da yazılabilir.

$$\begin{aligned}
f_{X_{(j)}}(x) &= \frac{n!}{(n-j)!(j-1)!} \\
&\sum_{i=0}^{n-j} (-1)^i \binom{n-j}{i} \left[\sum_{l=0}^{\infty} (l+1) b_{l+1} g(x) G(x)^l \right] \left[\sum_{k=0}^{\infty} b_k G(x)^k \right]^{j+i-1}
\end{aligned}$$

Kuvvet serinin kuvveti ile ilgili eşitlik yardımıyla (Gradshteyn, ve Ryzhik, 2002) $X_{(j)}$ sıra istatistiğinin oyf,

$$\begin{aligned}
f_{X_{(j)}}(x) &= \frac{n!}{(n-j)!(j-1)!} \\
&\sum_{i=0}^{n-j} (-1)^i \binom{n-j}{i} \left[\sum_{l=0}^{\infty} (l+1) b_{l+1} g(x) G(x)^l \right] \left[\sum_{k=0}^{\infty} a_k G(x)^k \right]
\end{aligned}$$

olarak yazılabilir. Bu eşitlikteki $a_0 = b_0^{j+i-1}$ ve $k \geq 1$ için

$a_m = (mb_0)^{-1} \sum_{k=1}^m [k(j+i) - m] b_k a_{m-k}$ şeklindedir. Bir dizi cebirsel işlemin ardından oyf,

$$\begin{aligned} f_{x_{(j)}}(x) &= \sum_{i=0}^{n-j} \sum_{k,l=0}^{\infty} \frac{n!}{(n-j)!(j-1)!} (-1)^i \binom{n-j}{i} (l+1) b_{l+1} a_k g(x) G(x)^{l+k} \\ &= \sum_{i=0}^{n-j} \sum_{k,l=0}^{\infty} \frac{n!}{(n-j)!(j-1)!} (-1)^i \binom{n-j}{i} \frac{(l+1) b_{l+1} a_k}{(l+k+1)} h(x)^{l+k+1} \end{aligned}$$

şeklinde elde edilebilir.

$$d_{i,k,l} = \frac{n!}{(n-j)!(j-1)!} (-1)^i \binom{n-j}{i} \frac{(l+1) b_{l+1} a_k}{(l+k+1)}$$

olarak alınsın. Sonuç olarak $X_{(j)}$ sıra istatistiğinin oyf,

$$f_{x_{(j)}}(x) = \sum_{i=0}^{n-j} \sum_{k,l=0}^{\infty} d_{i,k,l} h_{(l+k+1)}(x) \quad (4.56)$$

Bu eşitlikteki $h(x)$ şekil parametreleri $(l+k+1)$ ve α , ölçek parametresi θ olan üstelleştirilmiş Pareto dağılımının oyf'dur. Sonuç olarak *OBu-P* dağılımının sıra istatistiğinin oyf $h(x)$ 'in lineer kombinasyonu olarak ifade edilebilir.

4.2.11. *OBu-P*'nin en çok olabilirlik tahmincisi

$X_1, X_2, \dots, X_n$ *OBu-P* dağılımından örneklem olmak üzere, a , b , α ve θ parametrelerine göre log-olabilirlik fonksiyonu,

$$\begin{aligned} \log L(a, b, \alpha, \theta) &= n \log(ab\alpha) + nab\alpha \log \theta - (ab\alpha + 1) \sum_{i=1}^n \log x_i \\ &+ (a-1) \sum_{i=1}^n \log \left[1 - \left(\frac{x_i}{\theta} \right)^{-\alpha} \right] - (b+1) \sum_{i=1}^n \log \left\{ \left[1 - \left(\frac{x_i}{\theta} \right)^{-\alpha} \right]^a + \left(\frac{x_i}{\theta} \right)^{-\alpha a} \right\}. \end{aligned} \quad (4.57)$$

şeklinde ifade edilir. Bilindiği üzere parametrelerin olabilirlik tahmincileri log-olabilirlik fonksiyonunun parametrelere göre kısmi türevlerinin sıfıra eşitlenmesi ile elde edilir. Bu amaçla eşitliğin parametrelere göre kısmi türevleri aşağıdaki gibi elde edilmiştir.

$$\frac{\partial \log L}{\partial a} = \frac{n}{a} + nb\alpha \log \theta - b\alpha \sum_{i=1}^n \log x_i + \sum_{i=1}^n \log \left[1 - \left(\frac{x_i}{\theta} \right)^{-\alpha} \right] - (b+1) \sum_{i=1}^n \frac{\left[1 - \left(\frac{x_i}{\theta} \right)^{-\alpha} \right]^a \log \left[1 - \left(\frac{x_i}{\theta} \right)^{-\alpha} \right] - \alpha \log \left(\frac{x}{\theta} \right) \left(\frac{x}{\theta} \right)^{-a\alpha}}{\left[1 - \left(\frac{x_i}{\theta} \right)^{-\alpha} \right]^a + \left(\frac{x}{\theta} \right)^{-a\alpha}} = 0$$

$$\frac{\partial \log L}{\partial b} = \frac{n}{b} + na\alpha \log \theta - a\alpha \sum_{i=1}^n \log x_i - \sum_{i=1}^n \log \left\{ \left[1 - \left(\frac{x_i}{\theta} \right)^{-\alpha} \right]^a + \left(\frac{x_i}{\theta} \right)^{-a\alpha} \right\} = 0$$

$$\frac{\partial \log L}{\partial \alpha} = \frac{n}{\alpha} + nab \log \theta - ab \sum_{i=1}^n \log x_i + (a-1) \sum_{i=1}^n \frac{\left(\frac{x_i}{\theta} \right)^{-\alpha} \log \left(\frac{x_i}{\theta} \right)}{\left[1 - \left(\frac{x_i}{\theta} \right)^{-\alpha} \right]} - (b+1) \sum_{i=1}^n \frac{a \log \left(\frac{x_i}{\theta} \right) \left[1 - \left(\frac{x_i}{\theta} \right)^{-\alpha} \right]^{a-1} \left(\frac{x_i}{\theta} \right)^{-\alpha} - a \log \left(\frac{x_i}{\theta} \right) \left(\frac{x_i}{\theta} \right)^{a\alpha}}{\left(\frac{x_i}{\theta} \right)^{-a\alpha} + \left[1 - \left(\frac{x_i}{\theta} \right)^{-\alpha} \right]^a} = 0$$

$x \geq \theta$ olduğundan θ parametresinin en çok olabilirlik tahmincisi örneklemdeki en küçük değere karşılık gelir başka bir ifade ile örneklemin birinci sıralı istatistiği olan $x_{(1)}$ 'dir. Dağılımın diğer parametrelerin en çok olabilirlik tahminleri, yukarıda verilen denklemlerin eş zamanlı çözümü ile elde edilebilir. Bu amaçla denklemlerin açık çözümü mümkün olmadığından Newton-Raphson gibi nümerik yöntemlerin kullanılması gerekir.

Hipotez testi ve aralık tahmini için gözlemlenmiş bilgi matrisinin tersi alınarak elde edilen asimptotik varyans-kovaryans matrisi kullanılır. Gözlemlenmiş bilgi matrisinin tersi,

$$I(\hat{\beta}) = \begin{bmatrix} -\frac{\partial^2 \log L}{\partial a^2} & -\frac{\partial^2 \log L}{\partial a \partial b} & -\frac{\partial^2 \log L}{\partial a \partial \alpha} \\ -\frac{\partial^2 \log L}{\partial b \partial a} & -\frac{\partial^2 \log L}{\partial b^2} & -\frac{\partial^2 \log L}{\partial b \partial \alpha} \\ -\frac{\partial^2 \log L}{\partial a \partial \alpha} & -\frac{\partial^2 \log L}{\partial b \partial \alpha} & -\frac{\partial^2 \log L}{\partial \alpha^2} \end{bmatrix}^{-1} \quad (4.58)$$

şeklinde olup buradaki $\hat{\beta} = (\hat{a}, \hat{b}, \hat{\alpha})$, tahmin edilen parametre vektörüdür. (4.58)

matrisinin elemanları aşağıda ifade edildiği gibidir:

$$\frac{\partial^2 \log L}{\partial a^2} = -\frac{n}{a^2} - (b+1) \sum_{i=1}^n \frac{\left(\frac{x_i}{\theta}\right)^{aa} u_i^a \left\{ \log^2 u_i + 2\alpha \log\left(\frac{x_i}{\theta}\right) \log u_i + \alpha^2 \log^2\left(\frac{x_i}{\theta}\right) \right\}}{\left\{ \left(\frac{x_i}{\theta}\right)^{aa} u_i^a + 1 \right\}^2},$$

$$\frac{\partial^2 \log L}{\partial b^2} = -\frac{n}{b^2},$$

$$\begin{aligned} \frac{\partial^2 \log L}{\partial \alpha^2} = & -\frac{n}{\alpha^2} - (a-1) \sum_{i=1}^n \left[\log^2\left(\frac{x_i}{\theta}\right) (1-u_i) u_i^{-1} + \log^2\left(\frac{x_i}{\theta}\right) (1-u_i)^2 u_i^{-2} \right] \\ & - (b+1) \sum_{i=1}^n \left\{ \frac{-a \log^2\left(\frac{x_i}{\theta}\right) u_i^{a-1} (1-u_i) + (a-1) \log^2\left(\frac{x_i}{\theta}\right) u_i^{a-2} (1-u_i)^2 - a^2 \left(\frac{x_i}{\theta}\right)^{aa} \log^2\left(\frac{x_i}{\theta}\right)}{(1-u_i)^a + u_i^a} \right. \\ & \left. \left[\frac{a \log\left(\frac{x_i}{\theta}\right) u_i^{a-1} (1-u_i) - a \log\left(\frac{x_i}{\theta}\right) (1-u_i)^a}{[(1-u_i) + u_i^a]^2} \right] \left[\frac{a \log\left(\frac{x_i}{\theta}\right) u_i^{a-1} (1-u_i) - a \left(\frac{x_i}{\theta}\right)^{aa} \log\left(\frac{x_i}{\theta}\right)}{[(1-u_i) + u_i^a]^2} \right] \right\}, \end{aligned}$$

$$\frac{\partial^2 \log L}{\partial a \partial b} = n\alpha \log \theta - \alpha \sum_{i=1}^n \log x_i - \sum_{i=1}^n \frac{u_i^a \log u_i - \alpha \log\left(\frac{x_i}{\theta}\right) (1-u_i)^a}{u_i^a + (1-u_i)^a},$$

$$\begin{aligned} \frac{\partial^2 \log L}{\partial a \partial \alpha} = & nb \log \theta - b \sum_{i=1}^n \log x_i + \sum_{i=1}^n \log\left(\frac{x_i}{\theta}\right) (1-u_i) u_i^{-1} \\ & - (b+1) \sum_{i=1}^n \left\{ \frac{a \log^2\left(\frac{x_i}{\theta}\right) \alpha \left(\frac{x_i}{\theta}\right)^{-aa} + a \log\left(\frac{x_i}{\theta}\right) \log u_i u_i^{a-1} (1-u_i) + \log\left(\frac{x_i}{\theta}\right) u_i^{a-1} (1-u_i) - \log\left(\frac{x_i}{\theta}\right) (1-u_i)^a}{(1-u_i)^a + u_i^a} \right. \\ & \left. \left[\frac{a \log\left(\frac{x_i}{\theta}\right) u_i^{a-1} (1-u_i) - a \log\left(\frac{x_i}{\theta}\right) (1-u_i)^a}{[(1-u_i)^a + u_i^a]^2} \right] \left[\log u_i u_i^{a-1} - \alpha \log\left(\frac{x_i}{\theta}\right) (1-u_i)^a \right] \right\}, \end{aligned}$$

$$\frac{\partial^2 \log L}{\partial \alpha \partial b} = na \log \theta - a \sum_{i=1}^n \log x_i - \sum_{i=1}^n \frac{a \log\left(\frac{x_i}{\theta}\right) u_i^{a-1} (1-u_i) - a \log\left(\frac{x_i}{\theta}\right) \left(\frac{x_i}{\theta}\right)^{aa}}{(1-u_i)^a + u_i^a},$$

Yukarıdaki eşitliklerde ifade edilen u_i ,

$$u_i = 1 - \left(\frac{x_i}{\theta} \right)^{-\alpha}$$

şeklindedir.

