

ANKARA ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

YÜKSEK LİSANS TEZİ

İKİ BİLEŞENLİ KARMA ÜSTEL DAĞILIMLAR İÇİN PARAMETRE
TAHMİN YÖNTEMLERİ

Buse BÜYÜM

İSTATİSTİK ANABİLİM DALI

ANKARA
2017

Her hakkı saklıdır

TEZ ONAYI

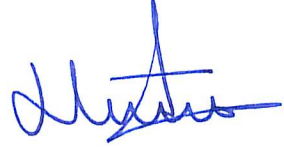
Buse BÜYÜM tarafından hazırlanan “İki Bileşenli Karma Üstel Dağılımlar İçin Parametre Tahmin Yöntemleri” adlı tez çalışması 13/06/2017 tarihinde aşağıdaki jüri tarafından oy birliği ile Ankara Üniversitesi Fen Bilimleri Enstitüsü İstatistik Anabilim Dalı’nda **YÜKSEK LİSANS TEZİ** olarak kabul edilmiştir.

Danışman : Prof. Dr. Mehmet YILMAZ
Ankara Üniversitesi İstatistik Anabilim Dalı



Jüri Üyeleri:

Başkan: Prof. Dr. Halil AYDOĞDU
Ankara Üniversitesi İstatistik Anabilim Dalı



Üye : Prof. Dr. Mehmet YILMAZ,
Ankara Üniversitesi İstatistik Anabilim Dalı



Üye : Yrd. Doç. Dr. Emel KIZILOK KARA
Kırıkkale Üniversitesi Aktüerya Bilimleri Bölümü



Yukarıdaki sonucu onaylarım.

Prof. Dr. Atila YETİŞEMİYEN
Enstitü Müdürü

ETİK

Ankara Üniversitesi Fen Bilimleri Enstitüsü tez yazım kurallarına uygun olarak hazırladığım bu tez içindeki bütün bilgilerin doğru ve tam olduğunu, bilgilerin üretilmesi aşamasında bilimsel etiğe uygun davrandığımı, yararlandığım bütün kaynakları atıf yaparak belirttiğimi beyan ederim.

06.07.2017



Buse BÜYÜM

ÖZET

Yüksek Lisans Tezi

İKİ BİLEŞENLİ KARMA ÜSTEL DAĞILIMLAR İÇİN PARAMETRE TAHMİN YÖNTEMLERİ

Buse BÜYÜM

Ankara Üniversitesi
Fen Bilimleri Enstitüsü
İstatistik Anabilim Dalı

Danışman: Prof. Dr. Mehmet YILMAZ

Bu çalışmanın öncelikli amacı, sonlu sayıda üstel dağılımın ağırlıklı toplamı ile elde edilen karma dağılımın parametre tahmin yöntemlerini incelemektir. İki bileşenli karma üstel dağılımın, yapısal olarak literatürde mevcut bulunan birçok karma dağılıma göre daha basit olması özelliği onun bu çalışmada kullanılmasına öncülük etmiştir. Çalışmada, öncelikle iki bileşenli karma üstel dağılımın parametrelerine ilişkin tahmin yöntemleri incelenmiş ve bu yöntemlerin etkinliği farklı örneklem hacimleri için simülasyon çalışması ile değerlendirilmiştir. Son olarak, literatürde kullanılan mevcut veriler ile bu çalışma için toplanan veri seti iki bileşenli karma üstel dağılım ile modellenmiş ve tahmin yöntemlerinin çeşitli kriterlere göre karşılaştırılması yapılmıştır.

Haziran 2017, 110 sayfa

Anahtar Kelimeler: Karma üstel dağılımlar, parametre tahmini, momentler yöntemi, en küçük kareler yöntemi, en çok olabilirlik yöntemi, EM algoritması.

ABSTRACT

Master Thesis

PARAMETER ESTIMATION METHODS FOR TWO-COMPONENT MIXED EXPONENTIAL DISTRIBUTIONS

Buse BÜYÜM

Ankara University
Graduate School of Natural and Applied Sciences
Department of Statistics

Supervisor: Prof. Dr. Mehmet YILMAZ

The primary aim of this study is to investigate parameter estimation methods of mixed distribution obtained by the weighted sum of a finite number of exponential distribution. Because of its simpler structure than many mixed distribution in the literature, two-component mixed exponential distribution is pioneered the use in this study. In here, parameter estimation methods for two-component mixed exponential distribution are primarily examined and the efficiency of these methods are evaluated with simulation study for different sample size. Finally, the data set, used in the literature and gathered for this study, are modeled by two-component mixed exponential daistribution.

June 2017, 110 pages

Key Words: Mixed exponential distributions, parameter estimation, method of moments, least squares method, maximum likelihood method, EM algorithm.

ÖNSÖZ ve TEŞEKKÜR

Çalışmamın her aşamasında önerileriyle beni yönlendiren, bana araştırma olanağı sağlayan, her zaman ilgisini ve alakasını gördüğüm danışman hocam Sayın Prof. Dr. Mehmet YILMAZ'a (Ankara Üniversitesi İstatistik Anabilim Dalı),

Çalışmam boyunca desteğini bir an olsun esirgemeyen ve zor günlerimde hep yanımda olup beni cesaretlendiren Sayın Araş. Gör. Mustafa Hilmi PEKALP'e (Ankara Üniversitesi İstatistik Anabilim Dalı),

İstatistik bilimini bana sevdiren ve öğrenmemde büyük katkıları olan Ankara Üniversitesi İstatistik Bölümü'ndeki bütün hocalarıma,

Hayatımın her aşamasında olduğu gibi çalışmalarım süresince benden maddi ve manevi desteklerini esirgemeyen ve her zaman arkamda durarak beni çalışmaya teşvik eden canım aileme içten teşekkürlerimi sunarım.

Bu tez çalışması, Ankara Üniversitesi Bilimsel Araştırmalar Projesi (BAP) kapsamında desteklenmiştir.

Buse BÜYÜM

Ankara, Haziran 2017

İÇİNDEKİLER

ETİK.....	i
ÖZET.....	ii
ABSTRACT	iii
ÖNSÖZ ve TEŞEKKÜR.....	iv
SİMGELER DİZİNİ	vii
ŞEKİLLER DİZİNİ	viii
ÇİZELGELER DİZİNİ	x
1. GİRİŞ	1
2. SONLU KARMA DAĞILIMLAR İLE İLGİLİ TEMEL KAVRAMLAR	7
2.1 Sonlu Karma Dağılım ve Olasılık Yoğunluk Fonksiyonları	7
2.2 Sonlu Karma Dağılımlarda Parametre Kümesi.....	9
2.3 Sonlu Karma Dağılımlardaki Temel Sorunlar	12
2.3.1 Belirlenebilirlik (Identifiability)	12
2.4 İki Bileşenli Karma Üstel Dağılım	16
2.4.1 İki bileşenli karma üstel dağılım için görsel örnekler	18
3. MATERYAL VE YÖNTEM.....	26
3.1 Momentler Yöntemi	26
3.2 En Çok Olabilirlik Yöntemi	28
3.2.1 EM Algoritması	28
3.2.1.1 EM Algoritmasının karma dağılımlarda genel işleyişi	32
3.2.1.2 EM Algoritmasının işleyişine örnekler	37
3.2.2 Newton-Raphson Yöntemi	46
3.3 En Küçük Kareler Yöntemi	47
3.4 Kolmogorov-Smirnov Testi	47
4. ARAŞTIRMA BULGULARI	49
4.1 İki Bileşenli Karma Üstel Dağılım İçin Momentler Yöntemi	49
4.1.1 İki bileşenli karma üstel dağılım parametreleri için α biliniyorken moment tahmin edicileri.....	50
4.1.2 İki bileşenli karma üstel dağılım parametreleri için α bilinmiyorken moment tahmin edicileri.....	51
4.2 İki Bileşenli Karma Üstel Dağılım İçin En Çok Olabilirlik Yöntemi.....	54

4.2.1 İki bileşenli karma üstel dağılım için EM algoritması.....	57
4.2.2 ML'nin sayısal çözümü için Newton Raphson Yöntemi.....	58
4.3 İki Bileşenli Karma Üstel Dağılım İçin En Küçük Kareler Yöntemi.....	60
5. SİMÜLASYON ÇALIŞMASI.....	61
6. UYGULAMA ÇALIŞMASI	75
7. SONUÇLAR	88
KAYNAKLAR	91
EKLER.....	96
EK 1 Proschan (1963) çalışmasından alınan veri seti çizelgesi	97
EK 2 Maguire, Pearson ve Wynn (1952) çalışmasından alınan veri seti çizelgesi	98
EK 3 Kuş (2007) çalışmasından alınan veri seti çizelgesi	99
EK 4 Dekking v.d. (2010) çalışmasından alınan veri seti çizelgesi	100
EK 5 AL-Jammal (2008) çalışmasından alınan veri seti çizelgesi	101
EK 6 Kumar ve Klefsjö (1992) çalışmasından alınan veri seti çizelgesi.....	102
EK 7 Nelson (1990) çalışmasından alınan veri seti çizelgesi	103
EK 8 Blischke (1995) çalışmasından alınan veri seti çizelgesi	104
EK 9 Kardiyoloji kliniğinde yatan kadın hastaların hastanede kalma sürelerine ilişkin veri seti çizelgesi	105
EK 10 Kardiyoloji kliniğinde yatan erkek hastaların hastanede kalma sürelerine ilişkin veri seti çizelgesi	106
EK 11 Beyin cerrahisi kliniğinde yatan kadın hastaların hastanede kalma sürelerine ilişkin veri seti çizelgesi	107
EK 12 Beyin cerrahisi kliniğinde yatan erkek hastaların hastanede kalma sürelerine ilişkin veri seti çizelgesi	108
ÖZGEÇMİŞ.....	109

SİMGELER DİZİNİ

χ^2 Ki-Kare Test İstatistiği

Kısaltmalar

Log-likelihood	Logaritması alınan olabilirlik
H	Hessian
MOM	Momentler Yöntemi
ML	En Çok Olabilirlik Yöntemi
MLE	En Çok Olabilirlik Tahmini
LS	En Küçük Kareler Yöntemi
LSE	En Küçük Kareler Tahmini
EM	Expectation-Maximization
RMSE	Hata Kareler Ortalamasının Kare Kökü
2MED	İki Bileşenli Karma Üstel Dağılım
KS	Kolmogrov-Smirnov

ŞEKİLLER DİZİNİ

Şekil 2.1 İki bileşenli karma üstel dağılımın histogramı ($\theta_1 = 10$ (koyu çizgi ile) $\theta_2 = 1$ ve $\alpha = 0,4$).....	18
Şekil 2.2 İki bileşenli karma üstel dağılımın histogramı ($\theta_1 = 10$ (koyu çizgi ile) $\theta_2 = 1$ ve $\alpha = 0,6$).....	19
Şekil 2.3 İki bileşenli karma üstel dağılımın histogramı ($\theta_1 = 10$ (koyu çizgi ile) $\theta_2 = 1$ ve $\alpha = 0,8$).....	19
Şekil 2.4 $\theta_1 = 10$ ve $\theta_2 = 1$ iken farklı α değerleri için iki bileşenli karma üstel dağılımın olasılık yoğunluk fonksiyonları.....	20
Şekil 2.5 $\theta_1 = 10$ ve $\theta_2 = 1$ iken farklı α değerleri için iki bileşenli karma üstel dağılımın dağılım fonksiyonları.....	20
Şekil 2.6 $\theta_1 = 10$ ve $\theta_2 = 1$ iken farklı α değerleri için iki bileşenli karma üstel dağılımın yaşam fonksiyonları.....	21
Şekil 2.7 $\theta_1 = 10$ ve $\theta_2 = 1$ iken farklı α değerleri için iki bileşenli karma üstel dağılımın bozulma oran fonksiyonları	21
Şekil 2.8 İki bileşenli karma üstel dağılımın histogramı ($\theta_1 = 0.8$ (koyu çizgi ile) $\theta_2 = 0.3$ ve $\alpha = 0,4$).....	22
Şekil 2.9 İki bileşenli karma üstel dağılımın histogramı ($\theta_1 = 0.8$ (koyu çizgi ile) $\theta_2 = 0.3$ ve $\alpha = 0,6$).....	22
Şekil 2.10 İki bileşenli karma üstel dağılımın histogramı ($\theta_1 = 0.8$ (koyu çizgi ile) $\theta_2 = 0.3$ ve $\alpha = 0,8$).....	23
Şekil 2.11 $\theta_1 = 0.8$ ve $\theta_2 = 0.3$ iken farklı α değerleri için iki bileşenli karma üstel dağılımın olasılık yoğunluk fonksiyonları.....	23
Şekil 2.12 $\theta_1 = 0.8$ ve $\theta_2 = 0.3$ iken farklı α değerleri için iki bileşenli karma üstel dağılımın dağılım fonksiyonları.....	24
Şekil 2.13 $\theta_1 = 0.8$ ve $\theta_2 = 0.3$ iken farklı α değerleri için iki bileşenli karma üstel dağılımın yaşam fonksiyonları.....	24
Şekil 2.14 $\theta_1 = 0.8$ ve $\theta_2 = 0.3$ iken farklı α değerleri için iki bileşenli karma üstel dağılımın bozulma oran fonksiyonları.....	25
Şekil 6.1 Beyin Cerrahisi Kliniği'nde yatan kadın hastaların hastanede kalma sürelerine ilişkin histogram.....	81
Şekil 6.2 Beyin Cerrahisi Kliniği'nde yatan erkek hastaların hastanede kalma sürelerine ilişkin histogram.....	82

Şekil 6.3 Beyin Cerrahisi Kliniği'nde yatan tüm hastaların hastanede kalma sürelerine ilişkin histogram.....	83
Şekil 6.4 Kardiyoloji Kliniği'nde yatan kadın hastaların hastanede kalma sürelerine ilişkin histogram.....	84
Şekil 6.5 Kardiyoloji Kliniği'nde yatan erkek hastaların hastanede kalma sürelerine ilişkin histogram.....	85
Şekil 6.6 Kardiyoloji Kliniği'nde yatan tüm hastaların hastanede kalma sürelerine ilişkin histogram.....	86



ÇİZELGELER DİZİNİ

Çizelge 3.1 Oyuncak örneği için EM algoritması ile elde edilen parametre değerleri....	41
Çizelge 3.2 Bir görüntüdeki nesneleri algılama problemi için EM algoritması ile elde edilen parametre değerleri.....	45
Çizelge 4.1 İki bileşenli karma üstel dağılımdan üretilen veri setleri ($\theta_1 = 5$, $\theta_2 = 3$ ve $\alpha = 0.6$).....	54
Çizelge 4.2 Veri setlerine ilişkin hesaplanan değerler.....	54
Çizelge 5.1 İki bileşenli karma üstel dağılıma ilişkin üretilen veri seti ($\theta_1 = 10$, $\theta_2 = 2$ ve $\alpha = 0.625$).....	61
Çizelge 5.2 MOM yöntemine göre veri setine ilişkin hesaplanan değerler.....	62
Çizelge 5.3 MOM parametre tahminleri	62
Çizelge 5.4 İki bileşenli karma üstel dağılıma ilişkin üretilen veri seti ($\theta_1 = 3$, $\theta_2 = 1$ ve $\alpha = 0.25$).....	62
Çizelge 5.5 MOM yöntemine göre veri setine ilişkin hesaplanan değerler	63
Çizelge 5.6 İki bileşenli karma üstel dağılıma ilişkin üretilen veri seti ($\theta_1 = 3$, $\theta_2 = 1$ ve $\alpha = 0.875$).....	63
Çizelge 5.7 MOM yöntemine göre veri setine ilişkin hesaplanan değerler	63
Çizelge 5.8 MOM parametre tahminleri	64
Çizelge 5.9 İki bileşenli karma üstel dağılıma ilişkin üretilen veri seti ($\theta_1 = 8$, $\theta_2 = 2$ ve $\alpha = 0.75$).....	64
Çizelge 5.10 MOM yöntemine göre veri setine ilişkin hesaplanan değerler	64
Çizelge 5.11 MOM parametre tahminleri	65
Çizelge 5.12 İki bileşenli karma üstel dağılıma ilişkin üretilen veri seti ($\theta_1 = 8$, $\theta_2 = 0.2$ ve $\alpha = 0.75$).....	65
Çizelge 5.13 MOM yöntemine göre veri setine ilişkin hesaplanan değerler	65
Çizelge 5.14 MOM parametre tahminleri	66

Çizelge 5.15 İki bileşenli karma üstel dağılıma ilişkin üretilen veri seti ($\theta_1 = 8$, $\theta_2 = 0.2$ ve $\alpha = 0.75$).....	66
Çizelge 5.16 MOM yöntemine göre veri setine ilişkin hesaplanan değerler	66
Çizelge 5.17 MOM, LS ve ML yöntemi ile elde edilen parametre tahminleri ve RMSE değerleri (n=20)	68
Çizelge 5.18 MOM, LS ve ML yöntemi ile elde edilen parametre tahminleri ve RMSE değerleri (n=40)	69
Çizelge 5.19 MOM, LS ve ML yöntemi ile elde edilen parametre tahminleri ve RMSE değerleri (n=100)	70
Çizelge 5.20 İki bileşenli karma üstel dağılıma ilişkin üretilen veri seti ($\theta_1 = 2$, $\theta_2 = 0.3$, $\alpha = 0.625$ ve n=200).....	71
Çizelge 5.21 MOM, ML ve LS yöntemi ile elde edilen parametre tahminleri ve (χ^2) değerleri ($\theta_1 = 2$, $\theta_2 = 0.3$ $\alpha = 0.625$ ve n=200).....	72
Çizelge 5.22 İki bileşenli karma üstel dağılıma ilişkin üretilen veri seti ($\theta_1 = 10$, $\theta_2 = 2$, $\alpha = 0.625$ ve n=200)	72
Çizelge 5.23 MOM, ML ve LS yöntemi ile elde edilen parametre tahminleri ve (χ^2) değerleri ($\theta_1 = 10$, $\theta_2 = 2$, $\alpha = 0.625$ ve n=200)	73
Çizelge 5.24 İki bileşenli karma üstel dağılıma ilişkin üretilen veri seti ($\theta_1 = 1$, $\theta_2 = 3$, $\alpha = 0.4$ ve n=200)	73
Çizelge 5.25 MOM, ML ve LS yöntemi ile elde edilen parametre tahminleri ve (χ^2) değerleri ($\theta_1 = 1$, $\theta_2 = 3$, $\alpha = 0.4$ ve n=200)	74
Çizelge 6.1 Veri seti 1 için ML parametre tahminleri, p ve KS değerleri.....	75
Çizelge 6.2 Veri seti 2 için ML parametre tahminleri, p ve KS değerleri.....	76
Çizelge 6.3 Veri seti 3 için ML parametre tahminleri, p ve KS değerleri.....	77
Çizelge 6.4 Veri seti 4 için ML parametre tahminleri, p ve KS değerleri.....	77
Çizelge 6.5 Veri seti 5 için ML parametre tahminleri, p ve KS değerleri (Makine 1).....	78

Çizelge 6.6 Veri seti 5 için ML parametre tahminleri, p ve KS değerleri (Makine 2).....	78
Çizelge 6.7 Veri seti 6 için ML parametre tahminleri, p ve KS değerleri.....	79
Çizelge 6.8 Veri seti 7 için ML parametre tahminleri, p ve KS değerleri.....	79
Çizelge 6.9 Veri seti 8 için ML parametre tahminleri, p ve KS değerleri.....	80
Çizelge 6.10 Beyin Cerrahisi Kliniği'nde yatan kadın hastalar ele alındığında modelleme sonucu elde edilen ML parametre tahminleri, p ve KS değerleri.....	81
Çizelge 6.11 Beyin Cerrahisi Kliniği'nde yatan erkek hastalar ele alındığında modelleme sonucu elde edilen ML parametre tahminleri, p ve KS değerleri.....	82
Çizelge 6.12 Beyin Cerrahisi Kliniği'nde yatan hastalar ele alındığında modelleme sonucu elde edilen ML parametre tahminleri, p ve KS değerleri	83
Çizelge 6.13 Kardiyoloji Kliniği'nde yatan kadın hastalar ele alındığında modelleme sonucu elde edilen ML parametre tahminleri, p ve KS değerleri.....	85
Çizelge 6.14 Kardiyoloji Kliniği'nde yatan erkek hastalar ele alındığında modelleme sonucu elde edilen ML parametre tahminleri, p ve KS değerleri.....	86
Çizelge 6.15 Kardiyoloji Kliniği'nde yatan hastalar ele alındığında modelleme sonucu elde edilen ML parametre tahminleri, p ve KS değerleri.....	87

1. GİRİŞ

Gerçek veriler alınarak yapılan tahmin ve modelleme problemlerinde, verinin genellikle tek bir kanuna sahip olduğu söylenir. Ancak ne var ki, bu elde edilen verilerin ya aynı dağılım ailesinin bir karması ya da kirli veriler olarak ortaya çıkması muhtemeldir. Örneğin, iki farklı risk taşıyan aynı hastalığa sahip bireylerin yaşam zamanları veya belirli bir birimde hastanede yatış yapan hastaların hastanede kalma süreleri, belirli özellikleri göz önüne alınan evli çiftlerin ilk çocuklarına sahip oluncaya kadar geçen süre, günlük yağın yağmurun şiddeti (m^2 ye düşen yağış miktarı), suçlunun aynı suçu (suçun türüne göre) işleyinceye kadar geçirdiği zaman, farklı sosyo kültüre sahip çiftlerin boşanma süreleri, farklı kalitelerde ama aynı işleve sahip parçaların ilk bozuluncaya kadar sistemdeki yaşam zamanları karma üstel dağılım ile modellenebilir.

Karma dağılımlarda parametre tahmini yapılırken iki dezavantaj ortaya çıkmaktadır. Bunlardan ilki, ne kadar fazla karma dağılım varsayımına gidilirse tahmini yapılacak parametre sayısı da bir o kadar artmaktadır. Bu da çoğu zaman doğrusal olmayan denklemlerin algoritmalar ile bilgisayar ortamında çözülebilme problemini ortaya çıkarmaktadır. Dolayısı ile böyle bir araştırmaya girilirken ya var olan algoritmalar ile ya da yeni bir algoritma ile problemi çözüme ulaştırmak gerekeceğini unutmamak gerekir. İkincisi, birbirine yakın ölçek parametresine sahip iki karma dağılımdan oluşan modelde sanki bir dağılımı gibi bir sonuç çıkabilir. Bu durumda, parametreler üzerinde gerekli ve probleme uygun koşullar konularak bu dezavantajı görmezden gelmek mümkün olabilir. Bu sorunu örnekle anlatmak gerekirse, biri 8 yıl diğeri 10 yıl dayanan aynı güce sahip ampullerden oluşan bir kutuyu habersizce satın alan kişi, ampullerin dayanma sürelerini belirlemek isteyebilir. Bu durumda kişi dayanma sürelerinin farklı iki dağılımdan geldiğini belirleyemeyebilir çünkü iç içe girmiş iki dağılımı birbirinden ayırt etmek zordur. Karma dağılımların tek bir dağılım olarak algılandığı duruma ilişkin günlük hayattan verilebilecek birçok örnek ile yukarıda bahsedilen örnek çoğaltılabilir. Karma dağılımlarda parametre tahminlerinde karşılaşılabilecek sorunlara ilişkin verilen örneğe ilaveten, sonlu karma dağılımların literatürde bulunan uygulama alanlarına bakıldığında da oldukça geniş bir alana sahip olduğu görülmektedir. Yapılan literatür

taraması sonucunda elde edilen tek boyutlu ve karma dağılımlara ilişkin bazı çalışmalara bu bölümde yer verilmektedir.

Chu ve Ke (2012), karma Weibull dağılımlarının parametreleri için en çok olabilirlik ve en küçük kareler yöntemlerini karşılaştırmış ve en küçük kareler yönteminin en çok olabilirlik yöntemine göre veri boyutu küçük olduğunda daha iyi sonuç verdiğini belirtmişlerdir.

Weinberg ve Finch (2012), Pareto dağılımının yüksek çözünürlüklü deniz gözetleme radarlarının sinyal işleme sürecinin modellenmesinde önemli rol oynadığı düşüncesiyle, pareto dağılımının karmasının da bu modelleme için etkili olacağını ileri sürmüşlerdir.

Petersen (2000), Pareto dağılımının parametrelerini tahmin etmek için korelasyon (ilişki) katsayılarına dayalı bir yöntem kullanmıştır.

Rosa ve Figueiredo (2010), çalışmalarında iki farklı Weibull ailesine sahip olan kırılğan materyallerin (cam, seramik vs.) hangi aileden geldiklerini belirlemek için en çok olabilirlik yöntemi yardımıyla parametre tahmini yapmışlardır.

Suhaila vd. (2011), Peninsular, Malezya Bölgesi'ndeki ana istasyonlardan alınan yağış oranlarına en uygun modellemeyi sağlayan dağılımın karma lognormal dağılım olduğunu ileri sürmüşlerdir.

Suhaila ve Jemain (2007) ise Peninsular, Malezya Bölgesi'ndeki yağış oranlarını modellemede optimal modeli elde edebilmek için karma gamma, karma Weibull ve karma üstel ile bu dağılımların tek boyutlularını önermişlerdir. Çalışmalarının sonucunda ise uyum iyiliği testine göre karmaların daha iyi sonuç verdiğini, karmaların arasında ise en uygun dağılımın karma Weibull dağılımı olduğunu belirtmişlerdir.

Chiodo (2012), güç (power) istatistikleriyle ilişkili rüzgâr hızlarını modellerken karma Weibull dağılımından faydalanmıştır.

Hung ve Chang (2011), karma Weibull dağılımının parametre tahmin sürecinde önemli olanın bileşen sayısının belirlenmesi olduğunu söylemiş ve bu sayıyı belirlemede kullanılan momentler yöntemi ve bulanık kümeleme algoritmasının bir düzenlemesi olan iki farklı yöntemi ele almışlardır.

Coroller vd. (2006), bakteri hücrelerinin dayanıklılığını Weibull dağılımının karmaları ile modellemişlerdir.

Anup vd. (2011), yüksek voltaj izolasyon güvenilirliğini hem Weibull dağılımını hem de Weibull karmalarını kullanarak değerlendirmişlerdir.

Nagode ve Fajdiga (2000), karma dağılımların bilinmeyen parametrelerinin tahmin edilmesi zor olduğundan, parametrelerin kabaca tahmin edilebilmesi için hem grafik yöntemini hem de optimizasyon tekniklerini içeren bir algoritma önermişlerdir.

Yukarıda değinilen çalışmalara ek olarak, literatür taraması sonucunda karma üstel dağılımın genellikle tıp alanında hasta gruplarını analiz etmekte, yaşam zamanlarını belirlemede, hastaların hastanede kalma sürelerini elde etmede ve bazı maliyetlerin veya gereksinimlerin belirlenmesinde, ekonomi alanında işsizliğin ve risk durumundaki kayıpların incelenmesinde, meteoroloji alanında günlük yağış miktarlarının modellenmesinde, bunların dışında ise özellikle makine ve parçalarının bozulma veya yaşam zamanlarının incelenmesinde kullanıldığı görülmektedir.

Tıp alanındaki uygulamalardan bazıları şunlardır:

Thomas (1966), sinir hücrelerinin boşalma zamanının dağılım modeli için karma üstel kullanmıştır. Suchndran ve Lachenbruch (1974), gebe kalma zamanları ile ilgili çalışmalarında karma üstellerin parametre tahmininde en çok olabilirlik yöntemini kullanmışlardır. Hougaard (1984) ve Aalen (1988), bireyler ya da gruplar arasında yaşam uzunluklarındaki farklılıkları test etmek için karma üstel dağılımdan yararlanmışlardır (Erol 1995). Marshall vd. (2004), C-ph modelini çeşitli hasta alt gruplarını ve farklı

hastalık aşamasındakilerin beklenen maliyetini önceden tahmin edebilmek için yaşlı hastaların kalış sürelerini ele alarak karma üstel dağılımı kullanmışlardır.

Taylor vd. (1999), hastaların yatak kullanım gereksinimleri belirleyebilmek için Markov modeli geliştirmiş ve bu modeli kurarken ele alınan hastaların ne kadar süreyle hastanede kaldığı bilinmediğinden beş değişkenli karma üstel dağılımı kullanmışlardır. Harrison (2001), hastanede yatılı kalan hastaların kalma sürelerinin en iyi iki ya da üç bileşenli karma üstel dağılım ile modellenebileceğini vurgulamıştır. Ayrıca, Harrison (2001) hastaların taburcu olma sürelerine göre hasta gruplarını üçe ayırmış ve iki farklı hastanede kalmakta olan hastaların kalma sürelerini karma üstel dağılım ile modellemiştir. Prodan ve Prodan, (2001), hasta bakım zamanlarını karma üstel dağılım ile modellemiş ve çalışmalarının sonunda kurdukları model ile hem yaşlı hastaların hem de hastane yöneticilerinin hastane kaynaklarını optimal şekilde kullanmalarını sağlamayı amaçlamışlardır. Marshall vd. (2005) ile Vasilakis ve Marshall (2005) çalışmalarında hastaların hastanede kalma sürelerini koşullu ve koşulsuz karma üstel dağılım yardımıyla incelenmişlerdir.

Ekonomi alanındaki uygulamalardan başlıcaları şunlardır:

Blossfeld vd. (1989), Hackman (1991), ekonomide işsizlik süresi verilerini modellerken karma üstel dağılımlarından faydalanmışlardır (Mosler ve Seidel 2001). Keatinge (1999), çalışmasında karma üstel dağılımın aktüerler açısından kullanımını sağlamak amacıyla riskte kayıplar üzerine eğilmiş ve bu kayıpların modellenmesi ile ilgilenmiştir.

Meteoroloji alanındaki uygulamalardan bazıları ise şu şekildedir:

Wilks (1998), sulak alanlara yakın olan yerlerde küçük miktarlardaki yağışları modellemede karma üstel dağılımın gamma dağılımına göre daha uygun olduğunu söylemiştir. Wan vd. (2005), karma üstel dağılımın yaz aylarındaki yağışları modellemede iyi sonuç verdiğini, gamma dağılımının ise Kanada'da kış aylarındaki yağışları modellemede daha uygun olduğunu belirtmişlerdir. Li vd. (2013), yirmi dört

hava istasyonu ve iki havzadan alınan veriler ile günlük yağış miktarlarını modellemeye çalışmışlar ve yağış miktarlarını modellerken ortalama, varyans ve uç değer istatistikleri ile sıklık dağılımını kullanmışlardır. Çalışma sonunda, yağış miktarlarının dağılım fonksiyonunun parametre sayısı ile ilişkili olduğu ve üç parametrelili karma üstel dağılımın yağış miktarlarını modellemede en iyi modellemeyi sağladığı belirtilmiştir. Fadhilah vd. (2007), Wilayah Persekutan bölgesindeki saatlik yağış miktarının modellenmesi için en iyi dağılımı belirlemeye çalışmışlar ve karma üstel dağılımın en iyi modellemeyi yapan dağılım olduğu sonucuna ulaşmışlardır.

Karma üstel dağılım ile yapılan diğer çalışmaların bazıları ise şu şekildedir:

Davis (1952), servis otobüsü bozulma zamanları uygulamasında normal ve üstel dağılımları; Merdenhull ve Harder (1958), verici-alıcı kullanım süreleri uygulamasında karma üstel dağılımın parametre tahmininde en çok olabilirlik yöntemini kullanmışlardır.

Everitt ve Hand (1981), valf kullanma süresi uygulamasında karma üstel dağılımlarda en çok olabilirlik yöntemini kullanmışlardır.

Rider (1961), iki bileşenli karma üstel dağılımın parametrelerini momentler yöntemini kullanarak tahmin etmiştir.

Perlstein vd. (2001), yaşam dağılımları karma üstel tarafından modellenen bir ürün partisi için optimal bozulma sürelerini belirlemek istemişlerdir. Kullandıkları yöntemde bozulma sürelerini ve garanti kapsam sürecindeki maliyetleri yansıtan toplam beklenen maliyet fonksiyonunun hatalarını en aza indirmeye çalışmışlardır. Bu beklenen maliyet fonksiyonu, karma üstel dağılım fonksiyonu ile bu dağılıma ilişkin parametreler kullanılarak oluşturulmuştur.

Wu vd. (2007), optimal bozulma sürelerini ve tamir edilemeyen ürünler için garanti süresini belirlemek amacıyla bir maliyet modeli oluşturmaya çalışmışlardır. Çalışmada

ele alınan ürünlerin bozulma zamanları hem karma üstel hem de karma Weibull dağılım ile modellenmiştir.

Mendoza-Rosas ve Cruz-Rayna (2009), farklı volkanik patlama olasılıkları tarafından karakterize edilen karma üstel dağılımları kullanarak bir sonraki patlamaların kümülatif sayılarını belirlemeye çalışmışlardır.

Çalışma aşağıdaki gibi düzenlenmiştir.

İkinci bölümde, genel olarak sonlu karma dağılımlar ile ilgili bazı temel kavramlar, sonlu karma dağılımların olasılık yoğunluk fonksiyonu, parametre kümesi, sonlu karma dağılımlarda belirlenememe sorunu ile çalışmanın konusunu oluşturan iki bileşenli karma üstel dağılım anlatılacaktır.

Üçüncü bölümde, sonlu karma dağılımların parametre tahmin yöntemlerinden olan momentler, en çok olabilirlik ve en küçük kareler yöntemleri ile en çok olabilirlik yönteminin nümerik çözüm yollarından olan EM Algoritması ve Newton-Raphson yöntemine değinilecektir.

