

**UKUROVA NİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ**

Ülkü ERİŐOĐLU

DOKTORA TEZİ

**YAŐAM SÜRESİ ANALİZİNDE KARMA VE KARIŐTIRILMIŐ
DAĐILIMLARIN KULLANILMASI**

İSTATİSTİK ANABİLİM DALI

ADANA, 2011

**UKUROVA NİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ**

DOKTORA TEZİ

Ülkü ERİŐOĐLU

**YAŐAM SÜRESİ ANALİZİNDE KARMA VE KARIŐTIRILMIŐ
DAĐILIMLARIN KULLANILMASI**

İSTATİSTİK ANABİLİM DALI

ADANA, 2011

**UKUROVA NİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ**

DOKTORA TEZİ

Ülkü ERİŐOĐLU

**YAŐAM SÜRESİ ANALİZİNDE KARMA VE KARIŐTIRILMIŐ
DAĐILIMLARIN KULLANILMASI**

İSTATİSTİK ANABİLİM DALI

ADANA, 2011

**ÇUKUROVA ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ**

**YAŞAM SÜRESİ ANALİZİNDE KARMA VE KARIŞTIRILMIŞ
DAĞILIMLARIN KULLANILMASI**

Ülkü ERİŞOĞLU

DOKTORA TEZİ

İSTATİSTİK ANABİLİM DALI

Bu Tez / / Tarihinde Aşağıdaki Jüri Üyeleri Tarafından
Oybirliği/Oyçokluğu ile Kabul Edilmiştir.

.....
Prof. Dr. Hamza EROL
DANIŞMAN

.....
Prof. Dr. Fikri AKDENİZ
ÜYE

.....
Doç. Dr. Aşır GENÇ
ÜYE

.....
Prof. Dr. Selahattin KAÇIRANLAR
ÜYE

.....
Prof. Dr. Sadullah SAKALLIOĞLU
ÜYE

Bu Tez Enstitümüz İstatistik Anabilim Dalında hazırlanmıştır.
Kod No:

Prof. Dr. İlhami YEĞİNGİL
Enstitü Müdürü

Bu Çalışma Ç. Ü. Araştırma Projeleri Birimi Tarafından Desteklenmiştir.
Proje No: FEF2009D7

Not: Bu tezde kullanılan özgün ve başka kaynaktan yapılan bildirişlerin, çizelge ve fotoğrafların
kaynak gösterilmeden kullanımı, 5846 sayılı Fikir ve Sanat Eserleri Kanunundaki hükümlere
tabidir.

ÖZ

DOKTORA TEZİ

YAŞAM SÜRESİ ANALİZİNDE KARMA VE KARIŞTIRILMIŞ DAĞILIMLARIN KULLANILMASI

Ülkü ERİŞOĞLU

ÇUKUROVA ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ
İSTATİSTİK ANABİLİM DALI

Danışman : Prof. Dr. Hamza EROL
Yıl: 2011, Sayfa: 153
Jüri : Prof. Dr. Hamza EROL
: Prof. Dr. Fikri AKDENİZ
: Prof. Dr. Selahattin KAÇIRANLAR
: Prof. Dr. Sadullah SAKALLIOĞLU
: Doç. Dr. Aşır GENÇ

Bu çalışmada yaşam analizinde karma ve karıştırılmış dağılım modellerinin kullanımı incelenmiştir. Özellikle heterojen yapıdaki yaşam sürelerinin modellenmesinde yetersiz kalan standart olasılık dağılımları yerine karma dağılım modelleri önerilmiştir. Çalışmada öncelikle yaşam analizi ile ilgili temel tanım ve fonksiyonlar verilmiştir. Yaşam sürelerinin modellenmesi için aynı dağılımın farklı parametrelerle karma modelleri, farklı iki dağılımın karma modelleri ve karıştırılmış genelleştirilmiş Üstel-Geometrik (EEG) dağılımının karma modeli oluşturulmuştur. Oluşturulan karma dağılım modellerindeki parametrelerin en çok olabilirlik tahminlerinin elde edilmesi için gerekli olan EM algoritmaları verilmiştir. Algoritmaların parametre tahminindeki etkinliği simülasyon çalışması ile ortaya konulmuştur. Ayrıca gerçek yaşam verileri kullanılarak veri modellemede önerilen karma dağılım modellerinin standart olasılık dağılımları karşısındaki üstünlükleri ortaya konmuştur.

Anahtar kelimeler: Karma dağılım, Karıştırılmış dağılım, Yaşam analizi, EM, Tahmin

ABSTRACT

PhD THESIS

THE USAGE OF MIXTURE AND MIXED DISTRIBUTIONS IN LIFETIME ANALYSIS

Ülkü ERİŞOĞLU

ÇUKUROVA UNIVERSITY
INSTITUTE OF NATURAL AND APPLIED SCIENCES
DEPARTMENT OF STATISTICS

Supervisor :Prof. Dr. Hamza EROL
Year: 2011, Pages: 153
Jury :Prof. Dr. Hamza EROL
:Prof. Dr. Fikri AKDENİZ
:Prof. Dr. Selahattin KAÇIRANLAR
:Prof. Dr. Sadullah SAKALLIOĞLU
:Assoc. Prof. Dr. Aşır GENÇ

In this study, examined the usage of mixture and mixed distributions in survival analysis. Especially, proposed mixture distribution models instead of the standard probability distributions of inadequate modeling of heterogeneous survival. The study primarily is given the basic definitions and functions about survival analysis. Modelling survival times formed mixture models with different parameters of the same distribution, mixture models of two different distribution, and mixture model of mixed extended exponential-geometric (EEG) distributions. The generated mixture distribution models the maximum likelihood estimations of the parameters are obtained by the EM algorithm. Effectiveness parameter estimation of algorithms are determined with simulation study. In addition, have been revealed the advantages opposite standard probability distributions of the proposed mixture distribution models the data modeling using real-life data.

Key Words: Mixture Distribution, Mixed Distribution, Survival Analysis, EM, Estimation

TEŞEKKÜR

Bu tez çalışması sürecinde değerli bilgi ve görüşlerini benimle paylaşan danışmanım sayın Prof. Dr. Hamza EROL'a; tez izleme komitesi üyeleri sayın Prof. Dr. Selahattin KAÇIRANLAR'a ve sayın Doç. Dr. Aşır GENÇ'e; hiçbir zaman yardımlarını esirgemeyen Çukurova Üniversitesi İstatistik Bölümü'nde görev yapan çalışma arkadaşlarıma teşekkürlerimi sunarım.

Ayrıca hayatım boyunca beni destekleyen anneme, babama, kız kardeşlerim Ülker GÖKAL'a, Aliye Nur GÖKAL'a, Damlanur GÖKAL'a ve hayatımın her anında hem bilimsel hem duygusal yanımda olan biricik sevdiğim eşim Murat ERİŞOĞLU'na sonsuz teşekkürlerimi sunarım.

İÇİNDEKİLER

SAYFA

ÖZ	I
ABSTRACT	II
TEŞEKKÜR	III
İÇİNDEKİLER	IV
TABLolar DİZİNİ	VII
ŞEKİLLER DİZİNİ	IX
1. GİRİŞ	1
2. ÖNCEKİ ÇALIŞMALAR	7
3. YAŞAM ANALİZİNDE TEMEL TANIM VE GÖSTERİMLER	19
3.1. Yaşam Analizindeki Önemli Fonksiyonlar	20
3.1.1. Yaşam Fonksiyonu	20
3.1.2. Olasılık Yoğunluk Fonksiyonu	22
3.1.3. Hazard Fonksiyonu	24
3.1.4. Yaşam Fonksiyonları Arasındaki İlişkiler	25
3.2. Yaşam Analizinde Kullanılan Bazı Önemli Dağılımlar	27
3.2.1. Üstel Dağılım	27
3.2.2. Weibull Dağılımı	30
3.2.3. Gamma Dağılımı	35
3.2.4. Lognormal Dağılımı	39
4. KARMA VE KARIŞTIRILMIŞ DAĞILIMLAR	45
4.1. Karma Dağılım Modelleri	45
4.1.1. Temel Tanım ve Gösterimler	47
4.1.2. Uygun Bileşen Sayısının Seçimi	51
4.1.3. Karma Dağılım Modellerinde Parametre Tahmini	53
4.1.3.1. Karma Dağılım Modellerinde Tamamlanmamış Veri Yapısı	53
4.1.3.2. EM Algoritması	54
4.2. Karıştırılmış Dağılımlar	57
4.2.1. Karıştırılmış Dağılımlarda Temel Gösterimler	57
4.2.2. Karıştırılmış Dağılımlarla İlgili Bazı Örnekler	59

5. YAŞAM ANALİZİNDE KARMA DAĞILIM MODELLERİ.....	65
5.1. Yaşam Sürelerinin Modellenmesinde Karma Dağılım Yaklaşımı	66
5.1.1. Karma Gamma Dağılım Modeli	66
5.1.2. Karma Lognormal Dağılım Modeli	70
5.1.3. Karma Weibull Dağılım Modeli	74
5.2. EM Algoritmasının Parametre Tahminindeki Etkinliği	78
5.3. Heterojen Yaşam Verilerinin Modellenmesinde Karma Dağılım Modelleri	80
5.3.1. Işınlanmış Farelerin Ölüm Zamanları Verisi	81
5.3.2. Akciğer Kanseri Verisi	85
5.4. Sonuçlar	90
6. YAŞAM ANALİZİNDE İKİ FARKLI DAĞILIMIN KARMASI	91
6.1. İki Farklı Dağılımın Karma Modelleri	91
6.1.1. Üstel-Gamma Karma Dağılımı	91
6.1.2. Üstel-Weibull Karma Dağılımı	95
6.1.3. Gamma-Weibull Karma Dağılımı	99
6.2. İki Farklı Dağılımın Karma Modelinde EM Algoritmasının Parametre Tahminindeki Etkinliği	103
6.3. Uygulama	105
6.3.1. Dış Sulama Aygıtlarının Yaşam Süreleri	106
6.3.2. Akciğer Kanseri Verisi	109
6.4. Sonuçlar	112
7. KARIŞTIRILMIŞ DAĞILIMLARIN KARMA MODELİ	113
7.1. Genelleştirilmiş Üstel-Geometrik Dağılım	113
7.2. EEG Dağılımının Karması	117
7.3. EM Algoritması İle Karma EEG Dağılımının Parametrelerinin En Çok Olabilirlik Tahminleri	120
7.4. Uygulamalar	124
7.4.1. Yaşam Yorgunluğu Veri Seti	124
7.4.2. Dış Sulama Aygıtlarının Yaşam Süreleri	127
7.5. Sonuçlar	130

8. SONUÇLAR VE ÖNERİLER.....	133
KAYNAKLAR.....	139
ÖZGEÇMİŞ.....	153

Tablo 5.1. İki bileşenli karma Gamma dağılım modelinde EM algoritması ile elde edilen parametrelerin en çok olabilirlik tahminlerinin ortalamaları ve standart hataları	78
Tablo 5.2. İki ve üç bileşenli karma Lognormal dağılım modelinde EM algoritması ile elde edilen parametrelerin en çok olabilirlik tahminlerinin ortalamaları ve standart hataları	79
Tablo 5.3. İki ve üç bileşenli karma Weibull dağılım modelinde EM algoritması ile elde edilen parametrelerin en çok olabilirlik tahminlerinin ortalamaları ve standart hataları	79
Tablo 5.4. Işınlanmış farelerin ölüm zamanları	81
Tablo 5.5. Işınlanmış farelerin ölüm zamanları verisinin modellenmesinde kullanılan dağılımların parametre tahminleri, $\ln L$, D_{hesap} , HKO ve $E(t)$ Değerleri	83
Tablo 5.6. Akciğer kanseri tanısı alan hastaların tanı konduktan sonraki yaşam süreleri	85
Tablo 5.7. Karma dağılım modellerinde bileşen sayısına göre $\ln L$ ve AIC değerleri	86
Tablo 5.8. Akciğer kanseri tanısı alan hastaların yaşam sürelerinin modellenmesinde kullanılan dağılımların parametre tahminleri, $\ln L$, D_{hesap} ve HKO Değerleri	88
Tablo 6.1. Üstel-Gamma karma dağılım modelinde EM algoritması ile elde edilen parametrelerin en çok olabilirlik tahminlerinin ortalamaları ve standart hataları	104
Tablo 6.2. Üstel-Weibull karma dağılım modelinde EM algoritması ile elde edilen parametrelerin en çok olabilirlik tahminlerinin ortalamaları ve standart hataları	104

Tablo 6.3. Gamma-Weibull karma dağılım modelinde EM algoritması ile elde edilen parametrelerin en çok olabilirlik tahminlerinin ortalamaları ve standart hataları	105
Tablo 6.4. Dış sulama aygıtlarının yaşam süreleri	106
Tablo 6.5. Dış sulama aygıtlarının yaşam süreleri veri seti için Üstel, Gamma ve Weibull dağılım modelleri ve Üstel-Gamma, Üstel-Weibull ve Gamma-Weibull karma dağılım modellerinin parametre tahminleri, D_{hesap} ve HKO değerleri.....	107
Tablo 6.6. Kanser tanısı alan 42 hastanın tanı konduktan sonraki yaşam süreleri..	109
Tablo 6.7. Kanser tanısı alan hastaların tanı konduktan sonraki yaşam sürelerinin modellenmesi için Üstel, Gamma ve Weibull dağılım modelleri ve Üstel-Gamma, Üstel-Weibull ve Gamma-Weibull karma dağılım modelleri ile modellenmesinde elde edilen parametre tahminleri, D_{hesap} ve HKO değerleri	110
Tablo 7.1. İki bileşenli EEG dağılımının karmasından 500 boyutlu 300 örneklemden elde edilen EM tahminlerinin ortalamaları ve standart hataları.....	123
Tablo 7.2. İki farklı tipteki 25 numunenin yaşam yorgunluğu.....	124
Tablo 7.3. Yaşam yorgunluğu veri seti için ilgilenilen dağılım modellerinde parametre tahminleri, log olabilirlik fonksiyonu ($\ln L$), AIC , D_{hesap} ve HKO değerleri.....	125
Tablo 7.4. Dış sulama aygıtlarının yaşam süreleri için ilgilenilen dağılım modellerinde parametre tahminleri, log olabilirlik fonksiyonu ($\ln L$), AIC , D_{hesap} ve HKO değerleri	128

Şekil 1.1. Yıllara göre içerisinde “survival analysis” terimi geçen çalışma sayıları.....	3
Şekil 3.1. Birikimli dağılım fonksiyonu ile sağ kalım fonksiyonu arasındaki ilişki.....	21
Şekil 3.2. Yaşam eğrisi için iki örnek	21
Şekil 3.3. Yoğunluk eğrisi için iki örnek	23
Şekil 3.4. Örnek hazard fonksiyon grafikleri	24
Şekil 3.5. Farklı λ parametre değerlerine göre Üstel dağılım için yoğunluk eğrileri	28
Şekil 3.6. Farklı λ parametre değerlerine göre Üstel dağılıma ait yaşam eğrileri ve hazard fonksiyonunun grafikleri.....	29
Şekil 3.7. Weibull dağılımında $\beta = 1$ ve farklı α parametre değerleri için yoğunluk eğrileri	32
Şekil 3.8. Weibull dağılımında $\beta = 1$ ve farklı α parametre değerleri için yaşam eğrileri	33
Şekil 3.9. Weibull dağılımında $\beta = 1$ ve farklı α parametre değerleri için hazard fonksiyonunun grafikleri	34
Şekil 3.10. Gamma dağılımında $\beta = 1$ ve farklı α parametre değerleri için yoğunluk eğrileri	36
Şekil 3.11. Gamma dağılımında $\alpha = 3$ ve farklı β parametre değerleri için yoğunluk eğrileri	37
Şekil 3.12. Gamma dağılımında $\beta = 1$ ve farklı α parametre değerleri için yaşam eğrileri	38
Şekil 3.13. Gamma dağılımında $\beta = 1$ ve farklı α parametre değerleri için hazard fonksiyonu eğrileri.....	39
Şekil 3.14. Lognormal dağılımında $\mu = 0$ ve farklı σ^2 parametre değerleri için yoğunluk eğrileri	40

Şekil 3.15. Lognormal dağılımında $\sigma^2 = 0.5$ ve farklı μ parametre değerleri için yoğunluk eğrileri	41
Şekil 3.16. Lognormal dağılımında $\sigma^2 = 0.5$ ve farklı μ parametre değerleri için yaşam eğrileri	42
Şekil 3.17. Lognormal dağılımında $\mu = 0$ ve farklı σ parametre değerleri için hazard fonksiyonu eğrileri.....	43
Şekil 4.1. (a) Yengeç verisi için uydurulan tek bileşenli normal dağılım (nokta nokta) ve iki bileşenli normal dağılım (düz çizgi) eğrileri. (b) Yengeç verisi için uydurulan iki bileşenli normal dağılım (düz çizgi) ve her bir bileşene ait normal dağılım eğrileri	46
Şekil 4.2. Roeder (1990)'in Galaxy veri seti için uydurulan 6 bileşenli karma normal dağılım (düz çizgi) ve tek bileşenli normal dağılımın (kesikli çizgi) yoğunluk eğrileri	47
Şekil 5.1. Farklı parametrelere sahip iki bileşenli karma Gamma dağılım modeli için yoğunluk eğrileri a) $\pi_1 = 0.3, \alpha_1 = 0.5, \beta_1 = 2, \pi_2 = 0.7, \alpha_2 = 5$ ve $\beta_2 = 2$ (düz çizgi) b) $\pi_1 = 0.4, \alpha_1 = 2, \beta_1 = 1, \pi_2 = 0.6, \alpha_2 = 10$ ve $\beta_2 = 1$ (kesik çizgi)	66
Şekil 5.2. Farklı parametrelere sahip iki bileşenli karma Gamma dağılım modeli için (i) yaşam fonksiyonu ve (ii) hazard fonksiyonu eğrileri a) $\pi_1 = 0.3, \alpha_1 = 0.5, \beta_1 = 2, \pi_2 = 0.7, \alpha_2 = 5$ ve $\beta_2 = 2$ (düz çizgi) b) $\pi_1 = 0.4, \alpha_1 = 2, \beta_1 = 1, \pi_2 = 0.6, \alpha_2 = 10$ ve $\beta_2 = 1$ (kesik çizgi).....	67
Şekil 5.3. Farklı parametrelere sahip iki ve üç bileşenli karma Lognormal dağılım modelleri için yoğunluk eğrileri a) $\pi_1 = 0.35, \mu_1 = 3, \sigma_1 = 0.5, \pi_2 = 0.65, \mu_2 = 4$ ve $\sigma_2 = 0.2$ (düz çizgi) b) $\pi_1 = 0.15, \mu_1 = 3.5, \sigma_1 = 0.5, \pi_2 = 0.3, \mu_2 = 4.5, \sigma_2 = 0.2$ ve $\pi_3 = 0.4, \mu_3 = 5, \sigma_3 = 0.05$ (kesik çizgi)	71

Şekil 5.4. Farklı parametrelere sahip iki ve üç bileşenli karma Lognormal dağılım modeli için (i) yaşam fonksiyonu ve (ii) hazard fonksiyonu eğrileri	
a) $\pi_1 = 0.35, \mu_1 = 3, \sigma_1 = 0.5, \pi_2 = 0.65, \mu_2 = 4$ ve $\sigma_2 = 0.2$ (düz çizgi)	
b) $\pi_1 = 0.15, \mu_1 = 3.5, \sigma_1 = 0.5, \pi_2 = 0.3, \mu_2 = 4.5, \sigma_2 = 0.2$ ve $\pi_3 = 0.4, \mu_3 = 5, \sigma_3 = 0.05$ (kesik çizgi)	72
Şekil 5.5. Farklı parametrelere sahip iki ve üç bileşenli karma Weibull dağılım modelleri için yoğunluk eğrileri	
a) $\pi_1 = 0.3, \alpha_1 = 2, \beta_1 = 100, \pi_2 = 0.7, \alpha_2 = 10$ ve $\beta_2 = 200$ (düz çizgi)	
b) $\pi_1 = 0.3, \alpha_1 = 5, \beta_1 = 50, \pi_2 = 0.2, \alpha_2 = 10, \beta_2 = 150, \pi_3 = 0.5, \alpha_3 = 20$ ve $\beta_3 = 250$ (kesik çizgi)	74
Şekil 5.6. Farklı parametrelere sahip iki ve üç bileşenli karma Weibull dağılım modeli için (i) yaşam fonksiyonu ve (ii) hazard fonksiyonu eğrileri	
a) $\pi_1 = 0.3, \alpha_1 = 2, \beta_1 = 100, \pi_2 = 0.7, \alpha_2 = 10$ ve $\beta_2 = 200$ (düz çizgi)	
b) $\pi_1 = 0.3, \alpha_1 = 5, \beta_1 = 50, \pi_2 = 0.2, \alpha_2 = 10, \beta_2 = 150, \pi_3 = 0.5, \alpha_3 = 20$ ve $\beta_3 = 250$ (kesik çizgi)	75
Şekil 5.7. Işınlanmış farelerinin ölüm zamanları için Gamma, Lognormal ve Weibull dağılımlarına ait P-P grafikleri	82
Şekil 5.8. Işınlanmış farelerinin ölüm zamanlarının modellenmesinde klasik istatistik dağılımları ile karma dağılım modellerinin grafiksel karşılaştırılması	84
Şekil 5.9. Karma Gamma, karma Lognormal ve karma Weibull dağılım modellerinde bileşen sayısına göre AIC değerlerinin değişimi	86
Şekil 5.10. Akciğer kanseri tanısı konan hastaların yaşam süreleri için P-P grafikleri	87
Şekil 5.11. Akciğer kanseri tanısı konan hastaların yaşam sürelerinin modellenmesinde klasik istatistik dağılımları ile karma dağılım modellerinin grafiksel karşılaştırılması	89

Şekil 6.1. Üstel-Gamma karma dağılım modelinde farklı parametre değerleri için yoğunluk eğrileri a) $\pi = 0.6, \lambda = 0.5, \alpha = 10$ ve $\beta = 1.5$ (düz çizgi) b) $\pi = 0.3, \lambda = 1, \alpha = 5$ ve $\beta = 0.9$ (kesik çizgi)	92
Şekil 6.2. Üstel-Gamma karma dağılım modelinde farklı parametre değerleri için (i) yaşam fonksiyonu ve (ii) hazard fonksiyonu eğrileri a) $\pi = 0.6, \lambda = 0.5, \alpha = 10$ ve $\beta = 1.5$ (düz çizgi) b) $\pi = 0.3, \lambda = 1, \alpha = 5$ ve $\beta = 0.9$ (kesik çizgi)	93
Şekil 6.3. Üstel-Weibull karma dağılım modelinde farklı parametre değerleri için yoğunluk eğrileri a) $\pi = 0.2, \lambda = 0.2, a = 5$ ve $b = 15$ (düz çizgi) b) $\pi = 0.6, \lambda = 0.125, a = 6$ ve $b = 4$ (kesik çizgi).....	96
Şekil 6.4. Üstel-Weibull karma dağılım modelinde farklı parametre değerleri için (i) yaşam fonksiyonu ve (ii) hazard fonksiyonu eğrileri a) $\pi = 0.2, \lambda = 0.2, a = 5$ ve $b = 15$ (düz çizgi) b) $\pi = 0.6, \lambda = 0.125, a = 6$ ve $b = 4$ (kesik çizgi).....	97
Şekil 6.5. Gamma-Weibull karma dağılım modelinde farklı parametre değerleri için yoğunluk eğrileri a) $\pi = 0.3, \alpha = 3, \beta = 0.5, a = 5$ ve $b = 8$ (düz çizgi) b) $\pi = 0.6, \alpha = 5, \beta = 2, a = 10$ ve $b = 5$ (kesik çizgi).....	100
Şekil 6.6. Gamma-Weibull karma dağılım modelinde farklı parametre değerleri için (i) yaşam fonksiyonu ve (ii) hazard fonksiyonu eğrileri a) $\pi = 0.3, \alpha = 3, \beta = 0.5, a = 5$ ve $b = 8$ (düz çizgi) b) $\pi = 0.6, \alpha = 5, \beta = 2, a = 10$ ve $b = 5$ (kesik çizgi)	101
Şekil 6.7. Diş sulama aygıtlarının yaşam sürelerine ait yaşam fonksiyonları	108
Şekil 6.8. Kanser tanısı alan hastaların tanı konduktan sonraki yaşam sürelerine ait yaşam fonksiyonları	111
Şekil 7.1. EEG dağılımında $\beta = 1$ ve farklı γ parametre değerlerine göre elde edilen yoğunluk eğrileri	115
Şekil 7.2. EEG dağılımında $\beta = 1$ ve farklı γ parametre değerlerine göre elde edilen (i) yaşam fonksiyonu ve (ii) hazard fonksiyonu eğrileri	116

Şekil 7.3. Farklı parametre değerleri için iki bileşenli EEG dağılımının karmasının olasılık yoğunluk fonksiyonları.....	118
Şekil 7.4. Karma EEG dağılımının modelinde farklı parametre değerlerine göre elde edilen (i) yaşam fonksiyonu ve (ii) hazard fonksiyonu eğrileri	119
Şekil 7.5. (a) Yaşam yorgunluğu verisinin Weibull dağılım modelinin yoğunluk eğrisi (nokta çizgi), EEG dağılım modelinin yoğunluk eğrisi (ince kırık çizgi), iki bileşenli Weibull dağılımın karma modelinin yoğunluk eğrisi (dairesel çizgi) ve iki bileşenli EEG dağılımın karma modelinin yoğunluk eğrisi (kalın çizgi) uyumunun grafiksel karşılaştırılması (b) Ampirik yaşam fonksiyonu (çokgen çizgi) ile Weibull dağılım modelinin yaşam fonksiyonunun (nokta çizgi), EEG dağılım modelinin yaşam fonksiyonunun (ince kırık çizgi), iki bileşenli Weibull dağılımın karma modelinin yaşam fonksiyonunun (dairesel çizgi) ve iki bileşenli EEG dağılımın karma modelinin yaşam fonksiyonunun (kalın çizgi) grafiksel karşılaştırmaları.....	126
Şekil 7.6. (a) Diş sulama aygıtlarının yaşam süreleri Weibull dağılım modelinin yoğunluk eğrisi(nokta çizgi), EEG dağılım modelinin yoğunluk eğrisi(ince kırık çizgi), iki bileşenli Weibull dağılımın karma modelinin yoğunluk eğrisi(dairesel çizgi) ve iki bileşenli EEG dağılımın karma modelinin yoğunluk eğrisi(kalın çizgi) uyumunun grafiksel karşılaştırılması (b) Ampirik yaşam fonksiyonu (çokgen çizgi) ile Weibull dağılım modelinin yaşam fonksiyonunun (nokta çizgi), EEG dağılım modelinin yaşam fonksiyonunun (ince kırık çizgi), iki bileşenli Weibull dağılımın karma modelinin yaşam fonksiyonunun (dairesel çizgi) ve iki bileşenli EEG dağılımın karma modelinin yaşam fonksiyonunun (kalın çizgi) grafiksel karşılaştırmaları.....	129

1. GİRİŞ

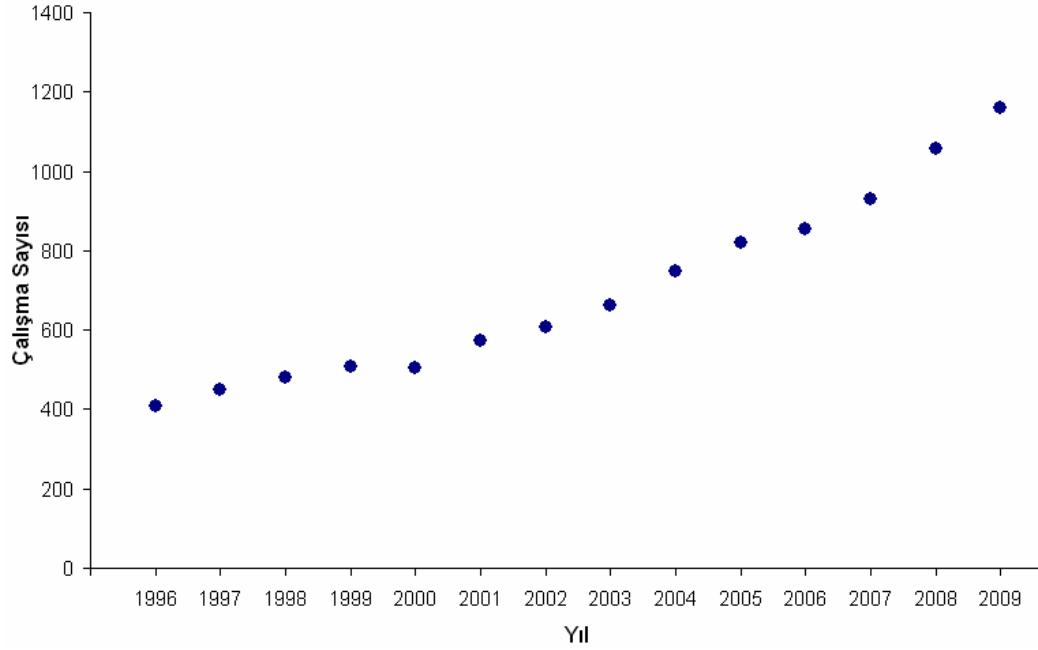
Teknoloji her geçen gün gelişmekte, sunduğu ürün ve hizmetler sayesinde hayatımızı kolaylaştırmaktadır. Özellikle sağlık alanında son yıllarda yapılan ciddi çalışmalar, yeni tedavi yöntemlerinin geliştirilmesi ve geliştirilen yöntemlerin insanlar üzerinde ne derece faydalı olabileceği üzerine çeşitli test çalışmaları üzerinde yoğunlaşmıştır. Sağlık alanı ile ilgili çalışmalarda yaşam analizi çok sık kullanılmaktadır (Klein ve Moeschberger 1997). Belirli bir hastalığa yakalanan bireylerin, hastalığın teşhisinden sonra uygulanan tedavi yöntemiyle ne kadar süre yaşayabileceğini tahmin etmek için ya da çeşitli tedavi yöntemlerinin yaşam süresine etkisini araştırmak için geliştirilmiş olan yöntemler bütünü “Yaşam Analizi” olarak adlandırılır (Johnson ve ark 1992, 1995). Yaşam analizi sadece insanlar üzerinde değil, aynı zamanda hayvanların kullanıldığı laboratuvar deneylerinde de uygulanabilir. Ayrıca mekanik parçaların ve elektronik eşyaların farklı şartlar altında ne kadar bozulmadan kalabileceği de yaşam analizi ile tahmin edilebilir. Yaşam analizinde sıkça geçen “başarısızlık” terimi şöyle açıklanabilir: Başarısızlık, araştırılan durumun denekte görülmesi durumudur. Canlılar için genelde ölüm veya hastalık, mekanik aletler için ise bozulma anlamına gelir. Bazı ürünler içinse performans yavaş yavaş düşer, ama yaşamın biteceği kesin değildir. Bu konuda yapılan çalışmalar özellikle son yüzyılın ikinci çeyreğinde kendini göstermektedir. 1958 yılında, yaşam fonksiyonlarının tahmini için çağrım-limit yöntemi Kaplan ve Meier (1958) tarafından geliştirilmiştir. Cox (1972) oransal hazard modelini geliştirmiştir ve Miller (1981) tarafından yaşam analizi için parametrik olmayan yöntemler uygulanmıştır.