4.2.12. *OBu-P*'nin en çok olabilirlik tahmincisi için benzetim çalışması

Bu bölümde, *OBu-P* dağılımı için EÇO yönteminin performansını değerlendirmek amacıyla bir benzetim çalışması yapılmıştır. *OBu-P* dağılımından rassal değişkenler üretmek için dağılımın kantil fonksiyonu kullanılmıştır. Her bir parametre değeri için örneklem büyüklükleri, $n=50$, 100 ve 500 olarak ele alınmıştır. Her bir durum için benzetim sonuçları 1000 kez tekrarlanarak elde edilmiştir. Elde edilen sonuçların değerlendirilmesi için, hata kareler ortalaması (*MSE*) ve yanlılık (*Bias*) kriterlerinden yararlanılmıştır.

Benzetim sonuçları Tablo 4.8.'de verilmiştir. Tablodaki sonuçlardan görüldüğü üzere örneklem büyüklüğü n arttıkça *MSE* ve *Bias* değerleri azalmaktadır. Bu, EÇO yönteminin, *OBu-P* dağılımının parametrelerini tahmin etmek için etkin bir şekilde uygulanabileceği anlamına gelmektedir.

Tablo 4.8. $OBu-P$ için MLE 'nin Bias ve MSE değerleri

n	a	b	α	BIAS				MSE			
				$\hat{a}$	$\hat{b}$	$\hat{\alpha}$	$\hat{\theta}$	$\hat{a}$	$\hat{b}$	$\hat{\alpha}$	$\hat{\theta}$
50	0,5	0,5	0,5	-0,0723	-0,1485	-0,1301	-0,0075	0,0218	0,3586	2,2202	0,0004
		0,5	1	-0,0770	-0,1429	-0,2300	-0,0032	0,0231	0,2453	3,2005	0,0001
		0,5	2	-0,0681	-0,1261	-0,4452	-0,0019	0,0188	0,2944	6,3371	0,0000
		1	0,5	-0,0447	-0,1349	-0,1108	-0,0016	0,0101	0,4553	0,2526	0,0000
		1	1	-0,0480	-0,1509	-0,1789	-0,0009	0,0099	0,4364	0,9482	0,0000
		1	2	-0,0440	-0,1479	-0,3103	-0,0004	0,0094	0,4415	2,5061	0,0000
	0,5	2	0,5	-0,0276	-0,0001	-0,2180	-0,0004	0,0049	0,7441	0,3174	0,0000
		2	1	-0,0261	-0,0112	-0,4327	-0,0002	0,0050	0,8103	1,2746	0,0000
		2	2	-0,0267	-0,0786	-0,8549	-0,0001	0,0050	0,9207	10,6907	0,0000
	100	0,5	0,5	-0,0450	-0,0947	-0,0327	-0,0017	0,0097	0,1912	0,1854	0,0000
		0,5	1	-0,0466	-0,0787	-0,0341	-0,0009	0,0096	0,1539	0,4484	0,0000
		0,5	2	-0,0455	-0,0688	-0,1040	-0,0004	0,0090	0,0840	2,0006	0,0000
		1	0,5	-0,0226	-0,0922	-0,0288	-0,0004	0,0041	0,1981	0,0651	0,0000
		1	1	-0,0266	-0,1094	-0,0274	-0,0002	0,0042	0,2076	0,2391	0,0000
		1	2	-0,0229	-0,0851	-0,1035	-0,0001	0,0042	0,1654	1,0007	0,0000
500	0,5	2	0,5	-0,0146	-0,0465	-0,0997	-0,0001	0,0021	0,4907	0,0944	0,0000
		2	1	-0,0172	-0,0394	-0,2162	-0,0001	0,0026	0,5030	0,4600	0,0000
		2	2	-0,0155	-0,0398	-0,4361	0,0000	0,0023	0,4987	1,7891	0,0000
	0,5	0,5	0,5	-0,0134	-0,0110	0,0016	-0,0001	0,0016	0,0087	0,0131	0,0000
		0,5	1	-0,0106	-0,0122	0,0049	0,0000	0,0013	0,0075	0,0466	0,0000
		0,5	2	-0,0126	-0,0146	0,0262	0,0000	0,0014	0,0076	0,1960	0,0000
		1	0,5	-0,0078	-0,0247	0,0018	0,0000	0,0008	0,0248	0,0116	0,0000
		1	1	-0,0060	-0,0159	-0,0067	0,0000	0,0007	0,0229	0,0425	0,0000
		1	2	-0,0083	-0,0161	-0,0037	0,0000	0,0008	0,0222	0,1617	0,0000
	0,5	2	0,5	-0,0050	-0,0354	-0,0102	0,0000	0,0004	0,1066	0,0159	0,0000
		2	1	-0,0043	-0,0323	-0,0294	0,0000	0,0005	0,1008	0,0641	0,0000
		2	2	-0,0039	-0,0245	-0,0656	0,0000	0,0005	0,1147	0,2668	0,0000

4.2.13. *OBu-P*'nin uygulamaları

Bu bölümde, *OBu-P* dağılımının performansını iki ayrı gerçek veri seti üzerinde incelenecektir.

Uygulama I

İlk veri seti Zea ve ark. (2012) ve Aljarrah ve ark. (2015) tarafından da kullanılmış olan mesane kanseri hastalarının hastalısız dönem sürelerini ifade eden 128 adet gözlemden oluşmaktadır. Veri seti Tablo 4.9'da verilmiştir.

Tablo 4.9. *Mesane Kanseri Verisi*

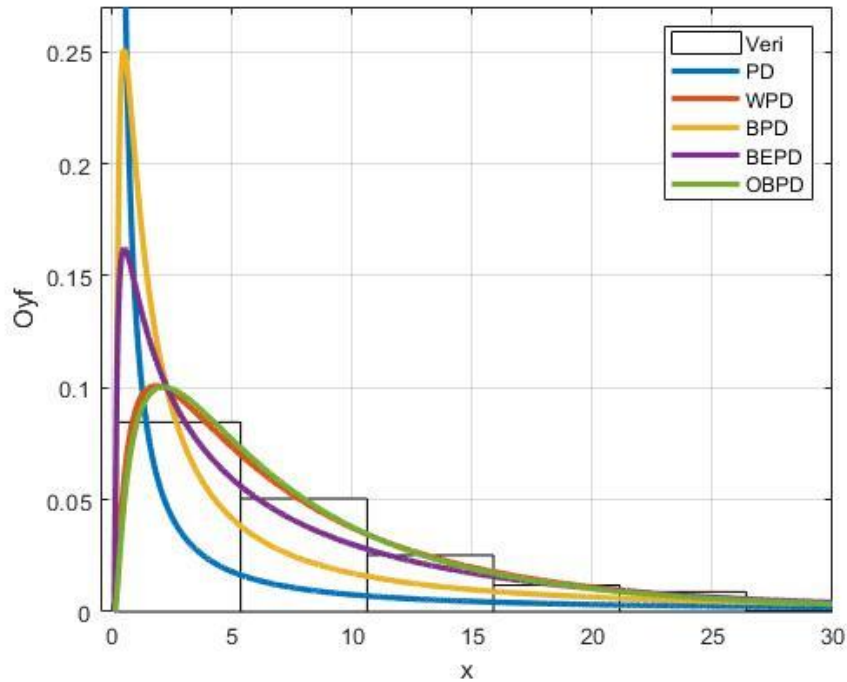
0,08	0,2	0,4	0,5	0,51	0,81	0,9	1,05	1,19	1,26
1,35	1,4	1,46	1,76	2,02	2,02	2,07	2,09	2,23	2,26
2,46	2,54	2,62	2,64	2,69	2,69	2,75	2,83	2,87	3,02
3,25	3,31	3,36	3,36	3,48	3,52	3,57	3,64	3,7	3,82
3,88	4,18	4,23	4,26	4,33	4,34	4,4	4,5	4,51	4,87
4,98	5,06	5,09	5,17	5,32	5,32	5,34	5,41	5,41	5,49
5,62	5,71	5,85	6,25	6,54	6,76	6,93	6,94	6,97	7,09
7,26	7,28	7,32	7,39	7,59	7,62	7,63	7,66	7,87	7,93
8,26	8,37	8,53	8,65	8,66	9,02	9,22	9,47	9,74	10,06
10,34	10,66	10,75	11,25	11,64	11,79	11,98	12,02	12,03	12,07
12,63	13,11	13,29	13,8	14,24	14,76	14,77	14,83	15,96	16,62
17,12	17,14	17,36	18,1	19,13	20,28	21,73	22,69	23,63	25,74
25,82	26,31	32,15	34,26	36,66	43,01	46,12	79,05		

OBu-P dağılımı Pareto, Weibull-Pareto (Aljarrah ve ark., 2015), beta-Pareto (Akinsete ve ark., 2008) ve beta-üstelleştirilmiş Pareto (Zea ve ark., 2012) dağılımları ile karşılaştırılmıştır. Bu dağılımlar sırasıyla PD, WPD, BPD ve BEPD olarak belirtilmektedir. Ele alınan dağılımların performans karşılaştırmaları Tablo 4.10.'da verilmiştir. Tahmin edilen parametrelere ilişkin standart hata değerleri de parantez içinde verilmiştir.

Tablo 4.10. Mesane Kanseri verisi için PD, WPD, BPD, BEPD ve OBU-P dağılımının parametre tahminleri ve kriterleri sonuçları

	PD	WPD	BPD	BEPD	OBu-P
	$\hat{\alpha} = 0,232 (0,021)$	$\hat{c} = 4,136 (0,118)$	$\hat{a} = 4,805 (0,055)$	$\hat{a} = 0,348 (0,097)$	$\hat{a} = 4,277 (0,338)$
	$\hat{\theta} = 0,080$	$\hat{\gamma} = 0,436 (0,088)$	$\hat{b} = 100,502 (0,251)$	$\hat{b} = 159831 (183,75)$	$\hat{b} = 6,674 (7,938)$
MLE		$\hat{k} = 0,077 (0,013)$	$\hat{k} = 0,011 (0,001)$	$\hat{k} = 0,051 (0,019)$	$\hat{\alpha} = 0,107 (0,028)$
		$\hat{\theta} = 0,080$	$\hat{\theta} = 0,080$	$\hat{\alpha} = 8,611 (2,093)$	$\hat{\theta} = 0,080$
				$\hat{\beta} = 0,080$	
-logL	539,591	407,370	480,446	432,410	406,965
AIC	1081,182	820,740	966,893	872,819	819,930
BIC	1084,026	829,273	975,425	884,197	828,463
K-S	0,425	0,043	0,217	0,142	0,039

Tablo 4.10.'dan görülebileceği gibi, OBU-P dağılımı en küçük negatif log-olabilirlik, AIC, BIC ve K-S istatistik değerlerine sahiptir, bu da OBU-P'nin diğer alternatif Pareto dağılımlarından daha iyi olduğunu göstermektedir. Diğer taraftan, Şekil 4.8, mesane kanseri verisi için PD, WPD, BPD ve BEPD'ye karşı OBU-P'nin modelleme yeteneğini görsel olarak gösterilmesi için elde edilmiştir. Grafikten de anlaşılacağı üzere OBU-R dağılımı ele alınan diğer dağılımların çoğundan daha iyi modelleme başarısı göstermiştir.