Dördüncü bölümde, üçüncü bölümde ele alınan parametre tahmin yöntemleri iki bileşenli karma üstel dağılım için ayrıca incelenecektir.

Beşinci bölümde, yapılan simülasyon çalışması sonuçlarına yer verilecek ve ayrıca momentler, en çok olabilirlik ve en küçük kareler yöntemlerinin parametre tahminindeki başarısı ki-kare test istatistiğine göre karşılaştırılacaktır.

Altıncı bölümde, simülasyon çalışmasına ek olarak literatürde sıkça karşılaşılan veri setlerinin iki bileşenli karma üstel dağılım ile modellenmesiyle elde edilen sonuçlara yer verilecektir. Ayrıca yine aynı bölümde çalışma için toplanan gerçek veri setine ilişkin modelleme sonuçlarının değerlendirilmesi de yapılacaktır.

Yedinci bölümde ise simülasyon ve yapılan modelleme çalışmaları sonucunda elde edilen bulgulara yer verilecektir.

2. SONLU KARMA DAĞILIMLAR İLE İLGİLİ TEMEL KAVRAMLAR

Bu bölümde genel olarak sonlu karma dağılımlara ilişkin temel kavramlardan bahsedilerek bazı özelliklere yer verilmiştir.

2.1 Sonlu Karma Dağılım ve Olasılık Yoğunluk Fonksiyonları

$\underline{\theta} = \underline{\theta}$ verilmişken $X_1 = x_1, X_2 = x_2, \dots, X_n = x_n$ gözlemine dayalı olarak $f(\underline{x}|\underline{\theta})$ koşullu olasılık (yoğunluk) fonksiyonu göz önüne alınsın. Burada, $\underline{\theta}$ ($k \times 1$) boyutlu rasgele parametre vektörünü göstermektedir. $H(\underline{\theta})$, $\underline{\theta}$ parametre vektörünün ortak dağılım fonksiyonunu göstermek üzere;

$$f(\underline{x}) = \int f(\underline{x}|\underline{\theta}) dH(\underline{\theta}) \quad (2.1)$$

eşitliği ile belirlenen f yoğunluk fonksiyonuna karma yoğunluk fonksiyonu, bu şekilde elde edilen dağılıma da karma dağılım denir. Bu durumda $H(\underline{\theta})$ karışım dağılımını ifade etmektedir. Burada $f(\underline{x}|\underline{\theta})$ koşullu olasılık yoğunluk fonksiyonu keyfi seçimli değildir. Öyle ki $f(\underline{x}|\underline{\theta})$,

$$\{f(\underline{\theta}): \underline{\theta} \in \Theta \subset R^k\} \quad (2.2)$$

gibi parametrik bir olasılık yoğunluk fonksiyonları sınıfına aittir.

Bu tanıma göre $f(\underline{x})$, $(k + n)$ boyutlu ortak olasılık yoğunluk fonksiyonunun marjinal olasılık yoğunluk fonksiyonu olarak görülebilir.

Özel olarak θ 'nın sonlu sayıda değerleri $\theta_1, \theta_2, \dots, \theta_r$ için $H(\underline{\theta})$ kesikli değerler alıyorsa yukarıdaki integral, toplam ile ifade edilebilen sonlu bir karma dağılımı vermektedir.

$$f(\underline{x}) = \sum_{i=1}^r H_i(\theta_i) f(\underline{x}|\underline{\theta}) \quad (2.3)$$

Burada $f_1, f_2, \dots, f_r$ fonksiyonları X 'in değeri aldığı D_x destek kümesi üzerinde birer olasılık yoğunluk fonksiyonudur. $\theta_1, \theta_2, \dots, \theta_r$ 'ler sırasıyla $f_1, f_2, \dots, f_r$ bileşen olasılık yoğunluk fonksiyonları ile ilgili parametre vektörleridir ve bunlar karma dağılımın da parametrelerinden bir kısmı olup genellikle,

$$\Theta = \begin{bmatrix} \theta_1 \\ \theta_2 \\ \vdots \\ \theta_r \end{bmatrix} \quad (2.4)$$

vektörü ile gösterilir. Karışım oranları $H_i(\theta_i)$ ya da $\alpha_i \geq 0, i=1,2,\dots,r; \sum_{i=1}^r \alpha_i=1$ olmak üzere karışım oranlarına ait vektör,

$$\alpha_i = \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_{r-1} \end{bmatrix} \quad (2.5)$$

şeklinde dir. Bu durumda karma dağılımın parametre vektörü,

$$\Phi = \begin{bmatrix} r \\ \alpha \\ \theta \end{bmatrix} \quad (2.6)$$

dır. Bileşen sayısını belirten parametre $r \in \{1,2, \dots\}$ olmak üzere bu parametre biliniyorsa karma dağılımın parametre vektörü aşağıdaki gibi olur.

$$\Phi = \begin{bmatrix} \alpha \\ \theta \end{bmatrix} \quad (2.7)$$

Karma dağılımda bileşen olasılık (yoğunluk) fonksiyonları olan $f_1, f_2, \dots, f_r$ aynı dağılım ailesinin elemanı ise karma dağılım,

$$f(\underline{x}, \Phi) = \sum_{i=1}^r \alpha_i f_i(\underline{x}; \theta_i) \quad (2.8)$$

biçiminde yazılır.

$F_1, F_2, \dots, F_r$ bileşen olasılık (yoğunluk) fonksiyonlarına karşılık gelen dağılım fonksiyonlarını gösterebilir. Bu durumda karma dağılımın dağılım fonksiyonu,

$$F(\underline{x}, \Phi) = \sum_{i=1}^r \alpha_i F_i(\underline{x}; \underline{\theta}_i) \quad (2.9)$$

dır.

Karma dağılımlarda bileşen dağılımların var olmaları halinde momentleri,

$$m_{j_i} = \int_{-\infty}^{\infty} x^j f_i(\underline{x}; \underline{\theta}_i) dx \quad i = 1, 2, \dots, r \quad j = 1, 2, \dots \quad (2.10)$$

olmak üzere karma dağılımın momentleri,

$$m_j = E(X^j) = \sum_{i=1}^r \alpha_i m_{j_i} \quad (2.11)$$

dır. Eğer bileşen dağılımlarına ait moment çıkaran fonksiyon varsa, karma dağılımın da moment çıkaran fonksiyonu vardır ve

$$M_i(t) = \int_{-\infty}^{\infty} e^{tx} f_i(\underline{x}; \underline{\theta}_i) dx, \quad -h < t < h \quad (2.12)$$

olmak üzere

$$M(t) = \sum_{i=1}^r \alpha_i M_i(t) \quad (2.13)$$

şeklindedir.

2.2 Sonlu Karma Dağılımlarda Parametre Kümesi

Sonlu karma dağılıma sahip bir X rasgele değişkeninin veya rasgele vektörünün olasılık yoğunluk fonksiyonu

$$f(\underline{x}) = \sum_{i=1}^r \alpha_i f_i(\underline{x}; \underline{\theta}_i) \quad (2.14)$$

dır. Burada r bileşen sayısını, α_i karışım oranlarını göstermektedir. $\alpha_1, \alpha_2, \dots, \alpha_{r-1}$ ($\alpha_r = 1 - \sum_{i=1}^{r-1} \alpha_i$) parametreleri ve sırasıyla $n_1, n_2, \dots, n_r$ boyutlu olan $\theta_1, \theta_2, \dots, \theta_r$ bileşen yoğunluklarının parametreleri bilinmediğinde, (2.6) eşitliğinde verilen parametre vektörü $p = 1 + r - 1 + \sum_{i=1}^r n_i$ boyutludur. Yani bilinmeyen parametre sayısı p 'ye eşit ve parametre kümesi Ω ($\Phi \in \Omega$) p boyutlu standart vektör uzayının bir alt kümesidir. ($\Omega \in R^p$)

Eğer $\underline{\theta}_1, \underline{\theta}_2, \dots, \underline{\theta}_r$ vektörlerinin bazıları aynı parametre vektörü içeriyorsa, Φ parametre vektörünün boyutu $p = 1 + r - 1 + \sum_{i=1}^r n_i$ şeklinde belirlenemez. Bu durumda aynı olan parametreler sadece bir kez yazılıp, Φ vektörünün boyutu belirlenmelidir.

Karma dağılıma ilişkin bileşen sayısı biliniyorsa ve $\underline{\theta}_1, \underline{\theta}_2, \dots, \underline{\theta}_r$ vektörleri aynı parametre içermiyorsa parametre vektörü $p = r - 1 + \sum_{i=1}^r n_i$ boyutludur. Karma oranları (α_i) bilindiğinde parametre vektörü $p = \sum_{i=1}^r n_i$ boyutludur. Bileşen sayısı (r) ve bileşen yoğunluklarının parametreleri bilindiğinde bilinmeyen parametreler $\alpha_1, \alpha_2, \dots, \alpha_r$ olup parametre vektörü $p = r - 1$ boyutludur. Bileşen yoğunlukları aynı dağılım ailesinin elemanları yani $i = 1, 2, \dots, r$ için,

$$f_i \in F = \{f(., Y): Y \in \Gamma \subset R^m\} \quad (2.15)$$

olduğunda $\theta_i \in \Gamma$, $i = 1, 2, \dots, r$ olmak üzere $\sum_{i=1}^r m_i = r \times m$ şeklinde elde edilir. Örneğin, aynı olmayan iki üstel dağılımdan oluşan bir karma dağılım için,

$$f(\underline{x}; \Phi) = \alpha f_1(x; \theta_1) + (1 - \alpha) f_2(x; \theta_2) = \alpha \frac{1}{\theta_1} e^{-x \frac{1}{\theta_1}} + (1 - \alpha) \frac{1}{\theta_2} e^{-x \frac{1}{\theta_2}}$$

$\alpha \in (0,1), \theta_i > 0 (i = 1,2), x > 0$ şeklindedir. Burada parametre vektörü üç boyutlu olup, parametre kümesi R^3 'ün bir alt kümesidir yani bilinmeyen parametre sayısı 3'tür. Parametre vektörü;

$$\Phi = \begin{bmatrix} \alpha \\ \theta_1 \\ \theta_2 \end{bmatrix} \in \Omega = [0,1] \times [0, \infty] \times [0, \infty] \subset R^3$$

şeklindedir. Karma üstel dağılıma benzer olarak aynı olmayan iki Bernoulli dağılımdan oluşan bir karma dağılım için,

$$f(\underline{x}; \Phi) = \alpha p_1^x (1 - p_1)^{1-x} + (1 - \alpha) p_2^x (1 - p_2)^{1-x}, x = 0,1$$

şeklindedir. Burada parametre vektörü yine üç boyutlu olup, parametre kümesi R^3 'ün bir alt kümesidir yani bilinmeyen parametre sayısı 3'tür. Parametre vektörü

$$\Phi = \begin{bmatrix} \alpha \\ p_1 \\ p_2 \end{bmatrix} \in \Omega = [0,1] \times [0,1] \times [0,1] \subset R^3$$

şeklindedir.

Yukarıda verilen karma dağılım örneklerinden farklı olarak aynı olmayan iki Weibull dağılımının karmasından oluşan bir karma dağılım için,

$$f(\underline{x}; \Phi) = \alpha \frac{\tau_1 x^{\tau_1-1}}{\beta_1^{\tau_1}} e^{(-\frac{x}{\beta_1})^{\tau_1}} + (1 - \alpha) \frac{\tau_2 x^{\tau_2-1}}{\beta_2^{\tau_2}} e^{(-\frac{x}{\beta_2})^{\tau_2}}$$

$\alpha \in (0,1), x > 0, \beta_i > 0$ ve $\tau_i > 0 (i = 1,2)$ şeklindedir. Burada ise parametre vektörü beş boyutlu olup, parametre kümesi R^5 'ün bir alt kümesidir yani bilinmeyen parametre sayısı 5'tir. Parametre vektörü;

$$\Phi = \begin{bmatrix} \alpha \\ \tau_1 \\ \tau_2 \\ \beta_1 \\ \beta_2 \end{bmatrix} \in \Omega = [0,1] \times [0, \infty] \times [0, \infty] \times [0, \infty] \times [0, \infty] \subset R^5$$

olarak gösterilir.

Sonlu karma dağılımda parametreler üç gruba ayrılır. Birincisi sadece sonlu karmaların bileşen sayısını gösteren r 'leri, ikincisi karışım oranları $H_i(\theta_i)$ 'leri içerir. (Çalışma boyunca karışım oranları α_i ile ifade edilecektir.) Üçüncüsü ise karma dağılımın bileşenlerinin parametre vektörünü ($\underline{\theta}_i$) içermektedir.

Bu üç grup içinde parametre tahmini için yöntemler bulunmaktadır. Ancak burada bazı karma dağılım türleri için parametre tahminini karmaşıktıran temel zorluklardan bahsetmek gerekir.

2.3 Sonlu Karma Dağılımlardaki Temel Sorunlar

2.3.1 Belirlenebilirlik (Identifiability)

Belli bir dağılım fonksiyonları sınıfına ait bir dağılımın tek olarak belirlenebilmesidir. Bileşen dağılımları aynı parametrik aileden gelen bir karma dağılım göz önüne alınsın.

$\mathcal{G} = \{G(x|\theta), \theta \in \Theta, x \in R^n\}$ karma dağılımı oluşturan n boyutlu dağılım fonksiyonunun bir sınıfı (ailesi) olsun. $\mathcal{G}$ sınıfından elde edilen sonlu karmaların sınıfı uygun dağılım fonksiyonlarının sınıfı $\mathcal{F}$ ile belirlenecektir.

$$\mathcal{F} = \{F(x): F(x) = \sum_{i=1}^r \alpha_i G(\underline{x}; \theta_i), \alpha_i > 0, \sum_{i=1}^r \alpha_i = 1, G(\cdot, \theta_i) \in \mathcal{G} \text{ bütün } i = 1, 2, \dots, r, \underline{x} \in R^n\}$$

Öyle ki bu sınıf, $\mathcal{G}$ sınıfının konveks bir kabuğudur. $\mathcal{G}$ sınıfından alınan $G_1, G_2, \dots, G_r$ 'lerin farklı elemanlar olduğu varsayılacaktır.

Bu durumda $\mathcal{F}$ 'nin belirlenebilir olması için gerek ve yeter koşul $\mathcal{F}$ 'nin elemanlarının farklı olmasıdır.

Tanım 2.1 (Identifiability) : F ve $\tilde{F}$, $\mathcal{F}$ sınıfının iki elemanı olmak üzere,

$$F = \sum_{j=1}^r \alpha_j G_j \quad (2.16)$$

ve

$$\tilde{F} = \sum_{j=1}^{\tilde{r}} \tilde{\alpha}_j \tilde{G}_j \quad (2.17)$$

şeklinde yazılabilirsin. $F \equiv \tilde{F}$ olabilmesi için gerek ve yeter şart $r = \tilde{r}$ ve $\alpha_j = \tilde{\alpha}_j$, $G_j = \tilde{G}_j$ $j = 1, 2, \dots, r$ dir. O zaman $\mathcal{F}$ belirlenebilir. (Yakowitz ve Spragins 1968)

ÖRNEK 2.1 $U\left(0, \frac{1}{2}\right) = \alpha U(0, \alpha) + (1 - \alpha)U(0, 1 - \alpha)$ şeklinde iki bileşenli düzgün dağılım karması verilsin.

$$2x = \alpha \left(\frac{x - 0}{\alpha} \right) + (1 - \alpha) \left(\frac{x - 0}{1 - \alpha} \right)$$

dır. Burada, $2x$ 'i oluşturabilecek birbirinden farklı çok sayıda α değeri bulunabilir. Bu durum belirlenememe sorununu ortaya çıkarmaktadır.

ÖRNEK 2.2 (Everitt, 1981) Karma oranı $\frac{1}{3}$ olan $(-1, 1)$ aralığındaki düzgün dağılım ile karma oranı $\frac{2}{3}$ olan $(-2, 2)$ aralığındaki düzgün dağılımın bir karması ele alınsın. O halde,

$$F(x) \sim \frac{1}{3} U(-1, 1) + \frac{2}{3} U(-2, 2)$$

şeklindedir. (Burada $U(a, b)$, aralığı (a, b) olan düzgün dağılımdır.) Burada,

$$F(x) = \frac{1}{3} \left(\frac{x+1}{2} \right) + \frac{2}{3} \left(\frac{x+2}{4} \right) = \frac{2x+3}{6}, -\frac{3}{2} \leq x \leq \frac{3}{2}$$

olarak yazılır.

Şimdi karma oranı $\frac{1}{2}$ olan $(-2,1)$ aralığındaki düzgün dağılım ile karma oranı $\frac{1}{2}$ olan $(-1,2)$ aralığındaki düzgün dağılımın bir karması ele alınsın ve dağılım fonksiyonu,

$$F(x) \sim \frac{1}{2} U(-2,1) + \frac{1}{2} U(-1,2)$$

olarak verilsin. O halde,

$$F(x) = \frac{1}{2} \left(\frac{x+2}{3} \right) + \frac{1}{2} \left(\frac{x+1}{3} \right) = \frac{2x+3}{6}, -\frac{3}{2} \leq x \leq \frac{3}{2}$$

şeklinde yazılır. Görüldüğü üzere iki farklı karma oranlarına sahip düzgün dağılım için dağılım fonksiyonları aynı çıkmaktadır. Bu durum belirlenememe sorununa yol açmaktadır. Yani farklı karmalar için birden fazla çözüm ortaya çıkmaktadır ve parametre tahmini zorlaşmaktadır.

ÖRNEK 2.3 İki bileşenli Bernoulli karması,

$$g_1(x) = \alpha p_1^x (1 - p_1)^{1-x} + (1 - \alpha) p_2^x (1 - p_2)^{1-x}, \quad x = 0,1$$

şeklinde verilsin. Bu karma dağılım için $\alpha = \frac{1}{4}$, $p_1 = \frac{4}{5}$, $p_2 = \frac{4}{9}$ olarak verilsin. Bu durumda,

$$g_1(0) = \left(\frac{1}{4} \right) \left(\frac{1}{5} \right) + \left(\frac{3}{4} \right) \left(\frac{5}{9} \right) = \frac{7}{15}$$

ve

$$g_1(1) = \binom{1}{4} \binom{4}{5} + \binom{3}{4} \binom{4}{9} = \frac{8}{15}$$

olur.

Üç bileşenli Bernoulli karması,

$$g_1(x) = \alpha_1 p_1^x (1 - p_1)^{1-x} + \alpha_2 p_2^x (1 - p_2)^{1-x} + \alpha_3 p_3^x (1 - p_3)^{1-x}, \quad x = 0,1$$

şeklinde verilsin. Bu karma dağılım için $\alpha_1 = \frac{1}{8}$, $\alpha_2 = \frac{2}{8}$, $\alpha_3 = \frac{5}{8}$ ve $p_1 = \frac{2}{3}$, $p_2 = \frac{8}{12}$, $p_3 = \frac{34}{75}$ olarak verilsin. Bu durumda,

$$g_2(0) = \binom{1}{8} \binom{1}{3} + \binom{2}{8} \binom{4}{12} + \binom{5}{8} \binom{41}{75} = \frac{7}{15}$$

ve

$$g_2(1) = \binom{1}{8} \binom{2}{3} + \binom{2}{8} \binom{8}{12} + \binom{5}{8} \binom{34}{75} = \frac{8}{15}$$

şeklinde bulunur. Her iki karma dağılım için bulunan olasılık fonksiyonu değerleri aynı olduğundan belirlenememe sorunu ortaya çıkmaktadır.

ÖRNEK 2.4 İki bileşenli iki farklı Bernoulli karmaları,

$$g_1(x) = \alpha p_1^x (1 - p_1)^{1-x} + (1 - \alpha) p_2^x (1 - p_2)^{1-x}, x = 0,1$$

$$g_2(x) = \alpha p_1^x (1 - p_1)^{1-x} + (1 - \alpha) p_2^x (1 - p_2)^{1-x}, x = 0,1$$

$g_1(x)$ için $\alpha = \frac{25}{26}$, $p_1 = \frac{1}{20}$, $p_2 = \frac{3}{8}$ ve $g_2(x)$ için $\alpha = \frac{1}{4}$, $p_1 = \frac{1}{8}$, $p_2 = \frac{1}{24}$ olarak verilsin. Bu durumda,

$$g_1(x) = \left(\frac{25}{26}\right) \left(\frac{1}{20}\right)^x \left(\frac{19}{20}\right)^{1-x} + \left(\frac{1}{26}\right) \left(\frac{3}{8}\right)^x \left(\frac{5}{8}\right)^{1-x}, x = 0,1$$

$$g_2(x) = \left(\frac{1}{4}\right) \left(\frac{1}{8}\right)^x \left(\frac{7}{8}\right)^{1-x} + \left(\frac{3}{4}\right) \left(\frac{1}{24}\right)^x \left(\frac{23}{24}\right)^{1-x}, x = 0,1$$

olur ve

$$g_1(0) = \left(\frac{25}{26}\right) \left(\frac{19}{20}\right) + \left(\frac{1}{26}\right) \left(\frac{5}{8}\right) = \frac{15}{16}$$

$$g_1(1) = \left(\frac{25}{26}\right) \left(\frac{1}{20}\right) + \left(\frac{1}{26}\right) \left(\frac{3}{8}\right) = \frac{1}{16}$$

$$g_2(0) = \left(\frac{1}{4}\right) \left(\frac{7}{8}\right) + \left(\frac{3}{4}\right) \left(\frac{23}{24}\right) = \frac{15}{16}$$

$$g_2(1) = \left(\frac{1}{4}\right) \left(\frac{1}{8}\right) + \left(\frac{3}{4}\right) \left(\frac{1}{24}\right) = \frac{1}{16}$$

olarak elde edilir. İki bileşenli iki farklı Bernoulli karmaları için farklı α ve $i = 1,2$ için p_i oranları ele alınarak bulunan olasılık fonksiyonu değerleri aynı olmaktadır. Bu durum belirlenememe sorunu olarak karşımıza çıkar.

2.4 İki Bileşenli Karma Üstel Dağılım

İki bileşenli karma üstel dağılımın (2MED) olasılık yoğunluk fonksiyonu,

$$f(x; \alpha, \theta_1, \theta_2) = \alpha f_1(x; \theta_1) + (1 - \alpha) f_2(x; \theta_2)$$

$$= \alpha \frac{1}{\theta_1} e^{-x \frac{1}{\theta_1}} + (1 - \alpha) \frac{1}{\theta_2} e^{-x \frac{1}{\theta_2}} \quad (2.18)$$

$\alpha \in (0,1)$, $\theta_i > 0$ ($i = 1,2$), $x > 0$ şeklinde tanımlıdır. Benzer biçimde dağılım fonksiyonu,

$$F(x; \alpha, \theta_1, \theta_2) = \alpha F_1(x; \theta_1) + (1 - \alpha) F_2(x; \theta_2)$$

$$= \alpha \left(1 - e^{-x \frac{1}{\theta_1}}\right) + (1 - \alpha) \left(1 - e^{-x \frac{1}{\theta_2}}\right) \quad (2.19)$$

Yaşam fonksiyonu,

$$\begin{aligned} S(x; \alpha, \theta_1, \theta_2) &= \alpha S_1(x; \theta_1) + (1 - \alpha) S_2(x; \theta_2) \\ &= \alpha e^{-x \frac{1}{\theta_1}} + (1 - \alpha) e^{-x \frac{1}{\theta_2}} \end{aligned} \quad (2.20)$$

ve bozulma oranı fonksiyonu,

$$h(x; \alpha, \theta_1, \theta_2) = \frac{\alpha h_1(x) S_1(x; \theta_1) + (1 - \alpha) h_2(x) S_2(x; \theta_2)}{\alpha S_1(x; \theta_1) + (1 - \alpha) S_2(x; \theta_2)}$$

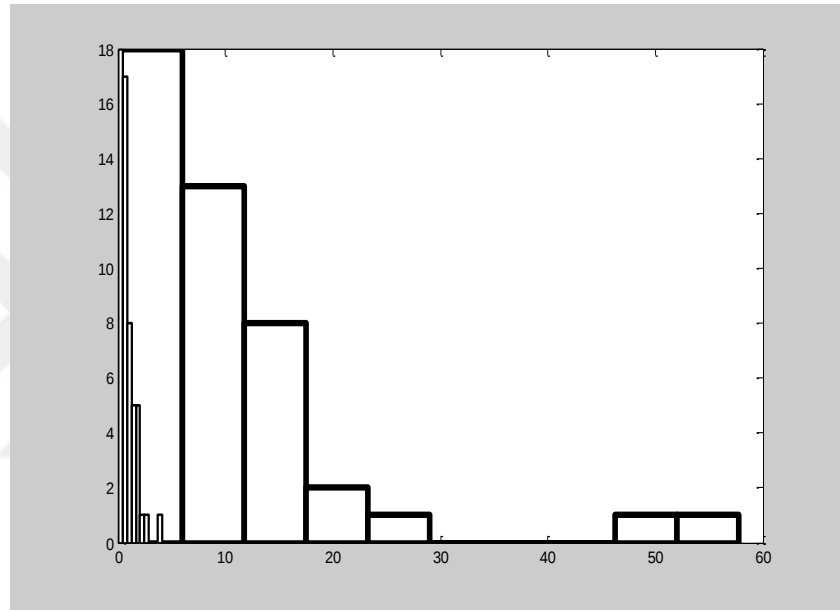
$$= h_1(x) \frac{\alpha S_1(x; \theta_1)}{\alpha S_1(x; \theta_1) + (1 - \alpha) S_2(x; \theta_2)} + h_2(x) \frac{(1 - \alpha) S_2(x; \theta_2)}{\alpha S_1(x; \theta_1) + (1 - \alpha) S_2(x; \theta_2)}$$

$$= h_1(x) w_1(x; \alpha, \theta_1, \theta_2) + h_2(x) w_2(x; \alpha, \theta_1, \theta_2) \quad (2.21)$$

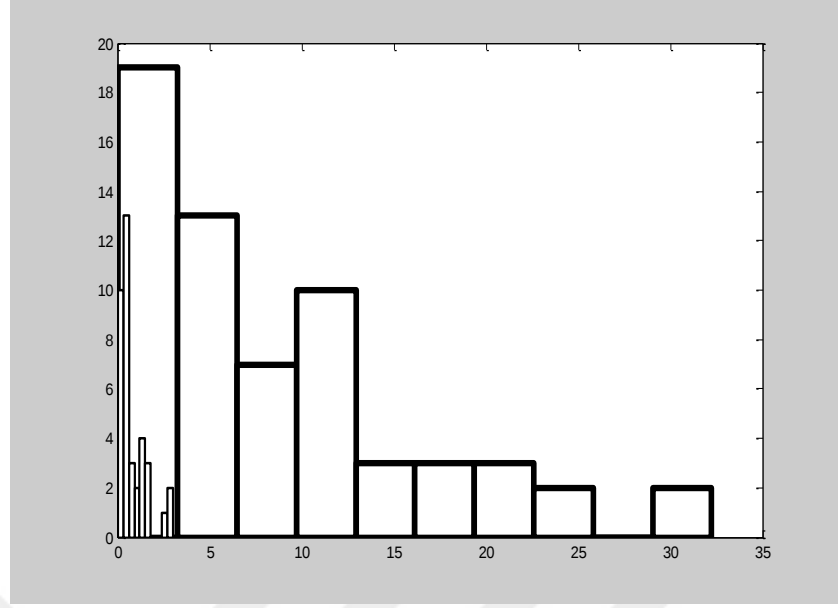
şeklinde tanımlanır. Burada $h_i(x) = \frac{1}{\theta_i}, w_1 + w_2 = 1$ 'dir.

2.4.1 İki bileşenli karma üstel dağılım için görsel örnekler

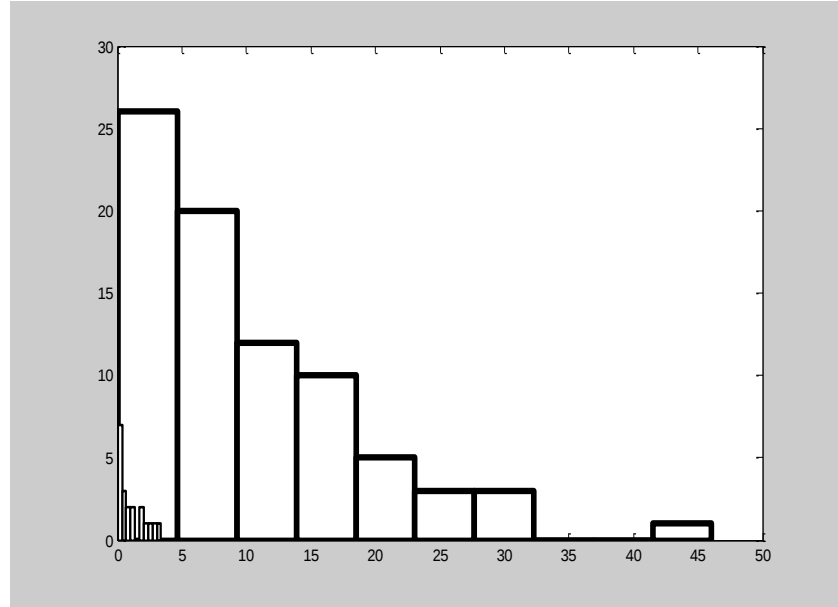
2MED'e ilişkin çeşitli parametre değerleri için olasılık yoğunluk fonksiyonu, yaşam fonksiyonu ve bozulma oranı fonksiyonu grafikleri aşağıda gösterilmektedir (Şekil 2.1-2.14).