Yaşam analizi çalışmalarında kullanılan parametrik yaklaşımlarda en önemli aşama, veri setini en uygun dağılım ile modellemektir. Yaşam analizi verilerinin modellenmesinde Üstel, Gamma, Lognormal, Weibull dağılımı gibi özellik gösteren dağılımlar yaygın olarak kullanılmaktadır. İstatistikte dağılımlar alanında yaşanan gelişmeler yaşam verilerinin modellenmesinde kullanılabilecek dağılımların sayısını artırmıştır. Özellikle karma (mixture) ve karıştırılmış (mixed) dağılım yaklaşımları ile elde edilen yeni dağılımlar yaşam verilerinin analizinde kullanılmaktadır.

Karma dağılım modelleri farklı uygulama alanlarında rassal olarak ortaya çıkan olayları istatistiksel olarak modellemek için kullanılır. Karma dağılım modelleri, esnek ve çok kullanışlı bir modelleme yöntemidir. Özellikle heterojen yapıdaki verilerin modellenmesinde standart olasılık dağılım modellerine göre daha kullanışlı ve daha yararlı olan karma dağılımların kullanımı yaygın bir hale gelmiştir. Standart olasılık dağılımları, frekans dağılımının tek modlu olduğu durumlarda kullanışlı iken karma dağılım modelleri frekans dağılımının iki veya daha fazla moda sahip olması hatta tek modlu olması durumunda bile verinin modellenmesinde daha kullanışlı ve daha yararlıdır. Bu nedenle karma dağılım modelleri hem teorik alanlarda hem de uygulamalı alanlarda gün geçtikçe önem kazanmaktadır.

Karıştırılmış dağılım, (Ambagaspitiya 1995, 1998) dağılımın parametresinin başka bir dağılıma sahip olduğu düşüncesi ile elde edilen yeni dağılıma verilen isimdir. Karıştırılmış dağılımların (Antzoulakos ve Chadjiconstantinidis 2004, Cekanavicius ve Roos 2004) elde edilmesinde genellikle üstel aileye ait olan dağılımların biçim ya da ölçek parametresinin farklı bir dağılıma sahip olduğu varsayılarak üretilir (Adamidis ve ark. 2005).

Thomson Reuters tarafından hazırlanan SCI, SCI-Expanded, SSCI ve AHCI indekslerince taranan çalışmalara ulaşılabilen ISI Web of KnowledgeSM (<http://apps.isiknowledge.com>) arama sayfasında arama kutucuğuna “survival analysis” terimi yazılıp arama yapılarak 1996 ile 2009 yılları arasında yıllara göre içerisinde yaşam analizi kavramı geçen çalışma sayıları Şekil 1.1’de verilmiştir.



Şekil 1.1. Yıllara göre içerisinde “survival analysis” terimi geçen çalışma sayıları

Bu tez çalışmasındaki genel amaç, yaşam verilerinin analizinde karma ve karıştırılmış dağılımların uygulanabilirliğinin ve etkinliğinin incelenmesidir. Çalışmada mevcut yöntemlere alternatif olabileceği düşünülen karma ve karıştırılmış dağılımların yaşam verilerinde kullanımına yönelik avantaj ve dezavantajların ortaya konması, veri yapısının özelliklerine göre hangi dağılımın hangi veri setinde kullanılabileceğinin araştırılması ve yaşam verisi analizinde kullanılan karma ve karıştırılmış dağılımların özelliklerinin araştırılması amaçlanmaktadır.

Çalışma kapsamında yapılan uygulamaların özetlendiği ve çalışmanın amacı ile öneminin belirtildiği bu bölümün ardından, ikinci bölümde yaşam analizinde karma ve karıştırılmış dağılımlarla ilgili önceki çalışmalar verilecektir.

Çalışmanın üçüncü bölümünde yaşam analizinde önemli olan olasılık yoğunluk fonksiyonu, yaşam fonksiyonu ve hazard fonksiyonu ile bu fonksiyonlar arasındaki ilişkiler tanımlanarak yaşam analizinde yaygın bir kullanıma sahip olan Üstel, Weibull, Gamma ve Lognormal dağılımları incelenecektir.

Çalışmanın dördüncü bölümünde karma ve karıştırılmış dağılımlarla ilgili temel gösterimler verilerek, karma dağılım modellerinde uygun bileşen sayısının seçimi, parametre tahmini, tamamlanmamış veri yapısı ve EM algoritması

anlatılacaktır. Bu bölümde birer karıştırılmış dağılım olan Negatif Binom, Pareto ve genelleştirilmiş Burr dağılımlarının elde edilişleri gösterilecektir.

Çalışmanın beşinci bölümünde özellikle heterojen yapıdaki yaşam sürelerinin modellenmesinde kullanılan karma Gamma dağılım modeli, karma Lognormal dağılım modeli ve karma Weibull dağılım modeli ile incelenecektir. İlgilenilen karma dağılım modellerindeki parametrelerin en çok olabilirlik tahminlerinin elde edilmesi için EM algoritmaları tanımlanarak, tanımlanan algoritmaların parametre tahminindeki etkinliği simülasyon çalışması ile gösterilecektir. Ayrıca literatürde kullanılan ve iyi bilinen iki gerçek heterojen yapıdaki yaşam süresi verisinin modellenmesinde, karma dağılım modellerinin standart olasılık dağılım modellerine göre üstünlüğü gösterilecektir.

Çalışmanın altıncı bölümünde heterojen yaşam sürelerinin modellenmesinde iki farklı dağılımın karmalarının kullanımı incelenecektir. Üstel-Gamma, Üstel-Weibull ve Gamma-Weibull karma dağılım modelleri ilgili temel gösterimler verilerek, modellerdeki parametrelerin en çok olabilirlik tahminlerinin elde edilmesi için gerekli EM algoritmaları tanımlanacaktır. Tanımlanan algoritmalarının parametre tahminindeki etkinliği simülasyon çalışması ile gösterilecektir. Heterojen yapıdaki iki gerçek verinin modellenmesinde iki farklı dağılımın karma modeli, standart olasılık dağılımları ile karşılaştırılacaktır.

Çalışmanın yedinci bölümünde heterojen yaşam verilerinin modellenmesi için Adamidis ve ark (2005) tarafından tanımlanan genelleştirilmiş Üstel-Geometrik (EEG) dağılımının karma dağılım modeli oluşturulacaktır. Çalışmada öncelikle dönüştürülmüş uç değer dağılımında α parametresinin Üstel dağılıma sahip bir rassal değişken olarak tanımlanması ile elde edilen EEG dağılımı incelenecektir. EEG dağılımının heterojen yaşam sürelerinin modellenmesinde kullanılabilmesi için karma EEG dağılım modeli tanımlanarak, parametrelerinin en çok olabilirlik tahminlerini için gerekli olan EM algoritması verilecektir. Çalışmada, tanımlanan EM algoritmasının parametre tahminindeki etkinliği simülasyon çalışması ile gösterilecektir. Ayrıca, iki bileşenli karma EEG dağılım modeli ile heterojen yapı gösteren iki gerçek veri seti modellenecektir. Gerçek veri setlerinin modellenmesinde

karma EEG dağılım modeli, tek bileşenli Weibull, tek bileşenli EEG ve karma Weibull dağılım modeli ile karşılaştırılacaktır.

Sonuç bölümünde çalışmadan elde edilen bulgular değerlendirilerek önerilerde bulunulacaktır.

2. ÖNCEKİ ÇALIŞMALAR

Farklı karakteristik özellikler taşıyan dağılımların bir arada kullanılabildiği karma dağılım modelleri veri modellemede standart olasılık dağılımlarına göre daha esnek ve kullanışlıdır. Farklı disiplinlerde çok sayıda araştırmacı, karma dağılım modelleri ile ilgili hem teorik hem de modellerin pratikteki uygulanabilirliği üzerinde çalışmıştır.

Karma dağılım modellerinde parametre tahmini üzerinde en fazla çalışılan konulardan biridir. Karl Pearson (1894) iki tek değişkenli normal dağılımın karmasındaki parametreleri tahmin ettiği çalışma bu alandaki ilk çalışma kabul edilmektedir. Karma dağılım modellerinin parametrelerinin tahminleri ile ilgili farklı birçok teknik geliştirilmiştir. Normal dağılımın istatistikteki önemi nedeniyle en fazla normal dağılımların karmasıyla ilgili çalışılmıştır. Parametre tahminini basitleştirmek için özellikle belli parametreler üzerinde yoğunlaşmıştır. Örneğin normal dağılımların karmalarında varyansların eşit veya biliniyor olması varsayımı altında karmanın parametreleri tahmin edilmeye çalışılmıştır. Bazı çalışmalarda ise karmayı oluşturan her bileşenin karma oranının bilindiği varsayılarak karmanın parametrelerinin tahminiyle ilgilenilmiştir.

20. yüzyılın ilk başlarında karmanın parametrelerinin tahmininde grafik tekniklerine dayalı yöntemler geliştirilmiştir. Bu yöntemler genellikle normal dağılımın karması ile ilgilidir. Bunun nedeni olarak normal dağılımın karmasında parametre tahmininin diğer karmaların parametre tahminlerine göre daha basit olması gösterilebilir. Buna rağmen genellikle grafik yöntemi ile karmanın parametrelerinin sadece kabaca tahminleri bulunabilmektedir. Grafik yöntemlerinde ampirik yoğunluklar ile karmanın bileşen yoğunlukları en az hata ile örtüşmedikçe modelin uyumu iyi değildir. Buna rağmen grafik teknikleri veri setinde bir karma yapının olup olmadığının incelenmesinde, karmadaki bileşen sayısına karar vermede yararlıdır. Grafik yöntemlerinin birkaçı normal olasılık grafiklerini içermektedir.

Harding (1949) ve Cassie (1954) parametre tahmininde standart normal dağılımın çeyrekliklerine karşı örneklemin grafiğinden yararlanmışlardır. Böyle bir algoritmada, tek bileşenli normal dağılımda bir doğru elde edilirken karma

modellerde lineer olamayan bir eğri ile karşılaşılır. Karma dağılımın parametrelerinin tahmininde eğrinin büküm noktaları temel alınır. Bu noktalar görsel olarak kolaylıkla belirlenebilir ve böylece karmanın yapısı ile ilgili önemli bilgiler elde edilir.

Kao (1959), elektron tüplerinin yaşam sürelerinin testinde iki bileşenli Weibull karma dağılım modelinin parametrelerini Weibull olasılık kağıdı ve grafik yöntemleri ile tahmin etmiştir.

Bhattacharya (1967), gruplandırılmış veriler için daha basit bir grafik yöntemi geliştirmiştir. Önerdiği yöntemde parametre tahminleri histogramdaki sınıf frekanslarının logaritmik oranının grafiği ile elde edilir. Bu teknikte karmanın farklı bileşenlerinin her biri için negatif eğimli bir dizi doğru çizilir. Parametre tahminlerinin doğruluğu örneklemin saçılım grafiğinde çizilen doğruların geçerliliğine bağlıdır.

Fowlkes (1979), iki bileşenli durumda büküm noktalarının hesaplanmasında düzeltilmiş Lojistik dağılım fonksiyonunun kullanılmasını önermiştir. Ancak bu yöntemle elde edilen parametre tahminlerinin doğruluğu şüphelidir.

Jiang ve Kececioğlu (1992), Weibull olasılık kağıdı (WPP) ile iki bileşenli Weibull dağılımının karmasını çeşitli şekillerde incelemişlerdir ve birikimli dağılım fonksiyonlarını 6 tipte (A-F) sınıflandırmışlardır. B, C ve D tipleri iki iyi karıştırılmış alt gruptan oluşan birikimli dağılım fonksiyonunu, F tipi iki çok iyi ayrılmış alt gruptan oluşan birikimli dağılım fonksiyonunu ve A ve E tipleri B, C ve D tipleri ile F tipi arasında oluşan birikimli dağılım fonksiyonuna sahiptir. Bu 6 tipteki birikimli dağılım fonksiyonuna Kao (1959)-Cran (1976) ve Jensen-Petersen (1982) grafik metotlarını uygulamışlardır. Kao (1959)-Cran (1976) metodunun A, B, C, E ve F tiplerindeki birikimli dağılım fonksiyonlarına uygulanamadığı ve D tipindeki birikimli dağılım fonksiyonuna uygulanamayacağını önermişlerdir. Teorikte iyi ayrılmış alt grupların karması yani F tipi gibi eğriler için Jensen-Petersen (1982) grafik metodunu geliştirmişlerdir. A ve E tipindeki eğrilere bu metodu uygulamışlardır. Ama B, C ve D tipleri gibi eğriler için bu metodun teorik temelini olmadığını göstermişlerdir.

Jiang ve Murthy (1995), iki bileşenli iki parametrelili Weibull dağılımının karmasında parametre tahminlerinin grafiksel yöntem ile elde edilmesini

göstermişlerdir ve iki gerçek veri seti üzerine uygulamışlardır. Jiang ve Kececioglu (1992)'deki sonuçlar ile karşılaştırmışlardır.

Jiang ve Murthy (1997), bir başarısızlık veri setine uygun modelin belirlenmesi için Weibull Olasılık Grafik metodunu incelemişlerdir. Diş sulama aygıtlarının yaşam süreleri veri setine iki bileşenli üç parametrelili Weibull dağılım modelini uyarlamış ve parametre tahminlerini Weibull Olasılık Grafik metodu ile elde etmişlerdir.

Momentler yöntemi karma dağılım modellerinde parametrelerin analitik tahminlerinde kullanılan yöntemlerden biridir. Bu yöntem herhangi bir olasılık yoğunluk fonksiyonuna uygulanabilir. Yöntemde parametre tahminleri örneklem momentlerinin dağılım momentlerine eşitlenmesi ile elde edilen denklemlerin çözümünden elde edilir. Karma dağılımın her bir bileşeninin parametrelerinin ve karma oranlarının tahminlerinde momentler yöntemi kullanılabilir. Birçok araştırmacı normal dağılımların karmasında momentler yöntemi üzerine çalışmıştır. Karma dağılım modellerinde momentler yöntemi ile parametre tahmini oldukça karmaşıktır. Pearson (1894) çalışmasında iki tek değişkenli normal dağılımın karmasının parametrelerinin tahmininde momentler yöntemini kullanmıştır. Böyle basit bir karma modelde dahi parametre tahmini için 9. dereceden bir polinomun çözülmesi gereklidir. Bundan dolayı çoğu araştırmacı momentler yöntemi ile parametre tahmininde, parametreler hakkında belirli varsayımlarda bulunarak problemi basit bir hale indirgemeye çalışmıştır. Pollard (1934) tek değişkenli üç bileşenli normal dağılımının karmasının parametrelerini momentler yöntemi ile tahmin etmiştir.

Normal dağılımın dışındaki diğer sürekli ve kesikli dağılımların karmasında momentler yöntemi ile parametre tahmini normal dağılımların karmasına göre daha az karmaşıktır. Özellikle Üstel dağılımların karmasında momentler yöntemi ile parametre tahmini basit eşitliklerin çözümüyle elde edilmektedir. Genellikle momentler yönteminde elde edilen eşitlikler aynı genel yaklaşımla çözülebilir. Çözümde değişim sadece bileşen yoğunluklarının özel biçimlerine bağlıdır. Karma dağılım modelinde k bileşen olması durumunda parametrelerin tahmini için elde edilen eşitlik sistemi k . dereceden bir polinoma indirgenir.

Üstel aileye ait dağılımların karmasında momentler yöntemi ile parametre tahmini normal dağılımların karmasına göre daha basit olmasına rağmen bu dağılımların karması ile ilgili çalışmalarda da genellikle iki bileşen durumu ile ilgilenilmiştir.

Rider (1961a) iki üstel dağılımın karmasında parametrelerin tahmini probleminde momentler yöntemi ile ilgilenmiştir. Yine aynı yıl içerisinde Rider (1961b) Poisson, Binom ve Weibull dağılımların iki bileşenli karmalarındaki parametrelerin tahmininde momentler yöntemini kullanmıştır.

Cohen (1963, 1965) iki Poisson dağılımının karmasını ele almış ve beklenen değer ile faktöriyel momentleri eşitleyerek parametre tahmini için basit bir formülasyon geliştirmiştir.

Blischke (1964) k bileşenli karma Binom dağılım modelindeki parametrelerin tahmininde momentler yöntemini kullanmıştır.

Cohen (1967), Pearson (1894) tarafından ele alınan yengeç veri setinin iki normal dağılımın karması ile modellenmesindeki parametre tahmin probleminin tek bir negatif kök için çözülecek kübik denklemi içeren bir iterasyon metodu ile çözülebileceğini göstermiştir. Bu yaklaşım tahminlerin sadece ilk dört momentten oluşturulan kübik eşitliğin sadece negatif bir köküne bağlı olması durumunda, bileşenlerin varyansları eşit kabul edilmesi koşulu ile çözüm önermektedir.

John (1970a) iki normal kitleden gelen gözlemlerin her birinin ait olduğu kitlenin belirlenmesinde momentler yöntemini kullanmıştır. Yine John (1970b) Gamma dağılımına sahip iki kitleden gelen gözlemlerin her birinin ait olduğu kitlenin belirlenmesinde momentler yöntemini kullanmıştır.

Falls (1970), iki bileşenli Weibull dağılımının 5 parametresini momentler yöntemi ile tahmin etmiştir.

Everitt ve Hand (1981) ile Titterington ve ark (1985) sonlu karma dağılım modellerinde parametre tahmininde momentler yöntemini incelemişlerdir.

Bilgisayar teknolojisindeki kazanımlar araştırmacıları karma dağılım modellerindeki parametrelerin tahmini için daha basit ve etkili alternatif yöntemler geliştirmeye yöneltmiştir. Tek bileşenli (standart) dağılımlarda parametre tahmini için en çok olabilirlik yöntemi ilk olarak Fisher (1912) tarafından önerilmiştir. En

çok olabilirlik tahminlerinin özelliklerinden dolayı çoğu araştırmacı bu yöntemi momentler yönteminden üstün tutmuştur. Dağılımın olabilirlik fonksiyonunu maksimum yapan parametre değerleri en çok olabilirlik tahminleri olarak tanımlanır. Olabilirlik fonksiyonunu maksimum yapan parametreleri bulmak için olabilirlik fonksiyonu her bir parametreye göre ayrı ayrı türevlenerek sıfıra eşitlenir. Elde edilen eşitliklerin çözümünden parametrelerin en çok olabilirlik tahminleri elde edilir.

Başlangıçta karmanın parametrelerinin en çok olabilirlik tahminlerinin elde edilmesindeki hesaplamaların zorluğundan dolayı yöntem ile sadece teorik olarak ilgilenilmiştir. Bu çalışmalarda karmanın bazı parametreleri üzerine yoğunlaşarak tahmin problemi daha basit bir hale dönüştürülmüştür.

Rao (1948) biyolojik sınıflandırma probleminde çoklu ölçümlerden yararlanma isimli çalışmasında iki tek değişkenli normal dağılımın karmasında varyansları eşit varsayarak parametre tahminlerini en çok olabilirlik yöntemi ile elde etmiştir. Day (1969), Rao'nun yöntemini çok boyutlu iki bileşenli normal karma dağılımda uygulamıştır.

Hasselblad (1966) tek değişkenli k tane normal dağılımın karmasından oluşan daha karmaşık yapıdaki karma normal dağılımın parametrelerinin tahmininde en çok olabilirlik yöntemi ile ilgilenmiştir.

Literatürde öncelikle normal dağılımların karması ile ilgilenilse de Üstel aileye ait dağılımların karmasında parametrelerin en çok olabilirlik tahminleri üzerine de çalışılmıştır.

Hasselblad (1969) Üstel aileye ait dağılımların karmasının parametrelerinin tahmininde en çok olabilirlik yöntemi ile ilgilenmiştir. Kaylan ve Harris (1981), Hasselblad (1969) tarafından önerilen yaklaşımı genelleterek Weibull dağılımlarının karmasının parametre tahminlerinin nasıl elde edileceğini göstermişlerdir. Bu çalışmada Newton iteratif yönteminden sonra ikinci bir aşama ekleyerek yerel maksimuma daha hızlı yakınsama sağlamışlardır.

Craigmille ve Titterington (1997), Gupta ve Miyawaki (1978)'de önerdikleri iki bileşenli düzgün dağılımın karmasını genişleterek üç bileşenli düzgün dağılım karma modelini önermişlerdir. Üç bileşenli düzgün dağılımın karma modelinin

parametre tahmininde momentler yöntemini ve en çok olabilirlik yöntemini kullanmışlardır. Momentler yönteminin ve en çok olabilirlik yönteminin parametre tahminindeki etkinliğini simülasyon çalışmalarıyla göstermişlerdir. Karma düzgün dağılımıyla ilgili hipotez testinden kısaca bahsetmişlerdir.

Ormoneit ve Tresp (1998), Ridolfi ve Idier (2002), karma dağılımın parametrelerini bir cezalandırılmış olabilirlik fonksiyonunu maksimize ederek tahmin etmişler ve olabilirlik fonksiyonunun dejenereliğinden kaçmak için ceza terimi kullanarak sayısal maksimizasyon için EM algoritmasının bir cezalandırılmış versiyonunu elde etmişlerdir.

Sultan ve ark (2007), iki ters Weibull dağılımının karma modelini önermişlerdir. İki bileşenli ters Weibull dağılımın karma modelinin yaşam fonksiyonunu ve hazard fonksiyonunu göstermişlerdir. Karma dağılımın parametrelerini en çok olabilirlik metodunda EM algoritmasını kullanarak tahmin etmişlerdir. Algoritmanın parametre tahminindeki etkinliği simülasyon çalışmasıyla göstermişlerdir.

Rufo ve ark (2009), yaşam dağılımlarının karma modelleri için parametrelerinin duyarlılığı konusunda çalışmışlardır. Çalışmalarında iki bileşenli Üstel dağılımın karma modelinde λ_1 ve λ_2 parametreleri için duyarlılık analizinin nasıl yapılacağını göstermişlerdir.

Razali ve Salih (2009), Weibull dağılımın karma modelini incelemişlerdir. Karma Weibull dağılımın karakteristik fonksiyonlarını farklı karma oranlarında grafiksel olarak incelemişlerdir.

Balakrishnan ve ark (2011), Birnbaum-Saunders (BS) dağılımına dayanan üç farklı karma modeli önermişlerdir. İlk model iki farklı BS dağılımın karma modelinden, ikinci model farklı parametrelili BS dağılımı ile uzun-yanlı (length-biased) BS dağılımın (LBS) karma modelinden ve üçüncü model aynı parametrelili BS dağılımı ile LBS dağılımın karma modelinden oluşmaktadır. Ayrıca bu üç karma modelin karakteristik özelliklerini incelemişlerdir. Üç karma modeli parametrelerini en çok olabilirlik yönteminde EM algoritmasını kullanarak tahmin etmişlerdir. Üç karma modeli enzim ve depresif durum veri setlerine uyarlamışlar ve model seçiminde K-S test istatistiğini kullanmışlardır. Enzim veri setine birinci karma

model daha uygun iken depresif durum veri setine ise üçüncü karma model daha uygun olduğunu göstermişlerdir.

En çok olabilirlik tahminlerinin elde edilmesinde kullanılan eşitlikler genelde kapalı formdadır ve bundan dolayı analitik olarak doğrudan çözülemezler. Bu eşitliklerin çözümünde iteratif yöntemler kullanılmaktadır. Bu yöntemlerde parametrelerin başlangıç değerlerinin seçimi, parametre tahminlerinin doğruluğu için son derece önemlidir.

Karmaların parametrelerinin tahmininde momentler ve en çok olabilirlik yöntemleri dışında farklı yöntemlerde kullanılmıştır. Kabir (1968), Üstel dağılımların karmasında parametre tahmini için yeni bir yöntem önermiştir. Bu yöntemde örnek uzağın sonlu aralığında tanımlanan bir örneklem istatistiği, momentler yöntemine benzer şekilde kurulan modelin beklenen değerine eşitlenir. Tallis ve Light (1968) momentler yönteminde düzenli tamsayı momentleri yerine kesirli momentlerin kullanımını incelemişler ve iki Üstel dağılımın karmasının parametrelerini bu yöntemle tahmin etmişlerdir.

Karmanın parametrelerinin tahmini için geliştirilen yöntemlerin bazılarında belli bir hata fonksiyonu minimize edilerek parametreler tahmin edilmeye çalışılmaktadır. Bu hata fonksiyonu karmanın ampirik ve teorik fonksiyonu arasındaki farkların kareleri toplamıyla ifade edilebilir. Bu yaklaşımın moment çıkartan fonksiyonu içerdiği Quandt ve Ramsey (1978) tarafından belirtilmiştir. Bu yaklaşımda karakteristik fonksiyonlar ise Kumar ve ark (1979) tarafından çalışılmıştır. Ayrıca Redner ve Walker (1984) birikimli dağılım fonksiyonunu kullandığı minimum χ^2 yöntemi bu yöntemler arasında gösterilebilir.