Şekil 4.8. Olasılık yoğunluk fonksiyonları ve histogram grafiği (Mesane kanseri)

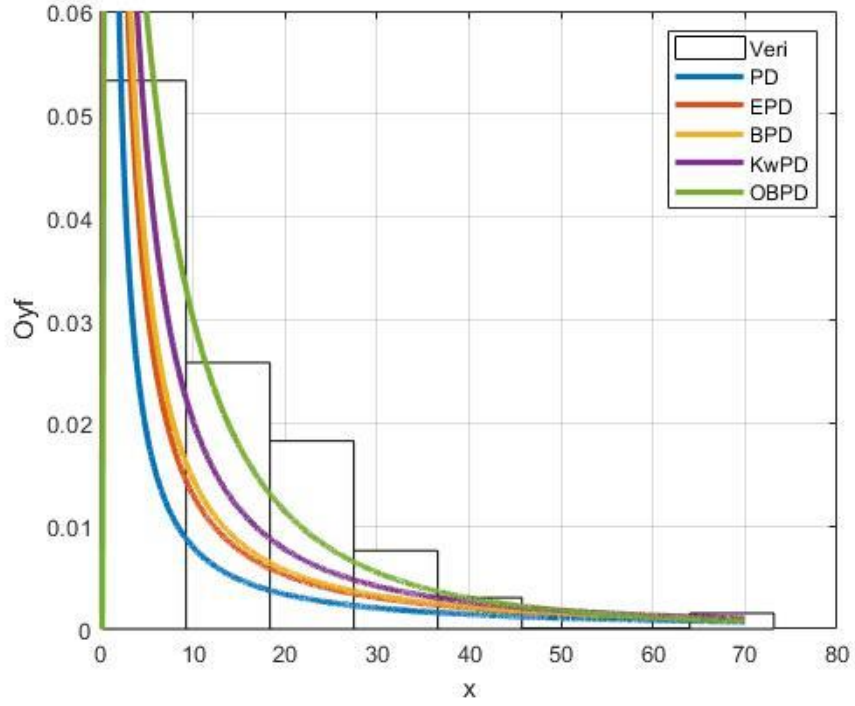
Uygulama II

İkinci uygulamada kullanılan veri seti Kanada'da bulunan Wheaton nehrine ait 1958-1984 yılları arasında gerçekleşen 72 adet taşkın (m^3/sn) gözleminden oluşmaktadır. Bu veri setini Akinsete ve ark. (2008) ve Bourguignon ve ark. (2013) çalışmalarında da kullanılmışlardır. *OBu-P* dağılımı PD, BPD, üstelleştirilmiş Pareto (EPD) ve Kumaraswamy Pareto (KwPD) (Bourguignon ve ark., 2013) dağılımları ile karşılaştırılmıştır. Negatif log-olabilirlik, AIC, BIC ve K-S istatistiklerinin sonuçları ile parantez içindeki standart hata değerleri Tablo 4.11.'de verilmiştir.

Tablo 4.11. Taşkın verisi için PD, EPD, BPD, KwPD ve *OBu-P* dağılımlarının parametre tahminleri ve kriterler sonuçları

	PD	EPD	BPD	KwPD	<i>OBu-P</i>
MLE	$\hat{\alpha} = 0,241$ (0,029)	$\hat{a} = 2,880$ (0,4911)	$\hat{a} = 3,147$ (0,499)	$\hat{a} = 2,855$ (0,337)	$\hat{a} = 3,366$ (0,361)
	$\hat{\theta} = 0,1$	$\hat{k} = 0,424$ (0,046)	$\hat{b} = 85,751$ (0,0001)	$\hat{b} = 85,847$ (60,421)	$\hat{b} = 68,017$ (191,48)
		$\hat{\beta} = 0,100$	$\hat{k} = 0,009$ (0,002)	$\hat{k} = 0,053$ (0,019)	$\hat{\alpha} = 0,054$ (0,037)
			$\hat{\theta} = 0,100$	$\hat{\beta} = 0,100$	$\hat{\theta} = 0,100$
-logL	303,949	287,300	283,700	271,200	250,899
AIC	609,897	578,600	573,400	548,400	507,797
BIC	612,160	583,125	580,188	555,188	514,586
K-S	0,341	0,199	0,175	0,170	0,125

Tablo 4.11.'de gözüktüğü gibi *OBu-P* dağılımı, taşkın verileri için tüm kriterlere göre en iyi performansı göstermiştir. Öte yandan önerilen dağılımın görsel olarak değerlendirilmesi için Şekil 4.9'te dağılımların oyf grafiği verilmiştir. *OBu-P* dağılımının modelleme yeteneği yine grafikte görülmektedir.



Şekil 4.9. Olasılık yoğunluk fonksiyonları ve histogram grafiği (Taşkın verileri)

Bu bölümde dört parametrelili $OBu-P$ dağılımı tanımlanmış olup dağılımın bir dizi istatistiksel özellikleri ele alınmıştır. $OBu-P$ dağılımının en çok olabilirlik tahmincileri elde edilmiş olup bu tahmincinin tutarlılığı benzetim çalışması yardımıyla elde edilmiştir. Ayrıca kullanılan iki gerçek veri seti yardımıyla, elde edilen yeni dağılımın literatürdeki bazı genelleştirilmiş Pareto dağılımlarından daha iyi sonuçlar verdiği gösterilmiştir. Sonuç olarak, $OBu-P$ dağılımı, sigorta verileri, hayatta kalma verileri ve aşırı yağış, aşırı rüzgâr ve aşırı kayıplar da dâhil olmak üzere aşırı olaylarla ilgili veriler için esnek bir dağılım modeli olarak sunulmaktadır.

4.3. OBU-LL Dağılımı ve İstatistiksel Özellikleri

Log-lojistik (LL) dağılımı ilk olarak nüfus artışını modellemek için Verhulst (1838) tarafından geliştirilen popüler bir lojistik dağılımdır. Ekonomi literatüründe daha çok Fisk (Fisk, 1961) dağılımı olarak adlandırılan LL dağılımı güvenilirlik, yaşam analizi, mühendislik, aktüerya gibi birçok alanda yaygın olarak kullanılmaktadır. Ayrıca, sansürlü veriler ile sıklıkla karşılaşılan güvenilirlik ve yaşam çözümlemesi alanlarında bu tarz verileri modellemede LL dağılımının oldukça uygun olduğu kabul görmektedir (Tahir ve ark., 2014). LL dağılımına sahip X rassal değişkeninin dağılım ve olasılık yoğunluk fonksiyonları sırasıyla (4.59) ve (4.60)'daki gibi ifade edilir:

$$G(t) = 1 - \left[1 + \left(\frac{t}{\beta} \right)^\gamma \right]^{-1}, \quad t > 0, \beta, \gamma > 0 \quad (4.59)$$

$$g(t) = \frac{\gamma}{\beta^\gamma} t^{(\gamma-1)} \left[1 + \left(\frac{t}{\beta} \right)^\gamma \right]^{-2} \quad (4.60)$$

Denklemlerde görülen β ve γ parametreleri sırasıyla LL dağılımının ölçek ve şekil parametreleridir. Son yıllarda modelleme başarısının artırılması amacıyla LL dağılımının genelleştirilmesi üzerine yoğun olarak çalışılmaktadır. Literatürde dağılımın birçok genelleştirilmiş hali bulunmaktadır. Örneğin, Santana ve ark. (2012) Kumaraswamy-G dağılımlar ailesi yönteminden (Cordeiro ve de Castro, 2011) yararlanarak Kumaraswamy-log-lojistik dağılımını önermişlerdir. Ramos ve ark. (2013), Zografos ve Balakrishnan (2009) tarafından geliştirilen Zografos-Balakrishnan-G yönteminden yararlanarak Zografos-Balakrishnan-LL dağılımını önermişler ve elde ettikleri yeni dağılımı göğüs kanseri verilerine uygulamışlardır. Gui (2013), Marshall ve Olkin'in (1997) önerdiği yöntemi kullanarak LL dağılımına yeni bir şekil parametresi ekleyerek Marshall-Olkin genişletilmiş LL dağılımını elde etmiştir. Beta-LL dağılımı Eugene'nin (2002) geliştirdiği dağılım genelleştirme yöntemi kullanılarak Lemonte (2014) tarafından önerilmiştir. Öte yandan Tahir ve ark. (2015) tarafından geliştirilen odd genelleştirilmiş üstelleştirilmiş-G ailesi yöntemi yardımıyla odd genelleştirilmiş üstelleştirilmiş LL dağılımı Rosaiah ve ark. (2016) tarafından önerilmiştir. Son olarak genişletilmiş-LL dağılımı ise Lima ve Cordeiro (2017) tarafından önerilmiş olup bu dağılımın genelleştirilmesinde Cordeiro ve ark. (2013) tarafından geliştirilen üstelleştirilmiş-genelleştirilmiş-G ailesi yöntemi kullanılmıştır.

Bu bölümde Odd-Burr Log-lojistik (*OBu-LL*) dağılımı *T-X* dağılımlar ailesi yöntemi temel alınarak geliştirilen yöntem yardımıyla (Alizadeh ve ark, 2017) tanımlanacaktır. Ayrıca, 4.1 ve 4.2 bölümlerinden farklı olarak, ömür sürelerinin analizinde kullanılmak üzere *OBu-LL* dağılımına dayalı yeni bir parametrik regresyon modeli önerilecek ve performansı araştırılacaktır.

4.3.1. *OBu-LL* dağılım ve olasılık yoğunluk fonksiyonları

(4.3) denkleminde $G(x)$ temel dağılımı yerine (4.59) ile ifade edilen LL dağılım fonksiyonu kullanılırsa *OBu-LL* dağılımının dağılım fonksiyonu;

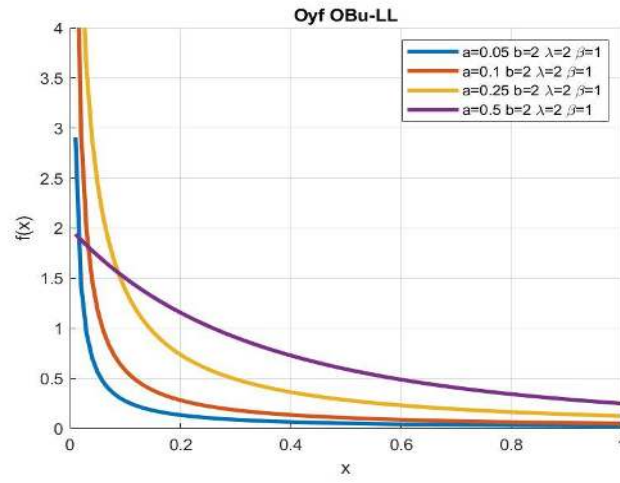
$$F(t) = 1 - \frac{\left[1 + \left(\frac{t}{\beta}\right)^\gamma\right]^{-ab}}{\left\{\left(1 - \left[1 + \left(\frac{t}{\beta}\right)^\gamma\right]^{-1}\right)^a + \left[1 + \left(\frac{t}{\beta}\right)^\gamma\right]^{-a}\right\}^b}, \quad a, b, \gamma, \beta > 0, t > 0, \quad (4.61)$$

(4.61) dağılım fonksiyonuna karşılık gelen olasılık yoğunluk fonksiyonu ise;

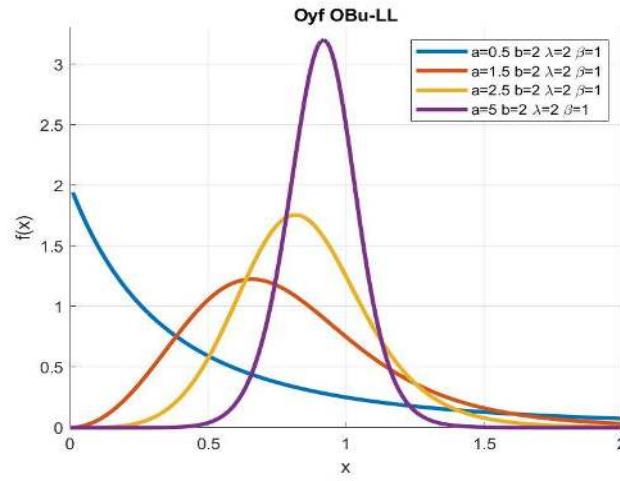
$$f(t) = \frac{ab\gamma t^{(\gamma-1)} \left[1 + \left(\frac{t}{\beta}\right)^\gamma\right]^{-(ab+1)} \left\{1 - \left[1 + \left(\frac{t}{\beta}\right)^\gamma\right]^{-1}\right\}^{(a-1)}}{\beta^\gamma \left\{\left(1 - \left[1 + \left(\frac{t}{\beta}\right)^\gamma\right]^{-1}\right)^a + \left[1 + \left(\frac{t}{\beta}\right)^\gamma\right]^{-a}\right\}^{(b+1)}} \quad (4.62)$$

şeklinde ifade edilir. Burada β ölçek parametresi, a , b ve γ şekil parametreleridir. Eşitlik (4.62) ile verilen oyf'na sahip olan T rassal değişkeni $T \sim OBU-LL(a, b, \gamma, \beta)$ şeklinde ifade edilir.