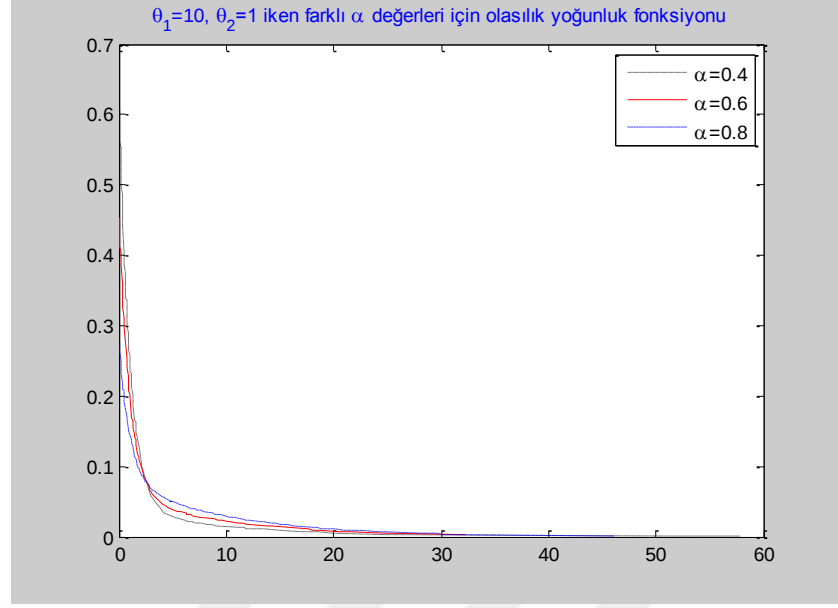
Şekil 2.1 İki bileşenli karma üstel dağılımın histogramı ($\theta_1 = 10$ (koyu çizgi ile) $\theta_2 = 1$ ve $\alpha = 0,4$)



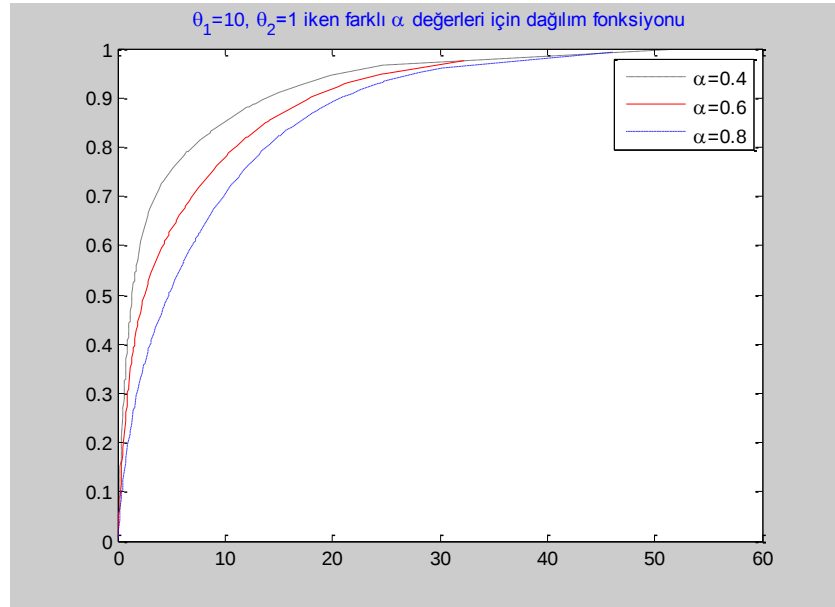
Şekil 2.2 İki bileşenli karma üstel dağılımın histogramı ($\theta_1 = 10$ (koyu çizgi ile) $\theta_2 = 1$ ve $\alpha = 0,6$)



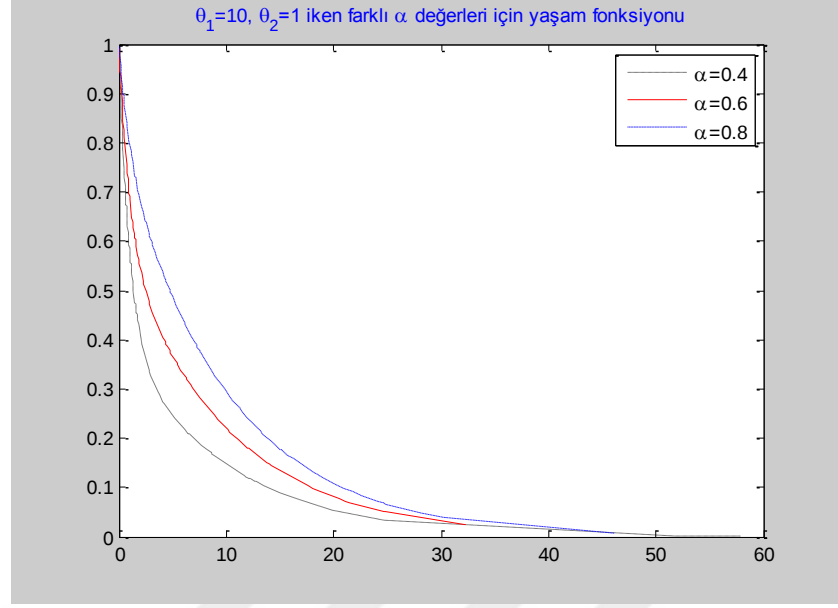
Şekil 2.3 İki bileşenli karma üstel dağılımın histogramı ($\theta_1 = 10$ (koyu çizgi ile) $\theta_2 = 1$ ve $\alpha = 0,8$)



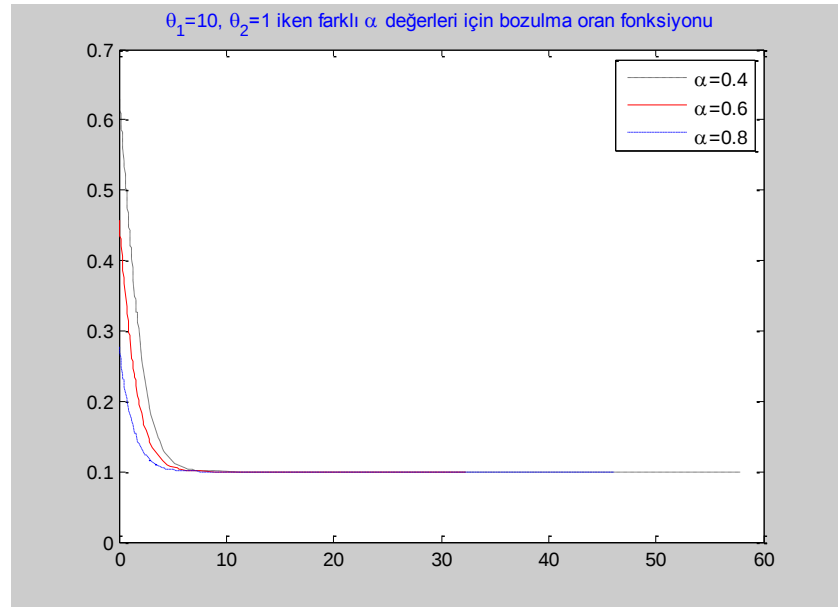
Şekil 2.4 $\theta_1 = 10$ ve $\theta_2 = 1$ iken farklı α değerleri için iki bileşenli karma üstel dağılımın olasılık yoğunluk fonksiyonları



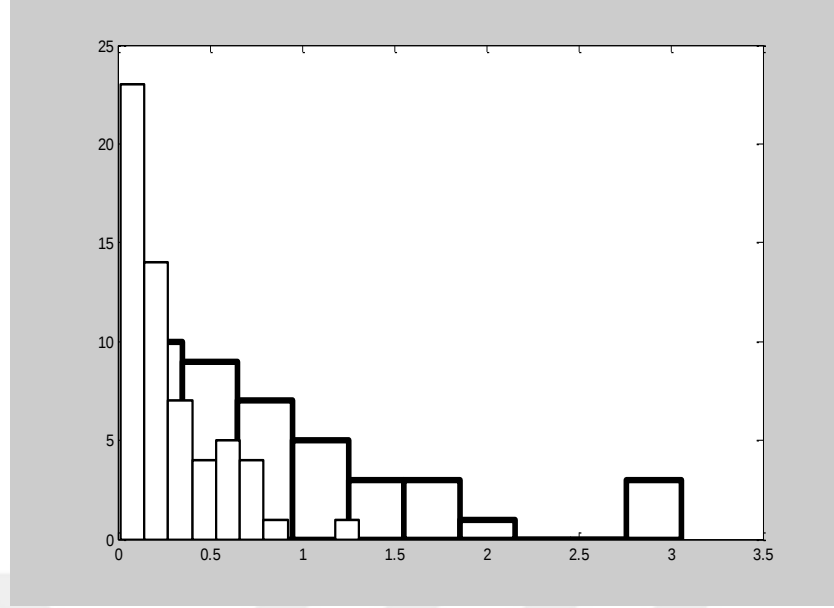
Şekil 2.5 $\theta_1 = 10$ ve $\theta_2 = 1$ iken farklı α değerleri için iki bileşenli karma üstel dağılımın dağılım fonksiyonları



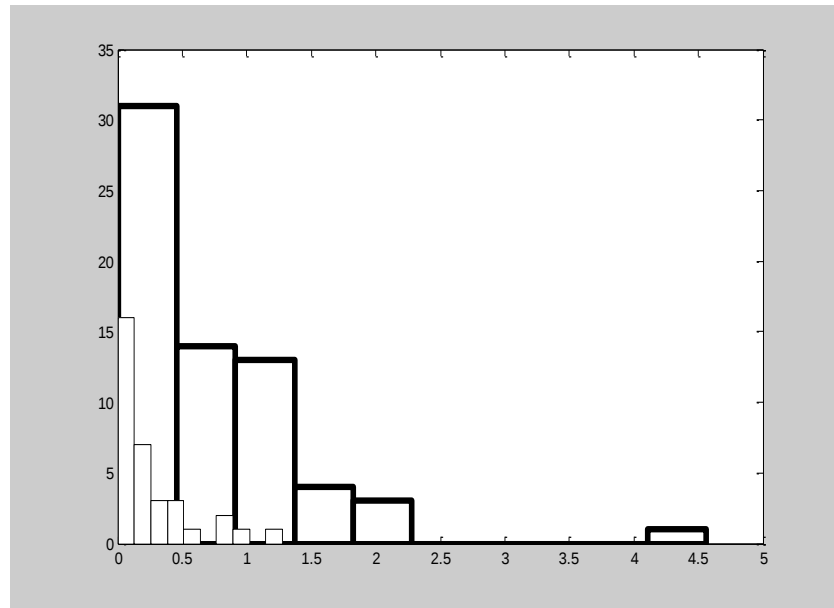
Şekil 2.6 $\theta_1 = 10$ ve $\theta_2 = 1$ iken farklı α değerleri için iki bileşenli karma üstel dağılımın yaşam fonksiyonları



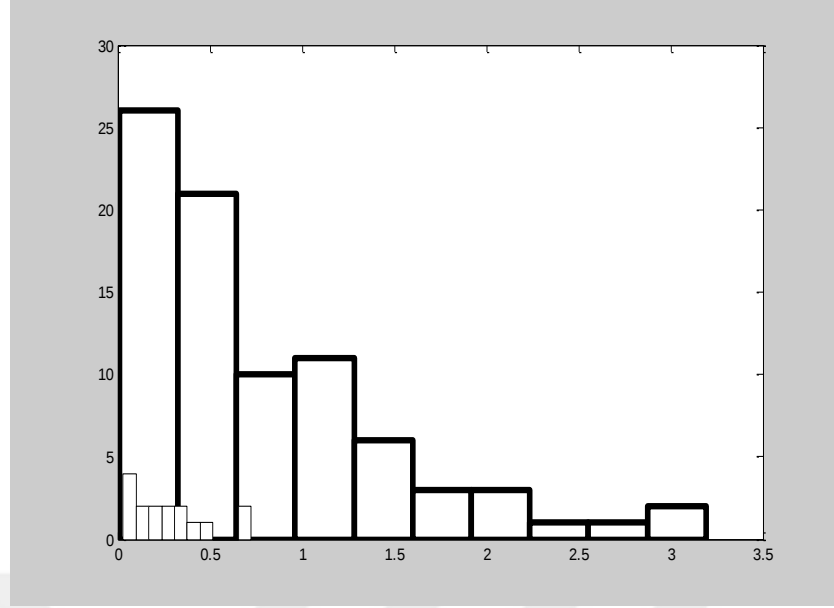
Şekil 2.7 $\theta_1 = 10$ ve $\theta_2 = 1$ iken farklı α değerleri için iki bileşenli karma üstel dağılımın bozulma oran fonksiyonları



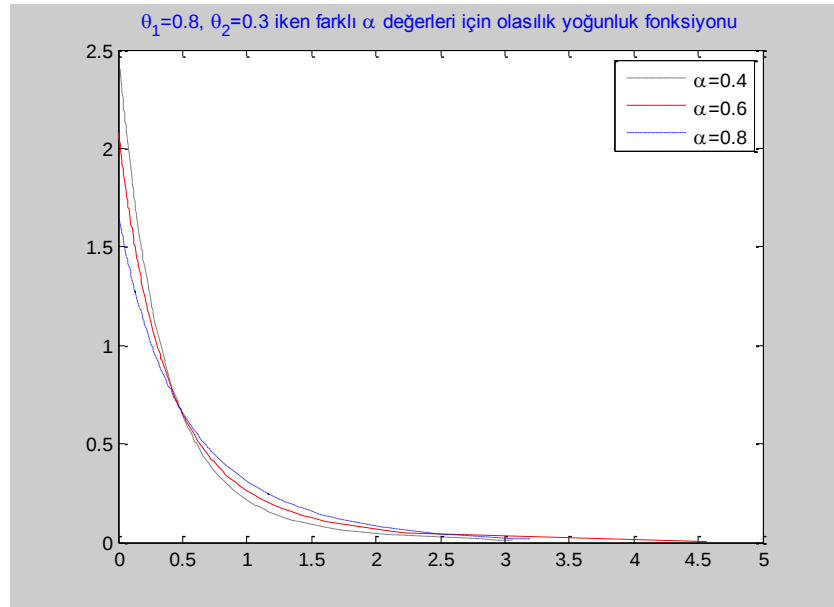
Şekil 2.8 İki bileşenli karma üstel dağılımın histogramı ($\theta_1 = 0.8$ (koyu çizgi ile) $\theta_2 = 0.3$ ve $\alpha = 0,4$)



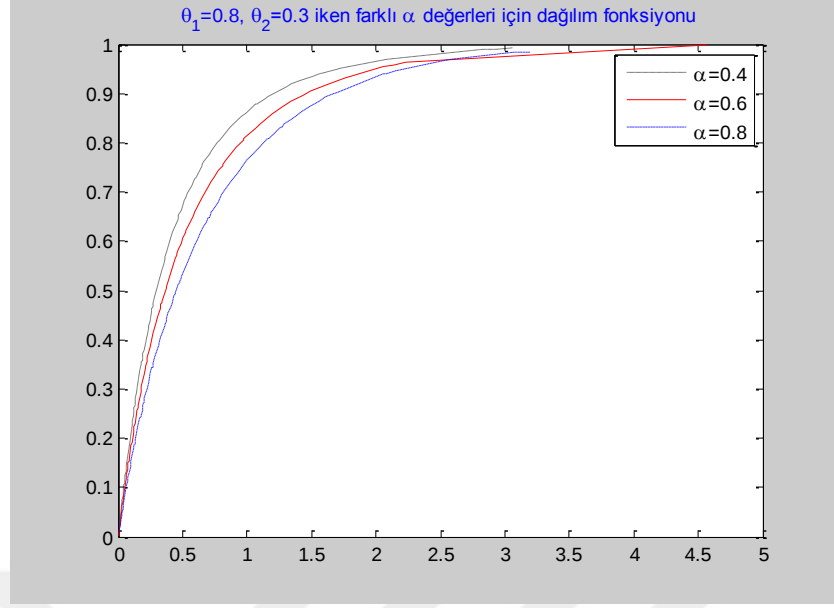
Şekil 2.9 İki bileşenli karma üstel dağılımın histogramı ($\theta_1 = 0.8$ (koyu çizgi ile) $\theta_2 = 0.3$ ve $\alpha = 0,6$)



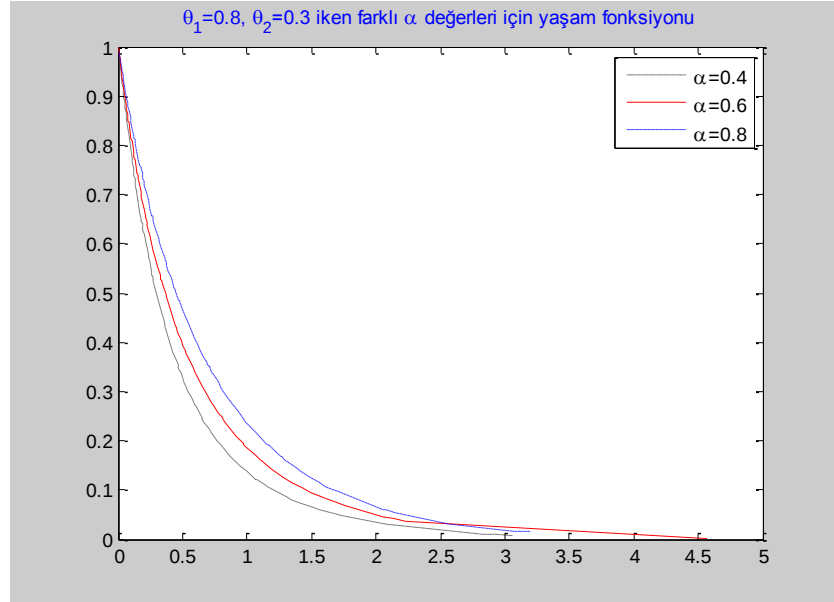
Şekil 2.10 İki bileşenli karma üstel dağılımın histogramı ($\theta_1 = 0.8$ (koyu çizgi ile) $\theta_2 = 0.3$ ve $\alpha = 0,8$)



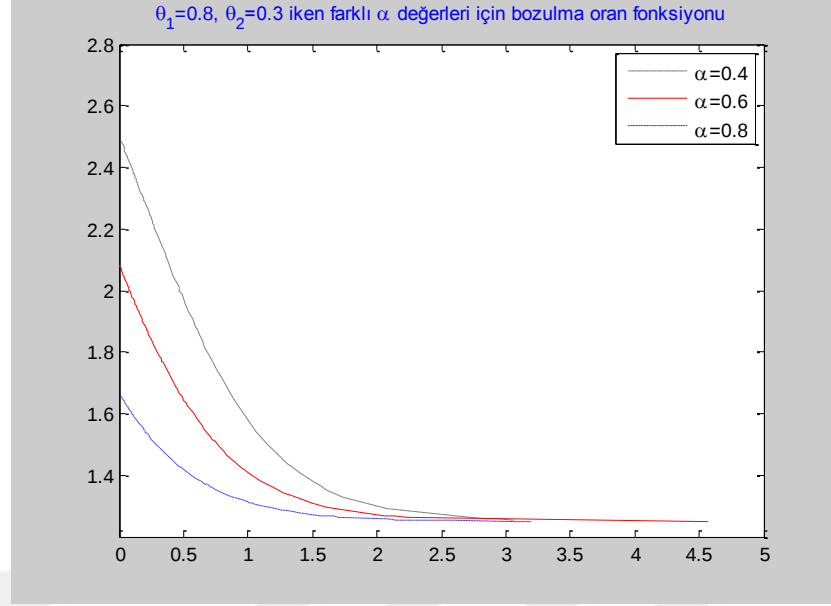
Şekil 2.11 $\theta_1 = 0.8$ ve $\theta_2 = 0.3$ iken farklı α değerleri için iki bileşenli karma üstel dağılımın olasılık yoğunluk fonksiyonları



Şekil 2.12 $\theta_1 = 0.8$ ve $\theta_2 = 0.3$ iken farklı α değerleri için iki bileşenli karma üstel dağılımın dağılım fonksiyonları



Şekil 2.13 $\theta_1 = 0.8$ ve $\theta_2 = 0.3$ iken farklı α değerleri için iki bileşenli karma üstel dağılımın yaşam fonksiyonları



Şekil 2.14 $\theta_1 = 0.8$ ve $\theta_2 = 0.3$ iken farklı α değerleri için iki bileşenli karma üstel dağılımın bozulma oran fonksiyonları

Verilen görsel örneklerde çizdirilen histogramlara bakıldığında, ölçek parametreleri arasındaki mesafe arttıkça ölçek parametresi daha küçük olan dağılımın parametrelerini tahmin etmek zorlaşmaktadır. Örneğin, $\theta_1 = 10$ ve $\theta_2 = 1$ iken farklı α değerleri için çizdirilen histogramlar incelendiğinde, α değerleri arttıkça θ_2 ölçek parametresine sahip üstel dağılım θ_1 ölçek parametresine sahip üstel dağılımın daha çok içine girmektedir. Yani, ortalamalar birbirinden uzak seçildiğinde iki farklı ortalamaya sahip karma üstel dağılım sanki tek bir dağılımı gibi algılanabilmekte ve α değerleri arttıkça da karma dağılımın ölçek parametresi daha büyük olan tek boyutlu üstel dağılımdan geldiği anlaşılabilmektedir.

Ortalamalar birbirine yaklaştıkça iki farklı ortalamaya sahip karma üstel dağılım için çizdirilen histogramlar için de yukarıda yapılan yoruma benzer bir yorum yapılabilir. Birbirine yakın ölçek parametrelerine sahip bu iki üstel dağılımı birbirinden ayırt etmek α değerleri arttıkça daha da zorlaşmaktadır. Söz konusu bu durum, gerçek hayatta üstel dağılım ile modellenmeye çalışılan bir problemde aslında iki farklı ortalamaya sahip olayların sanki tek boyutlu üstel dağılımdan geliyormuş gibi algılanmasına ve modellemede yanlış bir yol izlenmesine sebep olabilmektedir.

3. MATERYAL VE YÖNTEM

Bu bölümde genel olarak sonlu karma dağılımların parametre tahmin yöntemlerinden bahsedilecektir. Bunlar momentler yöntemi, en çok olabilirlik yöntemi ve en küçük kareler yöntemidir. Çalışmada, en çok olabilirlik yönteminde tahmin ediciler için analitik ifadeler elde edilemediğinde Expectation-Maximization (EM) Algoritması ve Newton-Raphson yönteminden faydalanılmıştır.

3.1 Momentler Yöntemi

Momentler yöntemi (MOM), karma dağılımların parametrelerinin hesaplanmasında en çok kullanılan yöntemlerden biridir. Bunun nedeni, en çok olabilirlik yöntemi oldukça uzun ve karmaşık çözümler gerektirirken, momentler yönteminde parametre tahmini yapılırken daha basit yöntemlerin kullanılmasıdır.

Bilinmeyen k tane parametre vektörü olduğu varsayılırsa momentler yönteminde ilk k tane kitle momenti örneklem momentlerine eşitlenir ve elde edilen denklem sisteminin çözümü parametrelerin tahminini verir. Bu tahminler, örneklem momentlerinin bir fonksiyonu olarak elde edilir. Örneklem momentleri kitle momentlerinin tutarlı bir tahmini olduğundan, parametre tahminleri de genellikle tutarlı olacaktır.

Herhangi bir karma dağılımın olasılık yoğunluk fonksiyonu,

$$f(x) = \alpha_1 f_1(x; \theta_1) + \alpha_2 f_2(x; \theta_2) + \dots + \alpha_r f_r(x; \theta_r) \quad (3.1)$$

şeklinde tanımlansın.

Burada,

$\underline{\theta}=(\theta_1 \quad \theta_2 \quad \dots \quad \theta_k)^T$ ve $\underline{\theta}_i = \begin{bmatrix} \underline{\theta}_1 \\ \underline{\theta}_2 \\ \vdots \\ \underline{\theta}_r \end{bmatrix}$ olmak üzere;

$$\Phi = \left\{ \underline{\alpha}, \Theta: \alpha_i, i = 1, 2, \dots, r \quad \underline{\theta}_i = \begin{bmatrix} \underline{\theta}_1 \\ \underline{\theta}_2 \\ \vdots \\ \underline{\theta}_r \end{bmatrix}, \Theta = \{\underline{\theta}_1, \underline{\theta}_2, \dots, \underline{\theta}_k\} \right\} \quad (3.2)$$

dır.

$\sum_{i=1}^r \alpha_i=1$ olmak üzere $\underline{\theta}_1, \underline{\theta}_2, \dots, \underline{\theta}_r$ ve $\alpha_1, \alpha_2, \dots, \alpha_{r-1}$ parametrelerini tahmin etmek için öncelikle kitle momentleri örneklem momentlerine eşitlenir. k parametre sayısı, r bileşen (karma) sayısı olmak üzere $r \times k + r - 1$ tane moment denklemi ve ham moment yazılır ($\sum_{i=1}^r \alpha_i=1$ koşulu yoksa $r \times k + r$ tane denklem yazılır).

$X_1, X_2, \dots, X_n$ örnekleme ele alındığında $\alpha_1, \alpha_2, \dots, \alpha_r$ ve $\underline{\theta}_1, \underline{\theta}_2, \dots, \underline{\theta}_r$ parametrelerini tahmin etmek için oluşturulan denklem sistemi $E[X^j] < \infty$ olmak üzere;

$$E[X^j] = \frac{1}{n} \sum_{i=1}^n X_i^j \quad j = 1, 2, \dots, r \times k + r - 1 \quad (3.3)$$

biçimindedir. Burada $E[X^j]$, j. kitle momentini ve $\frac{1}{n} \sum_{i=1}^n X_i^j$, j. örneklem momentini göstermektedir.

$$\widehat{\alpha}_r = 1 - \sum_{i=1}^{r-1} \widehat{\alpha}_i \quad (3.4)$$

Tahmin ediciler genellikle analitik yolla elde edilemediğinden denklem sisteminde örneklem momentinin değeri yazılarak nümerik yolla parametrelerin çözümü elde edilir.

3.2 En Çok Olabilirlik Yöntemi

En çok olabilirlik yöntemi (ML), olabilirlik fonksiyonunu maksimize eden parametre değerini bularak dağılımın parametrelerini tahmin eden bir yöntemdir.

$\underline{x} = \{x_1, x_2, \dots, x_n\}$ olasılık yoğunluk fonksiyonu $f(\underline{x}|\underline{\theta})$ olan birbirinden bağımsız ve aynı dağılımlı gözlemler olsun. $\underline{\theta} = (\theta_1, \theta_2, \dots, \theta_d)$ bu dağılıma ilişkin parametre kümesi olmak üzere $\underline{\theta}$ parametrelerinin olabilirlik fonksiyonu,

$$L(\underline{\theta}|\underline{x}) = \prod_{i=1}^n f(x_i|\underline{\theta}) \quad (3.5)$$

dır. Buradan,

$$\hat{\underline{\theta}} = \operatorname{argmax} L(\underline{\theta}|\underline{x}) \quad (3.6)$$

d boyutlu $\underline{\theta} = (\theta_1, \theta_2, \dots, \theta_d)$ parametresi için en çok olabilirlik tahmin edicisi genellikle olabilirlik fonksiyonunun logaritmasının bilinmeyen parametrelere göre ayrı ayrı kısmi türevlerinin sıfıra eşitlenmesi ile oluşturulan olabilirlik denklem sistemlerinin çözülmesi ile elde edilir. Bu denklem sistemleri genelde lineer olmayan ifadeleri içerdiği için analitik çözüme sahip değildir. Bu nedenle olabilirlik denklemlerinin çözümü için adımsal yöntemler kullanılır.

3.2.1 EM Algoritması

EM algoritması, olabilirlik fonksiyonunun türevinin alınmasının ve doğrudan maksimize edilmesinin zor olduğu ve olabilirlik fonksiyonunun kapalı çözüm formlarının mevcut olmadığı durumlarda kullanılan adımsal bir yöntemdir. EM algoritmasının her yinelemesinde, E (expectation) adımı ve M (maximization) adımı olarak adlandırılan iki adım bulunmaktadır. Bu yüzden bu algoritma EM algoritması olarak adlandırılır.

EM algoritması Dempster vd. tarafından 1977 yılında ortaya çıkarılmış ve en çok olabilirlik tahmin edicisinin yinelemeli yolla hesaplanmasında kullanılmıştır. Bu algoritma verilen veri setinin eksik veya veri setinde kayıp gözlemlerin bulunduğu durumlarda istatistiksel çıkarım yapabilmek için kullanılmaktadır. EM algoritması, tam veri verildiğinde en çok olabilirlik tahmin edicisinin karmaşıklığını daha basit düzeye indirgediğinden daha kullanışlıdır. Temel düşünce, verilen tam olmayan veri problemini, en çok olabilirlik tahmin edicisinin hesaplanmasının daha kolay olduğu tam veri problemi ile ilişkilendirmektir. Örneğin, seçilen tam veri problemi en çok olabilirlik tahmin edicisine kapalı bir çözüm formu sağlayabilir veya standart bilgisayar programları ile en çok olabilirlik tahmin edicisinin hesabını kolaylaştırabilir.

Parametrenin bazı başlangıç değerleri için EM algoritması çok yavaş yakınsamasına rağmen bu durumun üstesinden rasgele başlangıç değerleri ile başlanmasıyla gelinebilir ve en küçük adım sayısını sağlayan başlangıç değerlerinden biri en iyisi olarak alınır. Uygulamada en yaygını, adımlara karşılık gelen logaritması alınan olabilirlik (log-likelihood) fonksiyonunun grafiğindeki artışların gözlenerek yakınsama durumuna bakılmasıdır.

Dempster vd. (1977), EM algoritmasının her adımının yakınsayana kadar tam olmayan veriye ilişkin log-likelihood fonksiyonunun kesin olarak artacağını göstermişlerdir. Yakınsama oranı lineerdir ve bu oran gözlenen verinin bilgi oranına bağlıdır. Newton-Raphson yöntemine kıyasla EM algoritması her adımla birlikte log-likelihood değerlerini arttırmada karardır ve neredeyse herhangi bir başlangıç noktasında yerel maksimum noktasına yakınsar (Dennis ve Schnabel 1983). Ayrıca EM algoritması, her adımda olabilirlik eşitliklerine ilişkin Jacobian matrisinin hesaplanmasını gerektirmediğinden daha az hesaplama maliyetine sahiptir.

Tanım 3.1 Sonlu karma yoğunlukları,

$$f(\underline{x}|\theta) = \sum_{j=1}^k \alpha_j f_j(\underline{x}|\theta_j) \quad (3.7)$$

olarak tanımlansın.

Her bir $f_j(\underline{x}|\theta_j)$ yoğunluk fonksiyonu için $\sum_{j=1}^r \alpha_j=1$ ve $\alpha_j \geq 0$ dır. Burada α_j , j. bileşen yoğunluğuna $(f_j(\underline{x}|\theta_j))$ ait verinin olasılığı olarak düşünülebilir. Parametre vektörü,

$$\Theta = (\alpha_1, \alpha_2, \dots, \alpha_k; \theta_1, \theta_2, \dots, \theta_k) \quad (3.8)$$

dır. $X, \underline{x} = \{x_1, x_2, \dots, x_n\}$ Θ uzayında θ parametrelerine sahip dağılımdan üretilen gözlenen veriye ilişkin bir rasgele değişken olsun. $y = (y_1, y_2, \dots, y_n)$ gözlenemeyen veya kayıp veriyi göstermek üzere $z = (x, y)$ ise tam olan veri setini gösterebilir. Z rasgele değişkeninin ortak olasılık yoğunluk fonksiyonu,

$$f(z|\theta) = f(x, y|\theta) = f(y|x, \theta)f(x|\theta) \quad (3.9)$$

olup, tam veriye ait olabilirlik fonksiyonu,

$$L(\theta|z) = \prod_{i=1}^n f(z_i|\theta) = \prod_{i=1}^n f(x_i, y_i|\theta) \quad (3.10)$$

dır. Kayıp bilgi olan y bilinmediğinden, (3.10) eşitliği x ve θ verildiğinde y 'nin bir fonksiyonu olarak düşünülebilir. $L(\theta|x)$ ise tam olmayan verinin olabilirlik fonksiyonudur.

E Adımı: EM algoritmasının ilk adımı olan E adımı, gözlenen veri X ve parametrelerin mevcut değeri $\theta^{(m)}$ verildiğinde gözlenemeyen veriler Y 'lere ilişkin $\log L(\theta|x, y)$ fonksiyonunun beklenen değerinin bulunmasına yöneliktir.

$$Q(\theta, \theta^{(m)}) = E_y[\log L(\theta|x, y)] = \int \log(L(\theta|x, y)) f(y|x, \theta^{(m)}) dy \quad (3.11)$$

$\theta^{(m)}$, m. adımda beklenen değeri hesaplamak için kullanılan θ 'nın mevcut değerini göstermektedir.

E adımıdaki ana düşünce, X ve $\theta^{(m)}$ 'in verilen sabitler olduğu ve Y 'nin $f(y|x, \theta^{(m)})$ marjinal dağılıma sahip bir rasgele değişken olduğudur.

M Adımı: EM algoritmasının ikinci adımı olan M adımı, (3.11) eşitliğindeki ifadenin θ 'ya göre maksimize edilmesidir.

$$\theta^{(m+1)} = \operatorname{argmax}_{\theta} Q(\theta, \theta^{(m)}) \quad (3.12)$$

$\theta^{(m+1)}$, Θ uzayından seçilirse,

$$Q(\theta^{(m+1)}, \theta^{(m)}) \geq Q(\theta, \theta^{(m)}), \text{ her } \theta \in \Theta \text{ için} \quad (3.13)$$

$\theta^{(m)}$ başlangıç değeri ve X verildiğinde, EM algoritması E adımı ile başlar ve M adımı ile devam eder. Bu iki adım, tam olmayan veriye ilişkin log-likelihood fonksiyonun bir sonraki adım ile bir önceki adımdaki değerleri arasındaki fark ε gibi bir sayıdan daha küçük oluncaya kadar devam ettirilir.

$$L(\theta^{(m+1)}|x) - L(\theta^{(m)}|x) \leq \varepsilon \quad (3.14)$$

Aslında EM algoritması, verilen tam olmayan verinin olabilirlik fonksiyonu olan $L(\theta|x)$ 'i tam verinin olabilirlik fonksiyonu olan $L(\theta|z)$ açısından yeniden düzenler, bu iki fonksiyon arasındaki ilişkiden faydalanır ve tam veri problemi için EM algoritmasındaki daha basit hesaplamalara dayalı olarak bir en çok olabilirlik tahmin edicisi elde eder. Bazı durumlarda örneğin, Normal ve Poisson dağılımlarının karmalarında M adımı her zaman analitik çözüme sahiptir. Bu yüzden EM algoritmasının uygulaması kolaydır ve en çok olabilirlik tahmin edicisinin kapalı formu elde edilebilir. Karma üstel dağılımlarda ise M adımı genellikle analitik çözüme sahip olmadığından, ML tahminlerinin hesaplanmasında nümerik yollar tercih edilir (Ren 2011).

EM tahmini, hiçbir zaman daha kötüsünün olmayacağını garanti eder. Genellikle yöntem, $P(x|\theta)$ 'nin bir tepe noktası olduğunu söyler, fakat $P(x|\theta)$ birden fazla tepeye sahipse

EM, olabilirliğin mutlak maksimum noktasını bulamayabilir. Uygulamada, EM algoritmasını çoklu rasgele başlangıç noktaları ile başlatmak yaygındır ve en büyük olabilirliğe sahip olan θ , en son tahmin değeri olarak seçilir (Gupta ve Chen 2010).

3.2.1.1 EM algoritmasının karma dağılımlarda genel işleyişi

$x = (x_1, x_2, \dots, x_n)$ gözlem verilerinden elde edilen $f(x|\theta)$ için log-likelihood fonksiyonu;

$$\log(L(\theta|X)) = \log(\prod_{i=1}^n f(x_i|\theta)) = \sum_{i=1}^n \log\left(\sum_{j=1}^k \alpha_j f_j(x_i|\theta_j)\right) \quad (3.15)$$

Görüldüğü üzere $\log(L(\theta|X))$, karma yoğunluklarının toplamının logaritma fonksiyonu şeklindedir. Burada genel en çok olabilirlik işlemlerini kullanarak maksimizasyon yapılması oldukça zordur. EM algoritmasını uygulayabilmek için öncelikle kayıp bilginin $y = (y_1, y_2, \dots, y_n)$, $y_i \in (1, 2, \dots, k)$, $i = 1, 2, \dots, n$ olarak gösterilmesi gerekir. Eğer i . veri yani x_i , j . bileşenin yoğunluk fonksiyonuna $f_j(x|\theta_j)$, $j = 1, 2, \dots, k$ ait ise $y_i = j$ olarak gösterilir.

Tam veri için log-likelihood fonksiyonu;

$$\log(L(\theta|X, Y)) = \log(f(X, Y|\theta)) = \sum_{i=1}^n \log(f(y_i|x_i, \theta_{y_i})) \quad (3.16)$$

Başlangıç değeri $\theta^{(0)} = (\alpha_1^{(0)}, \dots, \alpha_k^{(0)}; \theta_1^{(0)}, \dots, \theta_k^{(0)})$ ve X verildiğinde marjinal olasılık yoğunluk fonksiyonu hesaplanabilir.

$$f(Y|X, \theta^{(0)}) = \prod_{i=1}^n f(y_i|x_i, \theta^{(0)}) \quad (3.17)$$

Burada,

$$f(y_i|x_i, \Theta^{(0)}) = \frac{f(y_i, x_i, \Theta^{(0)})}{f(x_i, \Theta^{(0)})} = \frac{f(x_i|y_i, \Theta^{(0)})f(y_i|\Theta^{(0)})}{f(x_i|\Theta^{(0)})} \quad (3.18)$$

dır. (3.7) eşitliğinden,

$$f(x_i|\Theta^{(0)}) = \sum_{j=1}^k \alpha_j f_j(x_i|\theta_i^{(0)}) \quad (3.19)$$

ve

$$f(y_i|\Theta^{(0)}) = \alpha_{y_i}^{(0)} \quad (3.20)$$

elde edilir. (3.19) ve (3.20) eşitlikleri (3.18) eşitliğinde kullanıldığında (3.17) eşitliği,

$$f(Y|X, \Theta^{(0)}) = \prod_{i=1}^n f(y_i|x_i, \Theta^{(0)}) = \prod_{i=1}^n \frac{f(x_i|y_i, \Theta^{(0)})\alpha_{y_i}^{(0)}}{\sum_{j=1}^k \alpha_j f_j(x_i|\theta_i^{(0)})} \quad (3.21)$$

şeklinde olur.