Shawky ve Bakoban (2009), iki bileşenli üstelleştirilmiş Gamma dağılımın karma modelini önermişlerdir. İki bileşenli üstelleştirilmiş Gamma dağılımın karma modelinin karakteristik fonksiyonlarını incelemişlerdir. Karma dağılımın parametre tahmininde en çok olabilirlik metodu ile LINEX kayıp fonksiyonu ve karesel hata kaybı altında Bayes metodunu kullanmışlardır. Simülasyon çalışmasıyla en çok olabilirlik ve Bayes metodunu karşılaştırmışlardır. Karşılaştırma sonucunda iki bileşenli Gamma dağılımın karma modelinin parametre tahmininde LINEX

fonksiyonu gibi bir kayıp fonksiyonu altında Bayes metodunun en çok olabilirlik metodundan daha iyi olduğunu belirtmişlerdir.

Sınıflama problemlerinde kuramsal olarak ilgi çeken bir diğer yöntemde Bayes tahmini olarak bilinir. Bu yöntemin temeli bir önsel olasılık yoğunluğudur ve $p(\theta | X^n)$ ile gösterilir. Burada $X^n = [\mathbf{x}_1, \dots, \mathbf{x}_n]$ gözlemlerin setidir. Bayes yöntemi tek bileşenli dağılımların parametrelerinin tahmininde kullanışlı olmasına rağmen karma dağılım modellerinde gözlemlenmemiş etiket vektörü ile ilgilenildiğinde karamanın parametrelerinin tahmininde kullanışlı değildir.

Kececioğlu ve Wang (1998), karma Weibull dağılım modelinin parametrelerinin tahmininde Bayes teoremi ile en küçük kareler yönteminin kombinasyonunu kullanmışlardır. Parametrelerin başlangıç değerlerinin grafik yöntemi ile elde edildiği bu yöntemin parametre tahmininde etkili bir yöntem olduğunu belirtmişlerdir.

Ling ve Pan (1998), karma dağılımların parametre tahminleri için optimizasyon yöntemini önermişlerdir. Optimizasyon yönteminde parametre tahminleri, Kolmogorov-Smirnov (K-S) test istatistiğini minimum yapmaya çalışarak gerçekleşmektedir. Bu yöntemde parametrelerin başlangıç değerleri grafik metodu ile elde edilmektedir. Bu yöntemi iki üç parametrelili Weibull dağılımların toplamına ve bazı yaşam dağılımlarına başarıyla uygulamışlardır. Sonuçlara dayanarak uygun modelin seçiminde ve parametre tahmininde yeni bir prosedür sunmuşlardır.

Jiang ve ark (1999), karma Weibull dağılım modeli ve karma ters Weibull dağılım modelinde karma oranların negatif olma durumunu incelemişlerdir. Simülasyonla üretilen bileşen sayısı bilinmeyen sistem başarısızlığı veri setine her iki modelin uygunluğunu göstermişlerdir.

Al-Hussaini ve ark (2000), Gompertz dağılımını bir büyüme modeli olarak kullanmışlardır. Heterojen bir popülasyonu iki bileşenli Gompertz yaşam zamanı modelinin sonlu karmasıyla temsil etmişlerdir. İki bileşenli Gompertz dağılımın karmasının parametrelerinin en çok olabilirlik ve Bayes tahminlerini yaşam ve hazard oran fonksiyonlarını elde etmişlerdir. Karma dağılımın parametrelerinin en

çok olabilirlik ve Bayes tahminlerini Monte Carlo simülasyon çalışması ile karşılaştırmışlardır.

Marin ve ark (2005), bilinmeyen sayıda bileşenli Weibull dağılımının karmasının parametrelerini tahmin etmede Bayes metodunun kullanımını göstermişlerdir. Ayrıca yaşam fonksiyonunun ve beklenen hazard fonksiyonunun Markov Zinciri Monte Carlo (MCMC) algoritmasıyla nasıl elde edildiğini göstermişlerdir.

Farklı birçok alanda geniş bir uygulamaya sahip olan karma dağılım modelleri ile ilgili sağlık alanında da oldukça fazla çalışma bulunmaktadır. Clark ve ark (1968) kan basıncı, Raper ve ark (1983) iki kitlenin plazma düzeyleri için normal karmalarda en çok olabilirlik yöntemini ve Thomas (1966) sinir hücrelerinin boşalma zamanının dağılım modeli için Üstel karma dağılımını kullanmıştır.

Suchindran ve Lachenbruch (1974) Üstel karmalarda gebe kalma zamanları ile ilgili uygulamasında en çok olabilirlik yöntemini, Masuyama (1977) arteritin başlama yaşı için Gamma karma dağılım modelinin parametrelerinin tahmininde momentler yöntemini kullanmıştır.

Marks ve Rao (1979), dış basınç kuvvetleri için normal karmalarda momentler yöntemini uygulamıştır.

Hougaard (1984), Aalen (1988) bireyler ya da gruplar arasında yaşam uzunluklarındaki farklılıkları test etmek için Üstel karmalardan yararlanmışlardır.

Farawell ve Sprott (1988), λ ve $\beta\lambda$ ortalamalı iki Poisson dağılımından oluşan karma dağılım modeli ile prematüre ventriküler kompleks veri setini modellemişlerdir. Modelin parametre tahmini ve analizi oldukça sade olmasına rağmen veri setini modellemede etkili olmuştur.

Bogardus ve ark (1989) bir boyutlu üç bileşenli normal karma dağılımı Arizona'da Gila nehri civarında yaşayan insanların kanlarındaki şeker oranını kontrol eden insülin hormonu dağılımının modellenmesinde kullanmışlardır.

Gordon (1990), kanserden ve diğer nedenlerden ölen Üstel dağılıma sahip iki gruptan oluşan kanser hastaları yığını analiz etmek için Üstel karma modelini kullanmıştır.

Shepherd ve ark (2000) tedaviler arasında DNA'ya zarar veren farklılıkları belirlemek için Gamma karma dağılımını kullanarak, parametreleri en çok olabilirlik yöntemiyle tahmin etmişlerdir.

Pavlic ve ark (2001) tedaviye cevap vermeyen hastaların olasılığını tahmin etmek için normal karma model kullanmışlardır.

Savanur ve ark (2002) tüberküloz hastalığının bulaşma riski için üç bileşenli normal karma dağılım modeli kullanmışlar ve bileşenlerin parametrelerini EM algoritması ile tahmin etmişlerdir.

Bebbington ve ark (2007), esnek Weibull (FW) dağılımı ile katkısı azaltılmış Weibull (RAW) dağılımının karmasını önermişlerdir. Çalışmalarında önerdikleri karma dağılımın karakteristik özelliklerini incelemişlerdir. Önerdikleri karma dağılım modelini genellikle yaşlanma ve ölüm sürelerinin modellendiği Gompertz-Makeham dağılımıyla karşılaştırmışlar ve karma dağılım modelini Gompertz-Makeham dağılımına göre daha esnek olduğunu göstermişlerdir.

Son dönemde karma dağılım modelleri ile ilgili çok sayıda yapılmış yüksek lisans ve doktora çalışması bulunmaktadır. Stephen (1993), sonlu karma dağılım modellerinde parametre tahmin yöntemlerini incelemiştir. Üstel dağılıma ait dağılımların karma modellerinde en çok olabilirlik ve momentler yönteminin incelendiği çalışma her iki tahmin yönteminin avantaj ve dezavantajlarını göstermiştir.

Patra (2000), yaşam analizi ve güvenilirlikte yeni yaklaşımlar isimli doktora çalışmasında iki değişkenli Üstel dağılımların karması ve Weibull dağılımlarının karma modeli ile ilgilenmiştir. Meeker ve LuValle (1995) tarafından tanımlanan veri setinin modellenmesinde iki değişkenli Üstel dağılımların karması ve Weibull dağılımlarının karmasını kullanmıştır.

Wondmagegnehu (2002), iki artan lineer bozulma oranlarının karma modeli ve Weibull dağılımların karma modelini incelemiştir. Artan Weibull bozulma oranı ile Üstel dağılım karma modelinin incelendiği çalışmada elde edilen hazard fonksiyonların özelliklerini göstermiştir.

Li (2003), tek değişkenli normal ve Gamma dağılımlarının karma modellerinin özelliklerini incelemiştir. Karma normal dağılım modelinde bileşenlere

ait standart sapmaların aynı veya farklı olması durumu ile karma Gamma dağılım modelinde bileşenlerin ölçek parametrelerinin aynı olması durumunu incelemiştir.

Liu (2003), Üstel ve Weibull dağılımlarının karma modellerinde tam ve sansürlü veri durumu için en çok olabilirlik tahmin edicilerinin özelliklerini incelemiştir. İki bileşenli Üstel ve iki bileşenli Weibull karma dağılım modellerinde parametrelerin en çok olabilirlik tahminlerinin elde edilmesinde kullanılan EM algoritmasının etkinliği simülasyon çalışması ile göstermiştir.

Barger (2006), iki ve üç bileşenli karma Üstel dağılım modellerinde parametre tahmini için EM algoritmalarını tanımlamıştır. Veri modellemede karma Üstel dağılım modelinin, standart tek bileşenli Üstel dağılım modeline göre daha başarılı olduğunu göstermiştir.

Ye (2006), iki bileşenli Üstel karma dağılım modelinde parametre tahmini ve parametrelerin anlamlılık testi üzerinde çalışmıştır. Sansürlü veri durumunda log olabilirlik fonksiyonları tanımlayarak parametre tahminlerinin elde edilişleri gösterilmiştir. İçerisinde iki grup barındıran bir gerçek veri setini ilgili karma dağılım modelleri ile modellemiş ve iki kitlenin farklılığının anlamlılığını test etmiştir.

Zhang (2008), yaşam analizinde yarı parametrik karma dağılım modellerinin kullanımını incelemiştir. Kouass-Singh yarı parametrik hazard fonksiyonu ve yarı parametrik yaşam fonksiyonunu baş ve boyun kanseri verisine uygulamıştır. Parametre tahmininde EM algoritmasının kullanıldığı çalışmada ayrıca EM algoritmasının etkinliği simülasyon çalışması ile göstermiştir.

Dağılımların birleştirilmesi yöntemlerinden biri de dağılımların parametresini farklı bir dağılımla karıştırmaktır. Bu yöntem ile basit bir forma sahip olan dağılımlardan daha kullanışlı ve daha esnek dağılımlar elde edilebilir (Douglas 1980). Karıştırılmış dağılımlarla ilgili çalışmalar 1920 yılına kadar gitmektedir. Greenwood ve Yule (1920), Poisson dağılımının parametresini Gamma dağılımı olarak ele almışlar ve Pascal (Negatif Binom) dağılımını elde etmişlerdir. Satterthwaite (1942), dağılımları karıştırma yöntemiyle genelleştirilmiş Poisson dağılımını tanımlayarak dağılımın özelliklerini incelemiştir. Feller (1943), dağılımların birleştirilmesi ile elde edilen bazı dağılımları karıştırılmış dağılım olarak tanımlamıştır. Gurland (1957) karıştırılmış dağılımların özelliklerini incelemiştir.

Adamidis ve Loukas (1998), azalan başarısızlık oranlı iki parametrelili Üstel-Geometrik (EG) dağılımını önermişlerdir. Bu dağılımın özelliklerini göstermiş ve parametrelerini en çok olabilirlik metodunda EM algoritmasını kullanarak tahmin etmişlerdir. Önerdikleri EG dağılımını iki veri setinin modellenmesinde Gamma ve Weibull dağılımlarıyla karşılaştırmışlardır. Karşılaştırma sonucunda önerdikleri dağılımın daha iyi sonuç verdiğini belirtmişlerdir.

Adamidis ve ark (2005), artan ve azalan başarısızlık oranlı iki parametrelili genelleştirilmiş Üstel-Geometrik (EEG) dağılımını önermişlerdir. EEG dağılımının karakteristik fonksiyonlarını ve güvenilirlik fonksiyonunu incelemişlerdir. Dağılımın parametrelerini en çok olabilirlik metodunda EM algoritmasını kullanarak tahmin etmişlerdir. Algoritmanın parametre tahminindeki etkinliği simülasyon çalışması ile gösterilmişlerdir. Önerdikleri EEG dağılımını iki veri setinin modellenmesinde Gamma ve Weibull dağılımlarıyla karşılaştırmışlardır. Karşılaştırma sonucunda önerdikleri dağılımın daha iyi sonuç verdiğini belirtmişlerdir.

Literatür incelendiğinde dağılımların karıştırılması ile ilgili son dönemde de çok sayıda çalışmanın bulunduğu görülmektedir. Murat ve Szynal (2000) ile Grubbström ve Tang (2006) karıştırılmış dağılımların momentlerinin hesaplanması ile ilgili çalışmışlardır. Liu ve ark (2002), Poisson-Gumbel karıştırılmış dağılımını ve uygulamasını incelemişlerdir. Karlis ve Xekalaki (2005), karıştırılmış Poisson dağılımını çalışmışlardır. Liu ve ark (2006), karıştırılmış Uç değer dağılımını ve dağılımın deniz bilimlerinde uygulamalarını çalışmışlardır.

Tahmasbi ve Rezaei (2008), azalan başarısızlık oranlı iki parametrelili Üstel-Logaritmik (EL) yaşam zamanı dağılımını önermişlerdir. EL dağılımının karakteristik fonksiyonlarını ve güvenilirlik fonksiyonunu incelemişlerdir. Dağılımın parametrelerini en çok olabilirlik metodunda EM algoritmasını kullanarak tahmin etmişlerdir. Algoritmanın parametre tahminindeki etkinliği simülasyon çalışması ile göstermişlerdir. Önerdikleri EL dağılımını üç veri setinin modellenmesinde EG, Üstel-Poisson (EP), Gamma ve Weibull dağılımlarıyla karşılaştırmışlardır. Karşılaştırma sonucunda önerdikleri dağılımın daha iyi sonuç verdiğini belirtmişlerdir.

3. YAŞAM ANALİZİNDE TEMEL TANIM VE GÖSTERİMLER

Yaşam analizi (survival analysis), başarısızlık zaman analizi (failure time analysis) ya da olay zamanı analizi (event time analysis) olarak da isimlendirilir ve ömür süreleri (lifetimes), bekleme zamanları (waiting times) ya da daha genel olarak tanımlanmış bir olayın meydana gelmesine kadar geçen zaman süresinden oluşan verilerin analizinde kullanılır. Yaşam analizi çalışmaya dahil olan birimlerin yaşam süreleri ile ilgilidir. Yaşam süresi, tanımlanan herhangi bir olayın başlaması ve sona ermesi arasında geçen süredir. Böyle birçok olay tanımlanabilir. Örneğin demografik bir araştırmada çalışmaya konu olan bireylerin yaşam süreleri incelenebildiği gibi evli çiftlerin evli kalma süreleri de analiz konusu olabilir. Yöntem sadece canlı birimler değil, cansız birimler içinde uygulanabilir. Herhangi bir aletin bozulma ya da işlevini göremez hale gelme süresi, farklı sıcaklık dereceleri altında çalışan motorların bozulma süreleri de incelenebilir. Yaşam analizinde genellikle, canlı ya da cansız bir nesnenin veya olayın yaşam süresi T rassal değişkeni ile gösterilir. Her bir birime ait yaşam süresi t , tanımı gereği daima sıfırdan büyük olur. Analizde temel olan bozulma zamanlarının veya yaşam sürelerinin incelenmesi olduğundan bu değişkenlerin duyarlı bir biçimde tanımlanması gerekir (Başar 1993). İlgilenilen olaya göre farklı şekillerde ortaya çıkabilen bu değişkenin duyarlı olarak ölçülebilmesi için,

- Başlangıç zamanı her birim için şüpheye yer vermeyecek şekilde tanımlanmalıdır.
- Ölçülen zaman için bir ölçek olmalıdır (Dakika, saat, gün, hafta, ay, yıl veya taşıt ömrü için kilometre gibi).
- Her birim için ömrün sona ermesi ya da başarısızlığın meydana geldiği an tamamen net olmalıdır (Cox ve Oakes 1984).

Yaşam süresinin dağılımı genellikle üç fonksiyon ile karakterize edilir.

- i. Yaşam fonksiyonu
- ii. Olasılık yoğunluk fonksiyonu
- iii. Hazard fonksiyonu

Uygulamalarda genellikle bu üç fonksiyon verilerin farklı yönlerini göstermek için kullanılır. Yaşam analizinde temel sorun, yaşam süresi hakkında çıkarsamalar yapabilmek için yaşam süresi değişkeninin uygun bir dağılımla temsil edilmesini sağlamaktır. Bu bölümde yaşam analizindeki temel üç fonksiyon ve bu fonksiyonlar arasındaki ilişkiler tanımlanacaktır.

3.1. Yaşam Analizinde Önemli Fonksiyonlar

3.1.1. Yaşam Fonksiyonu

Yaşam süresi rassal değişkeni T ile ifade edildiğinde yaşam süresi fonksiyonu $S(t)$ ile gösterilir. Bir birimin belli bir t zamanından daha fazla yaşaması olasılığına yaşam fonksiyonu denir ve

$$S(t) = P(T > t), \quad 0 < t < \infty \quad (3.1)$$

şeklinde ifade edilir. Yaşam fonksiyonu monoton azalan birikimli bir fonksiyondur ve

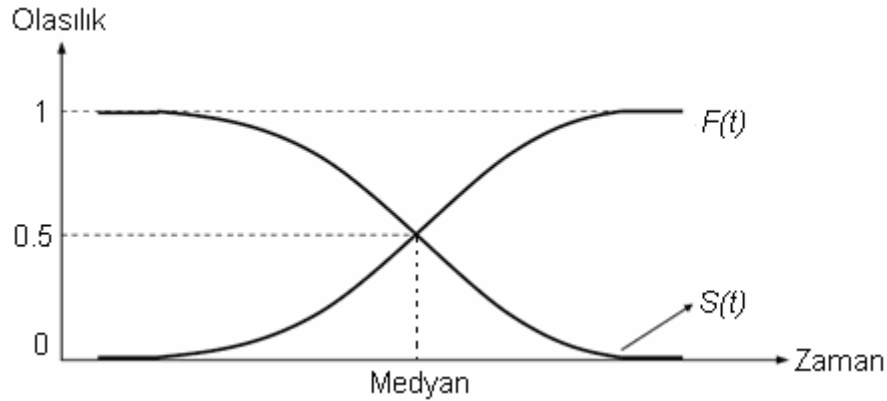
$$S(0) = \lim_{t \rightarrow 0} S(t) = 1 \quad (3.2)$$

$$S(\infty) = \lim_{t \rightarrow \infty} S(t) = 0 \quad (3.3)$$

zaman 0 olduğunda yaşam olasılığı 1 iken süre ilerledikçe bu olasılık düşer ve 0'a doğru gider (Leemis 1986). T 'nin birikimli dağılım fonksiyonu ise, t 'yi aşamayan yaşam süresinin ihtimalidir. Birikimli dağılım fonksiyonu

$$F(t) = P(T \leq t), \quad 0 < t < \infty \quad (3.4)$$

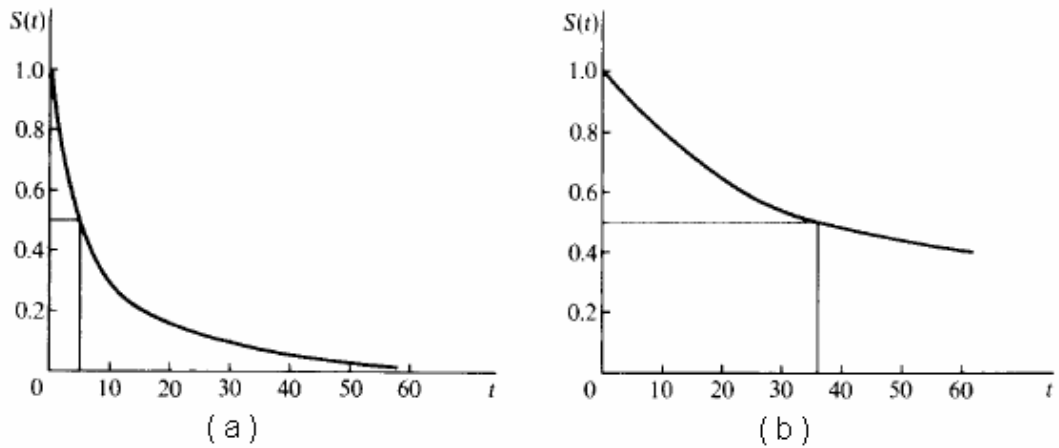
şeklinde gösterilir. Birikimli dağılım fonksiyonu ile yaşam fonksiyonu arasındaki ilişki Şekil 3.1'de gösterilmiştir.



Şekil 3.1. Birikimli dağılım fonksiyonu ile sağ kalım fonksiyonu arasındaki ilişki

Şekil 3.1'de görüldüğü gibi $F(t)$ artıyorken, $S(t)$ azalmaktadır. Bu iki fonksiyonun kesiştiği yer medyanı (ortanca) gösterir.

Yaşam fonksiyonu aynı zamanda birikimli yaşam oranı olarak da adlandırılır (Lee ve Wang 2003). Berkson (1942), yaşam fonksiyonlarının grafiksel gösterimlerinde yaşam eğrisi ismini kullanmıştır. Berkson'un (1942) tarafından örnek gösterilen farklı iki yaşam süresine ait yaşam eğrileri Şekil 3.2'de gösterilmiştir.



Şekil 3.2. Yaşam eğrisi için iki örnek (Berkson 1942)

Şekil 3.2.(a)'daki yaşam eğrisi düşük yaşam oranı ya da kısa yaşam süresini gösterirken Şekil 3.2.(b)'deki yaşam eğrisi yüksek yaşam oranı ya da uzun yaşam süresini göstermektedir.

Gözleme dayalı olarak yaşam fonksiyonu $S(t)$ genellikle uygulamalarda n birimlik bir örnekleme i , küçükten büyüğe doğru sıralanmış verilerde gözlem sırasını göstermek üzere

$$\hat{S}(t_{(i)}) = 1 - \frac{i}{n}, \quad i = 1, 2, \dots, n \quad (3.5)$$

$$\hat{S}(t_{(i)}) = 1 - \frac{i}{n+1}, \quad i = 1, 2, \dots, n \quad (3.6)$$

$$\hat{S}(t_{(i)}) = 1 - \frac{(i-0.5)}{n}, \quad i = 1, 2, \dots, n \quad (3.7)$$

$$\hat{S}(t_{(i)}) = 1 - \frac{(i-0.3)}{n+0.4}, \quad i = 1, 2, \dots, n \quad (3.8)$$

$$\hat{S}(t_{(i)}) = 1 - \frac{8i-3}{8i+2}, \quad i = 1, 2, \dots, n \quad (3.9)$$

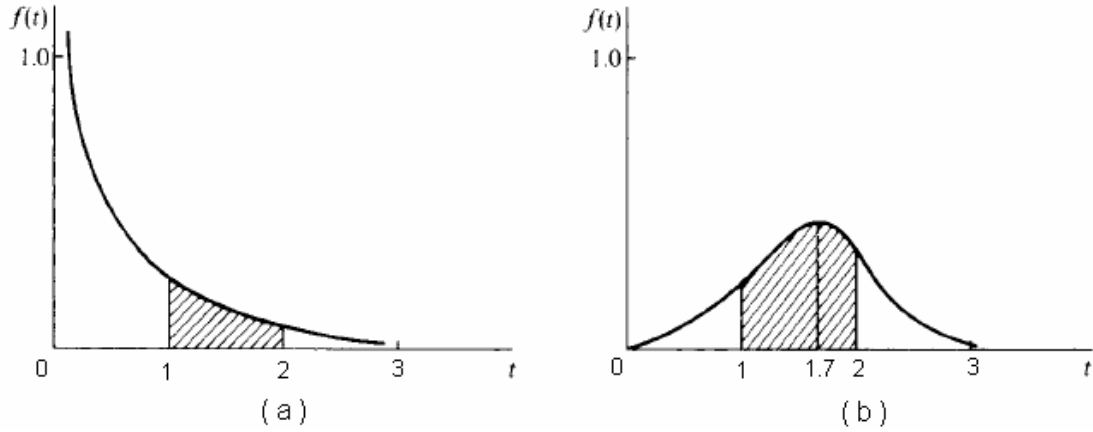
eşitliklerinden biri ile tahmin edilir.

3.1.2. Olasılık Yoğunluk Fonksiyonu

Yaşam süresi T rassal değişkeninin olasılık yoğunluk fonksiyonu, küçük bir zaman aralığında bir birimin başarısız olma olasılığının limitidir. Bu fonksiyon

$$f(t) = \lim_{\Delta t \rightarrow 0} \left[\frac{P(t \leq T \leq t + \Delta t)}{\Delta t} \right] \quad (3.10)$$

şeklindedir ve T rassal değişkeninin olasılık yoğunluk fonksiyonu t anında başarısızlığın yoğunluğu olarak yorumlanabilir (London 1988). Yaşam süresinin olasılık yoğunluk fonksiyonunun grafiği yoğunluk eğrisi olarak isimlendirilir. Şekil 3.3'de yoğunluk eğrilerine iki örnek verilmiştir.



Şekil 3.3. Yoğunluk eğrisi için iki örnek (Lee ve Wang 2003)

Şekil 3.3.(a)'daki yoğunluk eğrisi, başlangıçta yüksek başarısızlık oranına sahip iken zaman arttıkça başarısızlık oranının düştüğü durumu göstermektedir. Şekil 3.3.(b)'de yoğunluk eğrisi ise başarısızlık oranının zaman 1.7 değerine kadar arttığını, bu noktada en yüksek başarısızlık oranına ulaşıldığını ve bu noktadan sonra başarısızlık oranının zaman arttıkça düştüğü durumunu göstermektedir. Şekil 3.3'de $1 \leq t \leq 2$ zaman aralığındaki gölgeli alanlar başarısızlık oranını yani $P(1 \leq t \leq 2)$ olasılığını vermektedir.

$F(t)$ ve $S(t)$ zaman aralıkları ile ilgili olasılıklar olmasına rağmen $f(t)$ ise bir zaman noktası ile ilgilidir ve bir birimin t anında koşulsuz başarısız olma yoğunluğunu verir. Bir fonksiyonun olasılık yoğunluk fonksiyonu olabilmesi için

- i. $f(t)$, negatif olmayan bir fonksiyondur.

$$\begin{aligned} f(t) &\geq 0, & \forall t &\geq 0 \\ f(t) &= 0, & t &< 0 \end{aligned} \quad (3.11)$$

- ii. Yoğunluk eğrisi ile $-t$ ekseninde kalan alan 1'e eşittir. Yaşam süresi rassal değişkeni t sürekli ise

$$\int_0^{\infty} f(t) dt = 1 \quad (3.12)$$

özelliklerini taşıması gerekir.