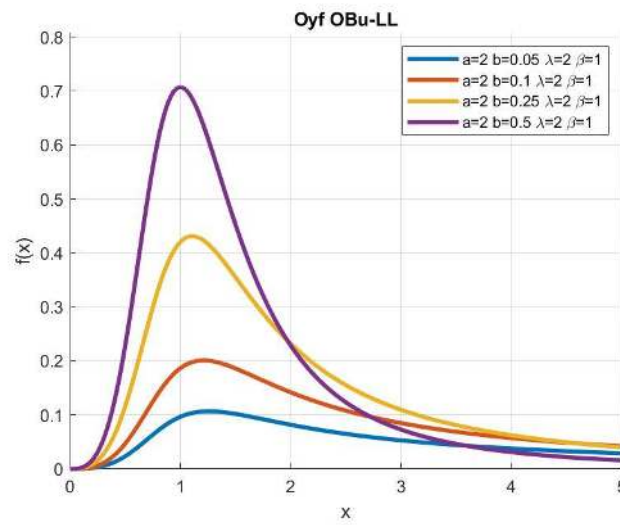
Farklı parametre değerleri için *OBu-LL* dağılımının oyf'nu grafikleri Şekil 4.10'da verilmiştir. Grafiklerden görülebileceği gibi, önerilen *OBu-LL* oyf, şekil parametrelerinin değerlerine bağlı olarak çok çeşitli şekiller, farklı kuyruk davranışları ve çarpıklık ve/veya basıklık sunmaktadır.



(a)



(b)



(c)

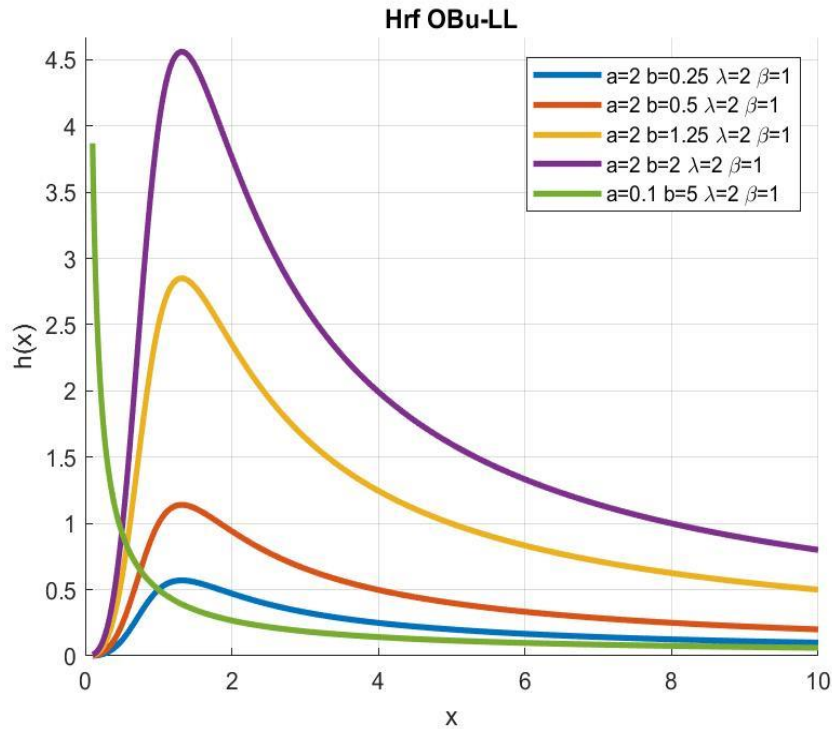
Şekil 4.10. Farklı parametre değerleri için OBU-LL dağılımının oyf grafiği

4.3.2. OBU-LL yaşam ve tehlike oranı fonksiyonları

OBu-LL dağılımının yaşam ve tehlike oranı fonksiyonları sırasıyla (4.63) ve (4.64)'te ifade edilmiştir.

$$S(t) = \frac{\left[1 + \left(\frac{t}{\beta}\right)^\gamma\right]^{-ab}}{\left\{1 - \left[1 + \left(\frac{t}{\beta}\right)^\gamma\right]^{-1}\right\}^a + \left[1 + \left(\frac{t}{\beta}\right)^\gamma\right]^{-a}}\right)^b, \quad a, b, \gamma, \beta > 0, t > 0, \quad (4.63)$$

$$h(t) = \frac{ab\gamma t^{(\gamma-1)} \left[1 + \left(\frac{t}{\beta}\right)^\gamma\right]^{-1} \left\{1 - \left[1 + \left(\frac{t}{\beta}\right)^\gamma\right]^{-1}\right\}^{(a-1)}}{\beta^\gamma \left\{1 - \left[1 + \left(\frac{t}{\beta}\right)^\gamma\right]^{-1}\right\}^a + \left[1 + \left(\frac{t}{\beta}\right)^\gamma\right]^{-a}} \quad (4.64)$$



Şekil 4.11. Farklı parametre değerleri için OBU-LL dağılımının tehlike oranı fonksiyonu grafiği

Farklı parametre değerleri için tehlike oranı fonksiyonunun grafikleri Şekil 4.11'de verilmiştir. Bu grafikten görüldüğü gibi, OBU-LL dağılımının tehlike oranı fonksiyonu,

monoton olarak azalan ve artan-azalan (tek modlu) gibi farklı şekillere sahiptir. *OBu-LL* dağılımının tehlike fonksiyonunun bu özelliğinden dolayı yaşam süresi analizinde kullanılabilir.

4.3.3. *OBu-LL*'nin kantil fonksiyonu

(4.61) denkleminde ifade edilen dağılım fonksiyonuna dayalı *OBu-LL* dağılımının kantil fonksiyonu (4.65)'teki gibi elde edilir:

$$F(t) = 1 - \frac{\left[1 + \left(\frac{t}{\beta}\right)^\gamma\right]^{-ab}}{\left\{\left(1 - \left[1 + \left(\frac{t}{\beta}\right)^\gamma\right]^{-1}\right)^a + \left[1 + \left(\frac{t}{\beta}\right)^\gamma\right]^{-a}\right\}^b},$$

$$\left(\left[1 - F(t)\right]^{-1/b} - 1\right)^{1/a} = \frac{1}{\left[1 + \left(\frac{t}{\beta}\right)^\gamma\right]^{-1}} - 1,$$

$$\left(\left[1 - F(t)\right]^{-1/b} - 1\right)^{1/a} = \left(\frac{t}{\beta}\right)^\gamma,$$

$$T = \beta \left\{ \left(\left[1 - F(t)\right]^{-1/b} - 1\right)^{1/a} \right\}^{1/\gamma} \quad (4.65)$$

OBu-LL rassal değişkeni $F(x) = u \sim U(0,1)$ alınarak (4.65) kantil fonksiyonu yardımıyla kolaylıkla üretilebilir.

4.3.4. *OBu-LL*'nin alternatif matematiksel gösterimi

Bu bölümde, *OBu-R* dağılımında olduğu gibi *OBu-LL*'nin dağılım ve olasılık yoğunluk fonksiyonlarının kuvvet serilerine dayalı alternatif gösterimleri elde edilecektir. Bir önceki bölümde izlenen yöntemlerin bu dağılım için aynı şekilde uygulanması sonucunda *OBu-LL*'nin dağılım ve olasılık yoğunluk fonksiyonları için alternatif gösterimler:

$$F(x) = \sum_{k=0}^{\infty} n_k G(x)^k = \sum_{k=0}^{\infty} n_k H_k(x) \quad (4.66)$$

$$f(x) = \sum_{k=0}^{\infty} n_{k+1} h_{k+1}(x) \quad (4.67)$$

şeklinde elde edilir. Burada $H_k(x)$ ölçek parametresi β ve şekil parametreleri k ve γ olan üstelleştirilmiş log-lojistik (*ELL*) dağılımı olup $m_k = \sum_{i=0}^{\infty} (-1)^i \binom{b}{i} \lambda_k$, $n_0 = 1 - m_0$ ve $n_k = -m_k$ şeklindedir. Önceki bölümlerde elde edilen dağılımlarda olduğu gibi (4.66) eşitliğine göre *OBu-LL*'nin dağılım fonksiyonu, temel alınan dağılımın ağırlıklandırılmış kuvvet serisi toplamı olarak ifade edilmiştir. Bu sonuç *OBu-LL* dağılımının momentlerinin elde edilmesinde kullanılacaktır.

4.3.5. *OBu-LL*'nin momentleri

Bu bölümde *OBu-LL* dağılımının moment fonksiyonu (4.66) eşitliği yardımıyla elde edilecektir. Bu amaçla üstelleştirilmiş log-lojistik dağılımının r 'inci momenti (4.68)'deki gibi ifade edilir:

$$\mu'_{ELL} = E_{ELL}(Z^r) = \lambda \theta^r B\left(1 - \frac{r}{\gamma}, \lambda + \frac{r}{\gamma}\right) \quad (4.68)$$

burada λ ve γ şekil parametreleri olup β ise ölçek parametresidir. Sonuç olarak (4.67) eşitliği ile ifade edilen ilişki yardımıyla *OBu-LL* dağılımının r 'inci momenti (4.69)'daki gibi ifade edilebilir (Ramos ve ark., 2013; Lemonte, 2014):

$$\begin{aligned} \mu'_{OBu-LL} &= E_{OBu-LL}(X^r) = \sum_{k=0}^{\infty} n_{k+1} \int_{\theta}^{\infty} x^r h_{(k+1)}(x) dx \\ &= \beta^r \sum_{k=0}^{\infty} n_{k+1} (k+1) B\left(1 - \frac{r}{\gamma}, (k+1) + \frac{r}{\gamma}\right), \quad r \leq \gamma \end{aligned} \quad (4.69)$$

$r=1$ olarak alındığında *OBu-LL* dağılımının birinci momenti yani ortalaması aşağıdaki eşitlik yardımıyla elde edilebilir:

$$\mu'_{OBu-LL} = \beta \sum_{k=0}^{\infty} n_{k+1} (k+1) B\left(1 - \frac{1}{\gamma}, (k+1) + \frac{1}{\gamma}\right).$$

4.3.6. *OBu-LL*'nin en çok olabilirlik tahmincisi

Önerilen *OBu-LL* dağılımının parametrelerinin tahmini için en çok olabilirlik yöntemi kullanılacaktır. Bu amaçla $X_1, X_2, \dots, X_n$, parametreleri a, b, γ ve β olan *OBu-LL* dağılımından örneklem olmak üzere, log-olabilirlik fonksiyonu,

$$\begin{aligned} \log L(a, b, \gamma, \beta) &= n \log \left(\frac{ab\gamma}{\beta^\gamma} \right) + (\gamma - 1) \sum_{i=1}^n \log(t_i) \\ &\quad - (ab + 1) \sum_{i=1}^n \log \left[1 + \left(\frac{t_i}{\beta} \right)^\gamma \right] + (a - 1) \sum_{i=1}^n \log \left\{ 1 - \left[1 + \left(\frac{t_i}{\beta} \right)^\gamma \right]^{-1} \right\} \\ &\quad - (b + 1) \sum_{i=1}^n \log \left\{ \left(1 - \left[1 + \left(\frac{t_i}{\beta} \right)^\gamma \right]^{-1} \right)^a + \left[1 + \left(\frac{t_i}{\beta} \right)^\gamma \right]^{-a} \right\} \end{aligned} \quad (4.70)$$