Buradan E adımı,

$$\begin{aligned} Q(\theta, \theta^{(0)}) &= E_{(Y|X, \Theta^{(0)})} \log(L(\theta|X, Y)) \\ &= \sum_{y \in Y} \log(L(\theta|X, y)) f(y|X, \Theta^{(0)}) \\ &= \sum_{y_1=1}^k \dots \sum_{y_n=1}^k \sum_{i=1}^n \log(\alpha_{y_i} f_{y_i}(x_i|\theta_{y_i})) f(y|X, \Theta^{(0)}) \\ &= \sum_{j=1}^k \sum_{i=1}^n \log(\alpha_j f_j(x_i|\theta_j)) \sum_{y_1=1}^k \dots \sum_{y_n=1}^k \delta_{(j, y_i)} f(y|X, \Theta^{(0)}) \end{aligned} \quad (3.22)$$

(3.22) eşitliğinde,

$$\delta_{(j,y_i)} = \begin{cases} 1, & j = y_i \\ 0, & d. d \end{cases} \quad (3.23)$$

ve

$$f(y|X, \Theta^{(0)}) = \prod_{z=1}^n f(y_z|x_z, \Theta^{(0)}) \quad (3.24)$$

olduğundan,

$$\begin{aligned} &= \sum_{y_1=1}^k \dots \sum_{y_n=1}^k \delta_{(j,y_z)} f(y|X, \Theta^{(0)}) \\ &= \sum_{y_1=1}^k \dots \sum_{y_n=1}^k \delta_{(j,y_z)} \prod_{z=1}^n f(y_z|x_z, \Theta^{(0)}) \\ &= \left[\sum_{y_1=1}^k \dots \sum_{y_{i-1}=1}^k \sum_{y_{i+1}=1}^k \dots \sum_{y_n=1}^k \prod_{z=1, z \neq i}^n f(y_z|x_z, \Theta^{(0)}) \right] f(j|x_i, \Theta^{(0)}) \\ &= \prod_{z=1, z \neq i}^n \underbrace{\left[\prod_{y_z=1}^k f(y_z|x_z, \Theta^{(0)}) \right]}_1 f(j|x_i, \Theta^{(0)}) \\ &= f(j|x_i, \Theta^{(0)}) \end{aligned} \quad (3.25)$$

dır.

(3.22) eşitliği,

$$\begin{aligned}
Q(\theta, \theta^{(0)}) &= \sum_{j=1}^k \sum_{i=1}^n \log(\alpha_j f_j(x_i | \theta_j)) f(j | x_i, \theta^{(0)}) \\
&= \sum_{j=1}^k \sum_{i=1}^n \log(\alpha_j) f(j | x_i, \theta^{(0)}) + \sum_{j=1}^k \sum_{i=1}^n \log(f_j(x_i | \theta_j)) f(j | x_i, \theta^{(0)}) \\
&= G(\alpha_j) + R(\theta_j)
\end{aligned} \tag{3.26}$$

olarak düzenlenebilir.

$G(\alpha_j)$ ve $R(\theta_j)$ bağımsızdır ve M adımında α_j ve θ_j 'e göre ayrı ayrı $Q(\theta, \theta^{(m)})$ maksimize edilir. $G(\alpha_j)$ 'nin maksimizasyonu $\sum_{j=1}^k \alpha_j = 1$ kısıtına bağlı olduğundan Langrange çarpanları yöntemi kullanılır.

$$\begin{aligned}
\Lambda(\alpha_j, \lambda) &= G(\alpha_j) + \lambda \left(\sum_{j=1}^k \alpha_j - 1 \right) \\
&= \sum_{j=1}^k \sum_{i=1}^n \log(\alpha_j) f(j | x_i, \theta^{(0)}) + \lambda (\sum_{j=1}^k \alpha_j - 1)
\end{aligned} \tag{3.27}$$

$$\frac{\partial \Lambda}{\partial \alpha_j} = \sum_{j=1}^k \sum_{i=1}^n \frac{f(j | x_i, \theta^{(0)})}{\alpha_j} + \sum_{j=1}^k \lambda = 0 \tag{3.28}$$

$$\frac{\partial \Lambda}{\partial \lambda} = \sum_{j=1}^k \alpha_j - 1 = 0 \tag{3.29}$$

(3.28) eşitliğinin her iki tarafı α_j ile çarpılırsa,

$$\sum_{j=1}^k \sum_{i=1}^n f(j|x_i, \Theta^{(0)}) + \sum_{j=1}^k \lambda \alpha_j = 0 \quad (3.30)$$

$$\sum_{j=1}^k (\sum_{i=1}^n f(j|x_i, \Theta^{(0)}) + \lambda \alpha_j) = 0 \quad (3.31)$$

$$\sum_{i=1}^n \sum_{j=1}^k f(j|x_i, \Theta^{(0)}) + \lambda = 0 \quad (3.32)$$

elde edilir.

$\sum_{j=1}^k f(j|x_i, \Theta^{(0)}) = 1$ olduğundan (3.32) eşitliği,

$$\sum_{i=1}^n 1 + \lambda = 0$$

$$\lambda = -n \quad (3.33)$$

şeklinde bulunur.

Diğer bir yandan (3.31) eşitliği,

$$\sum_{i=1}^n f(j|x_i, \Theta^{(0)}) + \lambda \alpha_j = 0$$

$$\alpha_j = -\frac{\sum_{i=1}^n f(j|x_i, \Theta^{(0)})}{\lambda} \quad (3.34)$$

dır.

$\lambda = -\frac{1}{n}$ olduğundan (3.34) eşitliğindeki karma yoğunluğu ağırlığı,

$$\alpha_j = \frac{1}{n} \sum_{i=1}^n f(j|x_i, \Theta^{(0)}) \quad (3.35)$$

$f(j|x_i, \Theta^{(0)})$, x_i ve parametrenin mevcut değeri $\Theta^{(0)}$ verildiğinde $y_i = j$. bileşenin marjinal yoğunluk fonksiyonunu verir.

$$f(j|x_i, \Theta^{(0)}) = \frac{\alpha_j^{(0)} f_j(x_i|\theta_j^{(0)})}{\sum_{l=1}^k w_l^{(0)} f_l(x_i|\theta_l^{(0)})}, j = 1, 2, \dots, k \quad (3.36)$$

Öte yandan $\theta = (\theta_1, \dots, \theta_k)$ parametrelerini tahmin etmek için $R(\theta_j)$ maksimize fonksiyonu gereklidir.

$$R(\theta_j) = \sum_{j=1}^k \sum_{i=1}^n \log(f_j(x_i|\theta_j)) f(j|x_i, \Theta^{(0)}) \quad j = 1, 2, \dots, k \quad (3.37)$$

$$\frac{\partial R(\theta_j)}{\partial \theta_j} = \sum_{j=1}^k \sum_{i=1}^n \frac{\partial \log(f_j(x_i|\theta_j))}{\partial \theta_j} f(j|x_i, \Theta^{(0)}) = 0 \quad (3.38)$$

Genellikle (3.38) eşitliğinin hesabı bileşen yoğunluklarının hesabı olan $f_j(x_i|\theta_j)$ 'ye bağlıdır ve $\theta_1, \dots, \theta_k$ parametrelerinin davranışına göre tahmin edilmelidir. En iyi durumda, her bir bileşenin yoğunluğu iyi bilinen bir dağılımdan gelir ve olasılık yoğunluk fonksiyonunun basit analitik bir formu olarak karşımıza çıkar. Bu formda θ_j 'nin kapalı hali doğrudan elde edilir. Fakat bazı durumlarda bileşen yoğunlukları çok karmaşık olduğundan (3.38) eşitliğini çözmek için genellikle nümerik yöntemlerin kullanılması gerekir.

3.2.1.2 EM algoritmasının işleyişine örnekler

A. Oyuncak Örneği (Gupta ve Chen 2010)

n sayıdaki çocuk içinde n adet oyuncak bulunan 4 farklı kutudan birini seçiyor. $Y = [Y_1 \ Y_2 \ Y_3 \ Y_4]^T$ olmak üzere $Y_i, i = 1, 2, 3, 4$ i . kutuyu seçen çocukların sayısını gösterebilir. Y 'nin rasgele histogramı “multinomial dağılım” olarak dağılır. Multinomial çocuk sayısı ($n \in N$) ve çocukların seçeceği 4 farklı kutuyu seçme olasılığı olmak üzere iki parametreye

sahiptir. Burada $\underline{p} = (p_1, p_2, p_3, p_4) \in [0,1]^4$ ve $p_1 + p_2 + p_3 + p_4 = 1$ dir. p_i ($i = 1,2,3,4$), i . kutunun herhangi bir çocuk tarafından seçilmesi olasılığıdır.

$$f_{Y|p}(\underline{y}) = \frac{n!}{y_1! y_2! y_3! y_4!} p_1^{y_1} p_2^{y_2} p_3^{y_3} p_4^{y_4}$$

Her bir çocuğun oyuncacı seçme olasılıkları aynıdır ve hiçbir çocuk kararsız değildir. Yani her çocuğun oyuncaklardan birini seçtiği biliniyorken, oyuncacın hangi olasılıkla seçildiği bilinmemektedir. Bu bilinmeyen olasılık θ parametresi ile gösterilmektedir. Burada θ parametresi her bir çocuk için aynıdır ve $\theta \in [0,1]$ dir. Bu gizli parametreye göre oyuncakların her bir çocuk tarafından seçilmesi olasılıkları,

$$[p_1 \ p_2 \ p_3 \ p_4]^T = \left[\frac{1}{2} + \frac{1}{4}\theta \quad \frac{1}{4}(1-\theta) \quad \frac{1}{4}(1-\theta) \quad \frac{1}{4}\theta \right]^T, \theta \in [0,1]$$

dır.

Buna göre θ verildiğinde $Y = (Y_1 \ Y_2 \ Y_3 \ Y_4)$ rasgele vektörünün olasılık fonksiyonu,

$$P(\underline{y}|\theta) = \frac{n!}{y_1! y_2! y_3! y_4!} \left(\frac{1}{2} + \frac{1}{4}\theta \right)^{y_1} \left(\frac{1-\theta}{4} \right)^{y_2} \left(\frac{1-\theta}{4} \right)^{y_3} \left(\frac{\theta}{4} \right)^{y_4}$$

şeklindedir.

Bu örnek için $\log(f_{Y|p}(\underline{y}))$ fonksiyonu doğrudan maksimize edilebilir, fakat burada amaç EM algoritmasını kullanarak θ 'nın en çok olabilirlik tahmin edicisini bulmaktır. EM algoritmasını kullanabilmek için tam veri olan X 'in ne olduğunu belirlemek gerekmektedir. Burada X şu şekilde belirlenmiştir. İçinde yine n tane oyuncak bulunan 1 numaralı kutunun çocukların seçimine sunulduğunu düşünelim ve bu kutunun seçilmesi olasılığının da baştan $\frac{1}{2}$ olduğu biliniyor olsun. Buna göre 1,1,2,3,4 kutularının seçilmesi olasılığını sırasıyla;

$$\left[\frac{1}{2} \quad \frac{1}{4}\theta \quad \frac{1}{4}(1-\theta) \quad \frac{1}{4}(1-\theta) \quad \frac{1}{4}\theta \right]^T, \theta \in [0,1]$$

vektörüyle ve i . kutudan seçilen oyuncak sayısını da X_i ile gösterelim. Buna göre $\underline{Y}$ rasgele vektörü aşağıdaki gibi ifade edilebilir.

$$\underline{Y} = T(\underline{X}) = \begin{bmatrix} \underbrace{X_1 + X_2}_{1 \text{ nolu kutudan seçilen oyuncak sayısı}} & X_3 & X_4 & X_5 \end{bmatrix}$$

$$P(\underline{x}|\theta) = \frac{n!}{\prod_{i=1}^5 x_i!} \left(\frac{1}{2}\right)^{x_1} \left(\frac{\theta}{4}\right)^{x_2+x_5} \left(\frac{1-\theta}{4}\right)^{x_3+x_4}$$

EM algoritması için Q fonksiyonunun maksimize edilmesi gerekmektedir.

$$Q^{(m+1)} = \operatorname{argmax}_{\theta \in [0,1]} Q(\theta|\theta^{(m)}) = \operatorname{argmax}_{\theta \in [0,1]} E_{\underline{x}|\underline{y},\theta^{(m)}} [\log(P(\underline{x}|\theta))]$$

Yukarıdaki eşitliği çözmek için sadece θ 'ya bağlı olan $\log(P(\underline{x}|\theta))$ ifadesine ihtiyaç vardır. Çünkü diğer terimler ilgilenilen θ üzerinden maksimize edilmesiyle ilişkisizdir. Böylece $P(\underline{x}|\theta)$ eşitliğinin logaritması alınır ve θ 'ya bağlı olmayan diğer terimler göz ardı edilirse,

$$\theta^{(m+1)} = \operatorname{argmax}_{\theta \in [0,1]} E_{\underline{x}|\underline{y},\theta^{(m)}} [(X_2 + X_5)\log\theta + (X_3 + X_4)\log(1 - \theta)]$$

$$= \operatorname{argmax}_{\theta \in [0,1]} E_{\underline{x}|\underline{y},\theta^{(m)}} (X_2) + E_{\underline{x}|\underline{y},\theta^{(m)}} (X_5)] \log\theta + [E_{\underline{x}|\underline{y},\theta^{(m)}} (X_3) + E_{\underline{x}|\underline{y},\theta^{(m)}} (X_4)] \log(1 - \theta)$$

Yukarıdaki maksimizasyon problemini çözmek için tam veri olan X 'lerin, tam veri olmayan Y 'ler üzerinden koşullandırılmasının beklenen değerine ihtiyaç vardır.

E adımı,

$$P(X_1 = x_1, X_2 = x_2, \dots, X_5 = x_5 | X_1 + X_2 = y_1, X_3 = y_2, X_4 = y_3, X_5 = y_4)$$

$$= \frac{P(X_1 = x_1, X_2 = y_1 - x_1, X_3 = y_2, X_4 = y_3, X_5 = y_4)}{P(X_1 + X_2 = y_1, X_3 = y_2, X_4 = y_3, X_5 = y_4)}$$

$$= \frac{\frac{n!}{x_1!(y_1-x_1)!y_2!y_3!y_4!} \left(\frac{1}{2}\right)^{x_1} \left(\frac{\theta}{4}\right)^{y_1-x_1} \left(\frac{1-\theta}{4}\right)^{y_2+y_3} \left(\frac{\theta}{4}\right)^{y_4}}{\frac{n!}{y_1!y_2!y_3!y_4!} \left(\frac{1+\theta}{2}\right)^{y_1} \left(\frac{1-\theta}{4}\right)^{y_2+y_3} \left(\frac{\theta}{4}\right)^{y_4}}$$

$$= \frac{y_1!}{x_1!(y_1-x_1)!} \left(\frac{\frac{1}{2}}{\frac{1+\theta}{2}}\right)^{x_1} \left(\frac{\frac{\theta}{4}}{\frac{1+\theta}{2}}\right)^{y_1-x_1}$$

$$(X_1 | \underline{Y}, \theta) \sim \text{Binom}\left(y_1, \frac{1}{1+\frac{\theta}{2}}\right)$$

$$(X_2 | \underline{Y}, \theta) \sim \text{Binom}\left(y_1, 1 - \frac{1}{1+\frac{\theta}{2}}\right)$$

$$E(X | \underline{Y}, \theta) = \left[y_1 \frac{2}{2+\theta} \quad y_1 \frac{\theta}{2+\theta} \quad \underbrace{y_2 \quad y_3 \quad y_4}_{\text{gözlenmiş}} \right]^T$$

şeklindedir.

M adımı,

$$Q(\theta | \theta^{(m)}) = \left(y_1 \frac{\theta^{(m)}}{2 + \theta^{(m)}} + y_4 \right) \log \theta^{(m)} + (y_2 + y_3) \log(1 - \theta^{(m)})$$

$$\frac{dQ}{d\theta} = (x_2 + x_5) \frac{1}{\theta} - \frac{1}{1-\theta} (x_3 + x_4) = 0$$

$$-\theta(x_2 + x_3 + x_4 + x_5) + x_2 + x_5 = 0$$

$$\theta^{(m+1)} = \frac{x_2 + x_5}{n - y_1 \frac{2}{2 + \theta}} = \frac{y_1 \frac{\theta^{(m)}}{2 + \theta^{(m)}} + y_4}{n - y_1 \frac{2}{2 + \theta^{(m)}}}$$

olarak elde edilir. Bu ifade, Excel yardımıyla başlangıç değeri $\theta^{(0)} = 0,5$ ve $\varepsilon = 10^{-6}$ alınarak çözdürülmüş ve çizelge 3.1’de adım adım elde edilen çözümler verilmiştir. Burada Y_i : i . kutuyu seçen çocukların sayısını göstermektedir. ($i = 1,2,3,4$)

Çizelge 3.1 Oyuncak örneği için EM algoritması ile elde edilen parametre değerleri

$y_1 = 5, y_2 = 2, y_3 = 4, y_4 = 1$			
i	$\theta^{(i)}$	$\text{Log}L^{(i)}$	$\text{Log}L^{(i)} - \text{Log}L^{(i-1)}$
0	0,5	-2,408239965	
1	0,25	-1,686170184	0,722069781
2	0,1985	-1,595840249	0,090329934
3	0,18748	-1,579601664	0,016238586
4	0,185092	-1,576246719	0,003354945
5	0,184573	-1,575525913	0,000720806
6	0,18446	-1,575369623	0,00015629
7	0,184436	-1,575335666	0,000034
8	0,18443	-1,575328285	0,00000738
9	0,184429	-1,575326681	0,0000016

Çizelge 3.1’de 10. adım ile 9. adımdaki olabilirlik fonksiyonu değerleri arasındaki farkın $\varepsilon = 10^{-6}$ değerinden küçük olduğu görülmektedir. Dolayısıyla ele alınan $\theta^{(0)} = 0,5$ başlangıç değeri için EM algoritması yöntemi ile elde edilen çözüm $\hat{\theta} = 0,184429 \cong 0,18$ olarak 10. iterasyonda bulunmuştur.

B. Bir Görüntüdeki Nesneleri Algılama Problemi (Moon 1996)

Bir görüntüde üç farklı nesne bulunmaktadır. Bunlar koyu yuvarlak nesneler, koyu kare nesneler ve parlak nesnelerdir. Burada,

x_1 : Koyu yuvarlak nesnelerin sayısını,

x_2 : Koyu kare nesnelerin sayısını,

x_3 : Parlak nesnelerin sayısını,

n : Toplam nesne sayısını ($x_1 + x_2 + x_3$) göstermektedir.

Görüntü için kullanılan sensörün koyu veya parlak nesneleri saptama olasılığı bilgisine sahibiz. Ancak sensörün yuvarlak veya kare nesneleri saptama olasılığı hakkında bir bilgiye sahip değiliz.

$$P(X_1 = x_1, X_2 = x_2, X_3 = x_3) = \frac{n!}{x_1! x_2! x_3!} p_1^{x_1} p_2^{x_2} p_3^{x_3}, p_1 + p_2 + p_3 = 1$$

Probleme kolaylık olması bakımından sensörün koyu yuvarlak nesneyi saptaması olasılığı, tüm mümkün durumlar {Koyu yuvarlak, Koyu kare, Parlak yuvarlak, Parlak kare} göz önüne alınarak $p_1 = \frac{1}{4}$ olarak belirlensin. $p_1 = \frac{1}{4}$ ve $p_1 + p_2 + p_3 = 1$ olduğundan $p_2 + p_3 = \frac{3}{4}$ dür.

$$P(X_1 = x_1, X_2 = x_2, X_3 = x_3) = \frac{n!}{x_1! x_2! x_3!} p_1^{x_1} p_2^{x_2} p_3^{x_3}, (x_1 + x_2 + x_3 = n; p_1 + p_2 + p_3 = 1)$$

$x_1 + x_2 + x_3 = n$ eşitliği bilindiğinden $x_3 = n - (x_1 + x_2)$ eşitliği kullanılırsa,

$$= \frac{n!}{x_1! x_2! (n - x_1 - x_2)!} p_1^{x_1} p_2^{x_2} (1 - (p_1 + p_2))^{n-x_1-x_2}$$

olur. Bu problem için,

Y_1 : Koyu nesnelerin sayısını,

Y_2 : Parlak nesnelerin sayısını,

göstermek üzere $y_1 = x_1 + x_2$ ve $y_2 = x_3$ şeklindedir. Buradan,

$$P(X_1 = x_1, X_2 = y_1 - x_1, X_3 = n - y_1) = \frac{n!}{x_1! (y_1 - x_1)! (n - y_1)!} p_1^{x_1} p_2^{y_1 - x_1} p_3^{n - y_1}$$

elde edilir.

$p_1 = \frac{1}{4}, p_2 = \frac{p+1}{4}, p_3 = \frac{2-p}{4}$ olarak alındığında,

$$P(X_1 = x_1 | Y_1 = y_1) = \frac{\frac{n!}{x_1! (y_1 - x_1)! (n - y_1)!} \left(\frac{1}{4}\right)^{x_1} \left(\frac{p+1}{4}\right)^{y_1 - x_1} \left(\frac{2-p}{4}\right)^{n - y_1}}{\frac{n!}{y_1! (n - y_1)!} \left(\frac{p+2}{4}\right)^{y_1} \left(\frac{2-p}{4}\right)^{n - y_1}}$$

$\left(\frac{p+2}{4}\right)^{y_1} = \left(\frac{p+2}{4}\right)^{x_1} \left(\frac{p+2}{4}\right)^{y_1 - x_1}$ olarak yazılırsa,

$$\begin{aligned} P(X_1 = x_1 | Y_1 = y_1) &= \frac{y_1!}{x_1! (y_1 - x_1)!} \left(\frac{\frac{1}{4}}{\frac{p+2}{4}}\right)^{x_1} \left(\frac{\frac{p+1}{4}}{\frac{p+2}{4}}\right)^{y_1 - x_1} \\ &= \binom{y_1}{x_1} \left(\frac{1}{p+2}\right)^{x_1} \left(1 - \frac{1}{p+2}\right)^{y_1 - x_1} \end{aligned}$$

olarak elde edilir.

$$X_1 \sim \text{Binom}\left(y_1, \frac{1}{p+2}\right) \text{ ya da } X_1 \sim \text{Binom}\left(y_1, \frac{p_1}{p_1 + p_2}\right)$$

Benzer şekilde,

$$X_2 \sim \text{Binom}\left(y_1, 1 - \frac{1}{p+2}\right) \text{ ya da } X_1 \sim \text{Binom}\left(y_1, 1 - \frac{p_1}{p_1+p_2}\right)$$

olduğu açıktır.

E adımı,

$$E(X_1|Y_1 = y_1, p) = y_1 \frac{1}{p+2} = y_1 \frac{p_1}{p_1+p_2}$$

$$E(X_2|Y_1 = y_1, p) = y_1 \frac{p+1}{p+2} = y_1 \frac{p_2}{p_1+p_2}$$

M adımı,

$$\log L = \log \binom{n}{x_1, x_2, x_3} + x_1 \log p_1 + x_2 \log p_2 + x_3 \log p_3$$

$$= \log \binom{n}{x_1, x_2, x_3} + x_1 \log \frac{1}{4} + x_2 \log \left(\frac{p+1}{4}\right) + x_3 \log \left(\frac{2-p}{4}\right)$$

$$\frac{d \log L}{dp} = x_2 \frac{1}{1+p} - x_3 \frac{1}{2-p} = 0$$

$$= x_2(2-p) - x_3(1+p) = 0$$

$$= -p(x_2 + x_3) + 2x_2 - x_3 = 0$$

$$\hat{p} = \frac{2x_2 - x_3}{x_2 + x_3}$$

X_2 yerine E adımıdaki beklenen değer yazılırsa,

$$\begin{aligned}\hat{p} &= \frac{2 \left[y_1 \left(1 - \frac{1}{2+p} \right) \right] - y_2}{y_1 \left(1 - \frac{1}{2+p} \right) + y_2} \\ &= \frac{2y_1(p+1) - (p+2)y_2}{y_1(p+1) + (p+2)y_2} = p^{(k+1)} \\ &= \frac{p^{(k)}(2y_1 - y_2) + 2(y_1 - y_2)}{p^{(k)}(y_1 + y_2) + (y_1 + 2y_2)} = p^{(k+1)}\end{aligned}$$

Bir görüntüdeki nesneleri algılama problemi için verilen teorik ifadeler, Excel yardımıyla başlangıç değeri $p^{(0)}=1$ ve $\varepsilon = 10^{-3}$ alınarak çözdürülmüş ve Çizelge 3.2’de elde edilen değerler gösterilmiştir.

Çizelge 3.2 Bir görüntüdeki nesneleri algılama problemi için EM algoritması ile elde edilen parametre değerleri

$y_1 = 10, y_2 = 5, x_1 = 6, x_2 = 4$ ve $x_3 = 5$			
$p_1 = \frac{1}{4}, p_2 = \frac{1}{2}, p_3 = \frac{1}{4}$			
i	$p^{(i)}$	$LogL^{(i)}$	$LogL^{(i)} - LogL^{(i-1)}$
0	$p^{(0)}=1$	630622,1732	-
1	$p^{(1)}=0,714286$	630622,4512	$LogL^{(1)} - LogL^{(0)} = 0,2779$
2	$p^{(2)} = 0,674419$	630622,4766	$LogL^{(2)} - LogL^{(1)} = 0,0254$
3	$p^{(3)} = 0,667954$	630622,4804	$LogL^{(3)} - LogL^{(2)} = 0,0098$
4	$p^{(4)} = 0,666881$	630622,4811	$LogL^{(4)} - LogL^{(3)} = 0,007$

Çizelge 3.2’de 4. adım ile 3. adımdaki olabilirlik fonksiyonu değerleri arasındaki farkın $\varepsilon = 10^{-3}$ değerinden küçük olduğu görülmektedir. Dolayısıyla ele alınan $p^{(0)}=1$ başlangıç değeri için EM algoritması yöntemi ile elde edilen çözüm $\hat{p} = 0,666881$ olarak bulunmuştur.

3.2.2 Newton-Raphson Yöntemi

n bilinmeyenli n tane lineer olmayan denklem sistemi,

$$f_1(x_1, x_2, \dots, x_n) = 0$$

$$f_2(x_1, x_2, \dots, x_n) = 0$$

$\vdots$

$$f_n(x_1, x_2, \dots, x_n) = 0$$

olarak verilsin. Bu sistem $\underline{f}(\underline{x}) = 0$ şeklinde yazılabilir. Burada $\underline{x}$ bağımsız değişkenlerden ve $\underline{f}, f_i(\underline{x})$ fonksiyonlarından oluşan bir vektördür.

$$\underline{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}, \underline{f}(\underline{x}) = \begin{bmatrix} f_1(x_1, x_2, \dots, x_n) \\ \vdots \\ f_n(x_1, x_2, \dots, x_n) \end{bmatrix} = \begin{bmatrix} f_1(\underline{x}) \\ \vdots \\ f_n(\underline{x}) \end{bmatrix}$$

Newton Raphson yöntemi ile lineer denklem sistemlerinin çözümü için Jacobian matrisine ihtiyaç vardır. Burada Jacobian matrisi aşağıdaki gibi gösterilmektedir.

$$J = \frac{\partial(f_1, f_2, \dots, f_n)}{\partial(x_1, x_2, \dots, x_n)} = \begin{bmatrix} \frac{df_1}{dx_1} & \dots & \frac{df_1}{dx_n} \\ \vdots & \ddots & \vdots \\ \frac{df_n}{dx_1} & \dots & \frac{df_n}{dx_n} \end{bmatrix} = \left[\frac{\partial f_i}{\partial x_j} \right]_{n \times n}$$

$\underline{x} = \underline{x}_0$ başlangıç noktası olarak alındığında, çözüm için ardışık adımlar şu şekilde hesaplanır.

$$\underline{x}_{n+1} = \underline{x}_n - J^{-1} \underline{f}(\underline{x}_n)$$

Burada yakınsama kriteri,

$$|\underline{x}_{n+1} - \underline{x}_n| < \varepsilon$$

olarak alınabilir. Burada ε yeterince küçük seçimli pozitif bir sayıdır.

Newton Raphson yönteminin zorluklarından biri $\frac{\partial f_i}{\partial x_j}$ ($i, j = 1, 2, \dots, n$) fonksiyonlarının her birinin Jacobian matrisinin içinde tanımlı olmasının gerekmesidir. Eşitlikler ve bilinmeyenler arttıkça, Jacobian matrisindeki ifadelerin sayısı artmakta ve çözüm elde etmek zorlaşmaktadır.

3.3 En Küçük Kareler Yöntemi

En küçük kareler yöntemi (LS), deneysel (ampirik) dağılım $\hat{F}$ ile parametrik dağılım F arasında bir regresyon ilişkisinin olduğu düşüncesine dayanır. Öyle ki, $x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$ sıralı gözlemlerine karşın $\hat{F}(x_{(i)}) \equiv \frac{i}{n+1}$ ($i = 1, 2, \dots, n$) ampirik dağılım fonksiyonu göz önüne alınarak,

$$Q(\underline{\Phi}) = \sum_{i=1}^n (\hat{F}(x_{(i)}) - F(x_{(i)}; \underline{\Phi}))^2 \quad (3.39)$$

ifadesini minimum yapan $\underline{\Phi}$ parametre vektörü belirlenmeye çalışılır.

3.4 Kolmogorov-Smirnov Testi

Kolmogorov-Smirnov (KS) testi, ele alınan bir örneklemin herhangi ya da belirli bir dağılımdan gelip gelmediğinin incelenmesinde kullanılan bir testtir. KS, ele alınan örneklemin deneysel dağılımı $(\hat{F}(X))$ ile H_0 varsayımı altında tanımlanan teorik dağılımın $(F(X))$ birbirine uygunluğunu test etmek için kullanılır.

$D_{max} = \text{maksimum}|F(X) - \hat{F}(X)|$ değeri, belirlenen anlamlılık düzeyi (α) için hesaplanan $D_{(\alpha)}$ kritik değerleri ile karşılaştırılır. Eğer,

1. $D_{max} < D_{(\alpha)}$ ise örneklem F gibi bir dağılımdan gelmektedir.
2. $D_{max} \geq D_{(\alpha)}$ ise örneklemin alındığı kitleyi F temsil etmemektedir.

Bu çalışmada KS testi, 6. bölümde incelenen veri setlerinin modellenmesinde iki bileşenli karma üstel dağılımın ne kadar başarılı olduğunu belirlemek için kullanılmıştır.



4. ARAŞTIRMA BULGULARI

Bu bölümde, çalışmada ele alınan 2MED için momentler, en çok olabilirlik ve en küçük kareler yöntemleri ile EM algoritması ve Newton Raphson yöntemlerinden bahsedilecektir.

4.1 İki Bileşenli Karma Üstel Dağılım İçin Momentler Yöntemi

(2.18) eşitliğinde örneklem momentleri kitle momentlerine eşitlenirse,

$$m_1 = \frac{\sum x_i}{n} = \alpha \theta_1 + (1 - \alpha) \theta_2 \quad (4.1)$$

$$m_2 = \frac{\sum x_i^2}{n} = 2\alpha \theta_1^2 + 2(1 - \alpha) \theta_2^2 \quad (4.2)$$

$$m_3 = \frac{\sum x_i^3}{n} = 6\alpha \theta_1^3 + 6(1 - \alpha) \theta_2^3 \quad (4.3)$$

Eşitlikler düzenlendiğinde,

$$m_1 = \alpha \theta_1 + (1 - \alpha) \theta_2 \quad (4.4)$$

$$\frac{m_2}{2} = \alpha \theta_1^2 + (1 - \alpha) \theta_2^2 \quad (4.5)$$

$$\frac{m_3}{6} = \alpha \theta_1^3 + (1 - \alpha) \theta_2^3 \quad (4.6)$$

dır (Rider 1961, Açığöz 2007, Yılmaz ve Buyum 2014, Yılmaz vd. 2014, Yılmaz ve Buyum 2015).

4.1.1 İki bileşenli karma üstel dağılım parametreleri için α biliniyorken moment tahmin edicileri

(4.4) eşitliği yoluyla elde edilen,

$$\theta_1 = \frac{m_1 - (1-\alpha)\theta_2}{\alpha} \quad (4.7)$$

(4.7) eşitliği (4.5) eşitliğinde yerine yazılırsa,

$$\theta_{1,2} = m_1 \pm \frac{1-\alpha}{\alpha} \sqrt{\frac{\alpha}{1-\alpha} \left(\frac{m_2}{2} - m_1^2 \right)} \quad (4.8)$$

θ parametresi için moment tahmin edicileri elde edilir.