3.1.3. Hazard Fonksiyonu

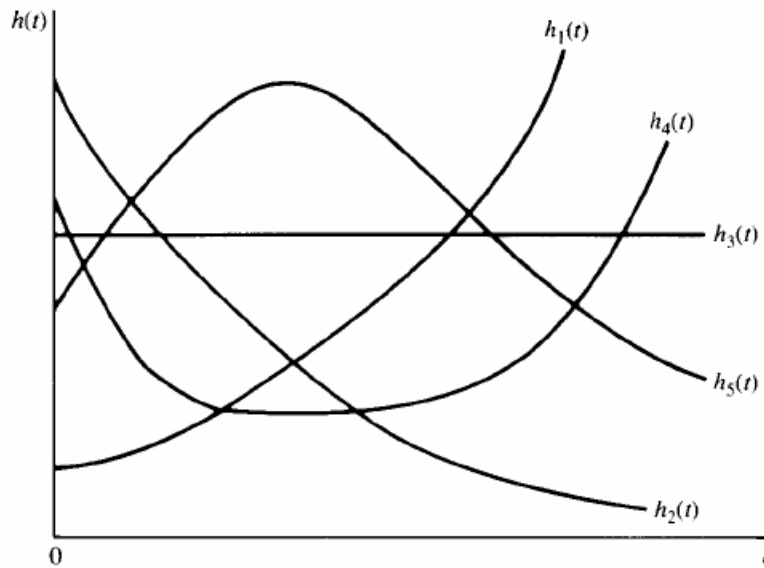
Hazard fonksiyonu T yaşam süresinin bir koşullu dağılımıdır ve genellikle $h(t)$ simgesiyle gösterilir. Bir birimin t zamanına kadar yaşadığı bilindiğinde belirli bir süre içinde (Δt) ölmesi olasılığına hazard (tehlike oranı) fonksiyonu denir. Yaşam süresi T rassal değişkeninin hazard fonksiyonu $h(t)$,

$$h(t) = \lim_{\Delta t \rightarrow 0} \left[\frac{P(t \leq T < (t + \Delta t) | T \geq t)}{\Delta t} \right] = \frac{f(t)}{S(t)} \quad (3.13)$$

şeklinde ifade edilir. Hazard fonksiyonu birikimli dağılım fonksiyonu cinsinden ise

$$h(t) = \frac{f(t)}{1 - F(t)} \quad (3.14)$$

olarak yazılır. Bu fonksiyon bir olasılık yoğunluk fonksiyonu değildir. Zamana göre artabilir, azalabilir veya sabit kalabilir. Şekil 3.4'de çeşitli türde hazard fonksiyonlarının grafiklerine yer verilmiştir.



Şekil 3.4. Örnek hazard fonksiyon grafikleri (Lee ve Wang 2003)

Şekil 3.4'deki $h_1(t)$ fonksiyonu artan hazard oranına sahiptir ve lösemi hastalarının tedaviye cevap vermeme riski artan hazard oranına örnek olarak gösterilebilir. Şekil 3.4'deki $h_2(t)$ fonksiyonu azalan hazard oranına sahiptir ve mermi ile yaralanan askerlerin ameliyat sonrası yaşam riski azalan hazard oranına örnek olarak gösterilebilir. Çünkü asıl tehlike ameliyattır ve ameliyat başarılı olursa hayati tehlike riski azalır. Şekil 3.4'deki $h_3(t)$ fonksiyonu sabit hazard oranına sahiptir ve ana ölüm riski kazalar olan sağlıklı 18–40 yaş aralığındaki insanların yaşam riski sabit hazard oranına örnek gösterilebilir. Şekil 3.4'deki $h_4(t)$ hazard fonksiyonu kuvvet eğrisi şeklindedir ve insan yaşamı bu hazard eğrisine sahiptir. Şekil 3.4'deki $h_5(t)$ hazard fonksiyonu başlangıçta artan daha sonra düşen hazard oranını göstermektedir. Şekil 3.4'deki $h_5(t)$ hazard fonksiyonuna örnek olarak tüberküloz hastalarının yaşam riski gösterilebilir. Tüberküloz hastalığının başlangıcında yaşam tehlikesi giderek artarken, tedavi sonrası yaşam riski giderek azalmaktadır.

Belirli bir süre içindeki ölüm olasılıklarının birikimli fonksiyonuna birikimli hazard fonksiyonu denir. Bu fonksiyon

$$H(t) = -\log(S(t)) = \int_0^t h(u)du \quad (3.15)$$

şeklinde gösterilir.

3.1.4. Yaşam Fonksiyonları Arasındaki İlişkiler

Yaşam fonksiyonu, olasılık yoğunluk fonksiyonu ve hazard fonksiyonu matematiksel olarak birbirleri cinsinden ifade edilebilirler. Yani, biri verildiğinde diğerlerine ulaşılabilir. Üç fonksiyon arasındaki matematiksel ilişki aşağıdaki adımlar kullanılarak kolaylıkla gösterilebilir.

1. Hazard fonksiyonu (3.13) de verildiği gibi olasılık yoğunluk fonksiyonunun yaşam fonksiyonuna bölünmesi ile elde edilir.

2. Birikimli dağılım fonksiyonunun türevi olasılık fonksiyonunu vereceğinden

$$f(t) = \frac{d}{dt}[1 - S(t)] = -S'(t) \quad (3.16)$$

eşitliği elde edilir.

3. (3.16) eşitliği (3.13) eşitliğinde yerine yazılırsa

$$h(t) = -\frac{S'(t)}{S(t)} = -\frac{d}{dt} \log S(t) \quad (3.17)$$

elde edilir.

4. (3.17) eşitliği 0'dan t 'ye integralenirse ve $S(0) = 1$ olacağından

$$-\int_0^t h(x) dx = \log S(t)$$
$$H(t) = -\log S(t) \quad (3.18)$$

$$S(t) = \exp[-H(t)] = \exp\left[-\int_0^t h(x) dx\right] \quad (3.19)$$

eşitlikleri elde edilir.

5. (3.13) ve (3.19) eşitliklerinden

$$f(t) = h(t) \exp[-H(t)] \quad (3.20)$$

eşitliği elde edilir.

Böylece $f(t)$ bilinirse yaşam fonksiyonu $f(t)$, $F(t)$ ve $S(t) = 1 - F(t)$ arasındaki ilişkiler yardımıyla tahmin edilir. Hazard fonksiyonu $h(t)$ ise, (3.13) eşitliği yardımı ile elde edilir. Eğer $S(t)$ biliniyorsa öncelikle (3.16) eşitliğinden $f(t)$ elde edilir. Daha sonra hazard fonksiyonu $h(t)$ elde edilir. Eğer $h(t)$ biliniyorsa $S(t)$ ve $f(t)$ sırasıyla (3.19) ve (3.20) eşitlikleri ile elde edilir. Böylelikle bu üç fonksiyondan biri verildiğinde diğer ikisine kolaylıkla ulaşılabileceği gösterilmiş oldu.

3.2. Yaşam Analizinde Kullanılan Bazı Önemli Dağılımlar

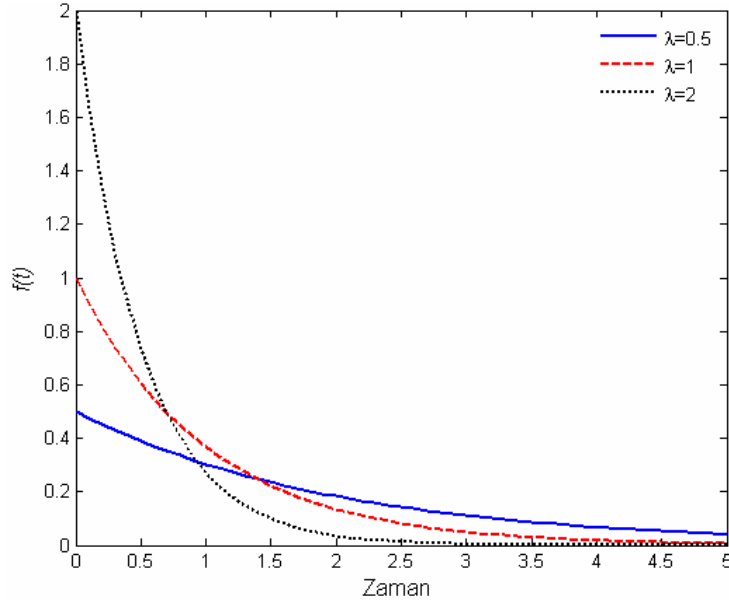
3.2.1. Üstel Dağılım

Üstel dağılım yaşam analizi çalışmalarında yaygın bir kullanıma sahip olan basit ve önemli bir dağılımdır. 1940'lı yılların sonlarında araştırmacılar elektronik sistem parçalarının yaşam sürelerini tanımlamak için Üstel dağılımı kullanmaya başladılar. Davis (1952), banka hesap ve defter hataları, bodro kontrol hataları, otomatik makine arızaları, radar seti bileşenlerindeki hatalar gibi örneklerde başarısızlıkların en iyi Üstel dağılım ile tanımlandığını belirtmişlerdir. Epstein ve Sobel (1953) çalışmalarında, tek sansürleme yapıldığında Üstel dağılımında parametre tahmini göstererek, normal dağılım yerine neden Üstel dağılımın seçilmesi gerektiğini göstermişlerdir. Aynı zamanda Epstein (1958) Üstel dağılım varsayımının yapılması için gerekçeleri ayrıntılı olarak ele almıştır. Üstel dağılım yaşam analizi ve istatistiğin diğer alanlarındaki önemini günümüze kadar sürdürmüştür.

Yaşam süresi T rassal değişkeni, λ parametresi ile Üstel dağılıma sahip ise olasılık yoğunluk fonksiyonu

$$f(t) = \begin{cases} \lambda e^{-\lambda t} & t \geq 0, \lambda > 0 \\ 0 & t < 0 \end{cases} \quad (3.21)$$

şeklindedir. Eşitlik (3.21)'den de görülebileceği gibi Üstel dağılımı tek parametre ile ifade edilir. Farklı λ parametre değerleri için olasılık yoğunluk fonksiyonu $f(t)$ 'nin grafikleri Şekil 3.5'de verilmiştir.



Şekil 3.5. Farklı λ parametre değerlerine göre Üstel dağılım için yoğunluk eğrileri

Üstel dağılımın birikimli dağılım fonksiyonu,

$$F(t) = 1 - e^{-\lambda t}, \quad t \geq 0 \quad (3.22)$$

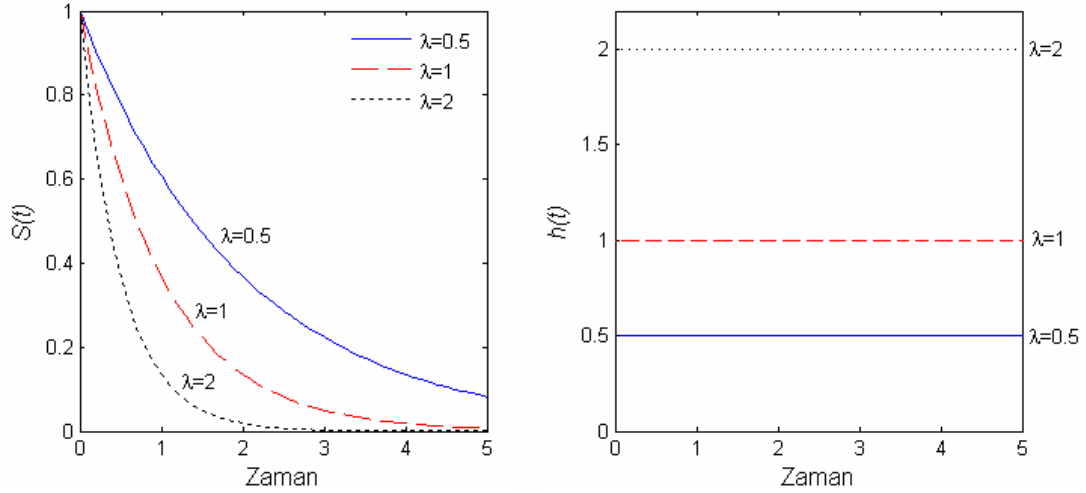
ve sağ kalım fonksiyonu

$$S(t) = e^{-\lambda t}, \quad t \geq 0 \quad (3.23)$$

eşitlikleri ile ifade edilir. Hazard fonksiyonu ise

$$h(t) = \frac{f(t)}{S(t)} = \frac{\lambda e^{-\lambda t}}{e^{-\lambda t}} = \lambda \quad (3.24)$$

şeklinde elde edilir. Farklı λ parametre değerlerine göre Üstel dağılım için elde edilecek yaşam eğrileri ve hazard fonksiyonunun grafikleri Şekil 3.6’da verilmiştir.



Şekil 3.6. Farklı λ parametre değerlerine göre Üstel dağılıma ait yaşam eğrileri ve hazard fonksiyonunun grafikleri

Şekil 3.6’daki Üstel dağılıma ait yaşam eğrileri incelendiğinde λ parametre değeri arttıkça yaşam olasılığının daha hızlı düştüğü görülmektedir. Üstel dağılımda hazard fonksiyonu sabittir. Üstel dağılımda λ parametresinin düşük değerler alması yaşam riskinin az olduğunu, büyük λ değerleri ise yaşam riskinin yüksek olduğunu gösterir.

Üstel dağılımı genellikle tamamen rassal başarısızlık desenlerinin modellenmesinde tercih edilmektedir. Üstel dağılımın yaşam analizindeki yaygın kullanımının bir nedeni belleksizlik özelliğidir (Lee ve Wang 2003). Belleksizlik özelliğine göre T rassal değişkeninin s ’den fazla yaşadığı bilindiğinde $s + c$ ’den fazla yaşama koşullu olasılığı, T rassal değişkeninin c ’den fazla yaşama olasılığına eşittir. Buna göre Üstel dağılımında belleksizlik özelliği,

$$P(T \geq s + c | T \geq s) = \frac{P(T \geq s + c, T \geq s)}{P(T \geq s)}$$

$$P(T \geq s + c | T \geq s) = \frac{P(T \geq s + c)}{P(T \geq s)}$$

$$P(T \geq s + c | T \geq s) = \frac{e^{-\lambda(s+c)}}{e^{-\lambda s}} = \frac{e^{-\lambda s} e^{-\lambda c}}{e^{-\lambda s}} = e^{-\lambda c}$$

$$P(T \geq s + c | T \geq s) = P(T \geq c) \quad (3.25)$$

şeklinde gösterilir. Üstel dağılıma ait beklenen değer $E(T)$, varyans $Var(T)$ ve medyan değeri sırasıyla aşağıda verilmiştir.

$$E(T) = \frac{1}{\lambda} \quad (3.26)$$

$$Var(T) = \frac{1}{\lambda^2} \quad (3.27)$$

$$Medyan = \frac{\ln 2}{\lambda} \quad (3.28)$$

3.2.2. Weibull Dağılımı

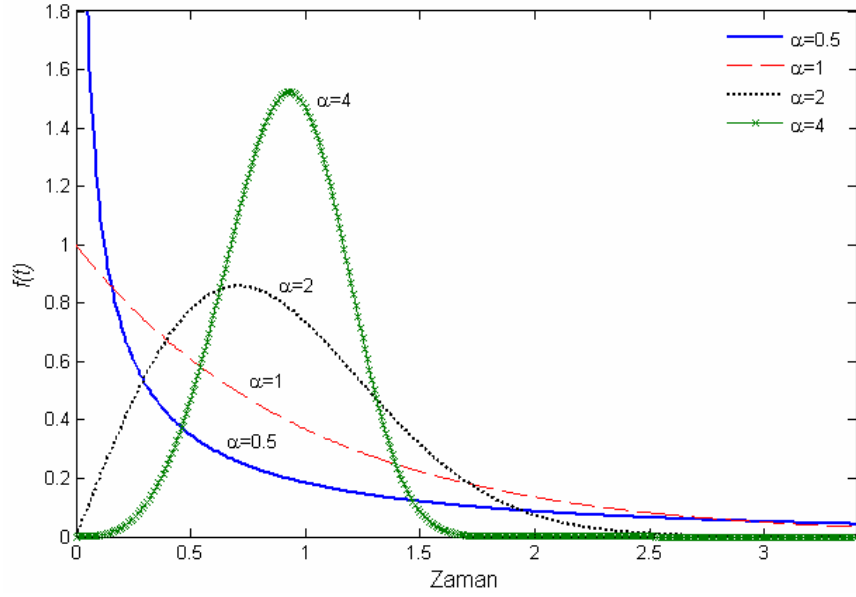
Weibull dağılımı ilk olarak 1939 yılında, dağılıma ismini veren İsveçli fizikçi, Walodi Weibull tarafından, materyallerin bozulma sürelerinin gösteriminde kullanılmıştır. Bu çalışma, modelin farklı karakterli birçok veri kümesi üzerindeki etkinliğini gösteren bir çalışmadır (Murthy ve ark, 2004). Walodi Weibull, 1951 yılında İngiltere’de yaptığı çalışmalarında, dağılımı; kalite kontrolünde, binaların ekonomik ömrü ve fabrika üretimi ile ilgili olarak kullanmıştır. Modeli, farklı bilim dallarına ait veri kümelerine uygulamış ve böylelikle farklı uygulama alanları üzerinde modelin çok yönlülüğünü göstermiştir. Benzer bir model, Rosen ve Rammeler tarafından, 1933 yılında, moleküller ile ilgili bir çalışmada kullanılmıştır. Weibull dağılımından söz eden en eski yayın, 1928’de Fisher ve Tippet tarafından yayınlanmıştır. Weibull; dağılımı öneren ilk insan olmamasına rağmen, modelin etkinliğinin ve uygulamalardaki çok yönlülüğünün vurgulanmasına en çok katkısı olan kişidir. Günümüzde çok geniş bir uygulama alanına sahip olan Weibull dağılımının, kalite kontrol ve güvenilirlik çalışmalarında kullanılması Kao tarafından savunulmuştur. 1958 yılında Kao, radyo lambalarının bozulma sürelerinin

gösteriminde, Weibull dağılımını kullanmıştır. Ayrıca Lieblein ve Zelen 1956 yılında bilyeli yatakların bozulma sürelerinin gösteriminde Weibull dağılımını kullanmıştır. Berretoni 1964 yılında, kimyasal aşınma testlerindeki (korozyon testi) çoğu uygulamaları, Weibull dağılımı ile tanımlamıştır. Weibull dağılımı, son dönemlerde bozulma süresi dağılımlarının en popüler üyesi haline gelmiştir. Plait (1962), Thompson (1968); bozulma durumunda dağılımın çok çeşitli uygulamalarını ele almıştır. Ayrıca elektronik parçalarda, yarı iletken cihazlarda, foto iletken hücrelerde, kondansatörlerde ve çeşitli biyolojik organizmalarda bozulma sürelerinin çözümlenmesinde Weibull dağılımı kullanılmıştır. Günümüzde Weibull dağılımı yaşam analizi uygulamalarının dışında; bekleme modellerinde, salgın hastalıkların sürme zamanının bulunmasında, deprem riskinin saptanmasında ve finansal uygulamalarda; yani rassal değişkenin pozitif değerler aldığı durumlarda kullanılabilirliğinden dolayı istatistiksel dağılımlar arasında popüler dağılımlardan biri haline gelmiştir.

Weibull dağılımı Üstel dağılımın genelleştirilmiş bir halidir. Bununla birlikte Üstel dağılıma benzemez. Weibull dağılımının hazard oranı sabit değildir ve bu nedenle Weibull dağılımı daha geniş bir uygulamaya sahiptir. Weibull dağılımı α ve β parametreleri ile karakterize edilir. Yaşam süresi T rassal değişkeni, α ve β parametreleri ile Weibull dağılımına sahip ise olasılık yoğunluk fonksiyonu

$$f(t, \alpha, \beta) = \frac{\alpha}{\beta} \left(\frac{t}{\beta} \right)^{\alpha-1} e^{-\left(\frac{t}{\beta} \right)^{\alpha}}, \quad t \geq 0 \quad (3.29)$$

ile verilir. Weibull dağılımında α parametresi biçim parametresidir ve dağılım eğrisinin biçimini belirler. Weibull dağılımının biçim parametresi $\alpha = 1$ alındığında Üstel dağılım, $\alpha = 2$ alındığında Rayleigh dağılımı elde edilir. Weibull dağılımında β parametresi ise ölçek parametresidir. Weibull dağılımında $\beta = 1$ ve farklı α parametre değerleri için olasılık yoğunluk fonksiyonu $f(t)$ 'nin grafikleri Şekil 3.7'de verilmiştir.



Şekil 3.7. Weibull dağılımında $\beta = 1$ ve farklı α parametre değerleri için yoğunluk eğrileri

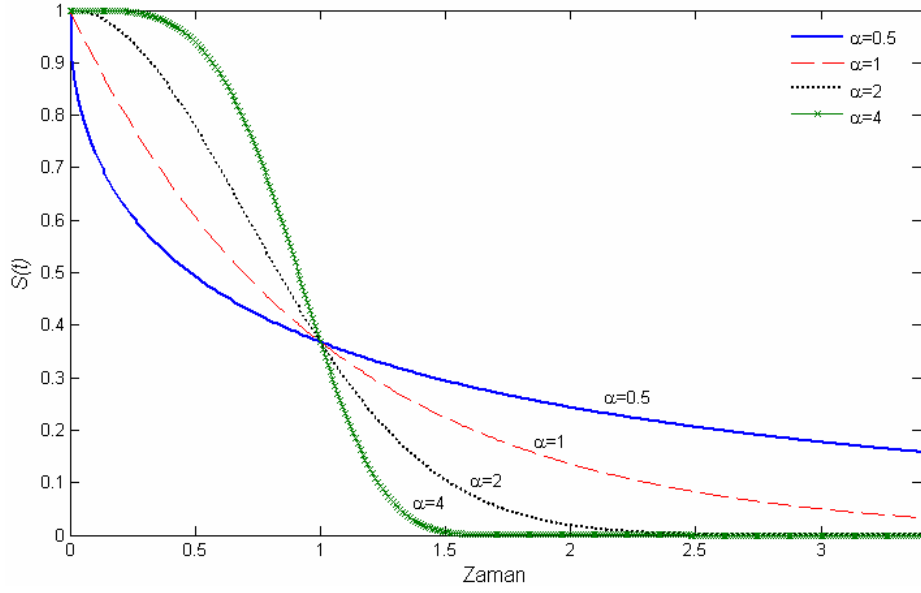
Weibull dağılımının birikimli dağılım fonksiyonu

$$F(t) = 1 - e^{-\left(\frac{t}{\beta}\right)^\alpha}, \quad t \geq 0 \quad (3.30)$$

ve yaşam fonksiyonu

$$S(t) = e^{-\left(\frac{t}{\beta}\right)^\alpha}, \quad t \geq 0 \quad (3.31)$$

olarak verilir. Weibull dağılımında $\beta = 1$ ve farklı α parametre değerleri için yaşam fonksiyonu $S(t)$ 'nin grafikleri Şekil 3.8'de verilmiştir.

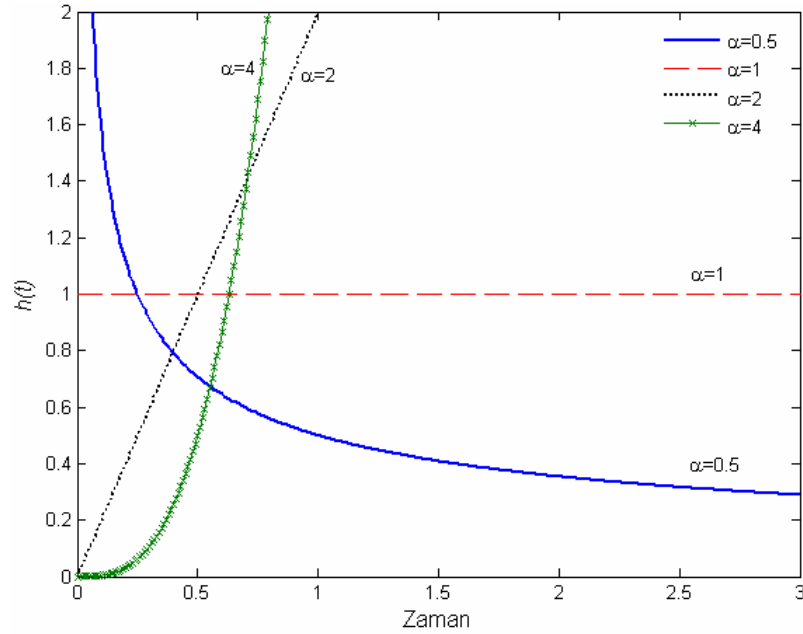


Şekil 3.8. Weibull dağılımında $\beta = 1$ ve farklı α parametre değerleri için yaşam eğrileri

Şekil 3.8 incelendiğinde biçim parametresi α 'nın değeri arttıkça yaşam eğrisinin sıfıra yaklaşma hızının arttığı açıkça görülmektedir. Weibull dağılımında hazard fonksiyonu,

$$h(t) = \frac{\alpha}{\beta} \left(\frac{t}{\beta} \right)^{\alpha-1}, \quad t \geq 0 \quad (3.32)$$

eşitliği ile verilir. Weibull dağılımında $\beta = 1$ ve farklı α parametre değerleri için hazard fonksiyonu $h(t)$ 'nin grafikleri Şekil 3.9'da verilmiştir.



Şekil 3.9. Weibull dağılımında $\beta = 1$ ve farklı α parametre değerleri için hazard fonksiyonunun grafikleri

Weibull dağılımının biçim parametresinin almış olduğu değerlere göre hazard fonksiyonu bazen monoton azalan, bazen monoton artan bazen de sabit olur. Şekil 3.9'da da görüldüğü gibi biçim parametresinin 1'den küçük olduğu durumlarda yani $\alpha < 1$ iken hazard fonksiyonu monoton azalan, biçim parametresinin 1'den büyük olduğu durumlarda yani $\alpha > 1$ iken hazard fonksiyonu monoton artan ve biçim parametresinin 1'e eşit olduğu durumlarda yani $\alpha = 1$ iken hazard fonksiyonu sabittir.

Gamma fonksiyonu

$$\Gamma(\alpha) = \int_0^{\infty} x^{\alpha-1} e^{-x} dx \quad (3.33)$$

olarak tanımlandığında Weibull dağılımında yaşam süresi T rassal değişkeninin r . beklenen değeri

$$E(T^r) = \int_0^{\infty} t^r \frac{\alpha}{\beta} \left(\frac{t}{\beta} \right)^{\alpha-1} e^{-\left(\frac{t}{\beta} \right)^{\alpha}} dt$$

$$E(T^r) = \beta^r \Gamma \left(1 + \frac{r}{\alpha} \right) \quad (3.34)$$

şeklinde elde edilir. Yaşam süresi T rassal değişkeni Weibull dağılımına sahip ise beklenen değeri (3.34) nolu eşitlikte $r = 1$ için

$$E(T) = \beta \Gamma \left(1 + \frac{1}{\alpha} \right) \quad (3.35)$$

şeklinde elde edilir. Yaşam süresi T rassal değişkeni Weibull dağılımına sahip ise varyansı

$$Var(T) = \beta^2 \left(\Gamma \left(1 + \frac{2}{\alpha} \right) - \Gamma^2 \left(1 + \frac{1}{\alpha} \right) \right) \quad (3.36)$$

eşitliği ile ifade edilir. Yaşam analizlerinde veri setleri tanımlanırken kullanılan medyan değeri Weibull dağılımında

$$Medyan = \beta (\ln 2)^{\frac{1}{\alpha}} \quad (3.37)$$

eşitliği ile tanımlanır.