şeklinde ifade edilir. Olabilirlik tahmincileri log-olabilirlik fonksiyonunun parametrelere göre aşağıda verilen kısmi türevlerinin sıfıra eşitlenmesi ile elde edilir.

$$\begin{aligned} \frac{\partial \log L}{\partial a} &= \frac{n}{a} - b \sum_{i=1}^n \log \left[1 + \left(\frac{t_i}{\beta} \right)^\gamma \right] + \sum_{i=1}^n \log \left\{ 1 - \left[1 + \left(\frac{t_i}{\beta} \right)^\gamma \right]^{-1} \right\} \\ &\quad - (b + 1) \sum_{i=1}^n \frac{\left\{ 1 - \left[1 + \left(\frac{t_i}{\beta} \right)^\gamma \right]^{-1} \right\}^a \ln \left\{ 1 - \left[1 + \left(\frac{t_i}{\beta} \right)^\gamma \right]^{-1} \right\} - \left[1 + \left(\frac{t_i}{\beta} \right)^\gamma \right]^{-a} \ln \left[1 + \left(\frac{t_i}{\beta} \right)^\gamma \right]}{\left(1 - \left[1 + \left(\frac{t_i}{\beta} \right)^\gamma \right]^{-1} \right)^a + \left[1 + \left(\frac{t_i}{\beta} \right)^\gamma \right]^{-a}}, \end{aligned}$$

$$\frac{\partial \log L}{\partial b} = \frac{n}{b} - a \sum_{i=1}^n \log \left[1 + \left(\frac{t_i}{\beta} \right)^\gamma \right] - \sum_{i=1}^n \log \left\{ \left(1 - \left[1 + \left(\frac{t_i}{\beta} \right)^\gamma \right]^{-1} \right)^a + \left[1 + \left(\frac{t_i}{\beta} \right)^\gamma \right]^{-a} \right\},$$

$$\begin{aligned}
\frac{\partial \log L}{\partial \gamma} &= n \left(\frac{1 - \gamma \log \beta}{\gamma} \right) + \sum_{i=1}^n \log(t_i) - (ab+1) \sum_{i=1}^n \frac{\left(\frac{t_i}{\beta}\right)^\gamma \ln\left(\frac{t_i}{\beta}\right)}{1 + \left(\frac{t_i}{\beta}\right)^\gamma} \\
&+ (a-1) \sum_{i=1}^n \ln\left(\frac{t_i}{\beta}\right) \left[1 + \left(\frac{t_i}{\beta}\right)^\gamma \right]^{-1} \\
&+ (b+1) \sum_{i=1}^n \frac{a \ln\left(\frac{t_i}{\beta}\right) \left\{ \left(\frac{t_i}{\beta}\right)^\gamma - \left[1 + \left(\frac{t_i}{\beta}\right)^\gamma \right]^a \left[1 - \left(1 + \left(\frac{t_i}{\beta}\right)^\gamma \right) \right]^a \right\}}{\left\{ \left(1 - \left[1 + \left(\frac{t_i}{\beta}\right)^\gamma \right]^{-1} \right)^a \left[1 + \left(\frac{t_i}{\beta}\right)^\gamma \right]^a + 1 \right\} \left[1 + \left(\frac{t_i}{\beta}\right)^\gamma \right]} \\
\frac{\partial \log L}{\partial \beta} &= -\frac{n\gamma}{\beta} + (ab+1) \sum_{i=1}^n \frac{\gamma \frac{t_i^\gamma}{\beta^{\gamma+1}}}{1 + \left(\frac{t_i}{\beta}\right)^\gamma} - (a-1) \sum_{i=1}^n \frac{\gamma \frac{t_i^\gamma}{\beta^{\gamma+1}} \left[1 + \left(\frac{t_i}{\beta}\right)^\gamma \right]^{-2}}{\left\{ 1 - \left[1 + \left(\frac{t_i}{\beta}\right)^\gamma \right]^{-1} \right\}} \\
&- (b+1) \sum_{i=1}^n \frac{a\gamma \left[1 + \left(\frac{t_i}{\beta}\right)^\gamma \right]^{-(a+1)} \frac{t_i^\gamma}{\beta^{\gamma+1}} - a\gamma \left\{ 1 - \left[1 + \left(\frac{t_i}{\beta}\right)^\gamma \right]^{-1} \right\}^{a-1} \frac{t_i^\gamma}{\beta^{\gamma+1}} \left[1 + \left(\frac{t_i}{\beta}\right)^\gamma \right]^{-2}}{\left\{ \left(1 - \left[1 + \left(\frac{t_i}{\beta}\right)^\gamma \right]^{-1} \right)^a + \left[1 + \left(\frac{t_i}{\beta}\right)^\gamma \right]^{-a} \right\}}
\end{aligned}$$

Parametrelerin en çok olabilirlik tahminleri, yukarıda verilen denklemlerin eş zamanlı çözümü ile elde edilebilir. Bu amaçla denklemlerin açık çözümü mümkün olmadığından Newton-Raphson gibi nümerik yöntemlerin kullanılması gerekir.

4.3.7. OBU-LL regresyon modeli

Bir bireyin yaşam süresini ya da herhangi bir ekipmanın bir parçasının kullanım süresini açıklayıcı değişken olarak birtakım etkenler etkileyebilmektedir. Dolayısıyla ömür süresi ile açıklayıcı değişkenler arasındaki bu ilişkinin belirlenmesi büyük bir öneme sahiptir. Bu amaçla bağımlı değişkenin $Y = \log(T)$ ve açıklayıcı değişken vektörünün $x = (x_1, x_2, \dots, x_p)^T$ olduğu regresyon modeli yaklaşımları kullanılmaktadır.

Dolayısıyla bağımlı değişken ile açıklayıcı değişkenler arasındaki ilişkinin belirlenebilmesi için Y rassal değişkeninin dağılımının belirlenmesine ihtiyaç duyulmaktadır. Sonuç olarak $T \sim OBU-LL(a, b, \gamma, \beta)$ olduğunda $Y = \log(T)$ rassal değişkeni Odd Burr lojistik dağılımına sahip olur ve $Y \sim OBU-L(a, b, \mu, \sigma)$ şeklinde ifade edilir ve buradaki $\mu = \ln(\beta)$ ve $\sigma = \gamma^{-1}$ olup elde edilen yeni dağılımın konum ve ölçek parametreleridir. Y rassal değişkeninin dağılım ve olasılık yoğunluk fonksiyonları sırasıyla,

$$F(y) = 1 - \frac{\left[1 + \exp\left(\frac{y - \mu}{\sigma}\right)\right]^{-ab}}{\left\{\left(1 - \left[1 + \exp\left(\frac{y - \mu}{\sigma}\right)\right]^{-1}\right)^a + \left[1 + \exp\left(\frac{y - \mu}{\sigma}\right)\right]^{-a}\right\}^b} \quad (4.71)$$

$$f(y) = \frac{ab \exp\left(\frac{y - \mu}{\sigma}\right) \left[1 + \exp\left(\frac{y - \mu}{\sigma}\right)\right]^{-(ab+1)} \left\{1 - \left[1 + \exp\left(\frac{y - \mu}{\sigma}\right)\right]^{-1}\right\}^{a-1}}{\sigma \left\{\left(1 - \left[1 + \exp\left(\frac{y - \mu}{\sigma}\right)\right]^{-1}\right)^a + \left[1 + \exp\left(\frac{y - \mu}{\sigma}\right)\right]^{-a}\right\}^{b+1}} \quad (4.72)$$

şeklinde dir. Dağılımın yaşam ve tehlike fonksiyonları da sırasıyla (4.73) ve (4.74)'te ifade edilmiştir:

$$S(y) = \frac{\left[1 + \exp\left(\frac{y - \mu}{\sigma}\right)\right]^{-ab}}{\left\{\left(1 - \left[1 + \exp\left(\frac{y - \mu}{\sigma}\right)\right]^{-1}\right)^a + \left[1 + \exp\left(\frac{y - \mu}{\sigma}\right)\right]^{-a}\right\}^b} \quad (4.73)$$

$$h(y) = \frac{ab \exp\left(\frac{y - \mu}{\sigma}\right) \left[1 + \exp\left(\frac{y - \mu}{\sigma}\right)\right]^{-1} \left\{1 - \left[1 + \exp\left(\frac{y - \mu}{\sigma}\right)\right]^{-1}\right\}^{a-1}}{\left\{\left(1 - \left[1 + \exp\left(\frac{y - \mu}{\sigma}\right)\right]^{-1}\right)^a + \left[1 + \exp\left(\frac{y - \mu}{\sigma}\right)\right]^{-a}\right\}} \quad (4.74)$$

$z = \frac{y - \mu}{\sigma}$ olmak üzere Z rassal değişkeninin olasılık yoğunluk fonksiyonu,

$$f(z) = \frac{ab \exp(z) [1 + \exp(z)]^{-(ab+1)} \left\{ 1 - [1 + \exp(z)]^{-1} \right\}^{a-1}}{\left\{ \left(1 - [1 + \exp(z)]^{-1} \right)^a + [1 + \exp(z)]^{-a} \right\}^{b+1}} \quad (4.75)$$

olarak elde edilir. Elde edilen dağılıma ilişkin *OBu-L* doğrusal regresyon modeli (4.76)'daki gibi tanımlanmıştır:

$$y_i = x_i^T \beta + \sigma z_i, \quad i = 1, 2, \dots, n \quad (4.76)$$

Bu modeldeki y_i i . birey için gözlemlenmiş log-yaşam zamanı ya da log-sansürleme zamanı, $(x_{i1}, x_{i2}, \dots, x_{ip})^T$ bilinen açıklayıcı değişken vektörü, $(\beta_1, \beta_2, \dots, \beta_p)$ tahmin edilecek regresyon katsayı vektörü ve z_i (4.75) eşitliği ile ifade edilen olasılık yoğunluk fonksiyonuna sahip hatalardır. z_i 'lerin bağımsız ve aynı dağılıma sahip oldukları varsayılmaktadır. Elde edilen yeni model farklı veri tiplerinin modellenmesinde kullanılabilecek yeni bir araç olarak literatüre sunulmuştur. Bu model lojistik dağılıma dayalı hızlandırılmış başarısızlık zamanı modelinin genelleştirilmiş halidir (Ortega ve ark., 2016).

n birimlik bağımsız gözlemlerden oluşan $(y_1, x_1), (y_2, x_2), \dots, (y_n, x_n)$ örnekleme ele alınsın. Burada $y_i = \min \{ \log(t_i), \log(c_i) \}$ şeklinde tanımlanıp i . birey için log-yaşam zamanını ya da log-sansürleme zamanını ifade eder. Burada sansürlemenin non-informatif olduğu varsayılmıştır. Yani gözlemlenmiş yaşam zamanları ile sansürleme zamanları birbirlerinden bağımsızdır. D ve C sırasıyla gözlemlenmiş ve sansürlenmiş yaşam sürelerinin kümeleri olsun. (4.76) modelindeki $\theta = (a, b, \sigma, \beta^T)^T$ parametre vektörü için log-olabilirlik fonksiyonu,

$$l(\theta) = \sum_{i \in D} l_i(\theta) + \sum_{i \in C} l_i^{(c)}(\theta)$$

şeklinde ifade edilir. Buradaki $l_i(\theta) = \log[f(y_i)]$ ve $l_i^{(c)}(\theta) = \log[S(y_i)]$ olup $f(y_i)$, (4.72) eşitliği ile verilen yoğunluk fonksiyonu ve $S(y_i)$, (4.73) eşitliği ile verilen yaşam fonksiyonudur. Elde edilen model için toplam log-olabilirlik fonksiyonu (3.77)'deki gibidir.

$$\begin{aligned}
l(\theta) = & r \log\left(\frac{ab}{\sigma}\right) + \sum_{i \in D} z_i - (ab+1) \sum_{i \in D} \log[1 + \exp(z_i)] \\
& + (a-1) \sum_{i \in D} \log\left\{1 - [1 + \exp(z_i)]^{-1}\right\} \\
& - (b+1) \sum_{i \in D} \log\left\{\left(1 - [1 + \exp(z_i)]^{-1}\right)^a + [1 + \exp(z_i)]^{-a}\right\} \\
& - ab \sum_{i \in C} \log[1 + \exp(z_i)] - b \sum_{i \in C} \log\left\{\left(1 - [1 + \exp(z_i)]^{-1}\right)^a + [1 + \exp(z_i)]^{-a}\right\}
\end{aligned} \tag{3.77}$$

Burada, $z_i = \frac{y - x_i^T \beta}{\sigma}$ ve r gözlemlenmiş yaşam süresinin sayısıdır. Bilindiği üzere parametrelerin en çok olabilirlik tahmincileri log-olabilirlik fonksiyonunun maksimize edilmesi ile elde edilir. Bu amaçla *MATLAB* programından yararlanılarak modelin en çok olabilirlik tahmincileri elde edilmiştir.