$\theta_1 > \theta_2$ durumunda,

$$\theta_1 = m_1 + \frac{1-\alpha}{\alpha} \sqrt{\frac{\alpha}{1-\alpha} \left(\frac{m_2}{2} - m_1^2 \right)} \quad (4.9)$$

$$\theta_2 = m_1 - \sqrt{\frac{\alpha}{1-\alpha} \left(\frac{m_2}{2} - m_1^2 \right)} \quad (4.10)$$

olmalıdır.

$\theta_2 \geq 0$ olup olmadığı kontrol edilirse,

$$m_1^2 \geq \frac{\alpha}{1-\alpha} \frac{m_2}{2} - \frac{\alpha}{1-\alpha} m_1^2 \quad (4.11)$$

(4.11) eşitliğinin her iki tarafı $(1 - \alpha)$ ile çarpılırsa;

$$m_1^2(1 - \alpha) \geq \alpha \frac{m_2}{2} - \alpha m_1^2 \quad (4.12)$$

$$m_1^2 \geq \alpha \frac{m_2}{2} \quad (4.13)$$

olduğundan θ_2 için tahmin değeri negatif değerli değildir. Burada çözümü elde edebilmek için aşağıda verilen iki koşulun varsayımı gereklidir.

$$1. \quad m_1^2 \geq \alpha \frac{m_2}{2}$$

$$2. \quad \frac{m_2}{2} \geq m_1^2$$

Her ne kadar teorik olarak bu ifadeler pozitif olsa da simülasyon yapıldığında sonuçlar tersi durum olarak çıkabilmektedir. Söz konusu koşulların, simülasyon sonucunda sağlanıp sağlanmadığına ilişkin örnekler 5. bölümde incelenecektir.

4.1.2 İki bileşenli karma üstel dağılım parametreleri için α bilinmiyorken moment tahmin edicileri

Yukarıda elde edilen (4.4), (4.5) ve (4.6) eşitlikleri kullanılarak,

$$\frac{m_3}{6} = \frac{m_2}{2} (\theta_1 + \theta_2) - \theta_1 \theta_2 m_1 \quad (4.14)$$

$$\frac{m_2}{2} = m_1(\theta_1 + \theta_2) - \theta_1 \theta_2 \quad (4.15)$$

eşitlikleri elde edilir.

$\theta_1 + \theta_2 = z_1$ ve $\theta_1 \theta_2 = z_2$ kabulü altında (4.14) ve (4.15) eşitlikleri birlikte çözülürse,

$$z_1 = \frac{\frac{m_3}{6} - \frac{m_1 m_2}{2}}{\frac{m_2}{2} - m_1^2} \quad (4.16)$$

ve

$$z_2 = \frac{\frac{m_1 m_3 - \frac{m_2^2}{2}}{6}}{\frac{m_2}{2} - m_1^2} \quad (4.17)$$

elde edilir.

Bulunan z_1 ve z_2 değerleri, tanımı gereği her zaman pozitif olmalıdır. θ_1 ve θ_2 'nin pozitif olması için z_1 ve z_2 'nin pozitif olması yeterlidir. Buna göre,

$$\text{i. } \frac{m_2}{2} > m_1^2 \text{ ise } \frac{m_3}{6} > \frac{m_1 m_2}{2} \text{ ve } \frac{m_3}{6} m_1 > \left(\frac{m_2}{2}\right)^2$$

Bu üç koşul düzenlenirse $\frac{1}{m_1} > \frac{m_1}{m_{2/2}} > \frac{m_{2/2}}{m_{3/6}}$ elde edilir.

$$\text{ii. } \frac{m_2}{2} < m_1^2 \text{ ise } \frac{m_3}{6} < \frac{m_1 m_2}{2} \text{ ve } \frac{m_3}{6} m_1 < \left(\frac{m_2}{2}\right)^2$$

Gerektirmeleri sonucu $\frac{1}{m_1} < \frac{m_1}{m_{2/2}} < \frac{m_{2/2}}{m_{3/6}}$ elde edilir.

Karma üstel dağılımın parametrelerini kolaylıkla tahmin edebilmek için bir eşitliğe daha ihtiyaç duyulmaktadır. Bu eşitlik, $\theta_1 - \theta_2 = \pm \sqrt{z_1^2 - 4z_2}$ şeklindedir. Eğer başta, $\theta_1 > \theta_2$ bilgisi varsa pozitif kısım çözüm için alınabilir. Buna göre, $\theta_1 + \theta_2 = z_1$ ve $\theta_1 - \theta_2 = \sqrt{z_1^2 - 4z_2}$ eşitliklerinden çözüme gidilir. Burada bir koşulun daha geldiğine dikkat edilmelidir. Bu da karekök ifadesindeki $z_1^2 - 4z_2$ 'nin pozitif tanımlı olmasıdır. Yukarıdaki (i) ve (ii) koşullarının bunun için yeterli olup olmadığına bakılacaktır. Ayrıca z_1 ve z_2 arasında örneklem momentleri açısından bir ilişki mevcuttur. Bu ilişki,

$$z_2 = m_1 z_1 + \frac{m_1 \frac{m_2}{2} - \left(\frac{m_2}{2}\right)^2}{\frac{m_2}{2} - m_1^2} \quad (4.18)$$

şeklindedir. Buradan,

$$z_2 = m_1 z_1 - \frac{m_2}{2} \quad (4.19)$$

dır.

$$z_1^2 - 4z_2 = (z_1 - 2m_1)^2 + 4\left(\frac{m_2}{2} - m_1^2\right) \quad (4.20)$$

olup (i) durumunda yani $\frac{m_2}{2} > m_1^2$ olduğunda $z_1^2 - 4z_2 \geq 0$ dır. (ii) durumunda ise yani $\frac{m_2}{2} < m_1^2$ olduğunda $(z_1 - 2m_1)^2 - 4(m_1^2 - \frac{m_2}{2})$ ifadesi z_1 'in konveks bir fonksiyonu olup $z_1^* = 2m_1$ de minimuma sahiptir ve bu değer negatif işaretlidir. Dolayısıyla (ii) koşulu tek başına yeterli değildir. (ii) koşuluna ek olarak reel çözüm olabilmesi için $z_1^2 \geq 4z_2$ olmalıdır. Ayrıca (ii) koşulu sağlansa bile $m_1^2 - \frac{m_2}{2} = \alpha(1 - \alpha)(\theta_1 - \theta_2)^2$ olduğundan (ii) ve $z_1^2 \geq 4z_2$ koşulları karma parametrelerinin aralığını garanti edemez (Çizelge 4.1, 4.2) (Yılmaz ve Buyum 2015).

Bu koşullar altında ($\theta_1 > \theta_2$ durumunda),

$$\widehat{\theta}_1 = \frac{z_1 + \sqrt{z_1^2 - 4z_2}}{2} \text{ ve } \widehat{\theta}_2 = \frac{z_1 - \sqrt{z_1^2 - 4z_2}}{2} \quad (4.21)$$

(4.21) eşitliğindeki ifadeler (4.4) eşitliğinde yerine konulduğunda,

$$\hat{\alpha} = \frac{m_1 - \widehat{\theta}_2}{\widehat{\theta}_1 - \widehat{\theta}_2} \quad (4.22)$$

olarak elde edilir.

Yukarıda verilen koşulların gerekliliği, $\theta_1 = 5, \theta_2 = 3$ ve $\alpha = 0,6$ olan iki bileşenli karma üstel dağılımdan üretilen 50 birimlik iki örneklem yardımıyla çizelge 4.1- 4.2’de açıklanmıştır (Yılmaz ve Buyum 2015).

Çizelge 4.1 İki bileşenli karma üstel dağılımdan üretilen veri setleri ($\theta_1 = 5, \theta_2 = 3$ ve $\alpha = 0,6$)

Veri Seti 1									
8,31	7,8	4,17	3,42	20,06	1,1	9,67	0,04	0,13	4,41
6,97	3,67	0,18	8,28	13,65	7,97	6,72	15,99	0,27	5,32
1,39	0,11	20,95	4,42	0,35	0,4	0,63	3,86	7,26	1,3
8,4	7,73	9,92	0,01	0,61	1,27	2,64	2,61	4,21	3,2
5,54	7,11	0,48	0,27	4,85	4,82	3,23	7,63	0,12	8,35
Veri Seti 2									
2,97	17,44	1,19	0,91	0,18	1,35	4,39	6,89	3,55	0,34
1,29	0,12	1,18	0,71	4,52	5,24	5,98	0,61	3,25	8,02
4,57	7,17	4,6	0,99	4,38	1,45	8,04	15,94	2,93	14,54
4,21	4,29	3,28	12,66	0,56	6,49	3,53	11,44	4,9	7,38
3,62	3,62	3,75	0,14	0,76	9,4	3,12	0,97	12,3	10,58

Çizelge 4.2 Veri setlerine ilişkin hesaplanan değerler

	$\frac{1}{m_1}$	$\frac{m_1}{m_2/2}$	$\frac{m_2/2}{m_3/6}$	z_1	z_2	$z_1^2 - 4z_2$	$\hat{\theta}_1$	$\hat{\theta}_2$	$\hat{\alpha}$
Veri Seti 1	0,20	0,21	0,23	18,81	70,16	73,05	13,68	5,13	-0,01
Veri Seti 2	0,21	0,23	0,26	8,76	21,47	-9,17	-	-	-

Çizelge 4.1-4.2 incelendiğinde, (ii) koşulunun her iki veri seti için de sağlandığı görülmektedir. Veri seti 1 için, (ii) koşulunun yanı sıra $z_1^2 - 4z_2 > 0$ koşulu da sağlanmasına rağmen θ_1 ve θ_2 parametrelerine ilişkin tahmin değerleri hesaplanabiliyorken α parametresine ilişkin hesaplanan tahmin değeri (0,1) aralığında elde edilememektedir. Veri seti 2 için ise $z_1^2 - 4z_2$ ifadesi negatif olduğundan parametrelere ilişkin tahmin değerleri hesaplanamamaktadır.

4.2 İki Bileşenli Karma Üstel Dağılım İçin En Çok Olabilirlik Yöntemi

2MED’e ilişkin logaritmik olabilirlik fonksiyonu aşağıdaki gibidir.

$$\log L = \sum_{j=1}^n \log \left[\sum_{i=1}^2 \alpha_i \frac{1}{\theta_i} \exp(-x_j/\theta_i) \right] \quad (4.23)$$

Maksimizasyon problemini çözebilmek için Langrange çarpanları yöntemi kullanılacaktır.

$$\log L = \sum_{j=1}^n \log \left[\sum_{i=1}^2 \alpha_i \frac{1}{\theta_i} \exp(-x_j/\theta_i) \right] - \lambda (\sum_{i=1}^2 \alpha_i - 1), \quad \sum_{i=1}^2 \alpha_i = 1 \quad (4.24)$$

$i = 1, 2$ için (4.24)'de verilen fonksiyonun α_i 'ye göre türevi alınıp sıfıra eşitlenirse,

$$\frac{\partial \log L}{\partial \alpha_i} = \sum_{j=1}^n \frac{\frac{1}{\theta_i} \exp(-x_j/\theta_i)}{\sum_{i=1}^2 \alpha_i \frac{1}{\theta_i} \exp(-x_j/\theta_i)} - \lambda = 0 \quad (4.25)$$

$$\Rightarrow \sum_{j=1}^n \frac{\frac{1}{\theta_i} \exp(-x_j/\theta_i)}{\sum_{i=1}^2 \alpha_i \frac{1}{\theta_i} \exp(-x_j/\theta_i)} = \lambda$$

elde edilir. (4.25) eşitliğinin her iki tarafı α_i ile çarpılıp i indisi üzerinden toplam alınırsa,

$$\sum_{j=1}^n \sum_{i=1}^2 \frac{\alpha_i \frac{1}{\theta_i} \exp(-x_j/\theta_i)}{\sum_{i=1}^2 \alpha_i \frac{1}{\theta_i} \exp(-x_j/\theta_i)} = \lambda \alpha_i \quad (4.26)$$

olup, $n = \lambda$ bulunur. $X_j = x_j$ gözlenmişken x_j nin i . bileşene ait olması olasılığı Bayes kuralına göre;

$$P(i|x_j) = \frac{\alpha_i \frac{1}{\theta_i} \exp(-x_j/\theta_i)}{\sum_{i=1}^2 \alpha_i \frac{1}{\theta_i} \exp(-x_j/\theta_i)} \quad (4.27)$$

olarak ifade edilirse

$$\hat{\alpha}_i = \frac{\sum_{j=1}^n P(i|x_j)}{n} \quad (4.28)$$

biçiminde elde edilir. (4.24) eşitliğinin θ_i ye göre türevi sıfıra eşitlenirse,

$$\frac{\partial \log L}{\partial \theta_i} = \sum_{j=1}^n \frac{-\frac{\alpha_i}{\theta_i^2} \exp\left(-\frac{x_j}{\theta_i}\right) + \frac{\alpha_i}{\theta_i} \frac{x_j}{\theta_i^2} \exp\left(-\frac{x_j}{\theta_i}\right)}{\sum_{i=1}^2 \alpha_i \frac{1}{\theta_i} \exp(-x_j/\theta_i)} = 0 \quad (4.29)$$

$$= \sum_{j=1}^n \frac{\frac{\alpha_i}{\theta_i} \left(\frac{x_j}{\theta_i^2}\right) \exp\left(-\frac{x_j}{\theta_i}\right) - \frac{\alpha_i}{\theta_i^2} \exp(-x_j/\theta_i)}{\sum_{i=1}^2 \alpha_i \frac{1}{\theta_i} \exp(-x_j/\theta_i)} = 0$$

$$= \sum_{j=1}^n \frac{\frac{1}{\theta_i} \left[\alpha_i \frac{x_j}{\theta_i^2} \exp\left(-\frac{x_j}{\theta_i}\right) - \frac{\alpha_i}{\theta_i} \exp(-\frac{x_j}{\theta_i}) \right]}{\sum_{i=1}^2 \alpha_i \frac{1}{\theta_i} \exp(-x_j/\theta_i)} = 0$$

$$= \frac{1}{\theta_i} \sum_{j=1}^n \frac{\left[\alpha_i \frac{x_j}{\theta_i^2} \exp\left(-\frac{x_j}{\theta_i}\right) \right]}{\sum_{i=1}^2 \alpha_i \frac{1}{\theta_i} \exp(-x_j/\theta_i)} - \frac{1}{\theta_i} \sum_{j=1}^n \frac{\left[\alpha_i \frac{1}{\theta_i} \exp\left(-\frac{x_j}{\theta_i}\right) \right]}{\sum_{i=1}^2 \alpha_i \frac{1}{\theta_i} \exp(-x_j/\theta_i)} = 0$$

$\underbrace{\sum_{i=1}^2 \alpha_i \frac{1}{\theta_i} \exp(-x_j/\theta_i)}_{\sum_{j=1}^n P(i|x_j)}$

$$= \frac{1}{\theta_i} \left[\sum_{j=1}^n x_j \frac{\alpha_i \frac{1}{\theta_i} \exp\left(-\frac{x_j}{\theta_i}\right)}{\sum_{i=1}^2 \alpha_i \frac{1}{\theta_i} \exp(-x_j/\theta_i)} - \frac{\theta_i \sum_{j=1}^n P(i|x_j)}{\theta_i} \right] = 0$$

$$= \frac{1}{\theta_i} \left[\frac{1}{\theta_i} \left\{ \sum_{j=1}^n x_j \frac{\alpha_i \frac{1}{\theta_i} \exp\left(-\frac{x_j}{\theta_i}\right)}{\underbrace{\sum_{i=1}^2 \alpha_i \frac{1}{\theta_i} \exp(-x_j/\theta_i)}_{P(i|x_j)}} - \theta_i \sum_{j=1}^n P(i|x_j) \right\} \right] = 0$$

$$= \frac{1}{\theta_i} \left[\frac{1}{\theta_i} \{ \sum_{j=1}^n x_j P(i|x_j) - \theta_i \sum_{j=1}^n P(i|x_j) \} \right] = 0$$

$$\sum_{j=1}^n x_j P(i|x_j) = \theta_i \sum_{j=1}^n P(i|x_j) \quad (4.30)$$

$$\hat{\theta}_i = \frac{\sum_{j=1}^n x_j P(i|x_j)}{\sum_{j=1}^n P(i|x_j)}, i = (1,2) \quad (4.31)$$

elde edilir. $P(2|x_j) = 1 - P(1|x_j)$ olduğu hatırlanırsa,

$$\hat{\theta}_1 = \frac{1}{n\hat{\alpha}_i} \sum_{j=1}^n x_j P(i|x_j) \quad (4.32)$$

$$\hat{\theta}_2 = \frac{1}{n(1-\hat{\alpha}_i)} \sum_{j=1}^n x_j (1 - P(i|x_j)) \quad (4.33)$$

eşitlikleri ile yeniden yazılabilir.

4.2.1 İki bileşenli karma üstel dağılım için EM algoritması

EM algoritması için aşağıdaki adımlar izlenir;

1. Başlangıç değerleri girilir. $(\alpha_i^{(0)}, \theta_i^{(0)})$, $(i = 1,2)$
2. $P(i|x_j)$: x_j verilmişken i . karma bileşene ait olma olasılığı hesaplanır.

$$P(i|x_j) = \frac{\alpha_i \frac{1}{\theta_i} \exp\left(-\frac{x_j}{\theta_i}\right)}{\sum_{i=1}^2 \alpha_i \frac{1}{\theta_i} \exp(-x_j/\theta_i)}$$

3. k . adımda $\hat{\alpha}_i$ 'nin alacağı değer,

$$\hat{\alpha}_i^{(k)} = \frac{\sum_{j=1}^n P(i|x_j)}{n}$$

şeklinde hesaplanır.

4. k . Adımda $\hat{\theta}_i$ 'nin alacağı değer,

$$\hat{\theta}_i^{(k)} = \frac{1}{n\hat{\alpha}_i} \sum_{j=1}^n x_j P(i|x_j), i = 1,2$$

olarak elde edilir.

5. $\hat{\alpha}_i$ ve $\hat{\theta}_i$ bulunduktan sonra $\log L$ fonksiyonunda yerine konulur ve fonksiyonun değeri bulunur. Yeterince küçük seçimli $\varepsilon > 0$ sayısı için,

$$\log L^{(k)} - \log L^{(k-1)} \leq \varepsilon$$

ise k . adımdaki değerler parametre tahminleri olarak kullanılır. Değilse Adım 2-5 yinelenir.

4.2.2 ML'nin sayısal çözümü için Newton Raphson Yöntemi

İki bileşenli karma üstel dağılımın parametrelerinin en çok olabilirlik tahmini için Newton Raphson uygulanırken (4.25) ve (4.29) eşitlikleri kullanılsın.

$$f_1 = \frac{\partial \log L}{\partial \alpha_i} = \sum_{j=1}^n \frac{\frac{1}{\theta_i} \exp(-\frac{x_j}{\theta_i})}{\sum_{i=1}^2 \alpha_i \frac{1}{\theta_i} \exp(-x_j/\theta_i)} - \lambda = 0 \quad (4.34)$$

$$f_2 = \frac{\partial \log L}{\partial \theta_1} = \sum_{j=1}^n \frac{-\frac{\alpha_i}{\theta_1^2} \exp(-\frac{x_j}{\theta_1}) + \frac{\alpha_i}{\theta_1} \frac{x_j}{\theta_1^2} \exp(-\frac{x_j}{\theta_1})}{\sum_{i=1}^2 \alpha_i \frac{1}{\theta_1} \exp(-x_j/\theta_1)} = 0 \quad (4.35)$$

$$f_3 = \frac{\partial \log L}{\partial \theta_2} = \sum_{j=1}^n \frac{-\frac{\alpha_i}{\theta_2^2} \exp(-\frac{x_j}{\theta_2}) + \frac{\alpha_i}{\theta_2} \frac{x_j}{\theta_2^2} \exp(-\frac{x_j}{\theta_2})}{\sum_{i=1}^2 \alpha_i \frac{1}{\theta_2} \exp(-x_j/\theta_2)} = 0 \quad (4.36)$$

olarak tanımlansın. Burada,

$$\begin{aligned} \underbrace{\begin{bmatrix} \alpha^{(m+1)} \\ \theta_1^{(m+1)} \\ \theta_2^{(m+1)} \end{bmatrix}}_{\Psi^{(m+1)}} &= \underbrace{\begin{bmatrix} \alpha^{(m)} \\ \theta_1^{(m)} \\ \theta_2^{(m)} \end{bmatrix}}_{\Psi^{(m)}} - \begin{bmatrix} \frac{df_1}{d\alpha} & \frac{df_1}{d\theta_1} & \frac{df_1}{d\theta_2} \\ \frac{df_2}{d\alpha} & \frac{df_2}{d\theta_1} & \frac{df_2}{d\theta_2} \\ \frac{df_3}{d\alpha} & \frac{df_3}{d\theta_1} & \frac{df_3}{d\theta_2} \end{bmatrix}^{-1} \begin{bmatrix} f_1(\Psi^{(m)}) \\ f_2(\Psi^{(m)}) \\ f_3(\Psi^{(m)}) \end{bmatrix} \\ \Psi^{(m+1)} &= \Psi^{(m)} - \begin{bmatrix} \frac{df_1}{d\alpha} & \frac{df_1}{d\theta_1} & \frac{df_1}{d\theta_2} \\ \frac{df_2}{d\alpha} & \frac{df_2}{d\theta_1} & \frac{df_2}{d\theta_2} \\ \frac{df_3}{d\alpha} & \frac{df_3}{d\theta_1} & \frac{df_3}{d\theta_2} \end{bmatrix}^{-1} \begin{bmatrix} f_1(\Psi^{(m)}) \\ f_2(\Psi^{(m)}) \\ f_3(\Psi^{(m)}) \end{bmatrix} \end{aligned} \quad (4.37)$$

şeklindedir. (4.37) eşitliği $\varepsilon > 0$ sayısı için,

$$|\Psi^{(m+1)} - \Psi^{(m)}| < \varepsilon$$

sağlanıncaya kadar devam ettirilir.

Newton-Raphson yönteminin avantajı hızlı yakınsamasıdır. Dezavantajı ise, parametrelerin sayısı fazla olduğunda Hessian (H) matrisinin tersinin bulunmasının zorlaşması ve yakınsamanın olmaması durumunda H matrisinin negatif tanımlı olmasıdır (Açıkgöz 2007).

4.3 İki Bileşenli Karma Üstel Dağılım İçin En Küçük Kareler Yöntemi

(3.39) denklemi ele alındığında,

$$Q(\alpha, \theta_1, \theta_2) = \sum_{i=1}^n (\hat{F}(x_{(i)}) - F(x_{(i)}; \alpha, \theta_1, \theta_2))^2$$

ifadesini minimum yapan $(\alpha, \theta_1, \theta_2)$ parametre vektörü belirlenmeye çalışılır. Bu optimizasyon probleminin çözümü için oluşacak denklem sistemi aşağıdaki gibidir.

$$\frac{\partial Q}{\partial \alpha} = \sum_{i=1}^n \left(\frac{i}{n+1} - \alpha \left(1 - e^{-x_{(i)} \frac{1}{\theta_1}} \right) - (1-\alpha) \left(1 - e^{-x_{(i)} \frac{1}{\theta_2}} \right) \right) \sum_{i=1}^n \left(1 - e^{-x_{(i)} \frac{1}{\theta_2}} - \alpha \left(1 - e^{-x_{(i)} \frac{1}{\theta_1}} \right) \right) = 0$$

$$\frac{\partial Q}{\partial \theta_1} = \sum_{i=1}^n \left(\frac{i}{n+1} - \alpha \left(1 - e^{-x_{(i)} \frac{1}{\theta_1}} \right) - (1-\alpha) \left(1 - e^{-x_{(i)} \frac{1}{\theta_2}} \right) \right) \left(\frac{\alpha x_{(i)}}{\theta_1^2} e^{-x_{(i)} \frac{1}{\theta_1}} \right) = 0$$

$$\frac{\partial Q}{\partial \theta_2} = \sum_{i=1}^n \left(\frac{i}{n+1} - \alpha \left(1 - e^{-x_{(i)} \frac{1}{\theta_1}} \right) - (1-\alpha) \left(1 - e^{-x_{(i)} \frac{1}{\theta_2}} \right) \right) \left(\frac{(1-\alpha)x_{(i)}}{\theta_2^2} e^{-x_{(i)} \frac{1}{\theta_2}} \right) = 0$$

Türev alındıktan sonra gelen ifadeler parametreye bağlı olduğundan çözümlerin elde edilmesi zordur. Bu nedenle nümerik yollara başvurulması gerekmektedir.

5. SİMÜLASYON ÇALIŞMASI

Bu kısımda, öncelikle 2. bölümde bahsedilen 2MED'in parametrelerini 4. bölümde bahsedilen MOM ile tahmin etmek için bir çalışma yapılmıştır. Burada, 2MED ile üretilen 40 gözlem ile parametre değerleri birbirine çok yakın, yakın, uzak ve çok uzak olacak şekilde seçilerek ve karma oranları her bir durum için değiştirilerek tahmin değerleri elde edilmeye çalışılmıştır. Ardından, çalışmada kullanılan yöntemlerden hangisinin parametre tahmininde daha başarılı olduğunu belirleyebilmek adına farklı örneklem büyüklükleri için simülasyon çalışması yapılmış ve hata kareler ortalaması kare köküne (RMSE) göre karşılaştırılmıştır. Son olarak ise 2MED'e ilişkin üretilen veri setleri yardımıyla parametre tahmin yöntemleri Ki-Kare test istatistiğine (χ^2) göre de karşılaştırılmıştır.

Gerçek parametre değerleri $\theta_1 = 10$, $\theta_2 = 2$ ve $\alpha = 0.625$ olan iki bileşenli karma üstel dağılıma ilişkin üretilen 40 gözlem değeri çizelge 5.1'de gösterilmiştir.

Çizelge 5.1 İki bileşenli karma üstel dağılıma ilişkin üretilen veri seti ($\theta_1 = 10$, $\theta_2 = 2$ ve $\alpha = 0.625$)

Gözlem	1	2	3	4	5	6	7	8
	12,86489	42,2078	3,56824	1,3881	5,61233	12,9275	2,64424	7,75514
Gözlem	9	10	11	12	13	14	15	16
	0,18203	7,64203	12,57203	13,30899	4,68376	5,13972	1,524725	19,490471
Gözlem	17	18	19	20	21	22	23	24
	0,34141	14,9511	3,49333	22,20045	16,3012	4,77239	9,87521	1,17522
Gözlem	25	26	27	28	29	30	31	32
	13,4103	0,21239	1,469521	2,551901	3,09831	2,50784	0,35569	2,831374
Gözlem	33	34	35	36	37	38	39	40
	1,44689	0,14629	2,262833	1,835730	2,28646	3,73205	0,55653	0,321897

Çizelge 5.2 MOM yöntemine göre veri setine ilişkin hesaplanan değerler

$m_1 = 6,641205$	$m_2/2 = 55,13087$	$m_3/6 = 478,9396$
$z_1 = 10,23143$	$z_2 = 12,81817$	$z_1^2 - 4z_2 = 53,40951$

Hesaplanan değerlere göre $\frac{1}{m_1} = 0,150575 > \frac{m_1}{m_2/2} = 0,120463 > \frac{m_2/2}{m_3/6} = 0,11511$ oranlar monoton azalmaktadır ve hesaplanan $z_1^2 - 4z_2$ değeri de pozitif bulunmuştur. Buna göre parametrelere ilişkin bulunan momentler tahmin değerleri çizelge 5.3'te gösterilmiştir.

Çizelge 5.3 MOM parametre tahminleri

$\hat{\theta}_1$	$\hat{\theta}_2$	$\hat{\alpha}$
8,769806	1,461625	0,708737

Gerçek parametre değerleri $\theta_1 = 3, \theta_2 = 1$ ve $\alpha = 0.25$ olan iki bileşenli karma üstel dağılıma ilişkin üretilen 40 gözlem değeri çizelge 5.4'de gösterilmiştir.

Çizelge 5.4 İki bileşenli karma üstel dağılıma ilişkin üretilen veri seti ($\theta_1 = 3, \theta_2 = 1$ ve $\alpha = 0.25$)

Gözlem	1	2	3	4	5	6	7	8
	5,061923	3,134541	0,374841	1,186258	0,225059	0,432430	0,889591	0,814989
Gözlem	9	10	11	12	13	14	15	16
	0,473359	0,831586	2,024277	3,483752	1,462756	1,111760	0,681885	0,469722
Gözlem	17	18	19	20	21	22	23	24
	1,994411	0,271994	1,002189	0,161800	0,192510	3,317741	1,615871	1,443968
Gözlem	25	26	27	28	29	30	31	32
	1,266431	1,654376	1,712446	0,800673	0,428703	0,856390	0,171051	1,468401
Gözlem	33	34	35	36	37	38	39	40
	0,246106	1,564676	2,390933	0,706507	2,738291	0,082804	1,403125	0,640586

Çizelge 5.5 MOM yöntemine göre veri setine ilişkin hesaplanan değerler

$m_1 = 1,269768$	$m_2/2 = 1,379362$	$m_3/6 = 1,374409$
$z_1 = 1,618641$	$z_2 = 0,675936$	$z_1^2 - 4z_2 = -0,08375$

Hesaplanan değerlere göre $\frac{1}{m_1} = 0,787546 < \frac{m_1}{m_2/2} = 0,920547 < \frac{m_2/2}{m_3/6} = 1,003604$ oranlar monoton artmaktadır. Hesaplanan $z_1^2 - 4z_2$ değeri de negatif olduğundan parametrelere ilişkin momentler tahmin değerleri bulunamaz.

Gerçek parametre değerleri $\theta_1 = 3, \theta_2 = 1$ ve $\alpha = 0.875$ olan iki bileşenli karma üstel dağılıma ilişkin üretilen 40 gözlem değeri çizelge 5.6'da gösterilmiştir.

Çizelge 5.6 İki bileşenli karma üstel dağılıma ilişkin üretilen veri seti ($\theta_1 = 3, \theta_2 = 1$ ve $\alpha = 0.875$)

Gözlem	1	2	3	4	5	6	7	8
	4,909768	0,019572	4,632858	2,885890	1,837935	7,826508	4,929157	4,682977
Gözlem	9	10	11	12	13	14	15	16
	1,828228	3,192809	0,706325	2,070339	2,466611	0,655215	8,157723	0,030919
Gözlem	17	18	19	20	21	22	23	24
	0,715494	7,574199	2,563539	0,467320	3,358953	1,714827	5,491864	5,305492
Gözlem	25	26	27	28	29	30	31	32
	0,253987	7,220071	1,241766	7,841088	0,952321	0,696135	1,906388	1,579058
Gözlem	33	34	35	36	37	38	39	40
	3,241651	10,624864	13,972618	0,811134	0,106486	0,232001	0,006565	0,787242

Çizelge 5.7 MOM yöntemine göre veri setine ilişkin hesaplanan değerler

$m_1 = 3,237447$	$m_2/2 = 10,45069$	$m_3/6 = 30,00261$
$z_1 = 126,1035$	$z_2 = 397,8026$	$z_1^2 - 4z_2 = 14310,87$

Hesaplanan değerlere göre $\frac{1}{m_1} = 0,308885 < \frac{m_1}{m_2/2} = 0,309783 < \frac{m_2/2}{m_3/6} = 0,348326$ oranlar monoton artmaktadır ve hesaplanan $z_1^2 - 4z_2$ değeri de pozitifdir. Buna göre bulunan parametrelere ilişkin momentler tahmin değerleri çizelge 5.8’de gösterilmiştir.

Çizelge 5.8 MOM parametre tahminleri

$\hat{\theta}_1$	$\hat{\theta}_2$	$\hat{\alpha}$
122,8658	3,237701	-2,1E-06

Bulunan $\hat{\alpha}$ değeri sıfıra çok yakın bir sayı olduğundan sıfır olarak alınabilir.

Gerçek parametre değerleri $\theta_1 = 8, \theta_2 = 2$ ve $\alpha = 0.75$ olan iki bileşenli karma üstel dağılıma ilişkin üretilen 40 gözlem değeri çizelge 5.9’da gösterilmiştir.

Çizelge 5.9 İki bileşenli karma üstel dağılıma ilişkin üretilen veri seti ($\theta_1 = 8, \theta_2 = 2$ ve $\alpha = 0.75$)

Gözlem	1	2	3	4	5	6	7	8
	6,79597	6,009654	7,582794	0,927202	3,770922	1,270207	5,677586	6,443097
Gözlem	9	10	11	12	13	14	15	16
	5,46821	1,645385	0,875803	7,693685	11,98617	2,819228	7,094189	5,665031
Gözlem	17	18	19	20	21	22	23	24
	11,2263	0,554910	1,692423	26,27425	4,430198	0,579792	0,418201	3,890905
Gözlem	25	26	27	28	29	30	31	32
	7,68553	7,681584	0,766614	3,133094	8,534752	4,721206	0,000610	0,172748
Gözlem	33	34	35	36	37	38	39	40
	0,03620	0,170272	0,378444	0,142074	0,009538	0,319463	0,968253	0,247881

Çizelge 5.10 MOM yöntemine göre veri setine ilişkin hesaplanan değerler

$m_1 = 4,144009$	$m_2/2 = 20,45744$
$\frac{m_2}{2} - m_1^2 = 3,284631555$	$m_1^2 - \alpha \frac{m_2}{2} = 1,829729$

α biliniyorken hesaplanan değerlere göre $\frac{m_2}{2} - m_1^2 > 0$ ve $m_1^2 - \alpha \frac{m_2}{2} > 0$ 'dır. Gerekli varsayımların sağlanması ile parametrelere ilişkin momentler tahmin değerleri çizelge 5.11'de gösterilmiştir.