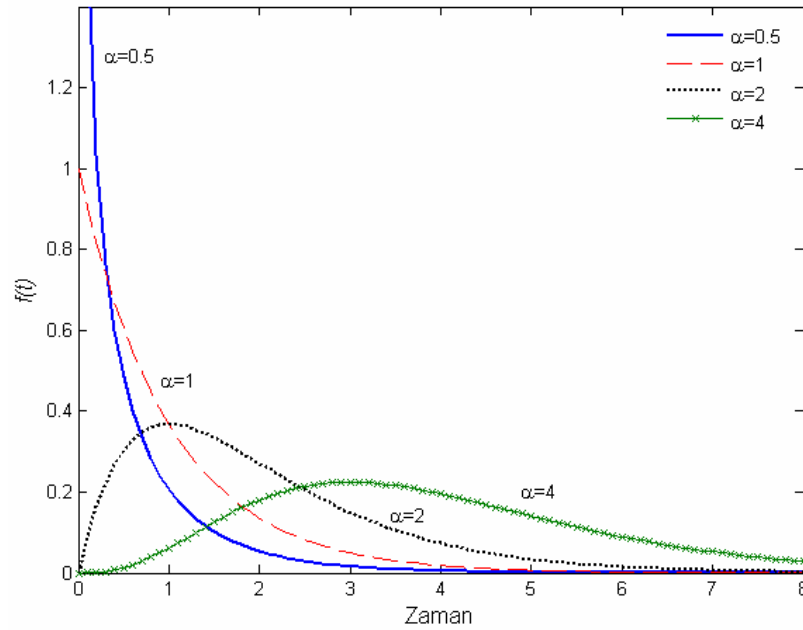
3.2.3. Gamma Dağılımı

Gamma dağılımı, Üstel ve Ki-kare dağılımını içeren bir dağılımdır. Gamma dağılımını Brown ve Flood (1947) kafeteryada cam bardakların kullanım süresini, Birnbaum ve Saunders (1958) materyallerin yaşam uzunluğunu Gamma dağılımı ile tanımlamışlardır. Gamma dağılımı özellikle endüstriyel güvenilirlik ve canlı yaşam

sürelerinin modellenmesinde sıklıkla kullanılan bir dağılımdır. Yaşam süresi T rassal değişkeni Gamma dağılımına sahip ise olasılık yoğunluk fonksiyonu

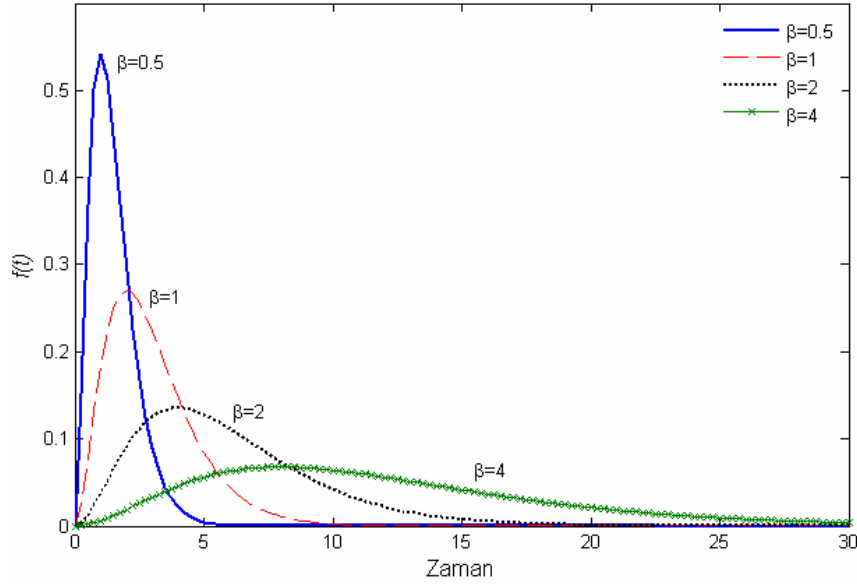
$$f(t; \alpha, \beta) = t^{\alpha-1} \frac{e^{-t/\beta}}{\beta^\alpha \Gamma(\alpha)}, \quad t > 0, \alpha, \beta > 0 \quad (3.38)$$

ile verilir. Gamma dağılımında $\beta = 1$ ve farklı α parametre değerleri için elde edilen yoğunluk eğrileri Şekil 3.10'da verilmiştir.



Şekil 3.10. Gamma dağılımında $\beta = 1$ ve farklı α parametre değerleri için yoğunluk eğrileri

Gamma dağılımında biçim parametresi $\alpha = 1$ alındığında Üstel dağılıma ulaşılır. Gamma dağılımında $\alpha = 3$ ve farklı β parametre değerleri için elde edilen yoğunluk eğrileri Şekil 3.11'de verilmiştir.



Şekil 3.11. Gamma dağılımında $\alpha = 3$ ve farklı β parametre değerleri için yoğunluk eğrileri

Gamma dağılımında biçim parametresi $\alpha = \nu/2$ ve ölçek parametresi $\beta = 2$ serbestlik derecesi ν olan Ki-kare dağılımına ulaşılır. Gamma dağılımının birikimli dağılım fonksiyonu

$$F(t) = \frac{\Gamma_x(\alpha)}{\Gamma(\alpha)}, \quad t > 0, \alpha > 0 \quad (3.39)$$

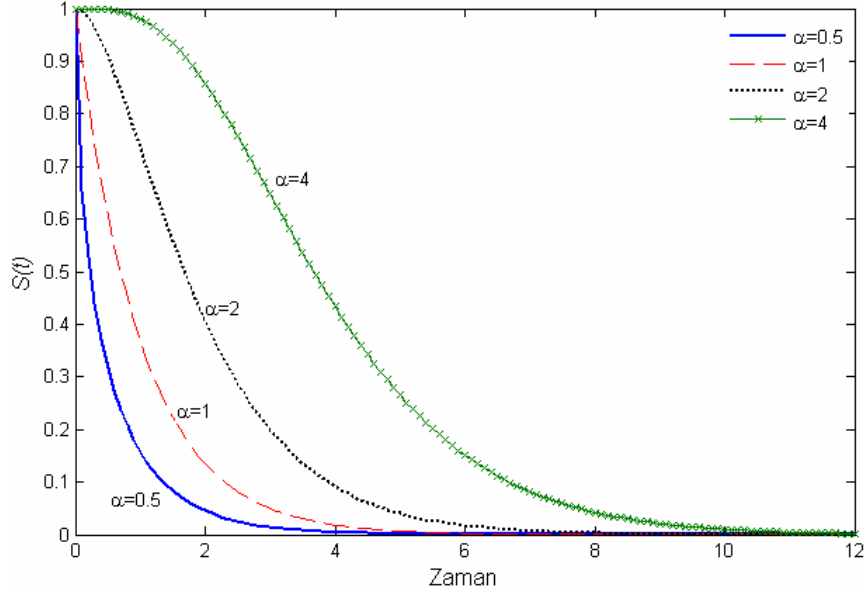
eşitliği ile tanımlanır. Eşitlikte yer alan $\Gamma_x(\alpha)$ tamamlanmamış gamma fonksiyonu olarak isimlendirilir ve

$$\Gamma_x(\alpha) = \int_0^x t^{\alpha-1} \exp(-t) dt \quad (3.40)$$

şeklinde hesaplanır. Gamma dağılımında yaşam fonksiyonu,

$$S(t) = 1 - \frac{\Gamma_x(\alpha)}{\Gamma(\alpha)}, \quad t > 0, \alpha > 0 \quad (3.41)$$

olarak verilir. Gamma dağılımında $\beta = 1$ ve farklı α parametre değerleri için elde edilen yaşam eğrileri Şekil 3.12’de verilmiştir.

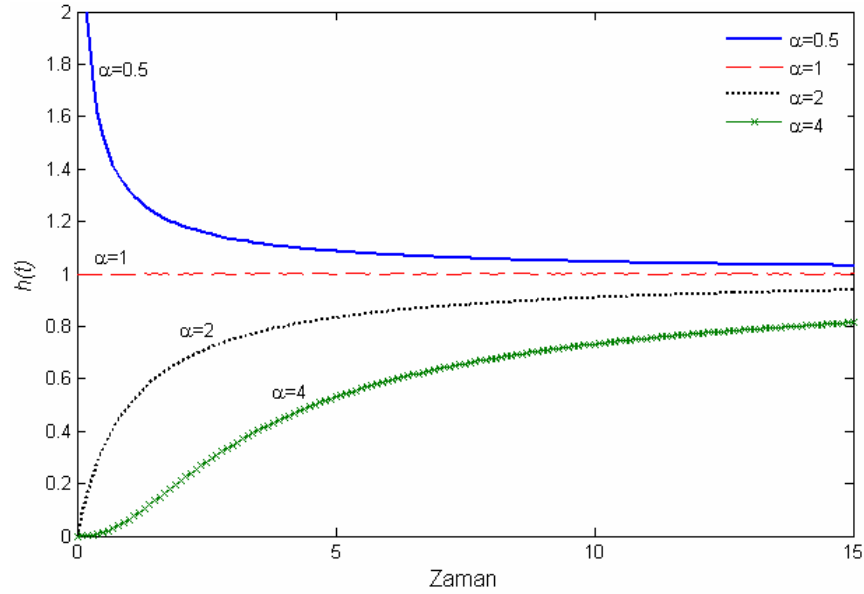


Şekil 3.12. Gamma dağılımında $\beta = 1$ ve farklı α parametre değerleri için yaşam eğrileri

Şekil 3.12 incelendiğinde Gamma dağılımında α parametre değeri küçüldükçe yaşam eğrisinin 0'a yaklaşma hızı artmaktadır. Gamma dağılımında hazard fonksiyonu $h(t)$

$$h(t) = \frac{t^{\alpha-1} e^{-\frac{t}{\beta}}}{\beta^{\alpha} (\Gamma(\alpha) - \Gamma_x(\alpha))} \quad (3.42)$$

olarak verilir. Gamma dağılımında $\beta = 1$ ve farklı α parametre değerleri için hazard fonksiyon eğrileri Şekil 3.13’de verilmiştir.



Şekil 3.13. Gamma dağılımında $\beta = 1$ ve farklı α parametre değerleri için hazard fonksiyonu eğrileri

Gamma dağılımında biçim parametresi $\alpha < 1$ ise hazard fonksiyonu monoton azalan, $\alpha > 1$ iken hazard fonksiyonu monoton artan ve $\alpha = 1$ iken hazard fonksiyonu sabittir. Gamma dağılımı için T rassal değişkeninin beklenen değeri ve varyansı sırasıyla,

$$E(T) = \alpha\beta \quad (3.43)$$

$$Var(T) = \alpha\beta^2 \quad (3.44)$$

şeklindedir.

3.2.4. Lognormal Dağılımı

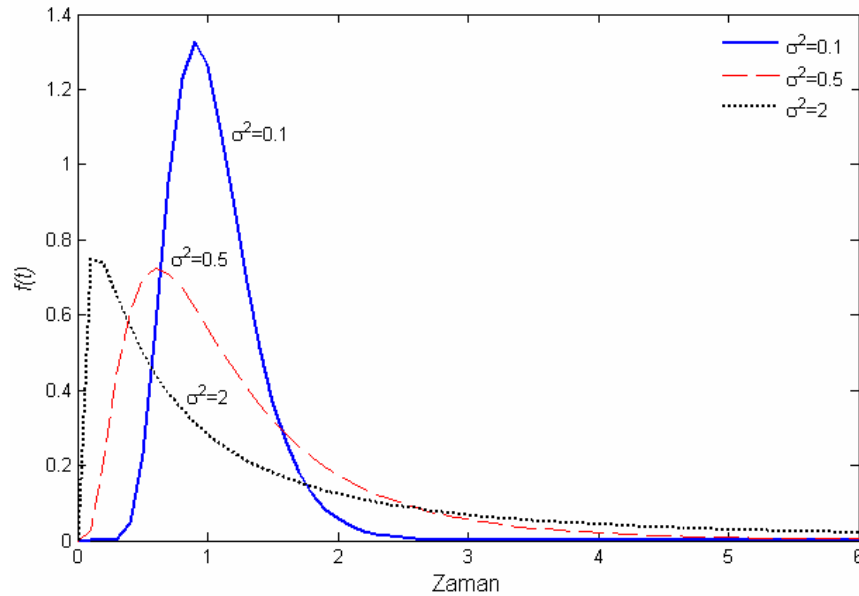
En basit şekliyle Lognormal dağılım, normal dağılıma sahip rassal değişkenin logaritmasının alınmasıyla oluşan yeni değişkenin dağılımı olarak tanımlanabilir. Lognormal dağılımı ilk olarak McAlister (1879) tarafından tanımlanmıştır. Lognormal dağılım birçok yönüyle çalışılan önemli bir dağılımdır. Gaddum (1945) biyoloji uygulamalarında, Boag (1949) kanser araştırmalarında Lognormal dağılımı

kullanmıştır. Aitchison ve Brown (1957) dağılımın özelliklerini, parametre tahminlerini ve ekonomi alanında kullanımını detaylı olarak tartışmışlardır. Sonraları araştırmacılar dağılımı Alzhemir hastalığının başlama yaşının gözlenmesinde, Alzhemir, Hodgkin ve kronik lösemi hastalarının yaşam sürelerinin modellenmesinde kullanmışlardır (Lee ve Wang 2003).

Yaşam süresi T rassal değişkeni Lognormal dağılıma sahip ise olasılık yoğunluk fonksiyonu,

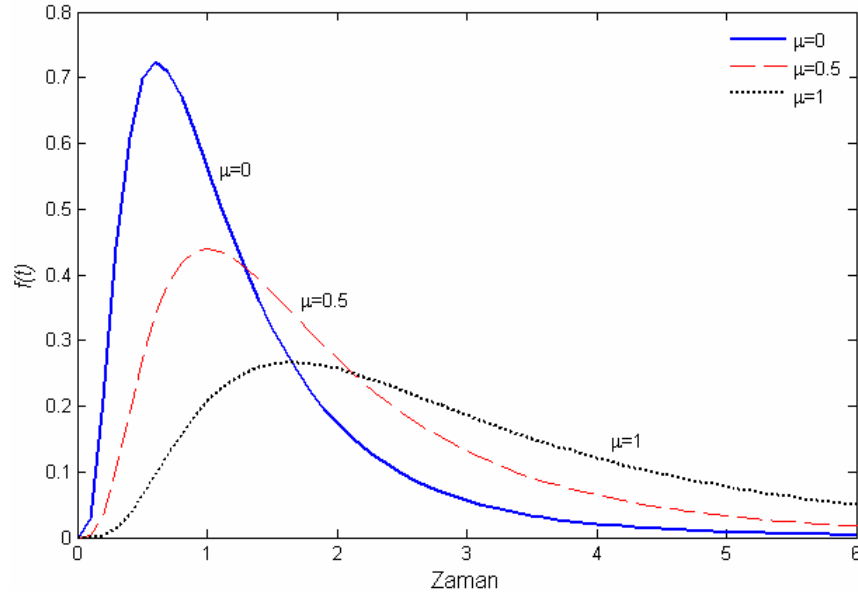
$$f(t) = \frac{1}{t\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\ln t - \mu}{\sigma}\right)^2}, \quad t > 0 \quad (3.45)$$

şeklinde verilir. Lognormal dağılımı konum parametresi μ ve ölçek parametresi σ ile karakterize edilir. Dağılımın esnekliğini göstermek için konum parametresi $\mu = 0$ ve σ^2 değeri sırasıyla 0.1, 0.5 ve 2 alındığında elde edilen yoğunluk eğrileri Şekil 3.14’de verilmiştir.



Şekil 3.14. Lognormal dağılımında $\mu = 0$ ve farklı σ^2 parametre değerleri için yoğunluk eğrileri

Lognormal dağılımda ölçek parametresi $\sigma^2 = 0.5$ ve konum parametresi μ sırasıyla 0, 0.5 ve 1 alındığında elde edilen yoğunluk eğrileri Şekil 3.15’de verilmiştir



Şekil 3.15. Lognormal dağılımda $\sigma^2 = 0.5$ ve farklı μ parametre değerleri için yoğunluk eğrileri

Lognormal dağılımda yaşam süresi T rassal değişkeni için birikimli dağılım fonksiyonu ve yaşam fonksiyonu sırasıyla

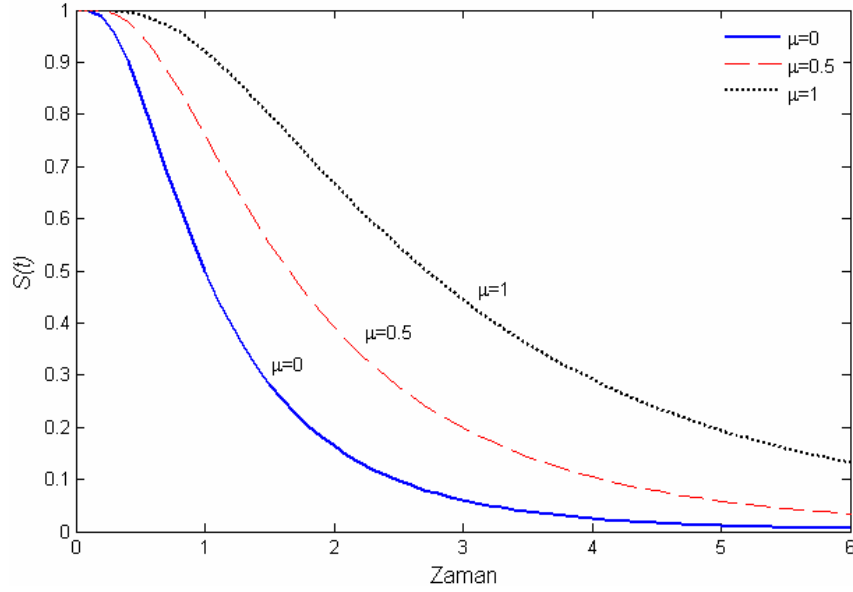
$$F(t) = \Phi\left(\frac{\ln t - \mu}{\sigma}\right) \quad (3.46)$$

$$S(t) = 1 - \Phi\left(\frac{\ln t - \mu}{\sigma}\right) \quad (3.47)$$

olarak verilir. Eşitliklerde yer alan Φ normal dağılımın birikimli dağılım fonksiyonudur ve

$$\Phi\left(\frac{\ln t - \mu}{\sigma}\right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{\ln t - \mu}{\sigma}} \exp\left(-\frac{u^2}{2}\right) du \quad (3.48)$$

şeklinde hesaplanır. Lognormal dağılımda ölçek parametresi $\sigma^2 = 0.5$ ve konum parametresi μ sırasıyla 0, 0.5 ve 1 alındığında elde edilen yaşam eğrileri Şekil 3.16'da verilmiştir.

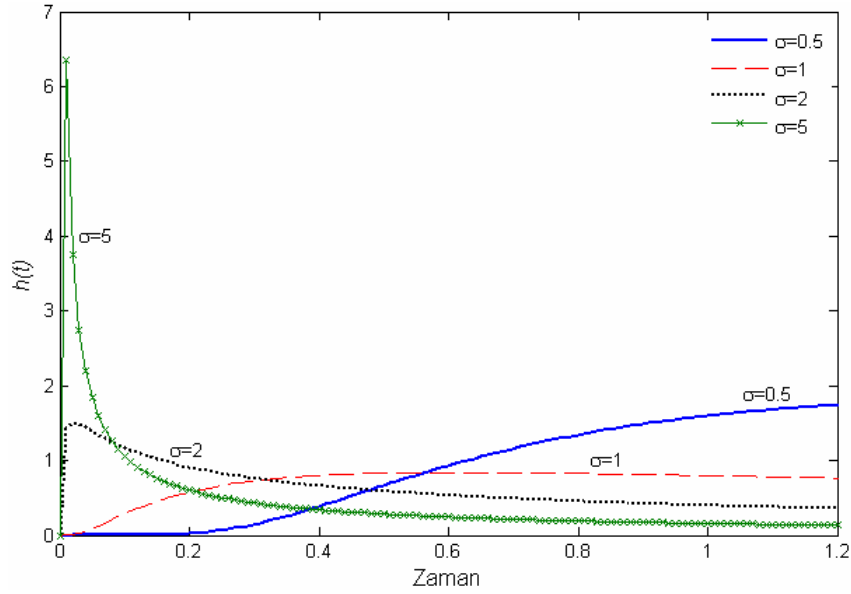


Şekil 3.16. Lognormal dağılımında $\sigma^2 = 0.5$ ve farklı μ parametre değerleri için yaşam eğrileri

Şekil 3.16 incelendiğinde Lognormal dağılımda konum parametresi μ 'nın değeri arttıkça yaşam olasılığı artmaktadır. Lognormal dağılımda hazard fonksiyonu

$$h(t) = \frac{\left(1/t\sigma\sqrt{2\pi}\right)\exp\left[-\frac{1}{2\sigma^2}(\ln t - \mu)^2\right]}{1 - \Phi\left(\frac{\ln t - \mu}{\sigma}\right)} \quad (3.49)$$

şeklinde elde edilir. Lognormal dağılımda konum parametresi $\mu = 0$ ve farklı σ parametre değerleri için elde edilen hazard fonksiyonlarının grafikleri Şekil 3.17'de verilmiştir.



Şekil 3.17. Lognormal dağılımında $\mu=0$ ve farklı σ parametre değerleri için hazard fonksiyonu eğrileri

Şekil 3.17 incelendiğinde Lognormal dağılımında farklı parametre değerleri için hazard fonksiyonun oldukça esnek olduğu görülmektedir. Lognormal dağılımı için T rassal değişkeninin beklenen değeri, varyansı ve medyan değeri sırasıyla,

$$E(T) = e^{\mu + \frac{\sigma^2}{2}} \quad (3.50)$$

$$Var(T) = (e^{\sigma^2} - 1)e^{2\mu + \sigma^2} \quad (3.51)$$

$$Medyan = e^{\mu} \quad (3.52)$$

şeklindedir.

4. KARMA VE KARIŞTIRILMIŞ DAĞILIMLAR

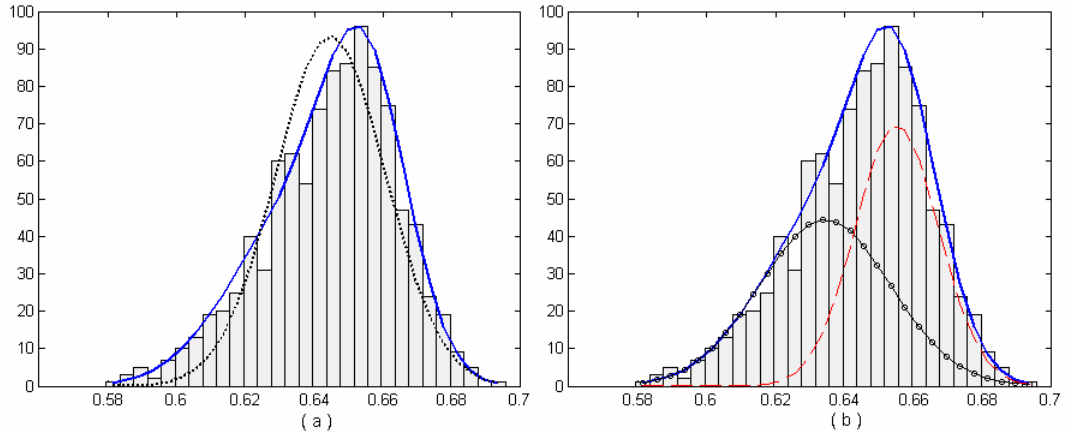
Yaşam analizinde değişkenliğin fazla olduğu yaşam sürelerinin analizinde yani heterojen yaşam süresi verilerin analizinde standart olasılık dağılımları yetersiz kalmaktadır. Bu nedenle son dönemlerde, klasik istatistik dağılımlarına göre daha esnek ve daha kullanışlı olan karma dağılım modellerinin heterojen yaşam sürelerinin analizinde kullanımı daha yaygın hale gelmiştir. Ayrıca dağılımı karakterize eden parametrelerden birinin rassal değişken olarak düşünülmesi ve bu parametre için dağılım varsayımı yapılmasıyla elde edilen çok sayıda yeni karıştırılmış dağılım yaşam sürelerinin modellenmesi için önerilmektedir. Bu bölümde son dönemde yaşam sürelerinin modellenmesinde kullanımı daha yaygın hale gelen karma ve karıştırılmış dağılımlar incelenecektir. Yaşam analizinde tek değişken içeren veri setlerinin modellenmesi ile ilgilenildiğinden burada karma dağılım modelleri sadece tek değişken durumu için ele alınacaktır.

4.1. Karma Dağılım Modelleri

İki ya da daha fazla bileşenden oluşan dağılımlar, karma dağılımlar olarak adlandırılır. Karma dağılım modelleri birçok alanda rasgelelik içeren doğal olayların farklı özellikleri hakkında toplanan ölçüm değerlerine istatistiksel olarak model oluşturmada matematiksel bir yaklaşım sağlar. Bu nedenlerden dolayı karma dağılım modellerinin hem teorik hem de uygulamalı birçok çalışma alanında önemi günden güne artmaktadır (Çalış 2005).

Karma dağılım modelleri ile ilgili ilk çalışma yaklaşık 120 yıl kadar önce ünlü biyometrici Karl Pearson tarafından yapılmıştır. Pearson (1894) yengeç verisi için birbirinden farklı iki tek değişkenli normal dağılımın karma modelini incelemiştir. Bu veriler Weldon (1892, 1893) tarafından ölçülmüştür. Weldon (1893) verilerin histogramı asimetric olduğundan ya da simetrik olmadığından dolayı, yengeç kitlesinin iki farklı türden oluştuğunu iddia etmiştir. Fakat matematiksel yönü zayıf olduğu için Karl Pearson'dan yardım istemiştir. Bu veri önce histogramda sınıf sayısı 29 alınarak bir tek normal dağılımla modellenmiştir. Daha sonra simetrik

olmayan yapı nedeniyle Şekil 4.1’de görüldüğü gibi normal dağılımların karması kullanılarak modellenmiştir.



Şekil 4.1.(a) Yengeç verisi için uydurulan tek bileşenli normal dağılım (nokta nokta) ve iki bileşenli normal dağılım (düz çizgi) eğrileri. (b) Yengeç verisi için uydurulan iki bileşenli normal dağılım (düz çizgi) ve her bir bileşene ait normal dağılım eğrileri

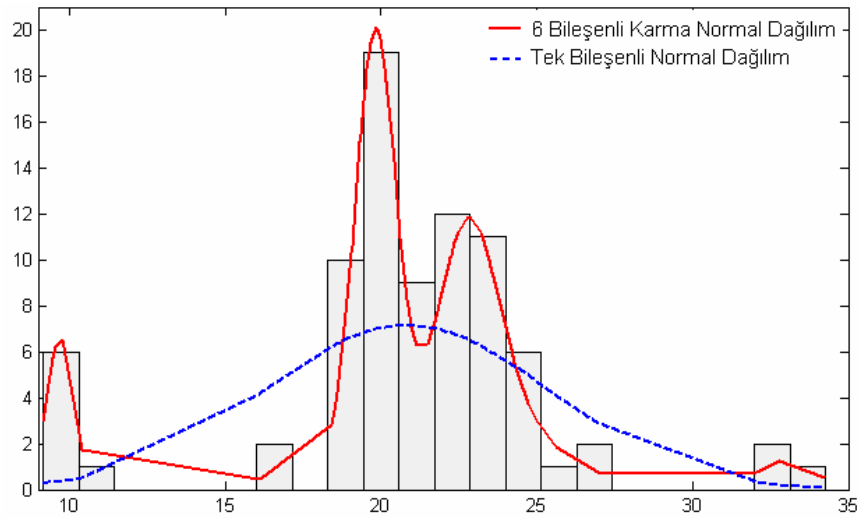
Yengeç verisi iki bileşenli normal dağılımların karması ile modellenirken kitle içerisinde farklı iki alt grup olduğu varsayılmış ve X yengeç vücut genişliği rassal değişkeni olmak üzere karma dağılım modelinin olasılık yoğunluk fonksiyonu

$$f(x) = \pi f_1(x) + (1 - \pi) f_2(x) \quad (4.1)$$

şeklinde tanımlanmıştır. Eşitlik (4.1)’de yer alan π , $\pi \in (0,1)$ olmak üzere birinci bileşene ait karma oranıdır. Modelde yer alan $f_1(x)$ ve $f_2(x)$ her bir alt grubun modellenmesinde kullanılan normal dağılımlara ait olasılık yoğunluk fonksiyonlarıdır.

Tahmin edilmesi gereken parametre sayısının fazla olması dezavantajı nedeni ile başlangıçta fazla ilgilenilmeyen karma dağılım modelleri, EM algoritmasının karma dağılım modellerindeki parametrelerinin tahminindeki etkinliği görüldükten sonra çok geniş bir uygulama alanına sahip olmuştur. Çünkü klasik olasılık dağılımları, frekans dağılımının tek modlu olduğu durumlarda kullanışlı iken karma dağılımlar frekans dağılımının iki veya daha fazla moda sahip olan hatta tek modlu

olması durumunda bile verinin modellenmesinde daha kullanışlı ve daha yararlıdır (Everitt ve Hand 1981). Karma dağılım modellerinin, klasik istatistik modellerine göre çok modlu verilerin modellenmesindeki başarısı Roeder (1990) tarafından tanımlanan Galaxy veri setinin modellenmesinde açıkça görülmektedir. Galaxy veri seti için uydurulan 6 bileşenli karma normal dağılım ve tek bileşenli normal dağılımın yoğunluk eğrileri Şekil 4.2’de verilmiştir.