4.3.8. *OBu-LL*'nin uygulamaları

Bu bölümde, tezde önerilen *OBu-LL* dağılımının ve bu dağılıma dayalı olarak elde edilen regresyon modelinin performansını incelemek amacıyla iki ayrı gerçek veri seti üzerinde uygulama yapılmış olup literatürde kullanılan dağılımlar ile karşılaştırması yapılmıştır.

Uygulama I

İlk olarak göğüs kanseri hastalarının yaşam sürelerinin olduğu 121 adet gözlemden oluşan veri seti ele alınacaktır. Bu veri seti daha önce Ramos ve ark. (2013) tarafından önermiş oldukları Zografos-Balakrishnan LL (ZBLL) dağılımının performansını değerlendirmek amacıyla kullanılmıştır. Kullanılan veri seti Tablo 4.12'de verilmiştir.

Tablo 4.12. Göğüs Kanseri Verisi

0,3	0,3	4	5	5,6	6,2	6,3	6,6	6,8	7,4
7,5	8,4	8,4	10,3	11	11,8	12,2	12,3	13,5	14,4
14,4	14,8	15,5	15,7	16,2	16,3	16,5	16,8	17,2	17,3
17,5	17,9	19,8	20,4	20,9	21	21	21,1	23	23,4
23,6	24	24	27,9	28,2	29,1	30	31	31	32
35	35	37	37	37	38	38	38	39	39
40	40	40	41	41	41	42	43	43	43
44	45	45	46	46	47	48	49	51	51
51	52	54	55	56	57	58	59	60	60
60	61	62	65	65	67	67	68	69	78
80	83	88	89	90	93	96	103	105	109
109	111	115	117	125	126	127	129	129	139
154									

Önerilen dağılımın performansını incelemek amacıyla Kumaraswamy-LL (KLL) (Cordeiro ve ark., 2012) ve beta-LL (BLL) (Lemonte, 2014) dağılımları kullanılmıştır. Dağılımların performanslarını karşılaştırmak amacıyla AIC, düzeltilmiş akaike bilgi kriteri (CAIC) ve BIC kriterleri kullanılmış olup parametre tahmin değerleri ile karşılaştırma kriterleri sonuçları Tablo 4.13’te verilmiştir.

Tablo 4.13. Göğüs kanseri verisi için KLL, BLL ve OBU-LL dağılımlarının parametre tahmin ve karşılaştırma kriterleri sonuçları

Dağılım	Parametre Tahmini				AIC	CAIC	BIC
<i>KLL</i> (a, b, γ, α)	33,698	23,048	0,336	0,044	1189,937	1190,282	1201,12
<i>BLL</i> (a, b, α, β)	0,364	0,732	53,251	3,368	1171,861	1172,206	1183,045
<i>OBuLL</i> (a, b, γ, β)	0,321	53,101	4,095	1210	1166,133	1166,4249	1177,316

Tablo 4.13’ten görüldüğü üzere karşılaştırma kriteri değerleri en düşük olan önerilen *OBu-LL* dağılımı olup bu da *OBu-LL* dağılımının ilgilenilen veri setini modellemede en başarılı dağılım olduğu anlamına gelmektedir.

Uygulama II

Bu bölümde elde edilen regresyon modelinin performansını incelemek amacıyla içerisinde açıklayıcı değişkeninde bulunduğu veri seti ele alınmıştır. Kullanılan veri seti Lawless (2013, S:335)’tan alınmış olup tutkaldan yapılmış bir yalıtkanın 3 farklı voltaj seviyesindeki dayanabilme sürelerini içermektedir. Veri seti Tablo 4.14.’te verilmiştir.

Veri setinde sansürlü veriler de yer almaktadır ve sansürlü verilerin oranı %10 olup sansürlü veriler yıldız ile işaretlenmiştir.

Tablo 4.14. *Voltaj Verisi*

Voltaj (kV)	Bozulma Zamanları									
52.5	4690	740	1010	1190	2450	1390	350	6095	3000	246
	1458	6200*	550	1690	745	1225	1480	245	600	1805
55	258	114	312	772	478	162	444	1464	132	1740*
	1266	300	2440*	520	1240	2600*	222	144	745	396
57,5	510	1000*	252	408	528	690	900*	714	348	546
	174	696	294	234	288	444	390	168	558	288

Önerilen yeni regresyon modelinin performansını değerlendirmek için alt model olan lojistik ve bozulma sürelerinin analizinde sıklıkla kullanılan log-Weibull regresyon modelleri de ele alınmıştır. Parametre tahmin ve değerlendirme kriterleri sonuçları Tablo 4.15’te verilmiştir.

Tablo 4.15. *Voltaj Verisi verisi için OBU-L, Lojistik ve Log-Weibull dağılımının parametre tahminleri ve kriter sonuçları*

Model	σ	β_0	β_1	a	b	AIC	CAIC	BIC	HQIC
OBu-L	0,3795	19,3520	-0,2310	1,0753	0,6110	135,834	136,945	146,305	139,930
Lojistik	0,5290	17,8853	-0,2073	-	-	164,260	164,689	170,544	166,718
Log-Weib	0,8452	22,0314	-0,2746	-	-	173,398	173,826	179,681	175,855

Tablo 4.15’te elde edilen sonuçlar incelendiğinde değerlendirme kriterlerinin sonuçları yeni önerilen regresyon modeli olan OBU-L için diğer karşılaştırılan modellere nazaran oldukça küçüktür. Yani yeni elde edilen regresyon modeli veri setini modelleme konusunda diğer ele alınan modellere göre oldukça başarılıdır.

Ayrıca elde edilen yeni modelin alt modele göre anlamlı derecede iyi modelleme başarısına sahip olup olmadığını değerlendirebilmek için OBU-L ve lojistik modellere olabilirlik oran testi uygulanmış olup elde edilen sonuç tablo 4.16’da verilmiştir.

Tablo 4.16. Olabilirlik oran testi sonuçları (Voltaj verisi)

Model	Hipotez	İstatistik	p-değeri
OBu-L ve Lojistik	$H0: a=b=1$ $H1: H0$ yanlış	32,5456	$p<0.0001$

Tablo 4.16'daki olabilirlik oran testi için elde edilen sonuçlar incelendiğinde p -değeri sıfıra çok yakındır. Bu da yeni modelin veriyi lojistik modelden daha iyi modellediğinin güçlü bir göstergesidir.

Kurulan regresyon modelindeki artıkların OBU-L dağılımına uygunluğunun test edilmesi için Martingale artıkları aşağıda gibi elde edilmiştir:

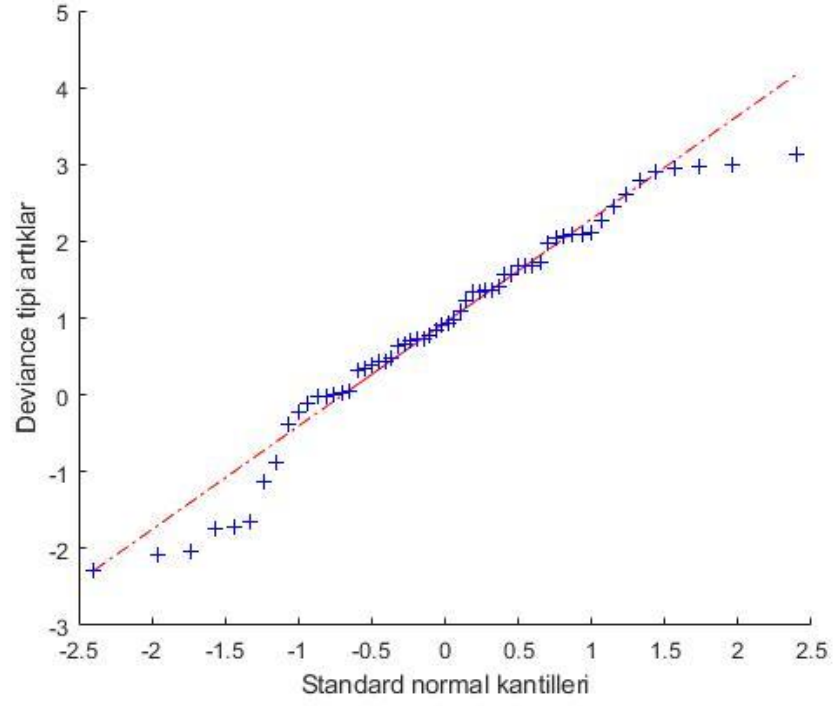
$$r_{Mi} = \delta_i - r_{Ci} = \delta_i + \log \hat{S}_i(t_i)$$

$$= \delta_i + \log \left\{ \frac{\left[1 + \exp\left(\frac{y - \hat{\mu}}{\hat{\sigma}} \right) \right]^{-\hat{a}\hat{b}}}{\left\{ \left(1 - \left[1 + \exp\left(\frac{y - \hat{\mu}}{\hat{\sigma}} \right) \right]^{-1} \right)^{\hat{a}} + \left[1 + \exp\left(\frac{y - \hat{\mu}}{\hat{\sigma}} \right) \right]^{-\hat{a}} \right\}^{\hat{b}}} \right\}$$

Martingale artıkları çarpık bir dağılım yapısına sahiptir. Bilindiği üzere artık analizinde normal dağılıma yakın artıklar tercih edilir. Dolayısıyla artık analizinde martingale artıklarından elde edilen ve simetrik bir yapıya sahip olan Deviance artıklarından yararlanılacak olup bu artıklar,

$$r_{Di} = \text{sign}(r_{Mi}) \left[-2 \{ r_{Mi} + \delta_i \log(\delta_i - r_{Mi}) \} \right]^{1/2}, \quad i = 1, \dots, n.$$

şeklinde hesap edilir. Uygulama için elde edilen artıkların standart $Q-Q$ plot grafiği aşağıda Şekil 4.20'deki gibi elde edilmiştir.



Şekil 4.12. *Deviance tipi artıkların Q-Q plot grafiği.*

Grafikten de görüldüğü üzere, uygulama için hataların *OBu-L* dağıldığı varsayımı sağlanmaktadır.

5. SONUÇ

Bu tez çalışmasında yaşam analizinde sıklıkla ele alınan konulardan olan yaşam süresinin modellenmesine ve log-yaşam süresi regresyon modeline katkı sağlaması amacıyla dağılım genelleştirme yöntemlerinden olan $T-X$ dağılımlar ailesi (Alzaatreh ve ark., 2013) yöntemini temel alan Odd-Burr-G (Alizadeh ve ark., 2016) metodu kullanılarak üç farklı dağılım genelleştirilmiştir.