Çizelge 5.11 MOM parametre tahminleri

$\hat{\theta}_1$	$\hat{\theta}_2$
5,190373	1,004918

Gerçek parametre değerleri $\theta_1 = 8, \theta_2 = 0.2$ ve $\alpha = 0.75$ olan iki bileşenli karma üstel dağılıma ilişkin üretilen 40 gözlem değeri çizelge 5.12'de gösterilmiştir.

Çizelge 5.12 İki bileşenli karma üstel dağılıma ilişkin üretilen veri seti ($\theta_1 = 8, \theta_2 = 0.2$ ve $\alpha = 0.75$)

Gözlem	1	2	3	4	5	6	7	8
	6,697942	3,040823	2,197411	2,062255	6,074840	1,857751	2,397243	0,801930
Gözlem	9	10	11	12	13	14	15	16
	17,057478	0,365596	11,944127	2,979266	2,594966	15,964862	3,033553	6,338430
Gözlem	17	18	19	20	21	22	23	24
	1,514699	8,935777	6,601960	4,314849	10,610713	53,577037	0,494656	14,305303
Gözlem	25	26	27	28	29	30	31	32
	8,905008	7,284671	11,602462	5,592612	1,900948	1,962300	0,001637	0,003627
Gözlem	33	34	35	36	37	38	39	40
	0,121345	0,030972	0,214845	0,081814	1,103640	0,001216	0,049279	0,298058

Çizelge 5.13 MOM yöntemine göre veri setine ilişkin hesaplanan değerler

$m_1 = 5,622948$	$\frac{m_2}{2} = 56,28718$
$\frac{m_2}{2} - m_1^2 = 24,66963579$	$m_1^2 - \alpha \frac{m_2}{2} = -10,59784079$

α biliniyorken hesaplanan değerlere göre $\frac{m_2}{2} - m_1^2 > 0$ fakat $m_1^2 - \alpha \frac{m_2}{2} < 0$ 'dır. Gerekli her iki varsayım sağlanmadığından θ_1 parametresine ilişkin momentler tahmin değeri

bulunabilirken θ_2 parametresine ilişkin bulunan tahmin değeri negatif çıkmaktadır. Dolayısıyla θ_2 parametresine ilişkin moment tahmin değeri bulunmamaktadır.

Çizelge 5.14 MOM parametre tahminleri

$\hat{\theta}_1$	$\hat{\theta}_2$
8,490562	-2,9799

Gerçek parametre değerleri $\theta_1 = 8, \theta_2 = 0.2$ ve $\alpha = 0.75$ olan iki bileşenli karma üstel dağılıma ilişkin üretilen 40 gözlem değeri çizelge 5.15’de gösterilmiştir.

Çizelge 5.15 İki bileşenli karma üstel dağılıma ilişkin üretilen veri seti ($\theta_1 = 8, \theta_2 = 0.2$ ve $\alpha = 0.75$)

Gözlem Sayısı	1	2	3	4	5	6	7	8
	1,721916	2,62334	0,08212	7,103111	10,16608	6,121808	0,816772	2,855012
Gözlem Sayısı	9	10	11	12	13	14	15	16
	6,709390	6,41851	7,85017	15,38088	11,053559	0,803342	3,976450	14,367029
Gözlem Sayısı	17	18	19	20	21	22	23	24
	2,010768	9,116160	2,643298	9,478706	4,522808	10,398749	14,780736	3,146373
Gözlem Sayısı	25	26	27	28	29	30	31	32
	12,301143	9,367208	2,808201	2,234863	8,997216	1,357151	0,187364	0,018661
Gözlem Sayısı	33	34	35	36	37	38	39	40
	0,246946	0,030527	0,428232	0,012476	0,196581	0,037457	0,449495	0,067872

Çizelge 5.16 MOM yöntemine göre veri setine ilişkin hesaplanan değerler

$m_1 = 4,822212$	$m_2/2 = 22,69518$
$\frac{m_2}{2} - m_1^2 = -0,55854521$	$m_1^2 - \alpha \frac{m_2}{2} = 6,232341$

α biliniyorken her iki parametreye ilişkin tahmin edicinin hesaplanmasında $\sqrt{\frac{\alpha}{1-\alpha}} \left(\frac{m_2}{2} - m_1^2 \right)$ ifadesi kullanılmaktadır. Dolayısıyla burada $\frac{m_2}{2} -$

$m_1^2 < 0$ olduğundan kökün içi negatif çıkmakta ve tahmin değerleri hesaplanamamaktadır.

Aşağıda 2MED parametrelerini MOM, LS ve ML ile tahmin etmek için karma dağılımın parametrelerine birbirinden uzak ve yakın olacak şekilde değerler verilmiş, karma oranının da düşük ve yüksek değerleri ile birlikte simülasyon çalışması yapılmıştır. Tahmin yöntemlerini karşılaştırmada RMSE ölçütü kullanılmış ve varsayılan parametre değerlerine göre hangi yöntemin daha iyi olduğu tartışılmıştır. Bu amaçla, çalışmada başlangıç değerleri kullanılarak 2MED parametrelerini ML ile tahmin ederken EM algoritması kullanılmış ve üç yöntem için tahminde bulunan bir MATLAB programı yazılmıştır.

Çizelge 5.17-5.19 'da sırasıyla $n = 20, 40$ ve 100 için $r = 1000$ tekrarlı simülasyon ile MOM, LS ve ML ile elde edilen parametre tahminlerini ve RMSE (parantez içinde) değerlerini göstermektedir.

$$RMSE_{\alpha_i} = \sqrt{\frac{\sum_{k=1}^r (\hat{\alpha}_k - \alpha_i)^2}{r}}$$

$$RMSE_{\theta_i} = \sqrt{\frac{\sum_{i=1}^r (\hat{\theta}_k - \theta_i)^2}{r}} \quad (i = 1, 2)$$

Çizelge 5.17 MOM, LS ve ML yöntemi ile elde edilen parametre tahminleri ve RMSE değerleri (n=20)

n	α	θ_1, θ_2	Parametre(α)			Parametre(θ_1)			Parametre(θ_2)		
			MOM	LSE	MLE	MOM	LSE	MLE	MOM	LSE	MLE
20	0.4	(10,1)	0.4625 (0.1426)	0.3647 (0.1872)	0.3003 (0.2008)	8.4145 (4.6951)	9.2826 (1.1427)	11.4051 (6.8530)	0.4954 (0.6383)	1.1427 (0.5926)	1.2969 (0.7033)
		(3,2)	0.5401 (0.2698)	0.4325 (0.2023)	0.2651 (0.2057)	4.3930 (5.8288)	2.1643 (2.2265)	5.6647 (4.0266)	0.8923 (1.2924)	2.5654 (1.0045)	1.5223 (0.7017)
		(0.8,0.3)	0.5447 (0.2425)	0.3309 (0.2193)	0.2578 (0.2130)	0.8555 (0.3478)	1.0895 (0.6238)	1.3284 (0.8833)	0.1627 (0.1864)	0.3008 (0.1321)	0.3070 (0.1148)
		(2,0.3)	0.4828 (0.1739)	0.3813 (0.1919)	0.2806 (0.2194)	1.8345 (0.8435)	2.1906 (1.1948)	2.7446 (1.7256)	0.1585 (0.1907)	0.3228 (0.1408)	0.3768 (0.1943)
	0.45	(10,1)	0.4805 (0.1337)	0.4198 (0.1945)	0.3501 (0.2133)	9.1633 (4.1289)	9.7134 (4.7509)	11.5160 (6.1693)	0.5539 (0.6487)	1.1657 (0.6979)	1.3455 (0.8206)
		(3,2)	0.5154 (0.2519)	0.4663 (0.2165)	0.2509 (0.2394)	4.1841 (3.1600)	2.2936 (2.2260)	5.5906 (3.7153)	1.0145 (1.2397)	2.8228 (1.3209)	1.6019 (0.6488)
		(0.8,0.3)	0.5579 (0.2219)	0.3597 (0.2375)	0.2794 (0.2442)	0.8840 (0.4225)	1.1000 (0.6347)	1.3055 (0.8269)	0.1742 (0.1862)	0.3088 (0.1361)	0.3136 (0.1291)
		(2,0.3)	0.5141 (0.1656)	0.4092 (0.2097)	0.3130 (0.2311)	1.8588 (0.8432)	2.1969 (1.1056)	2.6617 (1.6853)	0.1504 (0.1893)	0.3151 (0.1426)	0.3809 (0.2064)
	0.55	(10,1)	0.5125 (0.1400)	0.4984 (0.2030)	0.4213 (0.2438)	10.8430 (4.7665)	10.4300 (4.4257)	12.6346 (7.7415)	0.7921 (0.7677)	1.2564 (0.9401)	1.5240 (1.1800)
		(3,2)	0.5493 (0.2352)	0.5274 (0.2097)	0.2928 (0.3113)	4.2534 (2.8430)	2.4867 (2.2190)	5.7381 (3.9369)	1.0350 (1.2343)	2.9481 (1.3930)	1.6447 (0.7050)
		(0.8,0.3)	0.5535 (0.2066)	0.3807 (0.2894)	0.2905 (0.3178)	0.9730 (0.8963)	1.0841 (0.6281)	1.3628 (0.9096)	0.1779 (0.1812)	0.3206 (0.1432)	0.3258 (0.1331)
		(2,0.3)	0.5393 (0.1437)	0.4847 (0.2028)	0.3661 (0.2724)	2.0816 (0.9202)	2.2707 (1.0618)	2.7509 (1.7501)	0.1823 (0.1949)	0.3225 (0.1447)	0.4189 (0.2357)
	0.6	(10,1)	0.5086 (0.1617)	0.5491 (0.1990)	0.4382 (0.2749)	11.4268 (4.6431)	10.1646 (4.3182)	13.2679 (7.6480)	0.9186 (0.8052)	1.2650 (0.9742)	1.6227 (1.3183)
		(3,2)	0.5546 (0.2420)	0.5602 (0.2292)	0.2814 (0.3569)	4.4386 (4.7308)	2.3480 (2.1323)	5.8883 (4.0774)	1.0419 (1.2171)	3.0442 (1.4537)	1.6909 (0.6355)
		(0.8,0.3)	0.5721 (0.2141)	0.4628 (0.2802)	0.3003 (0.3511)	1.0116 (0.7135)	1.0589 (0.5701)	1.4331 (0.9933)	0.1896 (0.1946)	0.3181 (0.1555)	0.3581 (0.1571)
		(2,0.3)	0.5302 (0.1660)	0.5620 (0.2017)	0.3770 (0.3136)	2.2633 (0.9923)	2.1002 (0.9257)	2.9244 (1.8077)	0.2380 (0.1965)	0.3223 (0.1605)	0.4643 (0.3127)
	0.8	(10,1)	0.5665 (0.2929)	0.7203 (0.2350)	0.5421 (0.3815)	13.5847 (9.5814)	10.7144 (3.6990)	4.5558 (8.6339)	1.9439 (1.9996)	1.4611 (1.3164)	2.2906 (2.4606)
		(3,2)	0.5333 (0.3613)	0.7607 (0.3013)	0.3217 (0.5289)	4.6819 (4.1073)	2.3791 (1.8516)	6.0489 (4.2563)	1.1302 (1.1925)	2.7658 (1.4160)	1.7129 (0.7607)
		(0.8,0.3)	0.5653 (0.3265)	0.5772 (0.3752)	0.3215 (0.5238)	1.2354 (1.4786)	1.0467 (0.5469)	1.5699 (1.1134)	0.2413 (0.2025)	0.3701 (0.1690)	0.4088 (0.1967)
		(2,0.3)	0.5708 (0.2930)	0.7702 (0.1915)	0.4546 (0.4464)	2.8412 (4.2704)	2.0134 (0.6805)	3.3869 (2.4684)	0.4438 (0.3942)	0.3260 (0.1842)	0.6487 (0.5623)

Çizelge 5.18 MOM, LS ve ML yöntemi ile elde edilen parametre tahminleri ve RMSE değerleri (n=40)

n	α	θ_1, θ_2	Parametre(α)			Parametre(θ_1)			Parametre(θ_2)		
			MOM	LSE	MLE	MOM	LSE	MLE	MOM	LSE	MLE
40	0.4	(10,1)	0.4209 (0.1247)	0.3851 (0.1463)	0.3576 (0.1393)	9.8812 (4.0102)	10.5911 (4.2156)	10.6526 (4.3323)	0.6770 (0.6136)	1.0347 (0.4179)	1.1048 (0.4269)
		(3,2)	0.4700 (0.2928)	0.4303 (0.2114)	0.2693 (0.2254)	4.6190 (4.5694)	2.0270 (1.8010)	5.3505 (3.5081)	1.2812 (1.0284)	2.8956 (1.2125)	1.6942 (0.6026)
		(0.8,0.3)	0.4952 (0.2586)	0.3133 (0.2081)	0.2850 (0.2182)	0.9739 (1.6556)	1.0367 (0.6197)	1.2060 (0.7415)	0.2015 (0.1606)	0.3177 (0.1179)	0.3079 (0.1060)
		(2,0.3)	0.4818 (0.1635)	0.3798 (0.1620)	0.3496 (0.1622)	1.8767 (0.7605)	2.3354 (1.0374)	2.2609 (0.9677)	0.1792 (0.1757)	0.3150 (0.1167)	0.3368 (0.1347)
	0.45	(10,1)	0.4573 (0.1206)	0.4467 (0.1441)	0.4130 (0.1435)	10.1397 (3.4468)	10.4304 (3.8550)	10.6740 (4.0548)	0.7441 (0.6220)	1.0313 (0.4679)	1.1358 (0.4836)
		(3,2)	0.4900 (0.2907)	0.4885 (0.2180)	0.2870 (0.2558)	4.4075 (3.3746)	1.9637 (1.7483)	5.2814 (3.4724)	1.2414 (1.0560)	3.1104 (1.4070)	1.6909 (0.6180)
		(0.8,0.3)	0.5174 (0.2430)	0.3611 (0.2197)	0.3086 (0.2424)	0.9510 (1.0251)	0.9926 (0.5763)	1.1941 (0.7109)	0.2054 (0.1656)	0.3211 (0.1289)	0.3144 (0.1160)
		(2,0.3)	0.4795 (0.1527)	0.4262 (0.1728)	0.3738 (0.1850)	2.0650 (0.8259)	2.2980 (1.0004)	2.4165 (1.2273)	0.2100 (0.1778)	0.3158 (0.1243)	0.3554 (0.1669)
	0.55	(10,1)	0.4863 (0.1502)	0.5387 (0.1456)	0.4851 (0.1789)	11.2129 (3.9174)	10.0337 (3.3989)	11.1192 (4.0967)	1.0537 (0.7872)	1.0820 (0.6652)	1.2386 (0.7405)
		(3,2)	0.4976 (0.2807)	0.5328 (0.2273)	0.2967 (0.3289)	4.1961 (2.2543)	1.9796 (1.6510)	5.5654 (3.6826)	1.2492 (1.0613)	3.3292 (1.6620)	1.7088 (0.6197)
		(0.8,0.3)	0.5209 (0.2386)	0.4067 (0.2615)	0.3247 (0.3058)	1.0052 (1.1430)	1.0145 (0.5595)	1.2960 (0.8441)	0.2240 (0.1639)	0.3326 (0.1382)	0.3352 (0.1270)
		(2,0.3)	0.5107 (0.1666)	0.5291 (0.1588)	0.4321 (0.2286)	2.2757 (0.8701)	2.1739 (0.8088)	2.5727 (1.3259)	0.2708 (0.1949)	0.3232 (0.1315)	0.4036 (0.2358)
	0.6	(10,1)	0.5136 (0.1760)	0.5782 (0.1482)	0.5260 (0.1906)	11.5982 (4.6758)	10.1415 (3.2686)	11.2830 (4.5995)	1.2071 (0.9461)	1.1155 (0.7196)	1.3164 (0.9412)
		(3,2)	0.4828 (0.3064)	0.5621 (0.2537)	0.3114 (0.3622)	4.6543 (4.3766)	2.1445 (1.6745)	5.4112 (3.3611)	1.3545 (1.0122)	3.4537 (1.7732)	1.7569 (0.6476)
		(0.8,0.3)	0.5353 (0.2436)	0.4138 (0.3018)	0.3347 (0.3407)	0.9928 (0.5692)	1.0378 (0.5713)	1.3013 (0.8330)	0.2299 (0.1673)	0.3461 (0.1496)	0.3459 (0.1410)
		(2,0.3)	0.5416 (0.1723)	0.5740 (0.1581)	0.4660 (0.2421)	2.2438 (0.8041)	2.1022 (0.6974)	2.5026 (1.1846)	0.2924 (0.2210)	0.3227 (0.1400)	0.4222 (0.2715)
	0.8	(10,1)	0.5672 (0.3003)	0.7403 (0.1718)	0.6357 (0.2873)	13.2325 (5.3936)	10.6499 (3.0394)	12.9491 (6.7095)	2.4717 (2.4027)	1.6332 (1.4367)	1.9853 (2.0799)
		(3,2)	0.5019 (0.4136)	0.6815 (0.3387)	0.3352 (0.5249)	4.9972 (5.5706)	2.3517 (1.6363)	5.7884 (3.8574)	1.4097 (1.0538)	4.3959 (2.5369)	1.8413 (0.6725)
		(0.8,0.3)	0.5356 (0.3721)	0.6248 (0.3649)	0.4092 (0.4788)	1.2209 (1.5974)	1.0463 (0.5144)	1.3741 (0.8830)	0.3091 (0.2051)	0.3521 (0.1806)	0.3886 (0.2079)
		(2,0.3)	0.5722 (0.3135)	0.7879 (0.1634)	0.5481 (0.3681)	2.7554 (3.0496)	1.9914 (0.5278)	2.7426 (1.4304)	0.5524 (0.5064)	0.3151 (0.1746)	0.5947 (0.5166)

Çizelge 5.19 MOM, LS ve ML yöntemi ile elde edilen parametre tahminleri ve RMSE değerleri (n=100)

n	α	θ_1, θ_2	Parametre(α)			Parametre(θ_1)			Parametre(θ_2)		
			MOM	LSE	MLE	MOM	LSE	MLE	MOM	LSE	MLE
100	0.4	(10,1)	0.4161 (0.1202)	0.3950 (0.0990)	0.3878 (0.0863)	10.2227 (3.2827)	10.3790 (2.8106)	10.1511 (2.1438)	0.7951 (0.6147)	1.0186 (0.2716)	1.0392 (0.2399)
		(3,2)	0.4426 (0.3216)	0.1311 (0.3192)	0.3016 (0.2353)	5.1873 (8.1030)	6.6436 (3.8600)	4.8456 (2.9139)	1.4715 (0.8937)	2.0268 (0.4634)	1.7849 (0.5149)
		(0.8,0.3)	0.5063 (0.2610)	0.3164 (0.2221)	0.3472 (0.2132)	0.8530 (0.5232)	0.7893 (0.5495)	1.0330 (0.5275)	0.2185 (0.1435)	0.3448 (0.1236)	0.2959 (0.0958)
		(2,0.3)	0.4525 (0.1515)	0.3863 (0.1272)	0.3806 (0.1107)	1.9895 (0.6476)	2.2465 (0.7779)	2.1081 (0.5309)	0.2149 (0.1648)	0.3071 (0.0806)	0.3127 (0.0806)
	0.45	(10,1)	0.4444 (0.1245)	0.4580 (0.1009)	0.4378 (0.0908)	10.5008 (3.1014)	9.9054 (2.4765)	10.1262 (2.0331)	0.9045 (0.6686)	0.9864 (0.3167)	1.0531 (0.2794)
		(3,2)	0.4597 (0.3246)	0.1446 (0.3630)	0.3216 (0.2638)	5.6133 (8.9101)	6.5766 (3.8193)	5.0017 (3.3464)	1.4602 (0.9092)	2.0228 (0.5014)	1.7660 (0.5551)
		(0.8,0.3)	0.5335 (0.2550)	0.3254 (0.2518)	0.3677 (0.2245)	0.8437 (0.3131)	0.8076 (0.5592)	1.0335 (0.5114)	0.2203 (0.1471)	0.3608 (0.1304)	0.3074 (0.0972)
		(2,0.3)	0.4813 (0.1433)	0.4473 (0.1297)	0.4283 (0.1168)	2.0221 (0.5954)	2.1395 (0.6759)	2.0974 (0.5075)	0.2320 (0.1737)	0.3017 (0.0933)	0.3177 (0.0968)
	0.55	(10,1)	0.4915 (0.1440)	0.5422 (0.0945)	0.5264 (0.0959)	11.1286 (3.1865)	10.0351 (2.2258)	10.2925 (1.9281)	1.1523 (0.8599)	1.0075 (0.3693)	1.0611 (0.3499)
		(3,2)	0.4609 (0.3292)	0.1428 (0.4551)	0.3593 (0.3098)	4.9482 (5.8862)	6.6834 (3.9064)	4.9308 (3.2019)	1.5128 (0.8885)	2.0826 (0.5470)	1.7860 (0.5937)
		(0.8,0.3)	0.5307 (0.2494)	0.4147 (0.2649)	0.3964 (0.2675)	0.9685 (1.6934)	0.8961 (0.5068)	1.0793 (0.5486)	0.2479 (0.1498)	0.3552 (0.1329)	0.3267 (0.1176)
		(2,0.3)	0.5189 (0.1535)	0.5300 (0.1221)	0.5054 (0.1347)	2.1596 (0.6283)	2.0981 (0.5375)	2.1376 (0.5283)	0.2900 (0.2030)	0.3143 (0.1051)	0.3376 (0.1296)
	0.6	(10,1)	0.5259 (0.1618)	0.5950 (0.1002)	0.5762 (0.1056)	11.2773 (3.1413)	10.0218 (2.1005)	10.3418 (1.9035)	1.3062 (1.0354)	1.0390 (0.4103)	1.1177 (0.4446)
		(3,2)	0.4693 (0.3477)	0.1608 (0.4930)	0.3710 (0.3505)	4.7091 (4.0436)	6.6691 (3.9343)	4.9936 (3.2977)	1.5205 (0.9224)	2.0962 (0.6033)	1.7398 (0.6726)
		(0.8,0.3)	0.5428 (0.2531)	0.4798 (0.2693)	0.4307 (0.2922)	0.9876 (0.8397)	0.9418 (0.4940)	1.0937 (0.5536)	0.2634 (0.1582)	0.3445 (0.1376)	0.3314 (0.1356)
		(2,0.3)	0.5477 (0.1610)	0.5804 (0.1249)	0.5500 (0.1460)	2.2033 (0.6083)	2.0775 (0.4538)	2.1517 (0.4949)	0.3179 (0.2172)	0.3172 (0.1092)	0.3482 (0.1527)
	0.8	(10,1)	0.6083 (0.2669)	0.7678 (0.1198)	0.7353 (0.1673)	12.1512 (3.6631)	10.2851 (1.7799)	10.8005 (2.3219)	2.5468 (2.3293)	1.2415 (0.9560)	1.3926 (1.2081)
		(3,2)	0.4689 (0.4647)	0.1886 (0.6708)	0.4674 (0.4531)	5.5492 (7.8306)	6.6697 (3.9898)	4.9904 (3.1741)	1.6638 (0.9073)	2.1943 (0.8081)	1.7202 (0.8026)
		(0.8,0.3)	0.5063 (0.4148)	0.6668 (0.3396)	0.4882 (0.4279)	1.2956 (3.7787)	1.0138 (0.4934)	1.2197 (0.7392)	0.3611 (0.2151)	0.3518 (0.1840)	0.3708 (0.1986)
		(2,0.3)	0.5852 (0.3076)	0.7919 (0.1117)	0.6709 (0.2519)	2.5030 (0.9819)	1.9868 (0.3277)	2.3514 (0.8764)	0.5974 (0.4930)	0.3062 (0.1544)	0.4543 (0.3688)

Simülasyon çalışması sonucunda küçük örneklem ($n = 20$ ve $n = 40$) için tahminler incelendiğinde, ortalamalar birbirinden uzaklaştıkça MOM ve LS, RMSE ölçütüne göre daha iyi tahminler vermektedir. Ortalamalar birbirine yakın olduğunda LS, diğer iki yöntemden daha güçlü iken, ML 2. bileşen ortalamalarını (θ_2) daha iyi tahmin etmiştir.

Büyük örneklem ($n = 100$) için elde edilen tahminlere bakıldığında ML'nin diğer yöntemler içinde RMSE ölçütüne göre en iyisi olduğu görülmektedir. Diğer iki yönteme bakıldığında LS, MOM'a göre daha iyi tahminler vermektedir.

Yapılan simülasyon sonucunda, RMSE ölçütüne göre karşılaştırılan üç tahmin edici yöntemi, çeşitli parametre değerlerine göre üretilen veri setleri gruplandırılarak Ki-Kare test istatistiği (χ^2) ölçütüne göre de karşılaştırılmıştır.

İlk olarak, $\theta_1 = 2, \theta_2 = 0.3$ ve $\alpha = 0.625$ olan iki bileşenli karma üstel dağılıma ilişkin üretilen 200 gözlem değeri çizelge 5.20'de verilmiştir.

Çizelge 5.20 İki bileşenli karma üstel dağılıma ilişkin üretilen veri seti ($\theta_1 = 2, \theta_2 = 0.3$ $\alpha = 0.625$ ve $n=200$)

1,4284	0,0743	0,0548	0,8096	0,7858	0,9552	0,051	4,2071	2,7108	3,7647
4,2131	2,1437	0,7531	6,1499	0,6565	5,2991	1,2189	1,1571	0,578	0,8714
3,9505	2,0297	1,1602	0,125	0,2466	1,6125	0,1299	1,0445	0,8101	0,5574
0,0572	1,5761	0,4856	1,0254	0,2114	3,1629	1,7963	0,6663	0,2192	0,7282
2,2415	0,6743	0,7867	1,6519	0,4099	2,803	1,958	1,1528	1,8419	0,8385
0,3682	5,6113	3,617	2,3965	1,2327	3,8413	2,2132	1,4193	0,2382	1,6594
2,7051	0,9794	1,5798	0,0691	1,1715	1,1523	1,5784	0,5084	1,8836	1,5558
2,9508	1,1186	0,7975	5,7657	3,7612	1,664	0,2927	0,1929	1,3194	1,1619
0,5314	4,5902	1,0702	0,1268	1,6391	0,3494	1,1816	1,0932	1,6246	1,8257
2,0408	2,6931	1,6907	2,4243	4,2538	0,6951	0,484	1,8634	3,0221	3,3301
2,633	3,9516	0,2084	1,9563	2,657	1,6587	7,2911	0,7712	3,0455	0,231
0,9931	2,2018	0,6668	2,1781	2,2669	1,2009	1,1858	2,8418	3,7395	2,0412
13,4535	0,7144	2,4936	0,5448	1,3688	4,7687	3,2373	0,3123	1,185	0,4393
0,7085	0,2756	0,0689	0,8713	0,0391	0,0668	0,1004	0,2509	0,0744	0,6041
0,3749	0,0329	0,1798	0,1651	0,1955	0,0491	0,0206	0,4728	0,7891	0,0629
0,0438	0,0271	0,0813	0,3766	0,1227	0,2825	0,1485	0,0982	0,0976	0,0455
0,1682	0,1774	0,279	0,06	0,0214	0,1566	0,1894	0,109	0,3176	0,2557
0,0938	0,253	0,0169	0,1046	1,2473	0,765	0,2078	0,1447	0,0261	0,2049
0,7468	0,1979	0,1265	0,3006	0,2698	0,2722	0,4526	0,2528	0,0814	0,1315
0,4435	0,2731	0,0286	0,2447	1,0217	0,02	0,0951	0,1832	0,1655	0,1313

Çizelge 5.20'de verilen veri setine ilişkin MOM, ML ve LS tahmin edicileri ile bu değerler kullanılarak hesaplanan χ^2 değerleri ise çizelge 5.21'de gösterilmektedir.

Çizelge 5.21 MOM, ML ve LS yöntemi ile elde edilen parametre tahminleri ve (χ^2) değerleri ($\theta_1 = 2, \theta_2 = 0.3, \alpha = 0.625$ ve $n=200$)

MOM			MLE			LSE		
$\hat{\theta}_1$	$\hat{\theta}_2$	$\hat{\alpha}$	$\hat{\theta}_1$	$\hat{\theta}_2$	$\hat{\alpha}$	$\hat{\theta}_1$	$\hat{\theta}_2$	$\hat{\alpha}$
2,2510	0,3389	0,77	1,6187	0,1820	0,7586	1,6046	0,1693	0,7675
$\chi^2 = 2,91$			$\chi^2 = 0,51$			$\chi^2 = 1,85$		

Çizelge 5.21’de elde edilen χ^2 değerleri $\alpha = 0,05$ anlamlılık düzeyindeki tablo değeri $\chi^2_{0,95;2} = 5,99$ ile karşılaştırıldığında, bulunan her üç χ^2 değeri tablo değerinden küçük olduğundan bu değerlerin anlamlı olduğu, yani ele alınan veri setinin 2MED’e uygun olduğu sonucuna ulaşılmaktadır. Ayrıca her üç tahmin edici yöntemi kendi aralarında χ^2 değerlerine göre karşılaştırıldığında ise, $\chi^2_{MLE} = 0,51 < \chi^2_{LSE} = 1,85 < \chi^2_{MOM} = 2,91$ olduğundan, ML’nin MOM ve LS’ye göre daha etkili olduğu söylenebilir.

$\theta_1 = 10, \theta_2 = 2$ ve $\alpha = 0.625$ olan iki bileşenli karma üstel dağılıma ilişkin üretilen 200 gözlem değeri çizelge 5.22’de verilmiştir.

Çizelge 5.22 İki bileşenli karma üstel dağılıma ilişkin üretilen veri seti ($\theta_1 = 10, \theta_2 = 2, \alpha = 0.625$ ve $n=200$)

57,7936	54,4102	51,7556	47,2001	37,2856	35,1769	28,6799	26,9736	25,6422	24,6924
24,6545	20,7593	20,2031	19,7963	19,5662	19,4954	18,5346	17,4781	17,416	17,2713
17,0463	16,9539	15,4519	15,0645	14,8439	14,7199	13,7039	13,4428	13,2689	12,7917
12,7358	12,7214	12,6161	12,4706	12,3965	12,0793	12,0743	11,69	11,2075	10,9849
9,6924	9,3166	9,1813	9,1325	8,9205	8,8245	8,6511	8,5782	8,4639	8,3601
8,2554	8,2056	8,1415	7,7276	7,1542	6,9695	6,9181	6,4964	6,3973	6,1368
6,1066	6,1056	6,0804	6,0534	5,8039	5,6909	5,6388	5,6368	5,4594	5,3367
5,3176	5,2185	5,1474	5,0015	4,9623	4,9451	4,9227	4,8382	4,734	4,6998
4,6984	4,4212	4,3008	4,2239	4,185	4,0394	4,0309	3,9874	3,9806	3,9251
3,7583	3,6285	3,5307	3,4687	3,453	3,3653	3,327	3,2544	3,2205	3,1824
3,0788	3,0039	2,9934	2,9278	2,912	2,8511	2,8342	2,7728	2,5542	2,418
2,4054	2,2697	2,2053	2,1408	2,14	2,1078	2,079	2,0336	1,9792	1,9724
1,9374	1,9267	1,9079	1,8898	1,8029	1,7927	1,7497	1,7373	1,7352	1,6723
1,6566	1,6549	1,6281	1,5953	1,5472	1,5143	1,5139	1,513	1,4986	1,4874
1,4587	1,4281	1,3872	1,3787	1,3774	1,3671	1,3568	1,3257	1,2927	1,2921
1,2797	1,2658	1,2536	1,2077	1,204	1,1894	1,176	1,1693	1,1648	1,1195
1,1179	1,1149	1,075	1,0698	1,063	1,017	0,9971	0,9961	0,9927	0,9224
0,8776	0,8659	0,7986	0,7346	0,7228	0,6801	0,6724	0,6049	0,5935	0,5794
0,5363	0,516	0,4483	0,4081	0,4021	0,3694	0,3646	0,3589	0,3172	0,3125
0,2661	0,2608	0,2398	0,2134	0,2112	0,2055	0,1341	0,128	0,0829	0,0825

Çizelge 5.22’de verilen veri setine ilişkin MOM, ML ve LS tahmin edicileri ile bu değerler kullanılarak hesaplanan χ^2 değerleri ise çizelge 5.23’te gösterilmektedir.

Çizelge 5.23 MOM, ML ve LS yöntemi ile elde edilen parametre tahminleri ve (χ^2) değerleri ($\theta_1 = 10, \theta_2 = 2, \alpha = 0.625$ ve $n=200$)

MOM			MLE			LSE		
$\hat{\theta}_1$	$\hat{\theta}_2$	$\hat{\alpha}$	$\hat{\theta}_1$	$\hat{\theta}_2$	$\hat{\alpha}$	$\hat{\theta}_1$	$\hat{\theta}_2$	$\hat{\alpha}$
11,8824	2,4036	0,4484	11,8988	2,6428	0,4333	12,3744	2,6664	0,4162
$\chi^2 = 0,43$			$\chi^2 = 0,74$			$\chi^2 = 0,60$		

Buradan, tahmin edici yöntemleri için çizelge 5.23'te elde edilen χ^2 değerleri $\alpha = 0,05$ anlamlılık düzeyindeki tablo değeri olan $\chi^2_{0,95;2} = 5,99$ ile karşılaştırıldığında, yukarıdaki örneğe benzer olarak bulunan her üç χ^2 değeri tablo değerinden küçük olduğundan bu değerlerin anlamlı olduğu, dolayısıyla ele alınan veri setinin 2MED'e uygun olduğu sonucuna ulaşılmaktadır. Ayrıca her üç tahmin edici yöntemi kendi aralarında χ^2 değerlerine göre karşılaştırıldığında ise, $\chi^2_{MOM} = 0,43 < \chi^2_{LSE} = 0,60 < \chi^2_{MLE} = 0,74$ olduğundan, MOM, LS ve ML'ye göre parametreleri tahmin etmede daha etkindir.