Şekil 4.2. Roeder (1990)'in Galaxy veri seti için uydurulan 6 bileşenli karma normal dağılım (düz çizgi) ve tek bileşenli normal dağılımın (kesikli çizgi) yoğunluk eğrileri

Şekil 4.2 incelendiğinde Galaxy verisinin tek bileşenli normal dağılımla modellenmesi durumunda yapılacak hatanın, 6 bileşenli karma normal dağılımla modellenmesi durumunda yapılacak olan hatadan çok daha fazla olduğu açıkça görülmektedir.

4.1.1. Temel Tanım ve Gösterimler

Sonlu karma dağılım modellerinde kitle içerisinde g (≥ 2) tamsayı olmak üzere, g tane farklı alt grup veya sınıfın olduğu varsayımı yapılır. Sonlu karma dağılım modelinin olasılık yoğunluk fonksiyonu

$$f(t|\psi) = \pi_1 f_1(t|\theta_1) + \pi_2 f_2(t|\theta_2) + \dots + \pi_g f_g(t|\theta_g)$$

$$f(t|\psi) = \sum_{k=1}^g \pi_k f_k(t|\theta_k) \quad (4.2)$$

şeklinde tanımlanır. Burada $k = 1, \dots, g$ olmak üzere $f_k(t|\theta_k)$, k . alt grupta T rassal değişkeninin olasılık yoğunluk fonksiyonu, θ_k bilinmeyen parametre vektörü, π_k , $\pi_k \in (0,1)$ ve $\sum_{k=1}^g \pi_k = 1$ koşuluyla k . alt grubun kitle içerisindeki oranı ve $\psi = (\pi, \theta)$, $\pi = (\pi_1, \dots, \pi_g)$ ve $\theta = (\theta_1, \dots, \theta_g)$ bilinmeyen vektörlerini içeren karmanın bilinmeyen parametre vektörü olarak tanımlanır. Bu tanımlamalar altında örneğin Weibull dağılımların iki bileşenli karma dağılım modeli

$$f(t|\psi) = \pi_1 f_1(t|\theta_1) + \pi_2 f_2(t|\theta_2)$$

$$f(t|\psi) = \pi_1 f_1(t; \alpha_1, \beta_1) + \pi_2 f_2(t; \alpha_2, \beta_2) \quad (4.3)$$

veya daha açık bir gösterimle

$$f(t|\psi) = \pi_1 \frac{\alpha_1}{\beta_1} \left(\frac{t}{\beta_1} \right)^{\alpha_1-1} e^{-\left(\frac{t}{\beta_1} \right)^{\alpha_1}} + \pi_2 \frac{\alpha_2}{\beta_2} \left(\frac{t}{\beta_2} \right)^{\alpha_2-1} e^{-\left(\frac{t}{\beta_2} \right)^{\alpha_2}} \quad (4.4)$$

olarak yazılır. Sonlu karma dağılım modellerinde, $\pi_k \in (0,1)$ ve $\sum_{k=1}^g \pi_k = 1$ koşulundan dolayı eşitlik (4.4)'de verilen modelde π_2 yerine doğrudan $(1 - \pi_1)$ 'de yazılabilir.

Sonlu karma dağılım modeli için eşitlik (4.2)'de verilen fonksiyonun olasılık yoğunluk fonksiyonu olabilmesi için

$$i. \quad \forall t_i \text{ için } f(t|\psi) \geq 0$$

$$ii. \quad \int_A f(t|\psi) dt = 1$$

özelliklerini taşıması gerekir. Eşitlik (4.2)'de verilen fonksiyonu oluşturan $\pi_k \in (0,1)$ ve $f_k(t | \theta_k) \geq 0$ olduğundan dolayı $f(t | \psi) \geq 0$ olacağı aşikardır. Fonksiyon tanımlı olduğu aralıkta integrallenirse

$$\begin{aligned} \int_0^{\infty} f(t | \psi) dt &= \int_0^{\infty} \sum_{k=1}^g \pi_k f_k(t | \theta_k) dt \\ \int_0^{\infty} f(t | \psi) dt &= \sum_{k=1}^g \pi_k \int_0^{\infty} f_k(t | \theta_k) dt \end{aligned}$$

elde edilir. Sonlu karma dağılım modelinde alt bileşenlerdeki fonksiyonlar olasılık yoğunluk fonksiyonu olarak tanımlandığından $\int_0^{\infty} f_k(t | \theta_k) dt = 1$ olacaktır. Dolayısıyla

$$\int_0^{\infty} f(t | \psi) dt = \sum_{k=1}^g \pi_k = 1$$

olarak elde edilir. Böylece eşitlik (4.2)'de verilen fonksiyonun olasılık yoğunluk fonksiyonu olduğu gösterilmiş olur.

Karma dağılım modelleri için birikimli dağılım fonksiyonu,

$$\begin{aligned} F(t) &= P(T \leq t) = \int_0^t f(t | \psi) dt \\ F(t) &= \int_0^t \sum_{k=1}^g \pi_k f_k(t | \theta_k) dt = \sum_{k=1}^g \pi_k \underbrace{\int_0^t f_k(t | \theta_k) dt}_{=F_k(t|\theta_k)} \\ F(t) &= \sum_{k=1}^g \pi_k F_k(t | \theta_k) \end{aligned} \tag{4.5}$$

olarak tanımlanır. Eşitlikte yer alan $F_k(t | \theta_k)$ gösterimi ile, k . alt gruba ait birikimli dağılım fonksiyonu ifade edilmektedir. Karma dağılım modelleri için yaşam fonksiyonu ise

$$S(t) = 1 - F(t)$$

$$S(t) = 1 - \sum_{k=1}^g \pi_k f_k(t | \theta_k) \quad (4.6)$$

olarak veya

$$S(t) = P(T \geq t) = \int_t^{\infty} f(t | \psi) dt$$

$$S(t) = \int_t^{\infty} \sum_{k=1}^g \pi_k f_k(t | \theta_k) dt = \sum_{k=1}^g \pi_k \underbrace{\int_t^{\infty} f_k(t | \theta_k) dt}_{=S_k(t | \theta_k)}$$

$$S(t) = \sum_{k=1}^g \pi_k S_k(t | \theta_k) \quad (4.7)$$

şeklinde gösterilir. Eşitlikte yer alan $S_k(t | \theta_k)$ gösterimi ile, k . alt gruba ait yaşam fonksiyonu ifade edilmektedir. Yaşam analizinde önemli bir fonksiyonu olan hazard fonksiyonu karma dağılım modellerinde

$$h(t) = \sum_{k=1}^g \pi_k \frac{f_k(t | \theta_k)}{\sum_{k=1}^g \pi_k S_k(t | \theta_k)} \quad (4.8)$$

şeklinde tanımlanır. Yaşam süresi T rassal değişkeninin karma dağılım modellerindeki beklenen değeri

$$\begin{aligned}
E(T) &= \int_0^{\infty} t f(t|\psi) dt \\
E(T) &= \int_0^{\infty} t \sum_{k=1}^g \pi_k f_k(t|\theta_k) dt = \sum_{k=1}^g \pi_k \underbrace{\int_0^{\infty} t f_k(t|\theta_k) dt}_{=E_k(t|\theta_k)} \\
E(T) &= \sum_{k=1}^g \pi_k E_k(T|\theta_k)
\end{aligned} \tag{4.9}$$

eşitliği ile hesaplanır. Eşitlikte yer alan $E_k(T|\theta_k)$ gösterimi ile, k . alt grupta T rassal değişkeninin beklenen değeri ifade edilmektedir.

4.1.2. Uygun Bileşen Sayısının Seçimi

Sonlu karma dağılım modellerinde k . alt grubun modellenmesinde kullanılan dağılımın parametre sayısı r_k olmak üzere tahmin edilmesi gereken serbest parametre sayısı r

$$r = (g - 1) + \sum_{k=1}^g r_k \tag{4.10}$$

eşitliği ile elde edilir. Eşitlik (4.10)'nun sağında yer alan ilk parça tahmin edilmesi gereken alt gruplara ait karma oranların sayısıdır. Kitle içerisinde k alt grup olmasına karşın $\sum_{k=1}^g \pi_k = 1$ kısıdından dolayı birbirinden bağımsız olarak değer alabilecek $(k - 1)$ tane karma oran vardır. Sonlu karma dağılım modelinde her alt bileşenin iki parametrelili bir dağılım ile modellendiği düşünüldüğünde iki bileşen kullanıldığı durumda 5, üç bileşen kullanıldığı durumda 8, dört bileşen kullanıldığında 11 parametrenin tahmin edilmesi gerekmektedir. Modeldeki bileşen sayısının artması tahmin edilecek parametre sayısını da arttıracığından dolayı modeldeki bileşen sayısının ne olacağına karar vermek önemli bir problemidir

(McLachlan ve Krishnan 1997). Karma dağılım modellerinde, bileşen sayısının ne olacağına karar vermek için literatürde birçok yaklaşım bulunmaktadır. Bu yaklaşımlar içerisinde en fazla bilenenlerden biri Akaike bilgi kriteridir. Akaike bilgi kriteri (AIC),

$$AIC = -2 \ln L + 2r \quad (4.11)$$

şeklinde tanımlanır (Akaike 1974). Eşitlikte yer alan $\ln L$ log olabilirlik değeri ve r serbest parametre sayısıdır. Akaike bilgi kriterine göre en uygun bileşen sayısı k olmak üzere

$$AIC(k) \leq AIC(k+1), \quad k = 1, 2, \dots \quad (4.12)$$

eşitsizliğini sağlayan ilk k değeridir. Literatürde yaygın bir kullanıma sahip olan bir diğer bilgi kriteri ise Bayes Bilgi Kriteri (BIC)'dir (McLachlan ve Pell 2001). Bayes bilgi kriteri,

$$BIC = -2 \ln L + r \ln(n) \quad (4.13)$$

şeklinde tanımlanır. Eşitlikte yer alan $\ln(n)$, örneklem hacmi n değerinin doğal logaritmik değerini göstermektedir. Bayes bilgi kriterine göre en uygun bileşen sayısı,

$$BIC(k) \leq BIC(k+1), \quad k = 1, 2, \dots \quad (4.14)$$

eşitsizliğini sağlayan ilk k değeridir. Örneklem hacminin 8'den büyük olduğu durumlarda yani $n \geq 8$ için $2 < \ln(n)$ olduğundan dolayı ceza terimi olarak eklenen parçanın Bayes bilgi kriterinde, Akaike bilgi kriterine göre daha büyük olacağı açıktır. Bu nedenle örneklem hacminin fazla olduğu durumlarda Akaike bilgi kriteri bileşen sayısını fazla gösterme eğilimindedir. Soromenho (1993) ve Celeux ve

Soromenho (1996) çalışmalarında Akaike bilgi kriterinin karma dağılım modellerindeki bileşen sayısını olduğundan daha fazla tahmin etmeye meyilli olduğunu göstermişlerdir. Buna karşın Akaike bilgi kriteri, pratikte halen yaygın olarak kullanılmaktadır (Fraley ve Raftery 1998).

4.1.3. Karma Dağılım Modellerinde Parametre Tahmini

Karma dağılımın modellerinde parametrelerin en çok olabilirlik tahminlerinin elde edilmesinde kullanılan EM (Expectation Maximization) algoritması etkin bir yöntemdir (McLachlan ve Peel 2001, Dempster ve ark 1977). EM algoritması verinin tamamlanmamış veri olması durumunda en çok olabilirlik tahminlerinin elde edilmesi için genel bir istatistiksel yöntemdir (Servi 2009). EM algoritması döngüsel (iteratif) bir algoritmadır. EM algoritması E beklenti (Expectation) ve M maksimum yapma (Maximization) adımlarından oluşur. Bu adımlar belirli bir yakınsama kriteri sağlanana kadar ardışık olarak gerçekleştirilir.

4.1.3.1. Karma Dağılım Modellerinde Tamamlanmamış Veri Yapısı

Bağımsız ve aynı dağılımlı n tane rassal değişken $T_1, T_2, \dots, T_n$ ve rassal değişkenlerin gözlem değerleri sırasıyla $t_1, t_2, \dots, t_n$ ile gösterilsin. $T_1, T_2, \dots, T_n$ rassal değişkenlerin $f(t_i)$ olasılık yoğunluk fonksiyonu (4.2)'deki eşitlik ile verilen karma dağılım olasılık yoğunluk fonksiyonu olsun. Bu durumda

$$T_1, T_2, \dots, T_n \sim F(t | \psi) \quad (4.15)$$

burada $F(t | \psi)$, $f(t_i)$ olasılık yoğunluk fonksiyonuna karşılık gelen dağılım fonksiyonudur.

EM yapısında gözlem vektörü $\mathbf{t} = (t_1 \ t_2 \ \dots \ t_n)'$, gözlemlerle ilişkilendirilen $\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_n$ bileşen etiket vektörleri bulunmadığından dolayı

$\mathbf{t} = (t_1 \ t_2 \ \dots \ t_n)'$ verisi tamamlanmamış veri olarak düşünülür. Karma olabilirlik yaklaşımında her bir t_i gözleminin karmanın hangi bileşenine ait olduğu önemlidir. $i = 1, 2, \dots, n$ ve $k = 1, 2, \dots, g$ olmak üzere t_i gözlemi karma dağılım modelinin k . bileşeninde gözlemlenmişse i . gözlemin k . bileşende gözlemlenme olasılığı $z_{ik} = 1$ değerini alırken i . gözlemin diğer bileşenlerde olma olasılığı $k' \neq k$ ve $k' = 1, 2, \dots, g$ olmak üzere $z_{ik'} = 0$ olacaktır. Bu durumda $\mathbf{z}_k$ bileşen etiket vektörleri ile ilişkilendirilen $t_1, t_2, \dots, t_n$ gözlem verisi ile tamamlanmış veri olarak adlandırılır. Tamamlanmış veri,

$$t_c = [\mathbf{t}' \ \mathbf{z}_1' \ \dots \ \mathbf{z}_k' \ \dots \ \mathbf{z}_g'] \quad (4.16)$$

veya daha açık bir gösterimle

$$t_c = \begin{bmatrix} t_1 & z_{11} & \dots & z_{1k} & \dots & z_{1g} \\ t_2 & z_{21} & \dots & z_{2k} & \dots & z_{2g} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ t_i & z_{i1} & \dots & z_{ik} & \dots & z_{ig} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ t_n & z_{n1} & \dots & z_{nk} & \dots & z_{ng} \end{bmatrix} \quad (4.17)$$

şeklinde yazılır.

4.1.3.2. EM Algoritması

EM algoritmasında $\mathbf{z}_i = (z_{i1}, \dots, z_{ig})$ vektörü kayıp gözlem vektörü olarak düşünülür ve dağılımına ait olasılık yoğunluk fonksiyonu

$$f(\mathbf{z}_i) = \prod_{k=1}^g \pi_k^{z_{ki}}, \quad i = 1, 2, \dots, n \quad (4.18)$$

şeklindedir. Burada $\pi_k, \pi_k \in (0,1)$ ve $\sum_{k=1}^g \pi_k = 1$ koşuluyla k . sınıfın karma oranıdır.

Bu tanımlamalar altında, $\mathbf{z}_i$ bilindiğinde, T_i 'nin dağılımının olasılık yoğunluk fonksiyonu,

$$f(t_i | \mathbf{z}_i) = \prod_{k=1}^g (f_k(t_i | \theta_k))^{z_{ki}}, \quad i = 1, 2, \dots, n \quad (4.19)$$

ve T ve Z rassal değişkenlerinin ortak olasılık yoğunluk fonksiyonu

$$f(t_1, \dots, t_n, \mathbf{z}_1, \dots, \mathbf{z}_n) = \prod_{i=1}^n \prod_{k=1}^g (f_k(t_i | \theta_k))^{z_{ki}} \pi_k^{z_{ki}} \quad (4.20)$$

olmalıdır. Böylece, tamamlanmış veri için olabilirlik fonksiyonunun doğal logaritması,

$$\ln L(\pi, \theta | t_1, \dots, t_n, \mathbf{z}_1, \dots, \mathbf{z}_n) = \sum_{i=1}^n \sum_{k=1}^g z_{ki} \ln(\pi_k f_k(t_i | \theta_k)) \quad (4.21)$$

olarak elde edilir.

EM algoritmasının E adımında $E(z_{ki} | t_i)$, $\mathbf{z}_i$ 'nin tahmini olarak hesaplanması beklenir. $E(z_{ki} | t_i)$ elde etmek için,

$$f(z_{ki} | t_i) = \frac{(f(t_i; \theta_k))^{z_{ki}} \pi_k^{z_{ki}}}{\sum_{j=1}^g (f(t_i; \theta_j))^{z_{ji}} \pi_j^{z_{ji}}}, \quad i = 1, 2, \dots, n \quad (4.22)$$

hesaplanması gerekir. Böylece

$$E(z_{ki} | t_i) = \frac{\pi_k f(t_i | \theta_k)}{\sum_{s=1}^g \pi_s f_s(t_i | \theta_s)}, \quad i = 1, 2, \dots, n \quad (4.23)$$

eşitliği sonlu karma dağılım modeli için tahmin edilir. Tamamlanmış veri için olabilirlik fonksiyonunun doğal logaritmasının beklenen değeri,

$$E(\ln L) = \sum_{i=1}^n \sum_{k=1}^g E(z_{ki} | t_i) \ln(\pi_k f_k(t_i; \theta_k)) \quad (4.24)$$

şeklindedir.

EM algoritmasının M adımında, E adımında elde edilen $E(\ln L)$, $\sum_{k=1}^g \pi_k = 1$

koşulu altında maksimize edilir. Bunun için Langrange yöntemi ile oluşturulan $\ln \tilde{L}$ maksimize edilmeye çalışılır.

$$\ln \tilde{L} = \sum_{i=1}^n \sum_{k=1}^g E(z_{ki} | t_i) \ln(\pi_k f_k(t_i; \theta_k)) - \lambda \left(\sum_{k=1}^g \pi_k - 1 \right) \quad (4.25)$$

Eşitlik (4.25)'i maksimize eden parametreleri tahmin etmek için, sırasıyla tüm parametrelere göre ayrı ayrı $\ln \tilde{L}$ 'nin ilk türevleri alınarak sıfıra eşitlenir. k . alt grubun karma oranının tahmini,

$$\hat{\pi}_k = n^{-1} \sum_{i=1}^n E(z_{ki} | t_i), \quad k = 1, \dots, g \quad (4.26)$$

şeklinde yapılır.

EM algoritmasında E ve M adımları bir yakınsama kriteri sağlanıncaya kadar tekrarlanır. Yakınsama için durdurma kuralı olarak $\|\cdot\|$ uygun bir norm ve $\varepsilon > 0$ olmak üzere $\|\psi^{(k+1)} - \psi^{(k)}\| < \varepsilon$ olma koşulunu sağlayan bir ε değeri alınabilir veya

$\ln \tilde{L}^{(k+1)} - \ln \tilde{L}^{(k)}$ farkı oldukça küçükse ya da durağanlaşıyorsa algoritma sonlandırılır.

4.2. Karıştırılmış Dağılımlar

Dağılımların birleştirilmesi yöntemlerinden biri de dağılımların parametresini farklı bir dağılımla karıştırmaktır. Bu yöntem ile basit bir forma sahip olan dağılımlardan daha kullanışlı ve daha esnek dağılımlar elde edilebilir (Douglas 1980). Karıştırılmış dağılımlarla ilgili çalışmalar 1920 yılına kadar gitmektedir. Greenwood ve Yule (1920), Poisson dağılımının parametresini Gamma dağılımı olarak ele almışlar ve Pascal (Negatif Binom) dağılımını elde etmişlerdir. Satterthwaite (1942), dağılımları karıştırma yöntemiyle genelleştirilmiş Poisson dağılımını tanımlayarak dağılımın özelliklerini incelemiştir. Feller (1943), dağılımların birleştirilmesi ile elde edilen bazı dağılımları karıştırılmış dağılım olarak tanımlamıştır. Gurland (1957) karıştırılmış dağılımların özelliklerini incelemiştir.

Literatür incelendiğinde dağılımların karıştırılması ile ilgili son dönemde de çok sayıda çalışmanın bulunduğu görülmektedir. Murat ve Szynal (2000) ile Grubbström ve Tang (2006) karıştırılmış dağılımların momentlerinin hesaplanması ile ilgili çalışmışlardır. Liu ve ark (2002), Poisson-Gumbel karıştırılmış dağılımını ve uygulamasını incelemişlerdir. Adamidis ve ark (2005), uç değer dağılımını, Üstel dağılımla karıştırarak genelleştirilmiş Üstel-Geometrik dağılımını tanımlamışlardır. Karlis ve Xekalaki (2005), karıştırılmış Poisson dağılımını çalışmışlardır. Liu ve ark (2006), karıştırılmış uç değer dağılımını ve dağılımın deniz bilimlerinde uygulamalarını çalışmışlardır.

4.2.1. Karıştırılmış Dağılımlarda Temel Gösterimler

X rassal değişkeni y parametresi ile $F(x|y)$ dağılımına sahip olsun. Farz edelim ki y parametresi bir rassal değişken ve Y 'nin dağılım fonksiyonu

$G(y)$ olsun. Bu tanımlamalar sonrası eşitlik (4.27) ile elde edilen X rassal değişkeninin dağılım fonksiyonu

$$K(x) = \int F(x|y)dG(y) \quad (4.27)$$

kariştirilmiş dağılım olarak adlandırılır (Gurland 1957). Yukarıda tanımlanan eşitlik olasılık yoğunluk fonksiyonu olarak

$$k(x) = \int f(x|y)g(y)dy \quad (4.28)$$

şeklinde yazılır. Burada $g(y)$, karıştıran yoğunluk olarak isimlendirilir (Karlis ve Xekalaki 2005). $f(x|y)$ olasılık yoğunluk fonksiyonunda y parametresinin $g(y)$ olasılık yoğunluk fonksiyonuna sahip olmasıyla karıştırılması $f(x|y) \wedge_y g(y)$ gösterimi ile ifade edilir.

Karıştırma işleminin birleşme özelliği,

$$\left[f(x|y) \wedge_y g(y|\mu) \right] \wedge_\mu z(\mu) \text{ eşdeğerdir } f(x|y) \wedge_y \left[g(y|\mu) \wedge_\mu z(\mu) \right] \quad (4.29)$$

şeklinde ifade edilir. Burada dağılım parametreleri birbirinden bağımsızdır. Buna göre $f(x|y)$ fonksiyonunda y parametresinin $g(y|\mu)$ ile karıştırılmasıyla oluşan dağılımın μ parametresinin $z(\mu)$ fonksiyonu ile karıştırılmasıyla elde edilen yeni dağılım, $f(x|y)$ fonksiyonun y parametresinin, $g(y|\mu)$ fonksiyonunun μ parametresinin $z(\mu)$ fonksiyonu ile karıştırılmasıyla oluşan fonksiyon ile karıştırılmasıyla elde edilecek yeni dağılıma eşdeğerdir (Karlis ve Xekalaki 2005).

4.2.2. Karıştırılmış Dağılımlarla İlgili Bazı Örnekler

Bu kısımda, Poisson, Üstel ve Weibull dağılımlarının herhangi bir parametresinin Gamma dağılımına sahip bir rassal değişken olarak tanımlanması ile sırasıyla Negatif Binom, Pareto ve genelleştirilmiş Burr dağılımlarının elde edilişleri gösterilecektir.

T rassal değişkeni λ parametresi ile Poisson dağılımına sahip ve olasılık yoğunluk fonksiyonu $f(t | \lambda)$

$$f(t | \lambda) = \frac{\lambda^t e^{-\lambda}}{t!}, \quad t = 0, 1, \dots, \text{ ve } \lambda > 0 \quad (4.30)$$

şeklinde tanımlansın. Poisson dağılımının λ parametresi Gamma dağılımına sahip bir rassal değişken ve olasılık yoğunluk fonksiyonu $g(\lambda | r, p^{-1}(1-p))$

$$g(\lambda | r, p^{-1}(1-p)) = \frac{(1-p)^r}{p^r \Gamma(r)} \lambda^{r-1} e^{-\frac{(1-p)\lambda}{p}}, \quad \lambda > 0, \quad r, p^{-1}(1-p) > 0 \quad (4.31)$$

şeklinde tanımlansın. Bu durumda

$$\begin{aligned} f(t | r, p) &= \int_0^{\infty} f(t | \lambda) g(\lambda | r, p^{-1}(1-p)) d\lambda \\ &= \int_0^{\infty} \frac{\lambda^t e^{-\lambda}}{t!} \frac{(1-p)^r}{p^r \Gamma(r)} \lambda^{r-1} e^{-\frac{(1-p)\lambda}{p}} d\lambda \\ &= \frac{(1-p)^r}{t! p^r \Gamma(r)} \int_0^{\infty} \lambda^{t+r-1} e^{-\frac{\lambda}{p}} d\lambda \end{aligned}$$

integralinde $\lambda p^{-1} = a$ dönüşümü yapılırsa

$$\begin{aligned}
 f(t|r, p) &= \frac{(1-p)^r}{t! p^r \Gamma(r)} \int_0^\infty (ap)^{t+r-1} e^{-a} p da \\
 &= \frac{p^t (1-p)^r}{t! \Gamma(r)} \int_0^\infty a^{t+r-1} e^{-a} da
 \end{aligned}$$

elde edilir. Gamma fonksiyonu $\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx$ şeklinde tanımlandığından

$$f(t|r, p) = \frac{\Gamma(t+r)}{t! \Gamma(r)} p^t (1-p)^r, \quad t = 0, 1, \dots \quad (4.32)$$

elde edilir. Gamma fonksiyonunda α tamsayı olarak ele alındığında $\Gamma(\alpha) = (\alpha-1)!$ özelliği gereği eşitlik (4.32)'deki r parametresi de tamsayı olarak tanımlanırsa

$$\begin{aligned}
 f(t|r, p) &= \frac{(t+r-1)!}{t!(r-1)!} p^t (1-p)^r \\
 f(t|r, p) &= \binom{t+r-1}{r} p^t (1-p)^r
 \end{aligned} \quad (4.33)$$

şeklinde Negatif Binom dağılımı elde edilir. Negatif Binom dağılımı Poisson ve Gamma dağılımlarından hareketle elde edildiğinden dolayı aynı zamanda Gamma-Poisson bileşiği dağılımı olarak da isimlendirilir.