Öncelikle Rayleigh dağılımı temel dağılım olarak kullanılmış olup Odd-Burr Rayleigh dağılımı elde edilmiştir. Dağılım, olasılık yoğunluk, yaşam, tehlike oranı ve kantil fonksiyonları elde edilmiş olup momentler, koşullu momentler ve sıra istatistikleri gibi bir dizi istatistiksel özellikleri incelenmiştir. Dağılımın olasılık yoğunluk ve tehlike oranı fonksiyonunun grafikleri incelendiğinde iki modlu verilerin ve azalan artan yapıda tehlike oranı yapısına sahip veri tiplerinin modellenebileceği esnekliğe sahip bir dağılım olduğu görülmüştür. Ayrıca dağılımın basıklık ve çarpıklık ölçüleri de hesaplanmış olup sabit basıklık ve çarpıklık değerlerine sahip olan Rayleigh dağılıma göre oldukça esnek bir dağılımın elde edildiği sonucuna varılmıştır. Elde edilen yeni dağılımın parametrelerinin tahmini için en çok olabilirlik yöntemi uygulanmıştır. Bu yöntemin parametre tahmininde tutarlılığını göstermek amacıyla benzetim çalışması yapılmış olup benzetim çalışması sonucunda tahmincinin tutarlılığı ortaya konmuştur. $OBu-R$ dağılımının yaşam süresini modellemedeki başarısını ortaya koyabilmek için gerçek veri seti üzerindeki modelleme performansı incelenmiştir. Bu amaçla literatürde daha önce genelleştirilen Rayleigh dağılımları ve Weibull dağılımı ile yaşam süresi verilerini içeren gerçek veri seti uygulaması yapılmış ve sonuç olarak elde edilen dağılımın modelleme performansının karşılaştırılan dağılımlardan daha iyi olduğu sonucuna varılmıştır. Ayrıca Rayleigh dağılımı rüzgâr hızı analizi literatüründe de önemli bir yere sahiptir. Temel alınan dağılımın literatürde önem arz ettiği alandaki, başarısını da gösterebilmek amacıyla elde edilen yeni dağılımın performansı rüzgâr hızı verileri üzerinde de incelenmiştir. Bu amaçla rüzgâr literatüründe önemli yere sahip dağılımlar ile belirli bir bölgenin aylık rüzgâr hızı verileri kullanılarak $OBu-R$ dağılımının karşılaştırması yapılmıştır. Elde edilen sonuçlar incelendiğinde $OBu-R$ dağılımının, rüzgâr hızını modellemede, karşılaştırmada kullanılan diğer dağılımlardan daha başarılı olduğu sonucuna varılmıştır. Böylelikle yaşam verilerini modellemenin yanı sıra rüzgâr enerjisi literatürüne de rüzgâr hızını modelleme başarısı yüksek bir dağılım sunulmuştur.

Daha sonra daha çok taşkın gibi ekstrem verileri modellemede kullanılan Pareto dağılımı temel dağılım olarak alınmış ve Odd Burr Pareto dağılımı önerilmiştir. Dağılıma ait dağılım, olasılık yoğunluk, yaşam, tehlike oranı ve kantil fonksiyonları elde edilmiştir. Moment, tamamlanmamış moment, ortalamadan sapma ve sıra istatistikleri gibi istatistiksel özellikleri incelenmiştir. Bu dağılımın istatistiksel özelliklerinin incelenmesinin yanı sıra Burr XII, Lomax gibi dağılımlar ile olan ilişkileri de ele alınmıştır. Parametre tahmini için en çok olabilirlik yöntemi kullanılmış olup bu yöntemin tutarlılığı benzetim çalışması yardımıyla gösterilmiştir. Dağılımın yaşam süresi verilerindeki performansını ortaya koyabilmek adına yaşam verileri içeren gerçek bir veri seti uygulaması aracılığıyla daha önceden genişletilmiş Pareto dağılımları ile performans karşılaştırılması yapılmış olup *OBu-P* dağılımının daha iyi modelleme başarısı göstermiş olduğu sonucuna varılmıştır. Ayrıca elde edilen yeni dağılımın, temel alınan dağılımın sıklıkla kullanıldığı alanlardaki başarısını da inceleyebilmek için nehir taşkın verilerini ele alarak da bir uygulama yapılmıştır. Bu uygulama sonucunda da *OBu-P* dağılımının karşılaştırılan dağılımlara nazaran daha başarılı olduğu sonucuna varılmış ve bu alana da modellemede kullanılabilecek yeni bir enstrüman sunulmuştur.

Son olarak yaşam analizinde de önem arz eden klasik istatistiksel dağılımlardan olan log-lojistik dağılımı temel alınarak Odd Burr Log-Lojistik dağılımı önerilmiştir. Elde edilen bu yeni dağılımın yaşam süresini modellemedeki performansını gerçek veri seti kullanarak incelemek amacıyla literatürdeki genelleştirilmiş log-lojistik dağılımları ile karşılaştırılması yapılmış olup *OBu-LL* dağılımının daha iyi performansa sahip olduğu görülmüştür. Ayrıca *OBu-LL* dağılımını temel alan log-doğrusal regresyon modeli de elde edilmiştir. Regresyon modelinin performansını incelemek amacıyla sansürlü gözlemlerin de olduğu gerçek veri seti üzerinde uygulama yapılmıştır. Dağılımın alt modeli olan lojistik model ve literatürde yaygın kullanıma sahip log-Weibull modeli kullanılarak yapılan performans değerlendirmesinde, elde edilen yeni regresyon modelinin diğer modellere nazaran daha başarılı olduğu ortaya koyulmuş olup ayrıca yapılan olabilirlik oran testi ile de bu sonuç desteklenmiştir.

Yukarıda özetlenen sonuçlardan sonra bu çalışmayı daha ileriye götürebilecek çalışma önerileri:

- *OBu-R* ve *OBu-P* dağılımlarının sansürlü veriler üzerindeki performanslarının incelenmesi,

- *OBu-R* dağılımının farklı bölgelere ait aylık, yıllık rüzgâr hızını modellemedeki performansının incelenmesinin yanı sıra, belirli bir bölgenin rüzgâr gücü potansiyelinin tahminin de incelenmesi,
- Elde edilen dağılımlar için en çok olabilirlik tahmincisinin yanı sıra özellikle dayanıklı tahmincileri dikkate alarak alternatif tahmincilerin incelenmesi,
- Farklı dağılım genişletme yöntemlerinin incelenmesi ile alternatif farklı dağılımların ortaya konulması,

biçiminde ifade edilebilir.



KAYNAKÇA

- Akinsete A., Famoye F. ve Lee, C. (2008). The beta-Pareto distribution. *Statistics*, 42(6), 547-563.
- Alexander, C., Cordeiro, G. M., Ortega, E. M. M. ve Sarabia, J. M. (2012). Generalized beta-generated distributions. *Comput Stat Data Anal.*, 56, 1880–1897.
- Alizadeh, A., Bagheri, F. S., Alizadeh, M. ve Nadarajah, S. (2017). A new four-parameter lifetime distribution. *Journal of Applied Statistics*, 44(5), 767-797.
- Alizadeh M., Cordeiro, G. M., Nascimento A. D. C., Lima M. C. S. ve Ortega, E. M. M. (2017). Odd-Burr generalized family of distributions with some applications. *Journal of Statistical Computation and Simulation*, 87(2), 367-389.
- Aljarrah M. A., Famoye F. ve Lee, C. (2015). A new Weibull-Pareto distribution. *Commun. Stat. Theory Methods*, 44, 4077-4095.
- Altun G., Alizadeh M., Altun E. ve Ozel G. (2017). Odd Burr Lindley distribution with properties and applications. *Hacettepe Journal of Mathematics and Statistics*, 46(2), 255- 276.
- Alzaatreh, A., Famoye, F. ve Lee, C. (2012). Gamma-Pareto distribution and its applications, *Journal of Modern Applied Statistical Methods*, 11(1), 78-94.
- Alzaatreh, A., Lee, C. ve Famoye, F. (2013). A new method for generating families of continuous distributions. *Metron*, 71, 63–79.
- Arnold B. C. (2008). *Pareto and Generalized Pareto Distributions*, Chapter 7, Springer.
- Ata, N. (2010). *Orantısız tehlikeler için yaşam modelleri*. Doktora Tezi. Ankara: Hacettepe Üniversitesi, Fen Bilimleri Enstitüsü.
- Azzalini, A. (1985). A class of distributions which includes the normal ones. *Scand. J. Stat.*, 12, 171–178.
- Balay, İ. G. (2014). *Yaşam-stres modellerinde ve hızlandırılmış başarısızlık süresi modelinde çarpık dağılımlara dayalı dayanıklı parametre tahmini*. Doktora Tezi. Ankara: Ankara Üniversitesi, Fen Bilimleri Enstitüsü.
- Birnbaum, Z. W. ve Saunders, S. C. (1969). A new family of life distributions. *J. Appl. Probab.* 6,319–327.
- Bourguignon, M., Silva, R. B., Zea, L. M. ve Cordeiro, G. M. (2013). The Kumaraswamy Pareto distribution. *JSTA*, 12(2), 129-144.
- Bourguignon, M., Silva, R. B. ve Cordeiro, G. M. (2014). The Weibull-G family of probability distributions. *J Data Sci.* 12, 53–68.
- Bradley, D. H, Bradley E. L. ve Naftel, D. C. (1984). A Generalized Gompertz-Rayleigh Model as A Survival Distribution, *Mathematical Biosciences*, 70(2), 195-202.
- Bhattacharya, S. K. ve Tyagi, R. K. (1990). Bayesian survival analysis based on the Rayleigh model. *Trabajos de Estadística*, 5: 81-92.

- Caudill S. B. (2012). A partially adaptive estimator for the censored regression model based on a mixture of normal distributions. *Stat. Methods Appl.*, 21, 121–137.
- Collett, D. (2003). *Modelling Survival data in medical research*. 2.basım. Amerika: Chapman and Hall/CRC.
- Cooray K ve Ananda M. M. A. (2005) Modeling actuarial data with a composite lognormal-Pareto model. *Scand. Actuar. J.*, 5, 321–334.
- Cordeiro, M. G., Ortega, E. M. M. ve Nadarajah, S. (2010). The Kumaraswamy Weibull distribution with application to failure data. *Journal of the Franklin Institute*, 347(8), 1399-1429.
- Cordeiro, G. M. ve de Castro, M. (2011). A new family of generalized distributions. *J. Statist. Comput. Simul.*, 81, 883–898.
- Cordeiro, G. M., Gomes A. E., da-Silva, C. Q. ve Ortega, E. M. M. (2013). The beta exponentiated Weibull distribution. *Journal of Statistical Computation and Simulation*, 83(1), 114-138.
- Cordeiro, M. G., Ortega, E. M. M., Popovic, B. V. ve Pescim R. R. (2014). The Lomax generator of distributions: Properties, minification process and regression model. *Applied Mathematics and Computation*, 247, 465-486.
- Cordeiro, M. G., Ortega, E. M. M. ve Popovic, B. V. (2015). The gamma-Lomax distribution. *Journal of Statistical Computation and Simulation*, 85(2), 305-319.
- Cox, D. R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society, Series B*, 34, 187-220.
- de Pascoa, M. A. R., Ortega, E. M. M. ve Cordeiro, M. G. (2011). The Kumaraswamy generalized gamma distribution with application in survival analysis. *Statistical Methodology*, 8, 411-433.
- Eugene, N., Lee, C. ve Famoye, F. (2002). Beta-normal distribution and its applications. *Commun. Stat. Theory Methods*, 31, 497–512.
- Fisk, P.R. (1961). The graduation of income distribution. *Econometrica*, 29, 171–185.
- Galton, F. (1983). *Enquiries into Human Faculty and its Development*. London: Macmillan & Company.
- Godase, S. S., Shirke, D. T. ve Kashid, D. N. (2017). Tolerance intervals and confidence intervals for the scale parameter of Pareto-Rayleigh distribution, *Electronic Journal of Applied Statistical Analysis*, 10, 1, 29-49.
- Gomes A. E., da-Silva, C. Q., Cordeiro, M. G., ve Ortega, E. M. M. (2014). A new lifetime model: the Kumaraswamy generalized Rayleigh distribution. *Journal of Statistical Computation and Simulation*, 84(2), 290-309.
- Gradshteyn, I. S ve Ryzhik I. M. (2002). *Table of integrals, series, and products*. San Diego, CA: Academic Press.
- Gui W. 2013. Marshall-Olkin extended log-logistic distribution and its application in minification processes. *Appl Math Sci.*, 7, 3947-3961.