Son olarak $\theta_1 = 1, \theta_2 = 3$ ve $\alpha = 0.4$ olan iki bileşenli karma üstel dağılıma ilişkin üretilen 200 gözlem değeri çizelge 5.24'te verilmiştir.

Çizelge 5.24 İki bileşenli karma üstel dağılıma ilişkin üretilen veri seti ($\theta_1 = 1, \theta_2 = 3$ $\alpha = 0.4$ ve $n=200$)

1,4254	0,1912	0,2545	2,2131	0,6076	0,3275	0,9833	0,0504	0,183	0,3808
6,9083	0,4052	2,123	0,7055	0,1005	0,2091	0,6325	0,8258	0,7371	0,5693
0,5634	1,1416	0,6587	0,2072	0,1228	1,3749	1,3463	0,3192	0,3885	0,2217
0,096	2,5576	0,4742	0,3616	0,2908	1,1424	2,116	2,9219	0,1126	0,5205
0,823	0,1322	0,7845	0,3035	0,0468	0,2758	0,7964	5,0982	0,828	1,2349
0,0654	0,2894	0,5472	2,8947	0,9267	0,7167	0,024	6,3203	3,2436	1,4191
0,2255	2,9133	2,1754	1,3976	0,496	0,0097	0,469	1,0306	0,1618	1,2903
0,2208	3,363	2,0924	2,8304	0,5257	4,609	1,613	2,0198	1,286	3,5924
0,396	0,4377	2,0141	7,9683	0,2413	2,5217	1,8128	2,1003	1,9991	2,3568
4,6364	1,9061	2,009	4,6354	5,8397	4,5471	1,3865	6,1771	2,1493	0,0733
4,2456	0,6189	18,3063	2,9121	4,2408	0,0553	3,7822	3,0806	7,4313	0,9081
5,7764	1,338	0,7335	2,2893	0,3658	0,2313	0,2725	2,831	0,604	7,3954
8,7787	6,3896	2,9174	8,9955	1,4328	0,069	0,6467	7,192	5,0238	3,1066
0,7047	0,1665	1,4589	5,8029	1,9771	4,4546	0,6273	0,9931	6,9491	1,3515
2,0987	1,7349	8,0742	3,7285	2,9965	0,1914	3,7886	4,7749	8,7661	2,443
1,7029	2,9716	0,1938	4,2701	3,0925	6,6205	2,5278	4,0277	6,074	1,2822
3,5322	0,3691	4,258	3,7041	2,7697	1,7128	3,9301	5,0838	10,2673	5,3293
1,8304	0,3308	2,4443	6,0761	11,5557	3,7336	4,7893	3,5682	0,3295	0,4256
1,511	14,6756	0,1665	6,5338	1,4695	0,7412	0,0213	2,1916	0,6325	2,0138
3,3079	0,8591	1,8606	0,9737	5,164	0,1229	0,6375	0,3634	12,8504	0,1501

Çizelge 5.24’te verilen veri setine ilişkin MOM, ML ve LS tahmin edicileri ile bu değerler kullanılarak hesaplanan χ^2 değerleri ise çizelge 5.25’de gösterilmektedir.

Çizelge 5.25 MOM, ML ve LS yöntemi ile elde edilen parametre tahminleri ve (χ^2) değerleri ($\theta_1 = 1, \theta_2 = 3, \alpha = 0.4$ ve $n=200$)

MOM			MLE			LSE		
$\hat{\theta}_1$	$\hat{\theta}_2$	$\hat{\alpha}$	$\hat{\theta}_1$	$\hat{\theta}_2$	$\hat{\alpha}$	$\hat{\theta}_1$	$\hat{\theta}_2$	$\hat{\alpha}$
0,5186	2,9349	0,1983	0,4163	2,9313	0,1890	0,3853	2,9796	0,1887
$\chi^2 = 0,91$			$\chi^2 = 0,91$			$\chi^2 = 0,83$		

Buradan, tahmin edici yöntemleri için çizelge 5.25’te elde edilen χ^2 değerleri $\alpha = 0,05$ anlamlılık düzeyindeki tablo değeri olan $\chi^2_{0,95;2} = 5,99$ ile karşılaştırıldığında, yapılan diğer iki örnekte olduğu gibi bulunan her üç χ^2 değeri tablo değerinden küçük olduğundan yukarıda belirtilen parametre değerlerinden üretilen veri setinin de 2MED’e uygun olduğu anlaşılmaktadır. Ayrıca her üç tahmin edici yöntemi kendi aralarında χ^2 değerlerine göre karşılaştırıldığında ise, $\chi^2_{LSE} = 0,83 < \chi^2_{MOM} = 0,91 = \chi^2_{MLE} = 0,91$ olduğundan, LS’nin MOM ve ML’ye göre parametreleri tahmin etmede daha etkin olduğu görülmektedir.

6. UYGULAMA ÇALIŞMASI

Bu bölümde, bir önceki bölümde yapılan simülasyon çalışmasına ek olarak, hem literatürde yer alan uygun veri setleri hem de Tekirdağ ili üniversite hastanesinde Beyin Cerrahisi ve Kardiyoloji Bölümünde yatış yapan hastaların cinsiyet, yaş ve hastanede kalma sürelerini içeren gerçek bir veri seti kullanılarak modelleme çalışması yapılmıştır (Bu çalışmanın sonuçları, Ankara Üniversitesi Bilimsel Araştırma Projesi (BAP) kapsamında yayınlanmıştır). Giriş bölümündeki literatür taramasında bahsedilen veri türleri incelendiğinde, genellikle verinin yaşam zamanlarına ilişkin olduğu ve bu yaşam zamanlarının modellenmesinde sürekli bir dağılım ile kesikli veya sürekli bir dağılım bileşiminin (compound) kullanıldığı göze çarpmaktadır. Hâlbuki bu çalışmada kullanılan 2MED'in yapısının hem daha basit hem de uygulanabilirliğinin birçok bileşik dağılıma göre daha kolay olması, bu dağılımın modelleme çalışmasındaki başarısını merak konusu haline getirmiştir. Bu amaçla aşağıda literatür taramasında karşılan bazı çalışmalara ilişkin veri setleri ele alınarak 2MED ile modellenmeye çalışılmıştır (Yılmaz ve Buyum 2014, Yılmaz vd. 2014, Yılmaz ve Buyum 2015).

Veri Seti 1. Proschan (1963) tarafından elde edilen ve daha sonra birçok araştırmacının çalışmalarına konu olan 13 adet Boeing 720 uçağına ait havalandırma sistemlerinde meydana gelen bozulma zamanlarına ilişkin veri setinin modellenmesiyle sıkça karşılaşılmıştır. Proschan (1963) yaptığı çalışmada veri setini tek bileşenli üstel dağılım ile modellemiştir. Yapılan mevcut çalışmada ise, bu veri seti 2MED ile modellenmeye çalışılmıştır. Modelleme sonucunda elde edilen parametre tahmin değerleri, p ve KS değerleri çizelge 6.1'de gösterilmektedir.

Çizelge 6.1 Veri seti 1 için ML parametre tahminleri, p ve KS değerleri

MLE		
$\hat{\alpha}$	$\hat{\theta}_1$	$\hat{\theta}_2$
0.4315	46.6148	128.4549
KS istatistiğı		p değeri
0.0482		0.6881

Proschan (1963) tarafından üstel dağılım ile modellenen bu veri setinin 2MED ile modellenmesi sonucu elde edilen p ve KS değerleri göz önüne alındığında, tek bileşenli üstel dağılımın yerine 2MED'in model olarak önerilmesinin daha uygun olduğu söylenebilir.

Veri Seti 2. İlk olarak Maguire vd. (1952) tarafından elde edilen ve ilk kez Cox ve Lewis (1966) tarafından analiz edilen maden verisidir. Bu veri seti, 1875'den 1951 yılına kadar gerçekleşen, 10 ve daha fazla insanın ölümüyle sonuçlanan maden kazalarının gerçekleşme zamanlarını (gün olarak) içermektedir. Bir önceki veri seti gibi bu veri seti de birçok modelleme çalışmasında kullanılmıştır. Burada diğer araştırmacıların elde ettiği sonuçlar ile bu çalışmadan elde edilen sonuçlar karşılaştırıldığında önerilen dağılımın veri setini modellemede başarılı olduğu söylenebilir. Bu maden veri seti için elde edilen parametre tahmin değerleri, p ve KS değerleri çizelge 6.2'de gösterilmektedir.

Çizelge 6.2 Veri seti 2 için ML parametre tahminleri, p ve KS değerleri

MLE		
$\hat{\alpha}$	$\hat{\theta}_1$	$\hat{\theta}_2$
0.1757	592.0210	166.1348
KS istatistiği		p değeri
0.0578		0.8386

Yapılan modelleme sonucunda maden kazalarının, ML tahminlerine göre % 82 olasılıkla 1 yıldan daha kısa sürede gerçekleştiği sonucuna ulaşılmıştır.

Veri Seti 3. Kuş (2007) çalışmasında ele alınan, Kuzey Anadolu fay hattında gerçekleşen ve şiddeti 6 ve 6'dan daha büyük olan depremlerin gerçekleşme zamanlarına ait veri seti 2MED ile modellenmiştir ve elde edilen tahmin değerleri çizelge 6.3'de verilmektedir.

Çizelge 6.3 Veri seti 3 için ML parametre tahminleri, p ve KS değerleri

MLE		
$\hat{\alpha}$	$\hat{\theta}_1$	$\hat{\theta}_2$
0.5	2549.4	506.13
KS istatistiği		p değeri
0.0762		0.9971

Çizelge 6.3’de elde edilen tahminler incelendiğinde, Kuzey Anadolu fay hattında oluşan depremlerin gerçekleşme zamanlarının 2MED’e uygun olduğu görülmektedir. Ayrıca ML tahminlerine göre, gerçekleşme zamanlarının 0,5 karma oranı ile ortalama 7 yıl ile 0,5 karma oranı ile ortalama 1,5 yıl olan farklı iki üstel dağılımdan geldiği söylenebilir.

Veri Seti 4. Bir yazılım sistemi için ele alınan merkezi işlem birimlerinde meydana gelen bozulmalar arası geçen süreleri içermektedir (Dekking vd. 2010). Yapılan inceleme sonucunda veri setini modellemede 2MED’in kullanılabileceği sonucuna ulaşılmıştır. Yapılan analiz sonucu çizelge 6.4’de verilmektedir.

Çizelge 6.4 Veri seti 4 için ML parametre tahminleri, p ve KS değerleri

MLE		
$\hat{\alpha}$	$\hat{\theta}_1$	$\hat{\theta}_2$
0.3548	95.8706	965.3713
KS istatistiği		p değeri
0.0622		0.6499

Çizelge 6.4 incelendiğinde, örneğin 100 adet merkezi işlem birimi ele alındığında bunlardan yaklaşık olarak 35 tanesinde bozulmalar arası geçen sürenin yaklaşık 96 saat, geriye kalan 65 tanesinde ise bu sürenin 965 saat olduğu yorumu yapılabilir. Çizelge 6.4’deki sonuçlara bakıldığında, ele alınan bozulmalar arası geçen sürelerin modellenmesinde 2MED’in başarılı olduğu sonucu ortaya çıkmaktadır.

Veri Seti 5. Iraktaki bir dokuma fabrikasından elde edilen iki farklı makineye ait bozulma zamanlarını (saat olarak) ele alan Al-Jammal (2008), çalışmasında veri setini ölçek parametresi λ , şekil parametresi γ olan üstelleştirilmiş (exponentiated) üstel dağılım ile modellemiştir. Veri setinin 2MED ile modellenmesi sonucunda elde edilen tahminler Makine 1 ve 2 için ayrı ayrı çizelge 6.5-6.6’da gösterilmiştir.

Çizelge 6.5 Veri seti 5 için ML parametre tahminleri, p ve KS değerleri (Makine 1)

MLE		
$\hat{\alpha}$	$\hat{\theta}_1$	$\hat{\theta}_2$
0.4116	247.4955	25.3119
KS istatistiği		p değeri
0.1119		0.8071

Çizelge 6.6 Veri seti 5 için ML parametre tahminleri, p ve KS değerleri (Makine 2)

MLE		
$\hat{\alpha}$	$\hat{\theta}_1$	$\hat{\theta}_2$
0.3391	321.1119	74.5066
KS istatistiği		p değeri
0.1068		0.9305

Al-Jammal (2008) çalışmasında, veri setini üstelleştirilmiş üstel dağılım ile modellenmesi sonucu elde ettiği KS değerlerini Makine 1 için $KS = 0.1906$ ve Makine 2 için $KS = 0.1317$ olarak elde etmiştir. Çizelge 6.5-6.6 ile verilen değerlere bakıldığında her iki makine için bulunan KS değerlerinin üstelleştirilmiş üstel dağılım için bulunanlardan daha düşük olduğu görülmektedir. Yani KS ölçütüne göre bu veri setini modellemede 2MED’in üstelleştirilmiş (exponentiated) üstel dağılım yerine tercih edilmesinin daha uygun olduğu söylenebilir.

Veri Seti 6. Kumar ve Klefsjö (1992) yaptıkları çalışmada, LHD (Load-Haul-Dump) makinelerinin bozulma zamanları arasında geçen süreyi (saat olarak) içeren veri setini ele almışlardır. Burada aynı veri setinin 2MED ile modellenmesi sonucunda ulaşılan sonuçlar

çizelge 6.7’de gösterilmektedir.

Çizelge 6.7 Veri seti 6 için ML parametre tahminleri, p ve KS değerleri

MLE		
$\hat{\alpha}$	$\hat{\theta}_1$	$\hat{\theta}_2$
0.0987	23.7290	143.4531
KS istatistiği		p değeri
0.0358		0.9861

Bu veri setinde 151 adet LHD makinesine ilişkin bozulma zamanları bulunmaktadır. Buna göre çizelge 6.7, ML sonuçlarına göre yorumlandığında yaklaşık 15 tanesinde bozulma zamanları arasında geçen süre 20 saat üzerinde iken geriye kalan 136 tanesi için bu süre 100 saatin üzerinde gerçekleşmektedir. Buradan bozulma zamanları birbirinden oldukça farklı olan iki grup olduğu sonucuna ulaşılır ki bu da ele alınan veri setinin modellenmesinde KS istatistiğine göre 2MED’in oldukça iyi sonuç verdiğini gösterir.

Veri Seti 7. Nelson (1990), çalışmalarında yedi farklı voltaj seviyesinde yalıtım sıvısının dökülme zamanlarını ele almıştır. Burada dökülme zamanları 2MED ile modellenmeye çalışılmış ve elde edilen parametre tahmin değerleri ile p ve KS değerleri çizelge 6.8’de gösterilmektedir.

Çizelge 6.8 Veri seti 7 için ML parametre tahminleri, p ve KS değerleri

MLE		
$\hat{\alpha}$	$\hat{\theta}_1$	$\hat{\theta}_2$
0.3192	295.7731	6.1010
KS istatistiği		p değeri
0.1089		0.3058

ML tahminlerine göre elde edilen ortalama değerlerine bakıldığında dökülme zamanları birbirinden farklı olan iki grup ortaya çıkmıştır. Bu gruplardan birincisinde dökülme zamanlarının ikinci gruptaki zamanlardan oldukça yüksek olduğu görülmektedir.

Veri Seti 8. Blischke (1995) tarafından ele alınan elektronik parçalara ait bozulma zamanları 2MED ile modellendiğinde elde edilen parametre tahmin değerleri ile p ve KS değerleri çizelge 6.9’da verilmektedir.

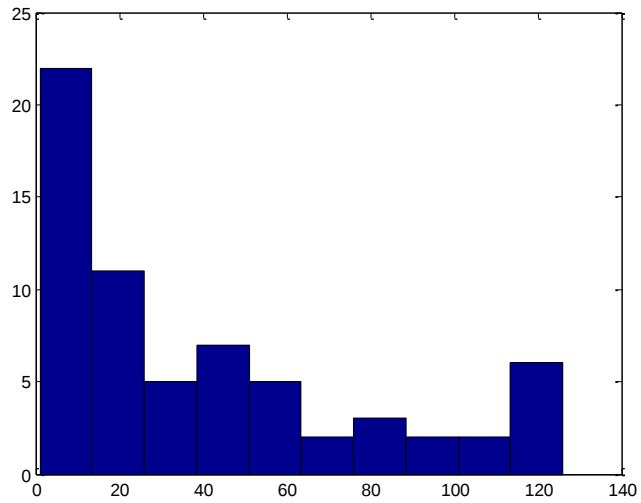
Çizelge 6.9 Veri seti 8 için ML parametre tahminleri, p ve KS değerleri

MLE		
$\hat{\alpha}$	$\hat{\theta}_1$	$\hat{\theta}_2$
0.8507	3.7577	0.1541
KS istatistiği		p değeri
0.0447		0.9512

Örneğin 100 adet elektronik parça ele alınırsa, ML sonuçlarına göre bu parçalardan yaklaşık 85 tanesinin 4 günde bir geri kalanlarının ise 1 günden daha kısa sürede bozulacağı tahmin edilmektedir.

Veri Seti 9. Bu bölümde, literatürde bulunan çalışmalara ek olarak Tekirdağ ili üniversite hastanesinin Beyin Cerrahisi birimine giriş yapan 500 hastaya ve Kardiyoloji Biriminde giriş yapan 410 hastaya ilişkin hastanede kalma süreleri ve cinsiyet bilgilerinin yer aldığı veri seti incelenmiştir. Çeşitli düzenlemeler yapıldıktan sonra hastaların birim ve cinsiyet bazında hastanede kalma sürelerinin karma üstel dağılıma uygun olduğu anlaşılmıştır. Ayrıca her iki birimde yatan kadın ve erkek hastaların hastanede kalma süreleri, hem birimler hem de cinsiyet değişkenleri için hangi dağılımın modellemede daha başarılı olduğunu göstermek amacıyla tek bileşenli üstel ve iki bileşenli karma üstel dağılım için ayrı ayrı modellenmiştir. Burada sadece her iki klinikte yatan hastaların hastanede kalma süreleri için elde edilen sonuçlara göre yorumlar yapılmıştır.

Öncelikle Beyin Cerrahisi Kliniği’ndeki hastalar cinsiyetlerine göre ardından cinsiyet ayrımı yapılmaksızın ele alınarak incelenecektir. İlk olarak Beyin Cerrahisi Kliniği’nde yatan kadın hastaların hastanede kalma sürelerine ilişkin histogram şekil 6.1’de verilmiştir.



Şekil 6.1 Beyin Cerrahisi Kliniği'nde yatan kadın hastaların hastanede kalma sürelerine ilişkin histogram

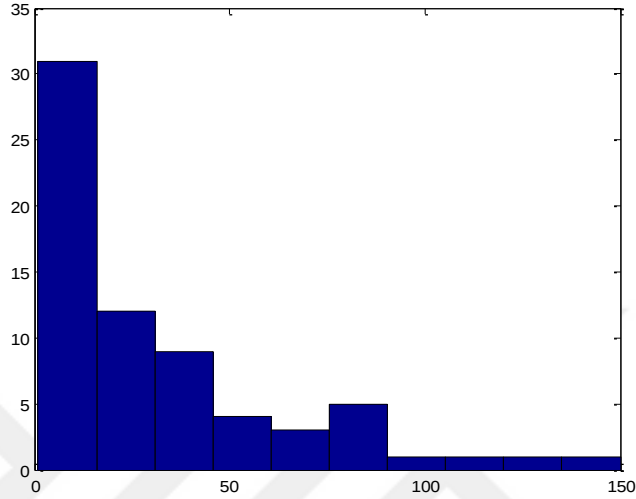
Şekil 6.1 incelendiğinde dağılımın yapısının üstel dağılıma uygun olduğu görülmektedir. Buna göre yapılan analiz sonucunda elde edilen parametre tahminleri, KS ve p değerleri çizelge 6.10'da gösterilmektedir.

Çizelge 6.10 Beyin Cerrahisi Kliniği'nde yatan kadın hastalar ele alındığında modelleme sonucu elde edilen ML parametre tahminleri, p ve KS değerleri

MLE		
$\hat{\alpha}$	$\hat{\theta}_1$	$\hat{\theta}_2$
0.0978	10.0396	43.4688
KS istatistiği		p değeri
0.0717		0.8685

Beyin cerrahisi kliniğinde yatan kadın hastalar için modelleme sonucunda elde edilen parametre tahmin değerlerine bakıldığında, 10 gün civarında hastanede kalanların 40 günden daha fazla hastanede kalanlardan daha az olduğu görülmektedir. Aynı veri seti tek bileşenli üstel dağılım ile modellendiğinde $\hat{\theta} = 40.2$, $p = 0.5143$ ve $KS = 0.0991$ bulunmuştur. Tek bileşenli üstel dağılım için bakıldığında ise 10 gün civarında hastanede kalan hasta grubu hiç göz önüne alınmamakta sadece Beyin Cerrahisi biriminde yatan kadın hastaların ortalama 40 gün hastanede kaldıkları yorumuna ulaşılmaktadır.

Kadın hastalara ilişkin incelemeden sonra beyin cerrahisi kliniğinde yatan erkek hastaların hastanede kalma süreleri incelenmiştir. Bu incelemeye ilişkin histogram şekil 6.2’de verilmiştir.



Şekil 6.2 Beyin Cerrahisi Kliniği’nde yatan erkek hastaların hastanede kalma sürelerine ilişkin histogram

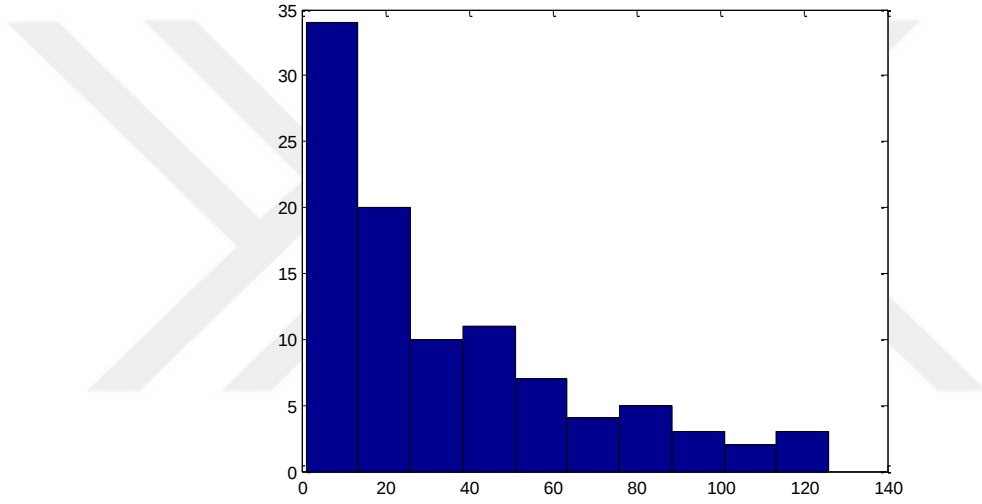
Şekil 6.2 ile gösterilen dağılıma bakıldığında erkek hastaların hastanede kalma sürelerinin iki farklı ortalamaya sahip karma üstel dağılıma uygun olduğu görülmektedir. Şekil 6.2’de verilen histogramın bir noktaya kadar azaldıktan sonra ani bir çıkış yapıp tekrar azalmaya başladığı görülmektedir. Bu durum, veri setinin 2MED ile modellendiğinde ortalama hastanede kalma sürelerinin birbirinden uzak olacağına ilişkin bir işarettir. 2MED ile modelleme sonucunda elde edilen parametre tahminleri, KS ve p değerleri çizelge 6.11’de gösterilmektedir.

Çizelge 6.11 Beyin Cerrahisi Kliniği’nde yatan erkek hastalar ele alındığında modelleme sonucu elde edilen ML parametre tahminleri, p ve KS değerleri

MLE		
$\hat{\alpha}$	$\hat{\theta}_1$	$\hat{\theta}_2$
0.2325	8.8409	37.4051
KS istatistiği		p değeri
0.0518		0.9888

Beyin Cerrahisi Kliniği'nde yatan erkek hastalar için bakıldığında, ML sonuçlarına göre ise yaklaşık 9 gün ila 37 gün hastanede kalan iki hasta grubu olduğu sonucuna ulaşılmaktadır. Aynı veri seti tek bileşenli üstel dağılım ile modellendiğinde $\hat{\theta} = 30.76$, $p = 0.4312$ ve $KS = 0.1035$ bulunmuştur. Üstel dağılım sonucuna göre erkek hastaların bu birimde kalma süreleri ortalama olarak 31 gündür.

Son olarak, beyin cerrahisi kliniğinde yatan tüm hastaların (cinsiyet ayrımı yapılmaksızın) hastanede kalma sürelerine ilişkin histogram şekil 6.3'de verilmiştir.



Şekil 6.3 Beyin Cerrahisi Kliniği'nde yatan tüm hastaların hastanede kalma sürelerine ilişkin histogram

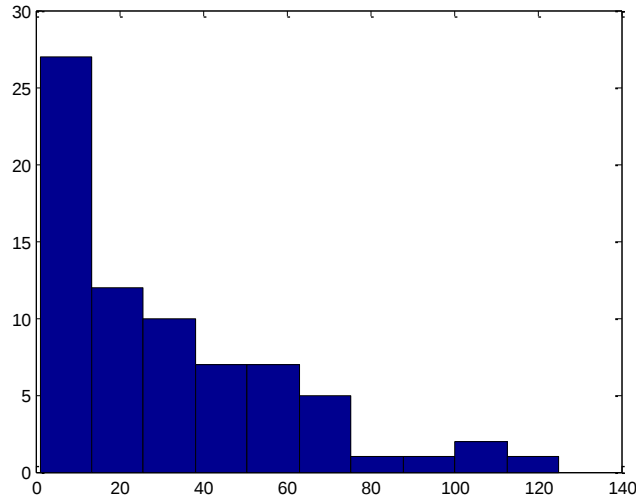
Modelleme sonucu elde edilen parametre tahminleri, KS ve p değerleri çizelge 6.12'de gösterilmektedir.

Çizelge 6.12 Beyin Cerrahisi Kliniği'nde yatan hastalar ele alındığında modelleme sonucu elde edilen ML parametre tahminleri, p ve KS değerleri

MLE		
$\hat{\alpha}$	$\hat{\theta}_1$	$\hat{\theta}_2$
0.0297	4.2594	34.8066
KS istatistiği		p değeri
0.0483		0.9663

ML sonuçlarına göre Beyin Cerrahisi biriminde yatan hastalara (cinsiyet ayrımı yapılmaksızın) ilişkin modelleme sonuçları incelendiğinde, biri 4 gün ortalama ile diğeri 35 gün ortalama ile iki farklı grup olduğu görülmektedir. Örneğin, bu birime 100 hasta geldiğinde yaklaşık 3 tanesi 4 gün ortalama ile hastanede kalırken, geri kalan 97 tanesi ise 30 günün üzerinde hastanede kalmaktadır. Aynı veri seti tek bileşenli üstel dağılım ile modellendiğinde $\hat{\theta} = 33.57$, $p = 0.8107$ ve $KS = 0.0625$ bulunmuştur. Yani üstel dağılıma göre yukarıda bahsedilen 2. grupta olduğu gibi hastalar ortalama olarak 30 günün üzerinde bu birimde kalmaktadır. Burada tek bileşenli üstel ve 2MED ile bulunan KS değerleri incelendiğinde, 2MED ile bulunan KS değerinin daha düşük olduğu görülmektedir. Bu da Beyin Cerrahisi kliniğinde yatan hastaların hastanede kalma sürelerinin modellenmesinde 2MED'in daha iyi olduğu sonucu ortaya çıkmaktadır.

Beyin cerrahisi kliniğindeki incelemenin ardından kardiyoloji kliniğindeki hastalar öncelikle cinsiyete göre ele alınmış daha sonra cinsiyet ayrımı yapılmaksızın genel olarak incelenmiştir. İlk olarak Kardiyoloji kliniğinde yatan kadın hastaların hastanede kalma süreleri ele alınmıştır ve histogram şekil 6.4'de verilmiştir.



Şekil 6.4 Kardiyoloji Kliniği'nde yatan kadın hastaların hastanede kalma sürelerine ilişkin histogram

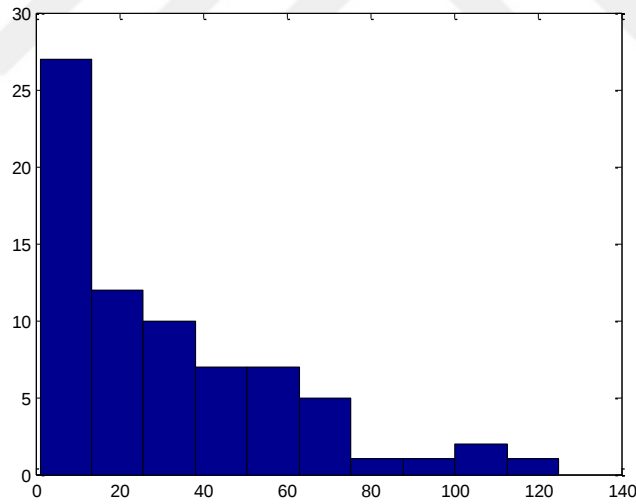
Elde edilen parametre tahminleri, KS ve p değerleri çizelge 6.13'de gösterilmektedir.

Çizelge 6.13 Kardiyoloji Kliniği'nde yatan kadın hastalar ele alındığında modelleme sonucu elde edilen ML parametre tahminleri, p ve KS değerleri

MLE		
$\hat{\alpha}$	$\hat{\theta}_1$	$\hat{\theta}_2$
0.0710	1.3039	33.8266
KS istatistiği		p değeri
0.0866		0.6124

Çizelge 6.13 incelendiğinde, ortalama 1 gün kalan hasta sayısından ziyade 30 günden fazla hastanede kalanların sayısının daha çok olduğu görülmektedir. Aynı veri seti tek bileşenli üstel dağılım ile modellendiğinde $\hat{\theta} = 31.52$, $p = 0.5357$ ve $KS = 0.0921$ bulunmuştur.

Kardiyoloji kliniğinde yatan erkek hastaların hastanede kalma sürelerine ilişkin histogram şekil 6.5'de verilmiştir.



Şekil 6.5 Kardiyoloji Kliniği'nde yatan erkek hastaların hastanede kalma sürelerine ilişkin histogram

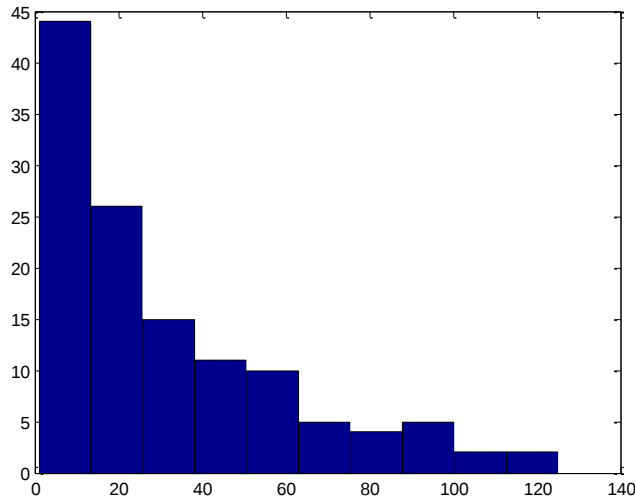
Burada histograma bakıldığında ilk olarak, Kardiyoloji Kliniği'nde yatan kadın ve erkek hastalara ilişkin histogramların birbirine benzer bir yapıda olduğu göze çarpmaktadır. Bu durum yapılacak yorumlamaların yani tahmin edilen ortalamaların erkek hastalar için elde edilen ortalama değerlerine yakın olacağına işaret etmektedir. Buna göre karma üstel dağılım ile elde edilen parametre tahmin, KS ve p değerleri aşağıda gösterilmektedir.

Çizelge 6.14 Kardiyoloji Kliniği'nde yatan erkek hastalar ele alındığında modelleme sonucu elde edilen ML parametre tahminleri, p ve KS değerleri

MLE		
$\hat{\alpha}$	$\hat{\theta}_1$	$\hat{\theta}_2$
0.1352	1.7295	37.8404
KS istatistiği		p değeri
0.0819		0.7041

Kardiyoloji biriminde yatan kadın hastaların sonucunda olduğu gibi erkek hastalar için 1-2 gün arasında kalan hasta sayısı, 30 günden daha fazla bu birimde kalan hasta sayısından daha düşüktür. Aynı veri seti tek bileşenli üstel dağılım ile modellendiğinde $\hat{\theta} = 32.96$, $p = 0.3095$ ve $KS = 0.1130$ bulunmuştur. Buna göre, üstel dağılım ile veri seti modellendiğinde erkek hastaların ortalama olarak 33 gün bu birimde kaldığı sonucuna ulaşılır.

Kardiyoloji Kliniği için son olarak hastaların (cinsiyet ayrımı yapılmaksızın) hastanede kalma sürelerine ilişkin histogram şekil 6.6'da verilmiştir.



Şekil 6.6 Kardiyoloji Kliniği'nde yatan tüm hastaların hastanede kalma sürelerine ilişkin histogram

Elde edilen parametre tahminleri, KS ve p değerleri çizelge 6.15'de gösterilmektedir.