T rassal değişkeni θ parametresi ile Üstel dağılıma sahip ve olasılık yoğunluk fonksiyonu $f(t|\theta)$

$$f(t|\theta) = \theta e^{-\theta t}, \quad t > 0, \theta > 0 \quad (4.34)$$

şeklinde tanımlansın. Üstel dağılımın θ parametresi Gamma dağılımına sahip bir rassal değişken ve olasılık yoğunluk fonksiyonu $g(\theta|\alpha, \beta)$

$$g(\theta | \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} \theta^{\alpha-1} e^{-\beta\theta}, \quad \theta > 0, \quad \alpha, \beta > 0 \quad (4.35)$$

şeklinde tanımlansın. Bu durumda

$$\begin{aligned} f(t | \alpha, \beta) &= \int_0^\infty f(t | \theta) g(\theta | \alpha, \beta) d\theta \\ &= \int_0^\infty \theta e^{-\theta t} \frac{\beta^\alpha}{\Gamma(\alpha)} \theta^{\alpha-1} e^{-\beta\theta} d\theta \\ &= \frac{\beta^\alpha}{\Gamma(\alpha)} \int_0^\infty \theta^\alpha e^{-\theta(t+\beta)} d\theta \end{aligned}$$

integralinde $\theta(t + \beta) = a$ dönüşümü yapılırsa

$$\begin{aligned} f(t | \alpha, \beta) &= \frac{\beta^\alpha}{\Gamma(\alpha)} \int_0^\infty \left(\frac{a}{t + \beta} \right)^\alpha e^{-a} \left(\frac{1}{t + \beta} \right) da \\ &= \frac{\beta^\alpha}{\Gamma(\alpha)(t + \beta)^{\alpha+1}} \int_0^\infty a^\alpha e^{-a} da \end{aligned}$$

elde edilir. Gamma fonksiyonu $\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx$ şeklinde tanımlandığından

$$f(t | \alpha, \beta) = \frac{\beta^\alpha \Gamma(\alpha + 1)}{\Gamma(\alpha)(t + \beta)^{\alpha+1}}$$

elde edilir. Gamma fonksiyonunda $\Gamma(\alpha + 1) = \alpha \Gamma(\alpha)$ olduğundan ilgili eşitlik

$$f(t | \alpha, \beta) = \frac{\beta^\alpha \alpha \Gamma(\alpha)}{\Gamma(\alpha)(t + \beta)^{\alpha+1}}$$

şeklinde yazılır ve gerekli sadeleştirmeler sonrası

$$f(t|\alpha, \beta) = \frac{\beta^\alpha \alpha}{(t + \beta)^{\alpha+1}}, \quad t > 0, \alpha, \beta > 0 \quad (4.36)$$

elde edilir. Elde edilen eşitlik Pareto dağılımının olasılık yoğunluk fonksiyonudur. Üstel dağılım iki parametrelidir

$$f(t|\theta, \lambda) = \theta e^{-\theta(t-\lambda)}, \quad t \geq \lambda, \theta, \lambda > 0 \quad (4.37)$$

olarak tanımlanarak benzer işlemler yapılırsa üç parametrelilik Pareto dağılımı

$$f(t|\alpha, \beta, \lambda) = \frac{\beta^\alpha \alpha}{(t + \beta - \lambda)^{\alpha+1}}, \quad t \geq \lambda, \alpha, \beta, \lambda > 0 \quad (4.38)$$

elde edilir.

Genelleştirilmiş Burr dağılımı, olasılık yoğunluk fonksiyonu

$$f(t|\varphi, \theta) = \varphi \theta t^{\varphi-1} e^{-\theta t^\varphi}, \quad t > 0, \varphi, \theta > 0 \quad (4.40)$$

şeklinde verilen Weibull dağılımında ölçek parametresi θ 'nın olasılık yoğunluk fonksiyonu eşitlik (4.35)'de verilen Gamma dağılımına sahip bir rassal değişken olarak tanımlanmasıyla elde edilir. Buna göre,

$$\begin{aligned} f(t|\varphi, \alpha, \beta) &= \int_0^\infty f(t|\varphi, \theta) g(\theta|\alpha, \beta) d\theta \\ &= \int_0^\infty \varphi \theta t^{\varphi-1} e^{-\theta t^\varphi} \frac{\beta^\alpha}{\Gamma(\alpha)} \theta^{\alpha-1} e^{-\beta \theta} d\theta \\ &= \frac{\varphi \beta^\alpha t^{\varphi-1}}{\Gamma(\alpha)} \int_0^\infty \theta^\alpha \underbrace{e^{-\theta t^\varphi - \beta \theta}}_{e^{-\theta(t^\varphi + \beta)}} d\theta \end{aligned}$$

integralinde $\theta(t^\varphi + \beta) = a$ dönüşümü yapılırsa

$$\begin{aligned} f(t | \varphi, \alpha, \beta) &= \frac{\varphi \beta^\alpha t^{\varphi-1}}{\Gamma(\alpha)} \int_0^\infty \left(\frac{a}{t^\varphi + \beta} \right)^\alpha e^{-a} \frac{da}{(t^\varphi + \beta)} \\ &= \frac{\varphi \beta^\alpha t^{\varphi-1}}{\Gamma(\alpha)} \frac{1}{(t^\varphi + \beta)^{\alpha+1}} \int_0^\infty a^\alpha e^{-a} da \end{aligned}$$

eşitliği elde edilir. Gamma fonksiyonu $\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx$ şeklinde tanımlandığından

$$f(t | \varphi, \alpha, \beta) = \frac{\varphi \beta^\alpha t^{\varphi-1}}{\Gamma(\alpha)} \frac{1}{(t^\varphi + \beta)^{\alpha+1}} \Gamma(\alpha + 1)$$

elde edilir. Gamma fonksiyonunda $\Gamma(\alpha + 1) = \alpha \Gamma(\alpha)$ olduğundan ilgili eşitlik

$$f(t | \varphi, \alpha, \beta) = \frac{\varphi \beta^\alpha t^{\varphi-1}}{\Gamma(\alpha)} \frac{1}{(t^\varphi + \beta)^{\alpha+1}} \alpha \Gamma(\alpha)$$

şeklinde yazılabilir. Burada gerekli sadeleştirmeler yapıldıktan sonra

$$f(t | \varphi, \alpha, \beta) = \frac{\varphi \alpha \beta^\alpha t^{\varphi-1}}{(t^\varphi + \beta)^{\alpha+1}}, \quad t > 0, \quad \varphi, \alpha, \beta > 0 \quad (4.40)$$

elde edilir. Bu eşitlikte $\beta = 1$ alındığında Burr dağılımı elde edilir.

5. YAŞAM ANALİZİNDE KARMA DAĞILIM MODELLERİ

Yaşam analizi, ilgilenilen değişkenin ortaya çıkıncaya kadar geçen sürelerden oluşan verilerin analizindeki istatistiksel tekniklerin bütünüdür. Yaşam analizi veri setleri genellikle Gamma, Lognormal ve Weibull dağılımları gibi klasik istatistik dağılımları ile temsil edilmiştir (Nelson 1982, Lawless 2003, Kleinbanm ve Klein 2005, Lee ve Wang 2003, Machin ve ark 2006). Bu klasik istatistik dağılım modellerine ilaveten, son zamanlarda yaşam süresi verilerinin modellenmesi için yeni modeller geliştirilmiştir. Özellikle heterojen yapıdaki verilerin modellenmesinde klasik istatistik dağılım modellerine göre daha kullanışlı ve daha yararlı olan karma dağılım modellerinin kullanımı yaygın bir hale gelmiştir. Chen ve ark (1985), kanser yaşam verilerinin analizinde iki bileşenli karma dağılım modelini kullanmışlardır. Qian (1994), akciğer kanseri çalışmalarında Weibull bileşenlerinin karmasına benzer bir karma dağılım modeli kullanmıştır. Angelis ve ark (1999), 1953 – 1992 yılları arasında Finlandiya’ya ait kanser kayıtlarından kolon kanseri ile ilgili derledikleri veri setinin analizinde karma dağılım modellerini kullanmışlardır. Marin ve ark (2005), bileşen sayısı bilinmeyen heterojen bir yaşam verisine karma Weibull modelinin uyumunu Bayes metoduyla nasıl yapılabileceğini göstermişlerdir.

Bu bölümde özellikle heterojen yapıdaki yaşam sürelerinin modellenmesinde kullanılan karma Gamma dağılım modeli, karma Lognormal dağılım modeli ve karma Weibull dağılım modeli incelenecektir. İlgilenilen karma dağılım modellerindeki parametrelerin en çok olabilirlik tahminlerinin elde edilmesi için EM algoritmaları tanımlanarak, tanımlanan algoritmaların parametre tahminindeki etkinliği simülasyon çalışması ile gösterilecektir. Ayrıca literatürde kullanılan ve iyi bilinen iki gerçek heterojen yapıdaki yaşam süresi verisinin modellenmesinde, karma dağılım modellerinin standart olasılık dağılım modellerine göre üstünlüğü gösterilecektir.

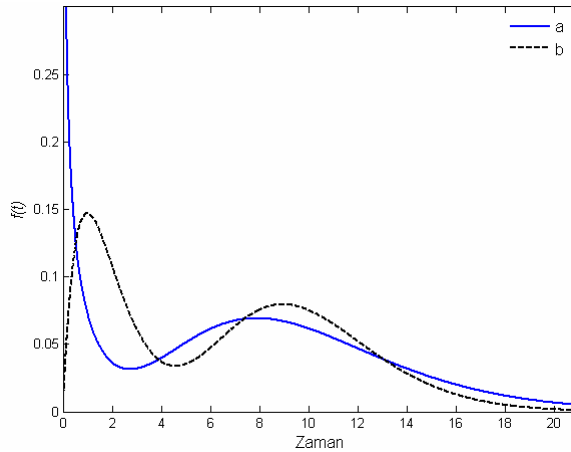
5.1. Yaşam Sürelerinin Modellenmesinde Karma Dağılım Yaklaşımı

5.1.1. Karma Gamma Dağılım Modeli

T rassal değişkeni $g \geq 2$ olmak üzere g bileşenli karma Gamma dağılımına sahip ise olasılık yoğunluk fonksiyonu

$$f(t | \psi) = \sum_{k=1}^g \pi_k t^{\alpha_k-1} \frac{e^{-t/\beta_k}}{\beta_k^{\alpha_k} \Gamma(\alpha_k)}, \quad t > 0, \alpha_k, \beta_k > 0 \quad (5.1)$$

şeklinde tanımlanır. Burada $k = 1, \dots, g$ olmak üzere $\pi_k, \pi_k \in (0,1)$ ve $\sum_{k=1}^g \pi_k = 1$ koşuluyla k . Gamma bileşeninin karma ağırlığı ve $\psi = (\pi, \alpha, \beta)$, $\pi = (\pi_1, \dots, \pi_g)$ ve $(\alpha, \beta) = ((\alpha_1, \beta_1), \dots, (\alpha_g, \beta_g))$ bilinmeyen vektörlerini içeren karma Gamma dağılım modelinde bilinmeyen parametre vektörü olarak tanımlanır. İki bileşenli karma Gamma dağılım modelinde farklı parametre değerlerine göre elde edilen yoğunluk eğrileri Şekil 5.1’de verilmiştir.



Şekil 5.1. Farklı parametrelere sahip iki bileşenli karma Gamma dağılım modeli için yoğunluk eğrileri a) $\pi_1 = 0.3, \alpha_1 = 0.5, \beta_1 = 2, \pi_2 = 0.7, \alpha_2 = 5$ ve $\beta_2 = 2$ (düz çizgi) b) $\pi_1 = 0.4, \alpha_1 = 2, \beta_1 = 1, \pi_2 = 0.6, \alpha_2 = 10$ ve $\beta_2 = 1$ (kesik çizgi)

Karma Gamma dağılım modelinde yaşam fonksiyonu, hazard fonksiyonu ve beklenen yaşam süresi sırasıyla

$$S(t | \psi) = \sum_{k=1}^g \pi_k \left(1 - \frac{\Gamma_x(\alpha_k)}{\Gamma(\alpha_k)} \right) \quad (5.2)$$

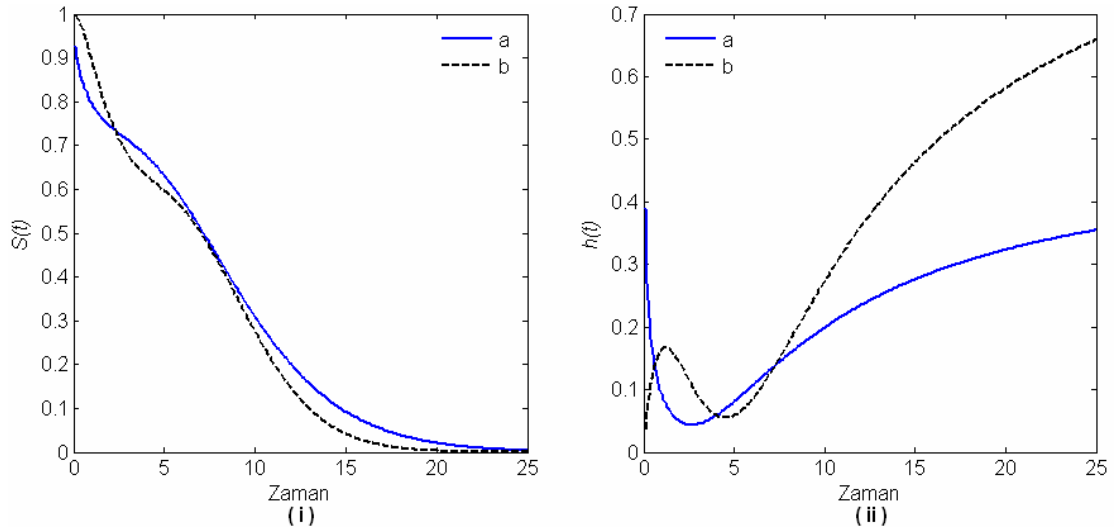
$$h(t) = \sum_{k=1}^g \pi_k t^{\alpha_k-1} \frac{e^{-t/\beta_k}}{\beta_k^{\alpha_k} \Gamma(\alpha_k) S(t | \psi)} \quad (5.3)$$

$$E(T) = \sum_{k=1}^g \pi_k \alpha_k \beta_k \quad (5.4)$$

şeklinde tanımlanır. Eşitlikte yer alan $\Gamma_x(\alpha)$ tamamlanmamış gamma fonksiyonu

olarak isimlendirilir ve $\Gamma_x(\alpha) = \int_0^x t^{\alpha-1} \exp(-t) dt$ eşitliği ile gösterilir. İki bileşenli

karma Gamma dağılım modelinde farklı parametre değerlerine göre elde edilen yaşam fonksiyonu ve hazard fonksiyonu eğrileri Şekil 5.2’de verilmiştir.



Şekil 5.2. Farklı parametrelere sahip iki bileşenli karma Gamma dağılım modeli için
 (i) yaşam fonksiyonu ve (ii) hazard fonksiyonu eğrileri a) $\pi_1 = 0.3$,
 $\alpha_1 = 0.5$, $\beta_1 = 2$, $\pi_2 = 0.7$, $\alpha_2 = 5$ ve $\beta_2 = 2$ (düz çizgi)
 b) $\pi_1 = 0.4$, $\alpha_1 = 2$, $\beta_1 = 1$, $\pi_2 = 0.6$, $\alpha_2 = 10$ ve $\beta_2 = 1$ (kesik çizgi)

Karma Gamma dağılım modelindeki parametrelerin en çok olabilirlik tahminlerini elde etmek için algoritmanın E adımında kayıp gözlem vektörünün elemanları

$$\hat{z}_{ki} = \frac{\pi_k f(t_i | \alpha_k, \beta_k)}{\sum_{s=1}^g \pi_s f_s(t_i | \alpha_s, \beta_s)} \quad (5.5)$$

şeklinde tahmin edilir. Eşitlikte yer alan $f(t_i | \alpha_k, \beta_k)$, k . Gamma bileşenine ait olasılık yoğunluk fonksiyonunu göstermektedir. Böylelikle karma Gamma dağılım modelinde tamamlanmış veri için log olabilirlik fonksiyonunun beklenen değeri

$$E(\ln L) = \sum_{i=1}^n \sum_{k=1}^g \hat{z}_{ki} \ln \left(\pi_k t_i^{\alpha_k - 1} \frac{e^{-t_i / \beta_k}}{\beta_k^{\alpha_k} \Gamma(\alpha_k)} \right) \quad (5.6)$$

şeklinde elde edilir. EM algoritmasının M adımında, E adımında elde edilen $E(\ln L)$,

$\sum_{k=1}^g \pi_k = 1$ koşulu altında maksimize edilir. Langrance metodu ile eşitlik (5.6)

düzenlenerek açık bir notasyonla

$$\begin{aligned} \ln \tilde{L} = & \sum_{i=1}^n \sum_{k=1}^g \hat{z}_{ki} \ln \pi_k + \sum_{i=1}^n \sum_{k=1}^g (\alpha_k - 1) \hat{z}_{ki} \ln t_i - \sum_{i=1}^n \sum_{k=1}^g \frac{1}{\beta_k} \hat{z}_{ki} t_i - \sum_{i=1}^n \sum_{k=1}^g \hat{z}_{ki} \alpha_k \ln \beta_k \\ & - \sum_{i=1}^n \sum_{k=1}^g \hat{z}_{ki} \ln \Gamma(\alpha_k) - \gamma \left(\sum_{k=1}^g \pi_k - 1 \right) \end{aligned} \quad (5.7)$$

şeklinde elde edilir. Eşitlikte yer alan γ Langrance çarpanı olarak isimlendirilir.

$\ln \tilde{L}$ fonksiyonunu maksimize eden parametreleri tahmin etmek için, sırasıyla tüm parametrelere göre ayrı ayrı $\ln \tilde{L}$ 'nin ilk türevleri alınarak sıfıra eşitlenir. Karma Gamma dağılım modelinde k . Gamma bileşeninin karma ağırlığının tahmini,

$$\hat{\pi}_k = n^{-1} \sum_{i=1}^n \hat{z}_{ki}, \quad k = 1, \dots, g \quad (5.8)$$

şeklinde yapılır. $\ln \tilde{L}$ fonksiyonu β_k 'ya göre türevlenip sıfıra eşitlenirse

$$\begin{aligned} \frac{\partial \ln \tilde{L}}{\partial \beta_k} &= \frac{1}{\beta_k^2} \sum_{i=1}^n \hat{z}_{ki} t_i - \frac{\alpha_k}{\beta_k} \sum_{i=1}^n \hat{z}_{ki} = 0, \quad k = 1, \dots, g \\ \beta_k &= \frac{1}{\alpha_k} \frac{\sum_{i=1}^n \hat{z}_{ki} t_i}{\sum_{i=1}^n \hat{z}_{ki}}, \quad k = 1, \dots, g \end{aligned} \quad (5.9)$$

şeklinde elde edilir. Elde edilen eşitlik, (5.7) nolu eşitlikte yerine yazılıp α_k 'ya göre türevlenip 0 eşitlenirse

$$\ln \alpha_k - \psi(\alpha_k) = \ln \left(\frac{\sum_{i=1}^n \hat{z}_{ki} t_i}{\sum_{i=1}^n \hat{z}_{ki}} \right) - \frac{\sum_{i=1}^n \hat{z}_{ki} \ln t_i}{\sum_{i=1}^n \hat{z}_{ki}}, \quad k = 1, \dots, g \quad (5.10)$$

elde edilir. Eşitlikte yer alan $\psi(\cdot)$ digamma fonksiyonu olarak isimlendirilir ve

$\psi(\cdot) = \frac{\Gamma'(\cdot)}{\Gamma(\cdot)}$ eşitliği ile tanımlanır. Eşitlik (5.10) kapalı formda olduğundan dolayı

doğrudan analitik olarak çözülemez. Eşitliğin çözümünde iteratif yöntemlerden Newton-Raphson yöntemi kullanılabilir. Newton-Raphson yönteminde $(r+1)$.

iterasyonda k . Gamma bileşeninin α_k ve β_k parametrelerinin en çok olabilirlik tahminleri $k = 1, \dots, g$ olmak üzere

$$\hat{\alpha}_{k,(r+1)} = \hat{\alpha}_{k,r} - \frac{\ln(\hat{\alpha}_{k,r}) - \psi(\hat{\alpha}_{k,r}) - \ln\left(\sum_{i=1}^n \hat{z}_{ki} t_i \left(\sum_{i=1}^n \hat{z}_{ki}\right)^{-1}\right) + \sum_{i=1}^n \hat{z}_{ki} \ln(t_i) \left(\sum_{i=1}^n \hat{z}_{ki}\right)^{-1}}{\frac{1}{\hat{\alpha}_{k,r}} - \psi'(\hat{\alpha}_{k,r})} \quad (5.11)$$

$$\hat{\beta}_k = \left(\hat{\alpha}_k \sum_{i=1}^n \hat{z}_{ki}\right)^{-1} \sum_{i=1}^n \hat{z}_{ki} t_i \quad (5.12)$$

ile elde edilir. Eşitlikte yer alan $\psi'(\cdot)$ digamma fonksiyonunun türevidir ve trigamma fonksiyonu olarak isimlendirilir. Digamma ve trigamma fonksiyonları yaklaşık olarak sırasıyla

$$\psi(\alpha) = \begin{cases} \ln(\alpha) - (1 + (1 - (1/10 - 1/(21\alpha^2)))/\alpha^2)/(6\alpha)/(2\alpha) & \alpha \geq 8 \\ \psi(\alpha + 1) - 1/\alpha & \alpha < 8 \end{cases} \quad (5.13)$$

$$\psi'(\alpha) = \begin{cases} (1 + (1 + (1 - (1/5 - 1/(7\alpha^2)))/\alpha^2)/(3\alpha))/(2\alpha)/\alpha & \alpha \geq 8 \\ \psi'(\alpha + 1) + 1/\alpha^2 & \alpha < 8 \end{cases} \quad (5.14)$$

şeklinde hesaplanır (Choi ve Wette 1969).

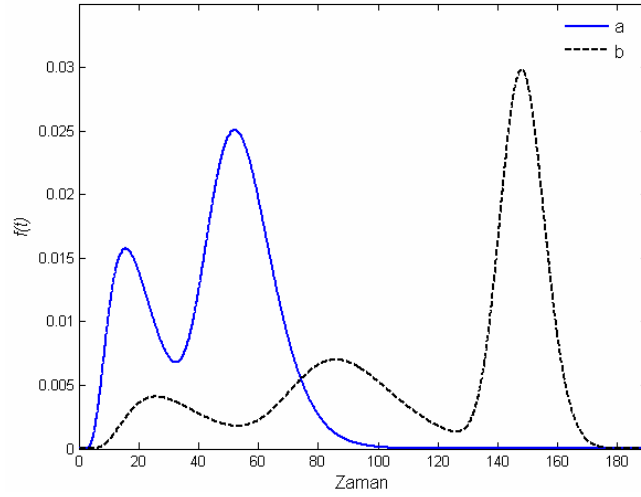
5.1.2. Karma Lognormal Dağılım Modeli

T rassal değişkeni $g \geq 2$ olmak üzere g bileşenli karma Lognormal dağılımına sahip ise olasılık yoğunluk fonksiyonu

$$f(t | \psi) = \sum_{k=1}^g \pi_k \frac{e^{-\frac{1}{2} \left(\frac{\ln t - \mu_k}{\sigma_k} \right)^2}}{t \sigma_k \sqrt{2\pi}}, \quad t > 0, \mu_k, \sigma_k > 0 \quad (5.15)$$

şeklinde tanımlanır. Burada $k = 1, \dots, g$ olmak üzere $\pi_k, \pi_k \in (0,1)$ ve $\sum_{k=1}^g \pi_k = 1$ koşuluyla k . Lognormal bileşenin karma ağırlığı ve $\psi = (\pi, \mu, \sigma)$, $\pi = (\pi_1, \dots, \pi_g)$ ve $(\mu, \sigma) = ((\mu_1, \sigma_1), \dots, (\mu_g, \sigma_g))$ bilinmeyen vektörlerini içeren karma Lognormal

dağılım modelinde bilinmeyen parametre vektörü olarak tanımlanır. Farklı parametrelere sahip iki ve üç bileşenli karma Lognormal dağılım modelleri için yoğunluk eğrileri Şekil 5.3’de verilmiştir.



Şekil 5.3. Farklı parametrelere sahip iki ve üç bileşenli karma Lognormal dağılım modelleri için yoğunluk eğrileri a) $\pi_1 = 0.35$, $\mu_1 = 3$, $\mu_2 = 4$, $\sigma_1 = 0.5$, $\pi_2 = 0.65$, ve $\sigma_2 = 0.2$ (düz çizgi) b) $\pi_1 = 0.15$, $\mu_1 = 3.5$, $\sigma_1 = 0.5$, $\pi_2 = 0.3$, $\mu_2 = 4.5$, $\sigma_2 = 0.2$ ve $\pi_3 = 0.4$, $\mu_3 = 5$, $\sigma_3 = 0.05$ (kesik çizgi)

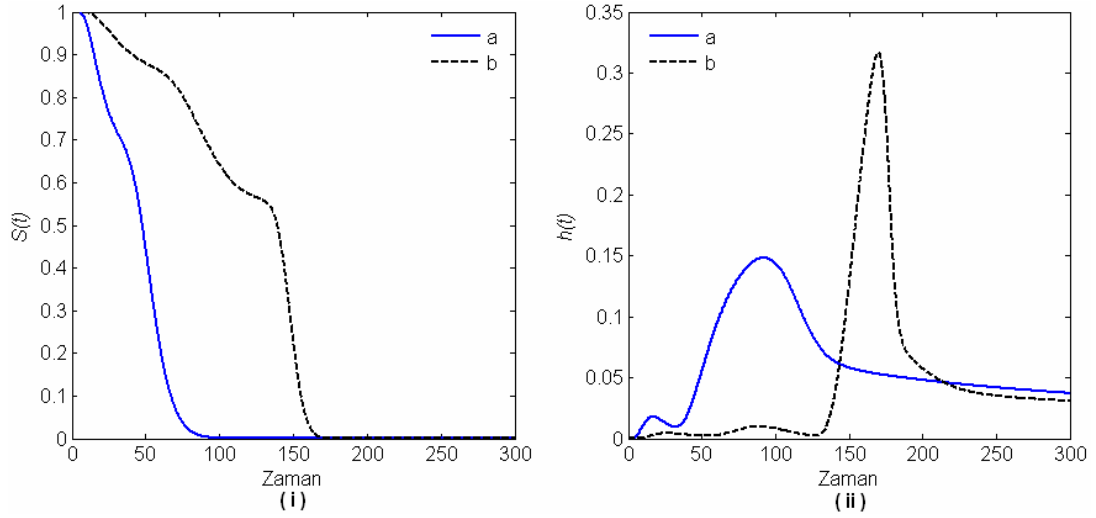
Karma Lognormal dağılım modelinde yaşam fonksiyonu, hazard fonksiyonu ve beklenen yaşam süresi sırasıyla

$$S(t | \psi) = \sum_{k=1}^g \pi_k \left(1 - \Phi \left(\frac{\ln t - \mu_k}{\sigma_k} \right) \right) \quad (5.16)$$

$$h(t) = \sum_{k=1}^g \pi_k \frac{e^{-\frac{1}{2} \left(\frac{\ln t - \mu_k}{\sigma_k} \right)^2}}{t \sigma_k \sqrt{2\pi} S(t | \psi)} \quad (5.17)$$

$$E(T) = \sum_{k=1}^g \pi_k e^{\mu_k + \frac{\sigma_k^2}{2}} \quad (5.18)$$

şeklinde tanımlanır. Farklı parametrelere sahip iki ve üç bileşenli karma Lognormal dağılım modelleri için yaşam fonksiyonu ve hazard fonksiyonu eğrileri Şekil 5.4’de verilmiştir.