- Jones, M. C. (2004). Families of distributions arising from distributions of order statistics. *Test*, 13(1), 1–43.
- Johnson, N. L. (1949). Systems of frequency curves generated by methods of translation. *Biometrika*, 36, 149–176.
- Johnson, N.L., Kotz, S. ve Balakrishnan, N. (1994). *Continuous Univariate Distributions*, cilt 1. 2. basım. New York: Wiley.
- Kantar Y. M. ve Usta I. (2015). Analysis of the upper-truncated Weibull distribution for wind speed. *Energy Convers. Manage.*, 96, 81–88.
- Kantar Y. M., Usta I., Arik I. ve Yenilmez I. (2018). Wind speed analysis using the extended generalized Lindley distribution. *Renewable Energy*, 118, 1024-1030.
- Kaplan, E.L. ve Meier. P. (1958). Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association*, 53, 457-481.
- Kleinbaum, D.G. ve Klein, M. (2010). *Survival Analysis: A Self-Learning Text*. 3. basım. New York: Springer.
- Lawless, J. F. (2003). *Statistical Models and Methods for Survival Data*. 2.basım. New Jersey: Wiley.
- Lee, C., Famoye, F. ve Olumolade, O. (2007). Beta-Weibull distribution: some properties and applications to censored data. *Journal of Modern Applied Statistical Methods*, 6(1), 173-186.
- Lee, C., Famoye, F. ve Alzaatreh, A. Y. (2013). Methods for generating families of univariate continuous distributions in the recent decades. *WIREs Computational Statistics*, 5, 219-238.
- Lee, E. T. ve Wang, J. W. (2003). *Statistical Methods for Survival Data Analysis*. 3.basım. New Jersey: Wiley.
- Lemonte, A. J. (2013). A new extended Birnbaum–Saunders regression model for lifetime modeling. *Computational Statistics and Data Analysis*, 64, 34-50.
- Lemonte, A. J. ve Cordeiro, M. G. (2013). An extended Lomax distribution. *Statistics*, 47(5), 800-816.
- Lemonte, A. J. (2014). The beta log-logistic distribution. *Braz. J. Probab. Stat.*, 28, 313–332.
- Lima, S. R. ve Cordeiro, G. M. (2017). The extended log-logistic distribution: properties and application. *Annals of the Brazilian Academy of Sciences*, 89(1), 3-17.
- Luguterah, A. (2016). Odd Generalized Exponential Rayleigh Distribution. *Advances and Applications in Statistics*, 48(1), 33-48.
- Mahmoudi E. (2011). The beta generalized Pareto distribution with application to lifetime data. *Math Comput Simul.*, 81, 2414-2430.
- Mahmoud, M. A. W. ve Ghazal, M. G. M. (2017). Estimations from the exponentiated Rayleigh distribution based on generalized Type-II hybrid censored data. 25, 71-78.

- Marshall, A. W. ve Olkin, I. (1997). A new method for adding a parameter to a family of distributions with application to the exponential and Weibull families. *Biometrika*, 84, 641-652.
- MirMostafaei, S. M. T. K., Mahdizadeh, M. ve Lemonte, A. J. (2017.). The Marshall-Olkin extended generalized Rayleigh distribution: Properties and applications, *Communications In Statistics-Theory And Methods*, 46(2), 653-671.
- Moors, J. J. (1988). A quantile alternative for kurtosis. *Statistician*, 37, 25–32.
- Mudholkar, G. S. ve Srivastava D. K. (1993). Exponentiated Weibull family for analyzing bathtub failure-rate data. *IEEE Trans Reliabil*, 42, 299–302.
- Nadarajah S. (2005). Exponentiated Pareto distributions. *Statistics*, 39(3), 255–260.
- Nadarajah, S., Bakouch, H., ve Tahmasbi, R. (2011). A generalized Lindley distribution. *Sankhyā: The Indian Journal of Statistics, Series B (2008-)*, 73(2), 331-359.
- Ortega, E. M. M., Cordeiro, M. G., ve Hashimoto, E. M. (2011). A Log-Linear Regression Model for the Beta-Weibull Distribution. *Communications in Statistics-Simulation and Computation*, 40(8), 1206-1235.
- Ortega, E. M. M., Cordeiro, M. G. ve Lemonte, A. J. (2012). A log-linear regression model for the beta-Birnbaum–Saunders distribution with censored data. *Computational Statistics and Data Analysis*, 56, 698-718.
- Ortega, E. M. M., Lemonte, A. J., Cordeiro, G. M. ve da Cruz, J. N. (2016). The odd Birnbaum–Saunders regression model with applications to lifetime data. *Journal of Statistical Theory and Practice*, 10, 780-804.
- Ouahabi, M. H., Benabdelouahab, F. ve Khamlichi, A. (2017). Analyzing wind speed data and wind power density of Tetouan city in Morocco by adjustment to Weibull and Rayleigh distribution functions. *Wind Engineering*, 41 (3), 174-184.
- Pearson, K. (1895). Contributions to the mathematical theory of evolution. II. Skew variation in homogeneous material. *Philos. Trans. Royal Soc. Lond. A*, 186, 343–414.
- Pescim R. R., ve Ortega, E. M. M., Cordeiro, M. G. ve Alizadeh M. (2017). A new log-location regression model: estimation, influence diagnostics and residual analysis, *Journal of Applied Statistics*, 44(2), 233-252.
- Potdar, K. ve Shirke, D. (2013). Inference for the parameters of generalized inverted family of distributions. *In Prod. Stat. Forum*, 6, 18–28.
- Ramos, M. W. A. R., Cordeiro, G. M., Marinho, P. R. D., Dias, C. R. B. ve Hamedani, G.G. (2013). The Zografos-Balakrishnan log-logistic distribution: properties and applications. *J. Stat. Theory Appl.*, 12, 225–244.
- Ramberg, J. S. ve Schmeiser, B. W. (1974). An approximate method for generating asymmetric random variables. *Commun. Assoc. Comput. Machinery*, 17,78–82.
- Ramberg, J. S., Tadikamalla, P. R., Dudewicz, E. J. ve Mykytka, E. F. (1979). A probability distribution and its uses in fitting data. *Technometrics*, 21, 201–214.

- Rashwan, N. I. (2016). A Note on Kumaraswamy Exponentiated Rayleigh distribution, *Journal Of Statistical Theory and Applications*, 15 (3), 286-29.
- Rosaiah, K., Rao, G. S., Sivakumar D. C. U. ve Kalyani, K. (2016). The odd generalized exponential log logistic distribution. *International. Journal of Mathematics and Statistics Invention (IJMSI)*, 4(5), 21-29.
- Saboor, A. ve Pogany, T., K. (2016). Marshall–Olkin gamma–Weibull distribution with applications. *Communications in Statistics - Theory and Methods*, 45(5), 1550-1563.
- Santana, T. V. F., Ortega, E. M. M., Cordeiro, G. M. ve Silva, G.O. (2012). The Kumaraswamy-log-logistic distribution. *J. Stat. Theory Appl.* 11, 265–291.
- Silva, G. O., Ortega, E. M. M., Cancho, V. G. ve Barreto, M. L. (2008). Log-Burr XII regression models with censored data. *Computational Stat. & Data Analysis*, 52, 3820-3842.
- Smith, R. L. ve Naylor, J. C. (1987). A comparison of maximum likelihood and Bayesian estimators for the three-parameter Weibull distribution. *Applied Statistics*, 36, 358-369.
- Tahir, M. H., Mansoor, M., Zubair, M. ve Hamedani, G. G. (2014). McDonald log-logistic distribution with an application to breast cancer data. *Journal of Statistical Theory and Applications*, 13(1), 65-82.
- Tahir. M. H., Cordeiro, G. M., Alizadeh, M., Mansoor, M., Zubair, M. ve Hamedani, G. G. (2015). The odd generalized exponential family of distributions with applications. *Journal of Statistical Distributions with Applications*, 2, 1-28.
- Tahir, M. H, Cordeiro, G. M., Alzaatreh, A., Mansoor, M. ve Zubair, M. (2016) A new Weibull–Pareto distribution: properties and applications. *Commun. Stat. Simul. Comput.* 45(10), 3548–3567.
- Verhulst, P.F. (1838). Notice sur la loi que la population suit dans son accroissement. *Correspondence mathématique et physique*, 10, 113–121.
- Zea, L. M, Silva R. B., Bourguignon M., Santos A. M. ve Cordeiro G. M. (2012). The beta exponentiated Pareto distribution with application to bladder cancer susceptibility. *International Journal of Statistics and Probability*, 1(2) 8–19.
- Zografos, K. ve Balakrishnan, N. (2009). On families of beta- and generalized gamma-generated distributions and associated inference. *Stat. Methodol.* 6, 344–362.

ÖZGEÇMİŞ

Adı Soyadı : İbrahim ARIK
Yabancı Dil : İngilizce
Doğum Yeri ve Yılı : Kütahya / 1984
E-Posta : iarik@anadolu.edu.tr

Eğitim ve Mesleki Geçmişi:

- 2014-2018, Doktora, Anadolu Üniversitesi, Fen Bilimleri Enstitüsü, İstatistik Anabilim Dalı,
- 2011-2014, Yüksek Lisans, Anadolu Üniversitesi, Fen Bilimleri Enstitüsü, İstatistik Anabilim Dalı,
- 2004-2009, Lisans, Kadir Has Üniversitesi, Fen Edebiyat Fakültesi, İstatistik ve Bilgisayar Bilimleri Bölümü (Tam Burslu).
- 2016- , Araş. Gör., Bilecik Şeyh Edebali Üniversitesi, Fen Edebiyat Fakültesi, İstatistik Bölümü,
- 2011-2016, Araş. Gör., Anadolu Üniversitesi, Fen Fakültesi, İstatistik Bölümü
- 2010-2011, Araş. Gör., Bilecik Şeyh Edebali Üniversitesi, Fen Edebiyat Fakültesi, İstatistik Bölümü,
- 2005-2010, Memur, SHÇEK.

Yayımları ve/veya Bilimsel/Sanatsal Faaliyetleri:

- Usta, I., Arık, I., Yenilmez, I. ve Kantar, Y. M. (2018). A new estimation approach based on moments for estimating Weibull parameters in wind power applications. *Energy Conversion and Management*, 164, 570-578.
- Kantar, Y. M., Usta, I., Arık, I. ve Yenilmez, I. (2018). Wind speed analysis using the Extended Generalized Lindley Distribution. *Renewable Energy*, 118, 1024-1030.
- Kantar, Y. M., Usta, I., Yenilmez, I. ve Arık, I. (2016). A Study on Estimation of Wind Speed Distribution by Using the Modified Weibull Distribution. *International Journal of Informatics Technologies*, 9(2).
- Kantar, Y. M., Arık, I., Usta, I. ve Yenilmez, I. (2016). Comparison of Some Estimation Methods of the two parameter Weibull Distribution for Unusual Wind Speed Data Cases. *International Journal of Informatics Technologies*, 9(2).
- Arık, I., Kantar, Y. M. ve Yenilmez, I. (2015). The Evaluation of Robust and Efficient Estimators for Log Logistic Distribution for Censored Data with without Outliers. *Journal of Scientific Research and Development*.

- Kantar Y. M. ve Arık, I. (2014). M-Estimation of Log Logistic Distribution Parameters with Outliers. *International Journal of Agricultural and Statistical Sciences*.
- Kantar, Y.M. ve Arık, I. (2014). The Use of the Data Transformation Techniques in Estimating the Shape Parameter of the Weibull Distribution for the Wind Speed. *International Journal of Energy Optimization and Engineering*, 3(3), 20-33.