Çizelge 6.15 Kardiyoloji Kliniği'nde yatan hastalar ele alındığında modelleme sonucu elde edilen ML parametre tahminleri, p ve KS değerleri

MLE		
$\hat{\alpha}$	$\hat{\theta}_1$	$\hat{\theta}_2$
0.0892	1.2096	34.2724
KS istatistiği		p değeri
0.0828		0.3435

Aynı şekilde Kardiyoloji Kliniği'ne 100 hastanın geldiği düşünüldüğünde, yaklaşık olarak 9 hastanın hastanede ortalama 1 gün kaldığı geriye kalan 91 hastanın ise ortalama 34 gün hastanede kaldığı söylenebilir. Aynı veri seti tek bileşenli üstel dağılım ile modellendiğinde $\hat{\theta} = 31.32$, $p = 0.0747$ ve $KS = 0.1137$ bulunmuştur. Üstel dağılım için elde edilen bu sonuçlara göre sadece yaklaşık olarak 31 gün Kardiyoloji biriminde yatış yapan hasta tek hasta grubu olduğu ortaya çıkmaktadır. Ayrıca bulunan KS değerleri birbirleriyle karşılaştırıldığında, 2MED'in hastanede kalma sürelerini modellemede tek bileşenli üstel dağılımına göre daha başarılı olduğu görülmüştür.

7. SONUÇLAR

Simülasyon çalışması sonucunda, küçük örneklem ($n = 20$ ve $n = 40$) ve birbirinden uzak seçilen ortalama değerleri ($(\theta_1 = 10, \theta_2 = 1)$ ve $(\theta_1 = 2, \theta_2 = 0.3)$) için tahminlere bakıldığında, RMSE ölçütüne göre karma oranı (α) ile 1.bileşen ortalaması (θ_1) parametrelerinin, MOM ve LS tarafından daha iyi tahmin edildiği; 2.bileşen ortalaması (θ_2) parametresinde ise LS tahminlerinin daha iyi sonuç verdiği görülmektedir. Yine küçük örneklemelerde birbirine yakın seçilen ortalama değerleri ($\theta_1 = 3, \theta_2 = 2$) ve ($\theta_1 = 0.8, \theta_2 = 0.3$) göz önüne alındığında ise α ve θ_1 parametreleri için MOM ve LS, θ_2 parametresi için ML tahminlerinin daha iyi sonuç verdiği gözlenmiştir. Birbirinden uzak seçilen ortalamalarda örneklem hacmi arttırıldığında ($n = 100$), ML ve LS tahminlerinin α , θ_1 ve θ_2 parametrelerini tahmin etmede MOM tahminlerine göre daha başarılı olduğu sonucu ortaya çıkmaktadır.

Genel olarak, uzak ortalamalar ele alındığında küçük örneklem hacmi için α ve θ_1 parametrelerinin tahmininde MOM ve LS tahmin edicileri daha başarılı iken, örneklem hacmi arttıkça bu parametrelerin tahmininde MOM tahmin edicilerinin yerini ML tahminlerinin aldığı görülmektedir. Benzer olarak, θ_2 parametresi için küçük örneklemelerde en iyi tahminleri veren LS'nin yerini de örneklem hacmi arttıkça ML almaktadır. Yani RMSE ölçütüne göre, ML tahmin edicileri, örneklem hacmi arttıkça birbirinden uzak seçilen ortalamalarda daha etkin olmaktadır. Yakın ortalamalar söz konusu olduğunda α ve θ_1 parametreleri için benzer yorumlar yapılabilir. θ_2 parametresi için küçük örneklemelerde en iyi sonucu veren ML iken, örneklem hacmi arttıkça LS tahminlerinin de iyileşmeye başladığı gözlenmiştir. Yani örneklem hacmi arttırıldığında, θ_2 için en iyi tahminler ML ve LS tarafından hesaplanmaktadır.

Yapılan simülasyon çalışmasının ardından, ele alınan üç parametre tahmin yöntemi χ^2 test istatistiğine göre karşılaştırıldığında, ortalamalar birbirine yaklaştıkça, LS için hesaplanan χ^2 değerlerinin genel olarak MOM için elde edilen χ^2 değerlerinden daha büyük olduğu görülmüştür. Ortalamalar birbirinden uzak seçildiğinde ise en iyi

sonucu veren tahmin edici yönteminin MOM olduğu ve onu sırasıyla LS ve ML'nin izlediği görülmüştür.

Ayrıca, 6. bölümde ayrıntılı bir şekilde yer verilen literatürde sıkça kullanılan veri setlerini modellemede 2MED'in oldukça iyi olduğu sonucuna ulaşılmıştır. Çalışmanın literatür tarama süresinde genel olarak karşılaşılan modelleme tekniklerinden birisi, sürekli bir dağılımın kesikli veya sürekli bir dağılım ile bileşiminin (compound) yapılmasıdır. Bu durum beraberinde karmaşık bir yapıda olan dağılımın parametrelerini tahmin etmede bir takım zorluklar çıkarmaktadır. Analitik olarak tahmin ediciler elde edilememekte ancak çeşitli yazılımlar ile algoritmaların geliştirilmesi ile çözüme gidilebilmektedir. Bu literatür taramasının sonucunda, çalışmalarda kullanılan veri setlerinin tek bileşenli dağılım ile modellenmeye çalışıldığı veya modelleme sırasında çalışmada da bahsedilen karmaşık modellerin önerildiği görülmüştür. Hâlbuki veri setinin yapısından dolayı 2MED'in çoğu zaman dikkate alınacak kadar iyi sonuçlar verdiği ve aslında tek bileşenli dağılım önerilen durumlarda 2MED'in veri setini modellemede daha etkin olduğu görülmüştür.

Bu sebeple, üstel dağılım özelliği gösteren veriler üzerinde öncelikle iki veya daha fazla bileşene sahip karma üstelin varlığının araştırılması önerilmektedir. Çünkü parametre tahminlerinde en çok bilinen üç yöntemin (ML, LS ve MOM) kolay uygulanabilirliği açısından, geliştirilen yazılımlar daha hızlı sonuç vermektedir.

Yapısının basit olması ve oldukça kullanışlı olması nedeniyle 2MED'in literatürde yer alan çalışmalarda kullanılan dağılımlara göre modellemede daha başarılı olması bu tez çalışmasının amacına ulaşmasını sağlamıştır.

Çalışmanın amacı doğrultusunda, Tekirdağ ili üniversite hastanesinden Beyin Cerrahisi birimine giriş yapan 500 hasta ve Kardiyoloji birimine giriş yapan 410 hasta için hastanede kalma süreleri ve cinsiyet bilgilerini içeren bir veri seti toplanmıştır. Bu veriyi toplamadaki amaç, hastanede kalma sürelerine de tek bileşenli üstel dağılımdan geliyormuş gibi davranılıp, ne kadar personel, ne kadar yatak

gerekeceğini planlamaktır. Hâlbuki cinsiyet, yaş, hangi birime geldiği, kaç kere geldiği, hastanın öyküsü gibi değişkenlerin de göz önüne alınıp hastanede kalma sürelerinin bu ayırt edici değişkenler kullanılarak modellenmesi ve yorumlanması gerekir. Ancak, planlamacılar veya yöneticiler, kuyruk teorisinde var olan üstel dağılım üzerinden gitmektedirler.

Karma modeller sadece hastaların hastanede kalma sürelerine ait verilerin modellenmesinde kullanılmamaktadır. Karma modellerin günlük hayatta üretim ve bekleme modellerinde kullanılabileceği alanlar da mevcuttur. Örneğin, bankada bekleme süreleri, süpermarketteki kasada bekleme süreleri, değişik parçalardan oluşan bir cihazın hazır hale gelinceye kadar geçen süre gibi verilerin modellenmesinde bile bir karma model karşımıza çıkabilir.

Bu çalışma en azından cinsiyet ve birime göre hastaların hastanede kalma sürelerinin bir karma modele uyduğunu ileri sürmüştür. İleride yukarıda bahsedilen değişkenlerin de içinde olduğu bir veri seti toplanarak modelleme yapılabilir. Yapılan bu çalışmada eğer hastalar gruplara ayrılmamış olsaydı, yaşam zamanları karma üstel dağılım yardımıyla belirlenebilir ve bu sayede hem yaşam zamanları birbirinden farklı olan gruplar kolay bir şekilde bulunabilir hem de o veri setinde kaç farklı grup olduğu ortaya çıkabilirdi.

KAYNAKLAR

- Açıkgöz, A. 2007. Sonlu karma dağılımlarda parametre tahmini. Doktora Tezi, Ankara Üniversitesi, Fen Bilimleri Enstitüsü, İstatistik Anabilim Dalı, 152, Ankara.
- AL-Jammal, Z.Y. 2008. Exponentiated exponential distribution as a failure distribution, Iraqi Journal of Statistical Science, 63.
- Blischke, W.R. and Murthy, D.N.P. 1995. Product warranty handbook. Marcel Dekker, Inc., New York.
- Chahkandi, M. and Ganjali, M. 2009. On some lifetime distributions with decreasing failure rate. Computational Statistics and Data Analysis, Vol. 53, 4433- 4440.
- Chiodo, E. 2012. Parameter estimation of mixed weibull probability distributions for wind speed related to power statistics. International Symposium on Power Electronics, Electrical Drives, Automation and Motion.
- Chu, Y.K. and Ke, J.C. 2012. Computation approaches for parameter estimation of weibull distribution. Mathematical and Computational Applications, Vol. 17, No. 1, 39–47.
- Coroller, L., Leguerinel, I., Mettler, E., Savy, N. and Mafart, P. 2006. General model, based on two mixed weibull distributions of bacterial resistance, for describing various shapes of inactivation curves. Applied and Environmental Microbiology, 6493–6502.
- Cox, D.R. and Lewis, P.A.W. 1966. The statistical analysis of series of events. Chapman and Hall. London:Methuen.
- Davis, D. J. 1952. An analysis of some failure data. Journal of The Ameican Statistical Association, Vol.47, No.258, 113-150.
- Davison, A.C. and Hinkley D.V. 1997. Bootstrap methods and their application, Cambridge Series in Statistical and Probabilistic Mathematics, 1 edition, 594 p.
- Dekking, F.M., Kraaikamp, C. and Lopuhaa, H.P. and Meester, L.E. 2010. A modern introduction to probability and statistics: understanding why and how. Springer Series in Statistics.
- Dempster, A.P., Laird, N.M. and Rubin, D.B. 1977. Maximum likelihood from incomplete data via the EM algorithm. Journal of the Royal Statistical Society. Series B (Methodological), Vol. 39, No. 1, 1–38.
- Dennis, J.E. and Schnabel, R.B. 1983. Numerical methods for unconstrained optimization and nonlinear equations. Prentice-Hall, Englewood Cliffs.

- Erol, H. 1995. Sonlu karma dağılım modelleri (sürekli tip-tek değişkenli), Doktora tezi, Çukurova Üniversitesi, Fen Bilimleri Enstitüsü, Adana.
- Everitt, B.S. and Hand, D.J. 1981. Finite mixture distributions. Chapman and Hall, 151, London.
- Fadhilah, Y., Zalina, M., Nguyen, V.T.V., Suhaila, S. and Zulkifli, Y. 2007. Fitting the best-fit distribution for the hourly rainfall amount in the Wilayah Persekutuan. Jurnal Teknologi, 46(C) Jun 2007: 49–58.
- Gupta, R.D. and Kundu, D. 2003. Closeness of gamma and generalized exponential distribution. Communications in Statistics. Vol. 32, Issue 4, 705–721.
- Gupta, M. R. and Chen, Y. 2010. Theory and use of the EM algorithm. Foundations and Trends in Signal Processing. Vol. 4, No. 3, 223–296.
- Gupta, R.D. and Kundu, D. 2000. Generalized exponential distribution: different method of estimations. Journal of Statistical Computation Simulation, Vol. 00, 1–22.
- Harrison, G.W. 2001. Implications of mixed exponential occupancy distributions and patient flow models for health care planning. Health Care Management Science 4: 37–45.
- Hasselblad, V. 1969. Estimation of finite mixtures of distributions from the exponential family. American Statistical Association Journal, 64, 1459–1471.
- Hung, W.L. and Chang Y.C. 2011. Comparison between method of moments and entropy regularization algorithm applied to parameter estimation for mixed-Weibull distribution. Journal of Applied Statistics Vol. 38, No. 12, 2709–2722.
- Jarrett, R.G. 1979. A note on the intervals between coal mining disasters. Biometrika 66 191–193.
- Keatinge, C.L. 1999. Modeling losses with the mixed exponential distribution. In Proceedings of the Casualty Actuarial Society, Volume 86, 654–698.
- Kumar, U. and Klefsjö, B. 1992. Reliability analysis of hydraulic systems of LHD machines using the power law process model. Reliability Engineering and System Safety, 35: 217–224.
- Kus, C. 2007. A new lifetime distribution. Computational Statistics & Data Analysis, 51: 4497 – 4509.
- Li, Z., Brissette F. and Chen J. 2013. The impacts of precipitation amount simulation on hydrological modeling in Nordic watersheds. Geophysical Research Abstracts, Vol. 15, EGU 2013–3799.

- Lin, C. and Ke, S. 2013. Estimation of $P(Y < X)$ for location-scale distributions under joint progressively type-II right censoring. *Quality Technology & Quantitative Management*, Vol. 10, No. 3, 339-352.
- Maguire, B.A., Pearson, E.S. and Wynn, A.H.A. 1952. The time intervals between industrial accidents. *Biometrika* 39, 168-180.
- Marshall, A.H., Vasilakis, C. and El-Darzi, E. 2005. Length of stay-based patient flow models: Recent developments and future directions. *Health Care Management Science*, 8, 213-220.
- Marshall, A.H., McClean S.I. and Millard, P.H. 2004. Addressing bed costs for the elderly: a new methodology for modelling patient outcomes and length of stay. *Health Care Management Science* 7, 27-33.
- McLachlan, G.J. and Krishnan, T. 2008. The EM algorithm and extensions. John Wiley & Sons, Inc., Hoboken, New Jersey.
- Mendoza-Rosas, A.T. and Cruz-Reyna S. 2009. A mixture of exponentials distribution for a simple and precise assessment of the volcanic hazard. *Nat. Hazards Earth Syst. Sci.*, 9, 425-431.
- Merdenhull, W. and Harder R.J. 1958. Estimation of parameters of mixed exponentially distributed failure time distributions from censored life test data. *Biometrika*, Vol.45, 504-520.
- Mirhossaini, S. M. and Dolati A. 2008. On a new generalization of the exponential distribution. *Journal of Mathematical Extension* Vol. 3, No. 1, 27-42.
- Moon, T.D. 1996. The expectation-maximization algorithm. *IEEE Signal Processing Magazine*. 1053-5888, 47-60.
- Mosler, K. and Seidel, W. 2001. Testing for homogeneity in an exponential mixture model. *Australian and New Zealand Journal of Statistics*, 43, 231-247.
- Nagode, M. and Fajdiga, M. 2000. An improved algorithm for parameter estimation suitable for mixed Weibull distributions. *International Journal of Fatigue* 22, 75-80.
- Nelson, W. 1990. Accelerated testing: statistical models, Test Plans and Data Analyses. John Wiley and Sons, New York.
- Oluyede, B. O., Huang, S. and Pararai, M. 2014. A new class of generalized dagum distribution with applications to income and lifetime data. *Journal of Statistical and Econometric Methods*, Vol.3, No.2, 125-151.

- Perlstein, D., Jarvis, W.H. and Mazzuchi, T.A. 2001. Bayesian calculation of cost optimal burn-in test durations for mixed exponential populations. *Reliability Engineering and System Safety* 72, 265-273.
- Petersen, J.L. 2000. Estimating the parameters of a pareto distribution. Technical report. University of Montana.
- Prodan A. and Prodan R. 2001. Stochastic simulation and modelling. Exchange of Technology and Know-how -New Techniques and Technologies for Statistics, 461-466, Crete Island. JRC-ISIS, European Commission.
- Proschan, F. 1963. Theoretical explanation of observed decreasing failure rate. *Technometrics* 5, 375-383.
- Ren, Y. 2011. The methodology of flowgraph models. PhD Thesis, Department of Statistics London School of Economics and Political Science.
- Rider, P. R. 1961. The method of moments applied to a mixture of two exponential distributions. *The Annals of Mathematical Statistics*, Vol. 32, No. 1, 143-147.
- Rodriguesa, J., Balakrishnan, N. and Cordeiro, G.M. 2011. A unified view on lifetime distributions. *Computational Statistics & Data Analysis*, Volume 55, Issue 12, 3311-3319.
- Rosa, L.G. and Figueiredo, I. 2010. Demonstration of the robustness of maximum likelihood estimation to distinguish mixtures of two weibull populations in failure data of brittle materials. *Materials Science Forum* Vol. 636-637, 1443-1450.
- Shawky, A.I. and Badr, M.M. 2012. Estimations and prediction from the inverse rayleigh model based on lower record statistics. *Life Science Journal* (9;1), 985-990.
- Suhaila, J. and Jemain, A.A. 2007. Fitting daily rainfall amount in Peninsular Malaysia using several types of exponential distributions. *Journal of Applied Sciences Research*, 3(10): 1027-1036.
- Suhaila, J., Ching-Yee, K., Fadhilah, Y. and Hui-Mean F. 2011. Introducing the mixed distribution in fitting rainfall data. *Open Journal of Modern Hydrology*, Vol. 1 No.2, 11-22.
- Sunitha Anup, S., Deshpande, A.S., Cheeran, A.N. and Mangalvedekar H.A. 2011. Reliability evaluation of high voltage insulation using Weibull distribution. *International Journal of Research and Reviews in Electrical and Computer Engineering (IJRRECE)* Vol. 1, No. 2, 55-62.
- Tahmasbi, R. and Rezaei, S. 2008. A two-parameter lifetime distribution with decreasing failure rate. *Computational Statistics and Data Analysis*, Vol. 52, 3889-3901.

- Taylor, G.J., McClean S.I. and Millard, P.H. 1999. Stochastic models of geriatric patient bed occupancy behaviour. *J.R. Statist. Soc. A.* 163, Part 1, 39–48.
- Teicher, H. 1960. On the mixture of distributions. *Ann. Math. Stat.*, 31, 55-73.
- Teicher, H. 1961. Identifiability of mixtures. *Annals of Math. Stats.* 32, 244-248.
- Teicher, H. 1963. Identifiability of finite mixtures. *Annals of Math. Stats.* 34, 1265-1269.
- Tian,Y., Tian, M. and Zhu, Q. 2014. Estimating a finite mixed exponential distribution under progressively type-II censored data. *Vol. 43(17)*, 3762-3776.
- Vasilakis, C. and Marshall, A. H. 2005. Modelling nationwide hospital length of stay: opening the black box. *Journal of the Operational Research Society*, 56, 862-869.
- Wan, H., Zhang, X. and Barrow E.M. 2005. Stochastic modelling of daily precipitation for Canada. *Atmosphere-Ocean.* 43:1, 23-32.
- Weinberg, G.W. and Finch, D. 2012. Analysis of a pareto mixture distribution for maritime surveillance radar. *Hindawi Publishing Corporation Journal of Electrical and Computer Engineering* Volume 2012, Article ID 382979, 6 pages.
- Wilks, D.S. 1998. Multisite generalization of a daily stochastic precipitation generation model. *J. Hydrol.* 210:178–191.
- Wu, C.C., Chou, C.Y. and Huang, C. 2007. Optimal burn-in time and warranty length under fully renewing combination free replacement and pro-rata warranty. *Reliability Engineering and System Safety* 92, 914–920.
- Yakowitz, S. J. and Spragins, J. D. 1968. On the identifiability of finite mixtures. *The Annals of Mathematical Statistics*, Vol. 39, No. 1, 209-214.
- Yilmaz, M. and Buyum, B. 2014. A study on modeling of phenomena air conditioning data. *Commun. Fac. Sci. Univ. Ank. Series A1*, Volume 63, Number 2, 109-118, ISSN 1303-5991.
- Yilmaz, M. Potas, N. and Buyum, B. 2014. A classical approach to modeling of coal mine data. *Chaos, Complexity and Leadership*. Chapter 7, Pages 65-73, Springer. DOI 10.1007/978-3-319-18693-1, ISSN 2213-8684.
- Yilmaz, M. and Buyum, B. 2015. Parameter estimation methods for two-component mixed exponential distributions. *İstatistik: Journal of The Turkish Statistical Association* Vol. 8, No. 3, 51–59. ISSN 1300-4077 | 15 | 3 | 51 | 59.

EKLER

- EK 1 Proschan (1963) çalışmasından alınan veri seti çizelgesi
- EK 2 Maguire, Pearson ve Wynn (1952) çalışmasından alınan veri seti çizelgesi
- EK 3 Kuş (2007) çalışmasından alınan veri seti çizelgesi
- EK 4 Dekking v.d. (2010) çalışmasından alınan veri seti çizelgesi
- EK 5 AL-Jammal (2008) çalışmasından alınan veri seti çizelgesi
- EK 6 Kumar ve Klefsjö (1992) çalışmasından alınan veri seti çizelgesi
- EK 7 Nelson (1990) çalışmasından alınan veri seti çizelgesi
- EK 8 Blischke (1995) çalışmasından alınan veri seti çizelgesi
- EK 9 Kardiyoloji kliniğinde yatan kadın hastaların hastanede kalma sürelerine ilişkin veri seti çizelgesi
- EK 10 Kardiyoloji kliniğinde yatan erkek hastaların hastanede kalma sürelerine ilişkin veri seti çizelgesi
- EK 11 Beyin cerrahisi kliniğinde yatan kadın hastaların hastanede kalma sürelerine ilişkin veri seti çizelgesi
- EK 12 Beyin cerrahisi kliniğinde yatan erkek hastaların hastanede kalma sürelerine ilişkin veri seti çizelgesi

EK 1 Proschan (1963) çalışmasından alınan veri seti çizelgesi

UÇAK KİMLİK NUMARALARI												
7907	7908	7909	7910	7911	7912	7913	7914	7915	7916	7917	8044	8045
194	413	90	74	55	23	97	50	359	50	130	487	102
15	14	10	57	320	261	51	44	9	254	493	18	209
41	58	60	48	56	87	11	102	12	5		100	14
29	37	186	29	104	7	4	72	270	283		7	57
33	100	61	502	220	120	141	22	603	35		98	54
181	65	49	12	239	14	18	39	3	12		5	32
	9	14	70	47	62	142	3	104			85	67
	169	24	21	246	47	68	15	2			91	59
	447	56	29	176	225	77	197	438			43	134
	184	20	386	182	71	80	188				230	152
	36	79	59	33	246	1	79				3	27
	201	84	27	15	21	16	88				130	14
	118	44	153	104	42	106	46					230
	34	59	26	35	20	206	5					66
	31	29	326		5	82	5					61
	18	118			12	54	36					34
	18	25			120	31	22					
	67	156			11	216	139					
	57	310			3	46	210					
	62	76			14	111	97					
	7	26			71	39	30					
	22	44			11	63	23					
	34	23			14	18	13					
		62			11	191	14					
		130			16	18						
		208			90	163						
		70			1	24						
		101			16							
		208			52							
					95							

EK 2 Maguire, Pearson ve Wynn (1952) çalışmasından alınan veri seti çizelgesi

0	1	1	2	2	3	4	4	4	6
7	10	11	12	12	12	13	15	15	16
16	16	17	17	18	19	19	19	20	20
22	23	24	25	27	28	29	29	29	31
31	32	33	34	34	36	36	37	40	41
41	42	43	45	47	48	49	50	53	54
54	55	56	59	59	61	61	65	66	66
70	72	75	78	78	78	80	80	81	88
91	92	93	93	95	95	96	96	97	99
101	108	110	112	113	114	120	120	123	123
124	124	125	127	129	131	134	137	139	143
144	145	151	154	156	157	176	182	186	187
188	189	190	193	194	197	202	203	208	215
216	217	217	217	218	224	225	228	232	233
250	255	275	275	275	276	286	292	307	307
312	312	315	324	326	326	329	330	336	345
348	354	361	364	368	378	388	420	431	456
462	467	498	517	536	538	566	632	644	745
806	826	871	952	1205	1312	1358	1630	1643	2366

EK 3 Kuş (2007) çalışmasından alınan veri seti çizelgesi

1163	3258	323	159	756	409
501	616	398	67	896	8592
2039	217	9	633	461	1821
4863	143	182	2117	3709	979



EK 4 Dekking v.d. (2010) çalışmasından alınan veri seti çizelgesi

30	50	55	4	263	21	232	129	296	2323
2323	143	1897	482	648	113	77	242	0	452
233	330	810	1755	2930	108	447	5509	5485	81
24	68	8	255	134	365	290	1064	1461	0
386	100	1160	115	108	422	227	197	357	1222
300	1783	843	3110	446	10	1864	9	88	180
65	193	193	543	529	860	12	1247	122	1071
4116	2	670	10	176	6	236	10	281	983
261	943	990	371	91	120	1146	58	79	31
16	160	707	1800	700	948	790	112	26	600
457	816	369	529	828	33	865	875	1082	6150
15	114	15	300	1351	748	379	1011	868	1435
245	22	3321	138	325	36	97	148	0	44
445	724	30	729	75	1045				

EK 5 AL-Jammal (2008) çalışmasından alınan veri seti çizelgesi

MAKİNE 1					
1	1	5	5.5	12.50	16.75
17.75	20.75	22.50	22.75	25	25
27.25	30.25	43.75	45	48	48.25
97.50	99.75	136.75	143.5	207.75	215
225.50	235	283.5	567	970.5	
MAKİNE 2					
3.5	6.5	10.5	23.25	23.5	43.5
69	70.5	75.5.	83.25	95.5	109.5
111.25	144	164	167.25	253	383.75
417.75	428.25	453	1215		

EK 6 Kumar ve Klefsjö (1992) çalışmasından alınan veri seti çizelgesi

LHD (Load-Haul-Dump) Makinelerinin Bozulma Zamanları Arasında Geçen Süre (saat)					
LHD1	LHD3	LHD9	LHD11	LHD17	LHD20
327	637	278	353	401	231
125	40	261	96	36	20
7	197	990	49	18	361
6	36	191	211	159	260
107	54	107	82	341	176
277	53	32	175	171	16
54	97	51	79	24	101
332	63	10	117	350	293
510	216	132	26	72	5
110	118	176	4	303	119
10	125	247	5	34	9
9	25	165	60	45	80
85	4	454	39	324	112
27	101	142	35	2	10
59	184	39	258	70	162
16	167	249	97	57	90
8	81	212	59	103	176
34	46	204	3	11	360
21	18	182	37	5	90
152	32	116	8	3	15
158	219	30	245	144	315
44	405	24	79	80	32
18	20	32	49	53	266
	248	38	31	84	
	140	10	259	218	
		311	283	122	
		61	150		
			24		

EK 7 Nelson (1990) çalışmasından alınan veri seti çizelgesi

Farklı Voltaj Seviyelerinde Yalıtım Sıvısının Dökülme Zamanları							
Voltaj (kV)	26	28	30	32	34	36	38
Zaman (Dakika)	5,79	68,85	7,74	0,27	0,19	0,35	0,09
	1579,52	108,29	17,05	0,4	0,78	0,59	0,39
	2323,7	110,29	20,46	0,69	0,96	0,96	0,47
		426,07	21,02	0,79	1,31	0,99	0,73
		1067,6	22,66	2,5	2,78	1,69	0,74
			43,4	3,91	3,16	1,97	1,13
			47,3	9,88	4,15	2,07	1,4
			139,07	13,95	4,67	2,58	2,38
			144,12	15,93	4,85	2,71	
			175,88	27,8	6,5	2,9	
			194,9	53,24	7,35	3,67	
				82,85	8,01	3,99	
				89,29	8,27	5,35	
				100,58	12,06	13,77	
				215,1	31,75	25,5	
					32,52		
					33,91		
					36,71		
					72,89		

EK 8 Blischke (1995) çalışmasından alınan veri seti çizelgesi

Elektronik Parçaların Bozulma Zamanları								
4,38	4,94	1,79	1,23	0,57	1,56	0,92	0,14	13,54
0,96	5,7	1,63	2,02	2,35	1,48	0,82	0,11	1,55
7,77	6,57	0,1	2,2	3,58	2,2	7,83	7,62	6,97
0,23	2,63	4,57	0,03	13,54	0,61	0,31	2,44	1,74
1,38	6,18	30,1	1,56	2,25	0,4	1,43	1,96	0,31
5,38	4,61	1,97	7,98	0,52	0,71	3,63	0,71	0,01
0,01	0,17	0,41	2,48	3,68	2,87	0,05	14,62	0,05
0,24	0,06	4,17	3,28	2,81	0,83	0,77	0,3	0,28
0,01	0,24	1,82	0,2	6,24	1,98	3,67	2,14	0,28
0,04	0,27	1,85	2,36	0,52	0,78	1,71	0,23	1,33
0,03	0,04	0,66	2,92	1,6	0,12	0,46	0,17	0,79
0,04	4,9	3,39	3,28	4,42	10,14	1,65	7,65	3,94
8,05	5,06	2,93	10,55	2,11	6,37	7,87	5,13	5,09
3,22	6,02	11,67	7,58	1,61	5,43	4,87	6,93	6,04
5,69								

EK 9 Kardiyoloji kliniğinde yatan kadın hastaların hastanede kalma sürelerine ilişkin veri seti çizelgesi

Hastanede Kalma Süreleri (gün)	
1	25
1	25
1	26
1	28
1	28
1	29
1	30
1	31
1	32
2	33
3	34
3	35
4	41
7	42
7	44
8	45
9	46
9	46
9	47
10	53
11	54
11	56
12	59
13	61
13	61
13	63
13	66
17	68
19	68
19	70
20	75
20	77
20	89
21	108
21	110
23	125
25	

EK 10 Kardiyoloji kliniğinde yatan erkek hastaların hastanede kalma sürelerine ilişkin veri seti çizelgesi

Hastanede Kalma Süreleri (gün)	
1	24
1	24
1	25
1	26
1	29
1	29
1	30
1	31
1	33
1	35
2	41
2	42
3	43
3	45
5	47
5	56
8	57
4	62
6	63
6	66
7	67
10	69
13	70
14	72
14	82
15	85
16	90
16	90
18	91
19	93
20	112
21	123
21	105
22	
23	
23	
24	

EK 11 Beyin cerrahisi kliniğinde yatan kadın hastaların hastanede kalma sürelerine ilişkin veri seti çizelgesi

Hastanede Kalma Süreleri (gün)	
1	25
1	25
1	26
1	28
1	28
1	29
1	30
1	31
1	32
2	33
3	34
3	35
4	41
7	42
7	44
8	45
9	46
9	46
9	47
10	53
11	54
11	56
12	59
13	61
13	61
13	63
13	66
17	68
19	68
19	70
20	75
20	77
20	89
21	108
21	110
23	125
25	

EK 12 Beyin cerrahisi kliniğinde yatan erkek hastaların hastanede kalma sürelerine ilişkin veri seti çizelgesi

Hastanede Kalma Süreleri (gün)	
1	21
1	23
1	24
1	24
1	25
1	26
2	31
3	32
4	32
4	33
4	39
5	39
6	41
6	41
6	42
7	47
7	51
7	54
7	58
8	61
8	67
8	74
9	77
10	80
11	81
11	82
11	84
12	95
12	120
13	125
14	150
17	
18	
19	
19	
19	
20	

ÖZGEÇMİŞ

Adı Soyadı : Buse BÜYÜM

Doğum Yeri : İstanbul

Doğum Tarihi : 01/08/1989

Medeni Hali : Bekar

Yabancı Dili : İngilizce

Eğitim Durumu (Kurum ve Yıl)

Lise : Mehmet Emin Resulzade Anadolu Lisesi (2003-2007)

Lisans : Ankara Üniversitesi Fen Fakültesi İstatistik Bölümü (2007-2012)

Yüksek Lisans: Ankara Üniversitesi Fen Bilimleri Enstitüsü İstatistik Anabilim Dalı (Eylül 2012-Haziran 2017)

Çalıştığı Kurum ve Yıl

T.C. Ekonomi Bakanlığı / Dış Ticaret Uzman Yardımcısı 10 Mart 2015 – (Devam)

Hakemli Dergi

Yilmaz, M. and **Buyum, B.** Parameter Estimation Methods For Two-Component Mixed Exponential Distributions. İstatistik: Journal of The Turkish Statistical Association Vol. 8, No. 3, December 2015, 51–59. ISSN 1300-4077 | 15 | 3 | 51 | 59.

Yilmaz, M. and **Buyum, B.** A Study On Modeling Of Phenomena Air Conditioning Data. Commun.Fac.Sci.Univ.Ank.Series A1, Volume 63, Number 2, 109-118 (2014), ISSN 1303-5991.

Yilmaz, M. and **Buyum, B.** An Econometric Approach To Tourism Demand: Evidence From Turkey. Commun.Fac.Sci.Univ.Ank.Series A1, Volume 64, Number 2, 99-109 (2015). DOI: 10.1501/Commua1_0000000737, ISSN 1303-5991.

Yılmaz, M. Potas, N. and **Buyum, B.** A Classical Approach to Modeling of Coal Mine Data. Chaos, Complexity and Leadership 2014. Chapter 7, 65-73, Springer. DOI 10.1007/978-3-319-18693-1, ISSN 2213-8684.

Uluslararası Kongrelerde Sunulan Bildiriler

- Yılmaz, M. ve Buyum, B. 2013. Turist Talebine Ekonometrik Bir Yaklaşım: Türkiye Örneği, Uluslararası 8. İstatistik Kongresi, Antalya, Türkiye.
- Yılmaz, M. ve Buyum, B. 2013. Parameter Estimation Methods for Two-Component Mixed Exponential Distributions, İstatistik Günleri Sempozyumu, Antalya, Türkiye.