Şekil 5.4. Farklı parametrelere sahip iki ve üç bileşenli karma Lognormal dağılım modeli için (i) yaşam fonksiyonu ve (ii) hazard fonksiyonu eğrileri
a) $\pi_1 = 0.35, \mu_1 = 3, \sigma_1 = 0.5, \pi_2 = 0.65, \mu_2 = 4$ ve $\sigma_2 = 0.2$ (düz çizgi)
b) $\pi_1 = 0.15, \mu_1 = 3.5, \sigma_1 = 0.5, \pi_2 = 0.3, \mu_2 = 4.5, \sigma_2 = 0.2$ ve $\pi_3 = 0.4, \mu_3 = 5, \sigma_3 = 0.05$ (kesik çizgi)

Karma Lognormal dağılım modelindeki parametrelerin en çok olabilirlik tahminlerini elde etmek için algoritmanın E adımında kayıp gözlem vektörünün elemanları

$$\hat{z}_{ki} = \frac{\pi_k f(t_i | \mu_k, \sigma_k)}{\sum_{s=1}^g \pi_s f_s(t_i | \mu_s, \sigma_s)}, \quad k = 1, \dots, g \quad (5.19)$$

şeklinde tahmin edilir. Eşitlikte yer alan $f(t_i | \mu_k, \sigma_k)$, k . Lognormal bileşenine ait olasılık yoğunluk fonksiyonunu göstermektedir. Böylelikle karma Lognormal dağılım modelinde tamamlanmış veri için log olabilirlik fonksiyonunun beklenen değeri

$$E(\ln L) = \sum_{i=1}^n \sum_{k=1}^g \hat{z}_{ki} \ln \left(\pi_k \frac{e^{-\frac{1}{2} \left(\frac{\ln t_i - \mu_k}{\sigma_k} \right)^2}}{t_i \sigma_k \sqrt{2\pi}} \right) \quad (5.20)$$

şeklinde elde edilir. EM algoritmasının M adımında, E adımında elde edilen $E(\ln L)$,

$\sum_{k=1}^g \pi_k = 1$ koşulu altında maksimize edilir. Langrange metodu ile eşitlik (5.20)

düzenlenerek açık bir notasyonla

$$\begin{aligned} \ln \tilde{L} = & \sum_{i=1}^n \sum_{k=1}^g \hat{z}_{ki} \ln \pi_k - \frac{1}{2} \sum_{i=1}^n \sum_{k=1}^g \hat{z}_{ki} \left(\frac{\ln t_i - \mu_k}{\sigma_k} \right)^2 - \sum_{i=1}^n \sum_{k=1}^g \hat{z}_{ki} \ln t_i \\ & - \sum_{i=1}^n \sum_{k=1}^g \hat{z}_{ki} \ln \sigma_k - \sum_{i=1}^n \sum_{k=1}^g \hat{z}_{ki} \ln \sqrt{2\pi} - \gamma \left(\sum_{k=1}^g \pi_k - 1 \right) \end{aligned} \quad (5.21)$$

şeklinde elde edilir. $\ln \tilde{L}$ fonksiyonunu maksimize eden parametreleri elde etmek için, sırasıyla tüm parametrelere göre ayrı ayrı $\ln \tilde{L}$ 'nin ilk türevleri alınarak sıfıra eşitlenir. Karma Lognormal dağılım modelinde k . Lognormal bileşeninin karma ağırlığının tahmini,

$$\hat{\pi}_k = n^{-1} \sum_{i=1}^n \hat{z}_{ki}, \quad k = 1, \dots, g \quad (5.22)$$

şeklinde yapılır. Karma Lognormal dağılım modelinde k . Lognormal bileşenine ait μ_k ve σ_k^2 parametrelerinin en çok olabilirlik tahminleri

$$\hat{\mu}_k = \left(\sum_{i=1}^n \hat{z}_{ki} \right)^{-1} \sum_{i=1}^n \hat{z}_{ki} \ln t_i, \quad k = 1, \dots, g \quad (5.23)$$

$$\hat{\sigma}_k^2 = \left(\sum_{i=1}^n \hat{z}_{ki} \right)^{-1} \sum_{i=1}^n \hat{z}_{ki} (\ln t_i - \hat{\mu}_k)^2, \quad k = 1, \dots, g \quad (5.24)$$

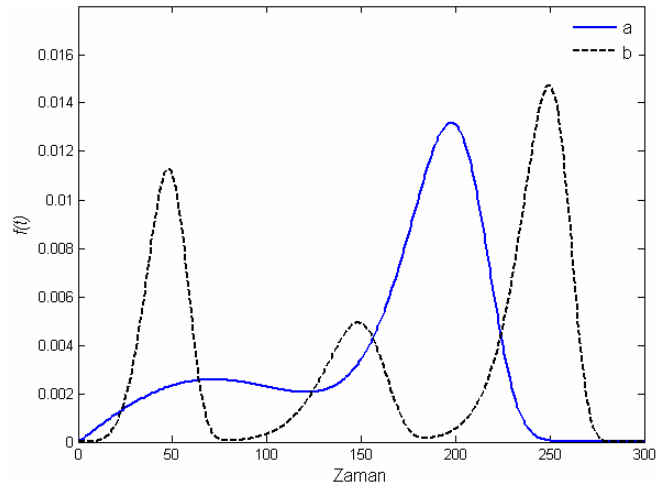
şeklinde elde edilir.

5.1.3. Karma Weibull Dağılım Modeli

T rassal değişkeni $g \geq 2$ olmak üzere g bileşenli karma Weibull dağılımına sahip ise olasılık yoğunluk fonksiyonu

$$f(t | \psi) = \sum_{k=1}^g \pi_k \frac{\alpha_k}{\beta_k} \left(\frac{t}{\beta_k} \right)^{\alpha_k - 1} e^{-\left(\frac{t}{\beta_k} \right)^{\alpha_k}}, \quad t > 0, \alpha_k, \beta_k > 0 \quad (5.25)$$

şeklinde tanımlanır. Burada $k = 1, \dots, g$ olmak üzere $\pi_k, \pi_k \in (0,1)$ ve $\sum_{k=1}^g \pi_k = 1$ koşuluyla k . Weibull bileşeninin karma ağırlığı ve $\psi = (\pi, \alpha, \beta)$, $\pi = (\pi_1, \dots, \pi_g)$ ve $(\alpha, \beta) = ((\alpha_1, \beta_1), \dots, (\alpha_g, \beta_g))$ bilinmeyen vektörlerini içeren karma Weibull dağılım modelinde bilinmeyen parametre vektörü olarak tanımlanır. Farklı parametrelere sahip iki ve üç bileşenli karma Weibull dağılım modelleri için yoğunluk eğrileri Şekil 5.5’de verilmiştir.



Şekil 5.5. Farklı parametrelere sahip iki ve üç bileşenli karma Weibull dağılım modelleri için yoğunluk eğrileri
a) $\pi_1 = 0.3, \alpha_1 = 2, \beta_1 = 100, \pi_2 = 0.7, \alpha_2 = 10$ ve $\beta_2 = 200$ (düz çizgi)
b) $\pi_1 = 0.3, \alpha_1 = 5, \beta_1 = 50, \pi_2 = 0.2, \alpha_2 = 10, \beta_2 = 150$ ve $\pi_3 = 0.5, \alpha_3 = 20, \beta_3 = 250$ (kesik çizgi)

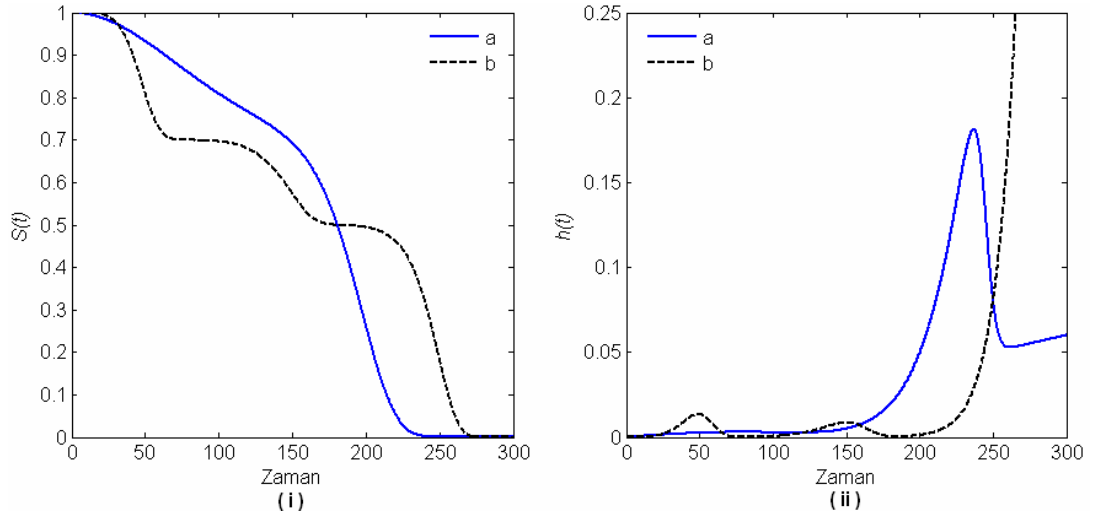
Karma Weibull dağılım modelinde yaşam fonksiyonu, hazard fonksiyonu ve beklenen yaşam süresi sırasıyla

$$S(t | \psi) = \sum_{k=1}^g \pi_k e^{-\left(\frac{t}{\beta_k}\right)^{\alpha_k}} \quad (5.26)$$

$$h(t) = \sum_{k=1}^g \frac{\pi_k}{S(t | \psi)} \frac{\alpha_k}{\beta_k} \left(\frac{t}{\beta_k}\right)^{\alpha_k-1} e^{-\left(\frac{t}{\beta_k}\right)^{\alpha_k}} \quad (5.27)$$

$$E(T) = \sum_{k=1}^g \pi_k \beta_k \Gamma\left(1 + \frac{1}{\alpha_k}\right) \quad (5.28)$$

şeklinde tanımlanır. Farklı parametrelere sahip iki ve üç bileşenli karma Weibull dağılım modelleri için yaşam fonksiyonu ve hazard fonksiyonu eğrileri Şekil 5.6'da verilmiştir.



Şekil 5.6. Farklı parametrelere sahip iki ve üç bileşenli karma Weibull dağılım modeli için (i) yaşam fonksiyonu ve (ii) hazard fonksiyonu eğrileri
a) $\pi_1 = 0.3$, $\alpha_1 = 2$, $\beta_1 = 100$, $\pi_2 = 0.7$, $\alpha_2 = 10$ ve $\beta_2 = 200$ (düz çizgi)
b) $\pi_1 = 0.3$, $\alpha_1 = 5$, $\beta_1 = 50$, $\pi_2 = 0.2$, $\alpha_2 = 10$, $\beta_2 = 150$, $\pi_3 = 0.5$,
 $\alpha_3 = 20$ ve $\beta_3 = 250$ (kesik çizgi)

Karma Weibull dağılım modelindeki parametrelerin en çok olabilirlik tahminlerini elde etmek için algoritmanın E adımında kayıp gözlem vektörünün elemanları

$$\hat{z}_{ki} = \frac{\pi_k f(t_i | \alpha_k, \beta_k)}{\sum_{s=1}^g \pi_s f_s(t_i | \alpha_s, \beta_s)}, \quad k = 1, \dots, g \quad (5.29)$$

şeklinde tahmin edilir. Eşitlikte yer alan $f(t_i | \alpha_k, \beta_k)$, k . Weibull bileşenine ait olasılık yoğunluk fonksiyonunu göstermektedir. Böylelikle karma Weibull dağılım modelinde tamamlanmış veri için log olabilirlik fonksiyonunun beklenen değeri

$$E(\ln L) = \sum_{i=1}^n \sum_{k=1}^g \hat{z}_{ki} \ln \left(\pi_k \frac{\alpha_k}{\beta_k} \left(\frac{t_i}{\beta_k} \right)^{\alpha_k - 1} e^{-\left(\frac{t_i}{\beta_k} \right)^{\alpha_k}} \right) \quad (5.30)$$

şeklinde elde edilir. EM algoritmasının M adımında, E adımında elde edilen $E(\ln L)$,

$\sum_{k=1}^g \pi_k = 1$ koşulu altında maksimize edilir. Langrange metodu ile eşitlik (5.30)

düzenlenerek açık bir notasyonla

$$\begin{aligned} \ln \tilde{L} = & \sum_{i=1}^n \sum_{k=1}^g \hat{z}_{ki} \ln \pi_k + \sum_{i=1}^n \sum_{k=1}^g \hat{z}_{ki} \ln \alpha_k + \sum_{i=1}^n \sum_{k=1}^g \hat{z}_{ki} (\alpha_k - 1) \ln t_i \\ & - \sum_{i=1}^n \sum_{k=1}^g \hat{z}_{ki} \alpha_k \ln \beta_k - \sum_{i=1}^n \sum_{k=1}^g \hat{z}_{ki} \left(\frac{t_i}{\beta_k} \right)^{\alpha_k} - \gamma \left(\sum_{k=1}^g \pi_k - 1 \right) \end{aligned} \quad (5.31)$$

şeklinde elde edilir. $\ln \tilde{L}$ fonksiyonunu maksimize eden parametreleri tahmin etmek için, sırasıyla tüm parametrelere göre ayrı ayrı $\ln \tilde{L}$ fonksiyonunun ilk türevleri alınarak sıfıra eşitlenir. Karma Weibull dağılım modelinde k . Weibull bileşeninin karma ağırlığının tahmini,

$$\hat{\pi}_k = n^{-1} \sum_{i=1}^n \hat{z}_{ki}, \quad k = 1, \dots, g \quad (5.32)$$

şeklinde yapılır. Karma Weibull dağılım modelinde k . Weibull bileşenine ait α_k ve β_k parametrelerinin en çok olabilirlik tahminleri için ise sırasıyla $\ln \tilde{L}$ fonksiyonu α_k ve β_k 'ya göre türevlenip sıfıra eşitlenir.

$$\frac{\partial}{\partial \alpha_k} = \sum_{i=1}^n \hat{z}_{ki} \left\{ \frac{1}{\alpha_k} - \ln \beta_k + \ln t_i - \left(\frac{t_i}{\beta_k} \right)^{\alpha_k} \ln \left(\frac{t_i}{\beta_k} \right) \right\} = 0 \quad k = 1, \dots, g \quad (5.33)$$

$$\frac{\partial}{\partial \beta_k} = \sum_{i=1}^n \frac{\hat{z}_{ki} \alpha_k}{\beta_k} \left\{ \left(\frac{t_i}{\beta_k} \right)^{\alpha_k} - 1 \right\} = 0 \quad k = 1, \dots, g \quad (5.34)$$

Eşitlik (5.34) düzenlendiğinde,

$$\beta_k = \left\{ \left(\sum_{i=1}^n \hat{z}_{ki} \right)^{-1} \left(\sum_{i=1}^n \hat{z}_{ki} t_i^{\alpha_k} \right) \right\}^{1/\alpha_k} \quad (5.35)$$

elde edilir. Eşitlik (5.33) kapalı formda olduğundan dolayı doğrudan analitik olarak çözülemez. Karma Weibull dağılım modelinde k . Weibull bileşeninin biçim parametresi α_k 'nın en çok olabilirlik tahminini elde etmek için öncelikle (5.33) nolu eşitlikte yer alan β_k ifadesi eşitlik (5.35)'teki gibi α_k cinsinden yazılır. Elde edilen denklem Newton-Raphson iteratif yöntemi ile çözüldüğünde karma Weibull dağılım modelinde k . Weibull bileşeninin biçim parametresi α_k 'nın $(r+1)$. iterasyondaki en çok olabilirlik tahminini

$$\hat{\alpha}_{k,(r+1)} = \hat{\alpha}_{k,r} + \frac{A_{k,r} + (1/\hat{\alpha}_{k,r}) - (C_{k,r}/B_{k,r})}{(1/\hat{\alpha}_{k,r}^2) + (B_{k,r}D_{k,r} - C_{k,r}^2)/B_{k,r}^2}, \quad k = 1, \dots, g \quad (5.36)$$

eşitliği ile elde edilir. Burada $A_{k,r} = \left(\sum_{i=1}^n \hat{z}_{ki} \right)^{-1} \sum_{i=1}^n \hat{z}_{ki} \ln t_i$, $B_{k,r} = \sum_{i=1}^n \hat{z}_{ki} t_i^{\hat{\alpha}_{k,r}}$,
 $C_{k,r} = \sum_{i=1}^n \hat{z}_{ki} t_i^{\hat{\alpha}_{k,r}} \ln t_i$ ve $D_{k,r} = \sum_{i=1}^n \hat{z}_{ki} t_i^{\hat{\alpha}_{k,r}} (\ln t_i)^2$ şeklinde tanımlanır.

5.2. EM Algoritmasının Parametre Tahminindeki Etkinliği

Karma dağılım modellerindeki parametrelerin en çok olabilirlik tahminlerinin elde edilmesi için tanımlanan EM algoritmalarının parametre tahminindeki etkinliği göstermek için örneklem genişliği 100 birim, tekrar sayısı 1000 olmak üzere simülasyonlar gerçekleştirilmiştir. Simülasyon çalışmasında yoğunluk eğrileri sırasıyla Şekil 5.1, Şekil 5.3 ve Şekil 5.5’de verilen karma Gamma, karma Lognormal ve Karma Weibull dağılımına sahip kitleler için parametre tahminleri gerçekleştirilmiştir. Tahminler arasındaki fark 10^{-4} ’den daha az olduğu zaman yakınsamanın olduğu varsayılmıştır.

İki bileşenli karma Gamma dağılım modeline sahip kitleler için gerçekleştirilen simülasyon sonrası EM algoritması ile elde edilen parametrelerin en çok olabilirlik tahminlerinin ortalamaları ve standart hataları Tablo 5.1’de verilmiştir.

Tablo 5.1. İki bileşenli karma Gamma dağılım modelinde EM algoritması ile elde edilen parametrelerin en çok olabilirlik tahminlerinin ortalamaları ve standart hataları

Parametreler	Kitle	Parametre Tahmini		Kitle	Parametre Tahmini	
		Ortalama	Std. Hata		Ortalama	Std. Hata
π_1	0.30	0.3047	0.0019	0.40	0.4018	0.0018
α_1	0.50	0.5021	0.0019	2.00	2.1384	0.0151
β_1	2.00	2.1134	0.0393	1.00	1.0126	0.0125
π_2	0.70	0.6953	0.0019	0.60	0.5982	0.0018
α_2	5.00	5.1274	0.0307	10.00	10.3456	0.0532
β_2	2.00	2.0223	0.0124	1.00	0.9928	0.0049

İki ve üç bileşenli karma Lognormal dağılım modeline sahip kitleler için gerçekleştirilen simülasyon sonrası EM algoritması ile elde edilen parametrelerin en çok olabilirlik tahminlerinin ortalamaları ve standart hataları Tablo 5.2’de verilmiştir.

Tablo 5.2. İki ve üç bileşenli karma Lognormal dağılım modelinde EM algoritması ile elde edilen parametrelerin en çok olabilirlik tahminlerinin ortalamaları ve standart hataları

Parametreler	Kitle	Parametre Tahmini		Kitle	Parametre Tahmini	
		Ortalama	Std. Hata		Ortalama	Std. Hata
π_1	0.35	0.3493	0.0019	0.15	0.1524	0.0010
μ_1	3.00	2.9899	0.0041	3.50	3.4954	0.0057
σ_1	0.50	0.4843	0.0023	0.50	0.4910	0.0026
π_2	0.65	0.6507	0.0019	0.30	0.2977	0.0016
μ_2	4.00	4.0003	0.0009	4.50	4.5043	0.0021
σ_2	0.20	0.1960	0.0008	0.20	0.1887	0.0013
π_3				0.55	0.5499	0.0015
μ_3				5.00	4.9951	0.0015
σ_3				0.05	0.0513	0.0006

İki ve üç bileşenli karma Weibull dağılım modeline sahip kitleler için gerçekleştirilen simülasyon sonrası EM algoritması ile elde edilen parametrelerin en çok olabilirlik tahminlerinin ortalamaları ve standart hataları Tablo 5.3’de verilmiştir.

Tablo 5.3. İki ve üç bileşenli karma Weibull dağılım modelinde EM algoritması ile elde edilen parametrelerin en çok olabilirlik tahminlerinin ortalamaları ve standart hataları

Parametreler	Kitle	Parametre Tahmini		Kitle	Parametre Tahmini	
		Ortalama	Std. Hata		Ortalama	Std. Hata
π_1	0.30	0.2961	0.0016	0.30	0.2994	0.0015
α_1	2.00	2.1064	0.0064	5.00	5.1873	0.0262
β_1	100.00	99.0050	0.0986	50.00	50.0388	0.0643
π_2	0.70	0.7039	0.0016	0.20	0.2010	0.0013
α_2	10.00	10.2530	0.0473	10.00	10.8048	0.0750
β_2	200.00	199.8378	0.0925	150.00	150.2172	0.1588
π_3				0.50	0.4996	0.0016
α_3				20.00	20.6329	0.0787
β_3				250.00	250.0025	0.0637

Simülasyon sonucunda elde edilen ortalama parametre tahminleri incelendiğinde EM algoritmasının karma dağılım modellerindeki farklı parametre değerlerinin tahmininde oldukça başarılı olduğu görülmektedir. Standart hataların oldukça küçük olması EM algoritmasının sayısal olarak tutarlı olduğunu vurgular.

Simülasyon sonuçlarına göre EM yaklaşımının farklı karma oranlarında da iyi çalıştığı görülmektedir.

5.3. Heterojen Yaşam Verilerinin Modellenmesinde Karma Dağılım Modelleri

Heterojen yaşam verilerinin modellenmesinde karma dağılım modellerinin kullanımını göstermek için literatürde kullanılan ve iyi bilinen iki gerçek yaşam verisi kullanılacaktır. Jiang ve Murthy (1995) tarafından tanımlanan ışınlanmış farelerinin ölüm zamanlarından oluşan veri setinde bileşen sayısının 2 olduğu bilinmekte iken Dirican (2004) tarafından tanımlanan akciğer kanseri hastalarının yaşam sürelerinden oluşan veri setinde bileşen sayısı bilinmemektedir. Akciğer kanser verisinin karma dağılım modelleri ile modellenmesi için öncelikle bileşen sayı Akaike Bilgi Kriterine göre belirlenmiştir.

Karma dağılım modellerinin klasik istatistik modelleri ile karşılaştırılmasında log olabilirlik değeri ($\ln L$), Kolmogorov-Smirnov test istatistiği ve hata kareler ortalaması değerleri kullanılacaktır. Karma dağılım modellerinin klasik istatistik modelleri ile karşılaştırılmasında log olabilirlik değeri büyük olan model daha başarılı model olarak tanımlanırken diğer iki kritere göre ise karşılaştırma değeri düşük olan model daha başarılı modeldir.

$F(t)$ verilerin modellenmesinde kullanılan teorik dağılımın birikimli dağılım fonksiyonu ve $G(t)$ gözleme dayalı birikimli dağılım fonksiyonu olmak üzere

$$D_n = \sup |F(t) - G(t)| \quad (5.37)$$

istatistiğe Kolmogorov-Smirnov test istatistiği denir (Guenther 1975). Gözleme dayalı birikimli dağılım fonksiyonu $G(t)$ 'nin oluşturulması için pratikte birçok formül bulunmaktadır. Bu çalışmada daha yaygın bir kullanıma sahip olması nedeni ile $G(t)$, küçükten büyüğe doğru sıralanmış verilerde i gözlem sırasını göstermek üzere

$$G(t_i) = \frac{i - 0.3}{n + 0.4}, \quad i = 1, \dots, n \quad (5.38)$$

eşitliği ile hesaplanmıştır. Kolmogorov-Smirnov test istatistiğinin hesaplanan değeri için anlamlılık değeri yaklaşık olarak

$$p = P(D_n \geq D_{hesap}) = 2 \sum_{i=1}^{\infty} (-1)^{i-1} e^{-2i^2 n D_{hesap}^2} \quad (5.39)$$

şeklinde hesaplanır (Borovkov 1984). Elde edilen anlamlılık değeri p , I. Tip hata olasılığı α 'dan büyük ise yani $p > \alpha$ ise verinin modellenmesi için önerilen teorik dağılımın verinin modellenmesi için uygun olduğuna karar verilir.

Karşılaştırmada kullanılacak diğer kriter Hata Kareler Ortalaması,

$$HKO = \frac{1}{n-d} \sum_{i=1}^n (S(t_i) - \hat{S}(t_i))^2 \quad (5.40)$$

eşitliği ile hesaplanır. Eşitlikte yer alan $\hat{S}(t_i)$ gözleme dayalı yaşam fonksiyonu ve d serbest parametre sayısı olarak isimlendirilir.

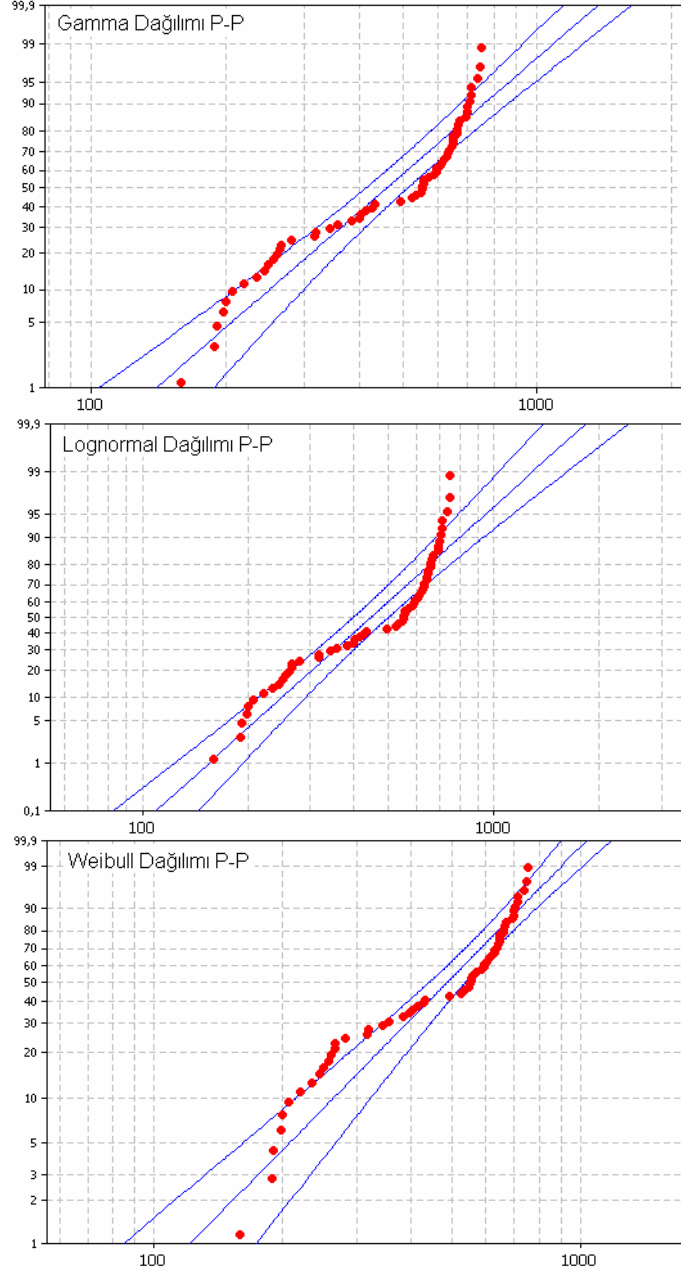
5.3.1. Işınlanmış Farelerin Ölüm Zamanları Verisi

Jiang ve Murthy (1995), ışınlanmış farelerinin ölüm zamanlarını ölüm nedenleri bakımından iki alt gruba ayırmıştır. Işınlanmış farelerin ölüm zamanlarının karma dağılımlarla modellenmesinde, karma modeldeki bileşen sayısı bu nedenle iki kabul edilecektir. Işınlanmış farelerin ölüm zamanları Tablo 5.4'de verilmiştir.

Tablo 5.4. Işınlanmış farelerin ölüm zamanları (Jiang ve Murthy 1995)

159	189	191	198	200	207	220	235	245	250	256	261	265	266	280
317	318	343	356	383	399	403	414	428	432	495	525	536	549	554
557	558	571	586	594	596	605	612	621	628	631	636	643	647	648
649	661	663	666	670	695	697	700	705	712	713	738	748		

Işınlanmış farelerin ölüm zamanlarının heterojen yapıda olduğunu göstermek için, klasik istatistik dağılımları ile oluşturulan P-P grafikleri Şekil 5.7’de verilmiştir.



Şekil 5.7. Işınlanmış farelerinin ölüm zamanları için Gamma, Lognormal ve Weibull dağılımlarına ait P-P grafikleri

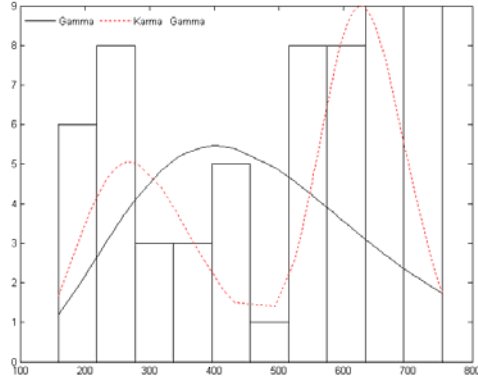
Şekil 5.7 incelendiğinde, ıřınlanmış farelerin ölüm zamanlarının heterojen yapıda olduđu görölmektedir.

Işınlanmış farelerin ölüm zamanlarının modellenmesinde kullanılan dağılımlara ait parametre tahminleri, log olabilirlik değerleri ($\ln L$), Kolmogorov-Smirnov test istatistikleri (D_{hesap}), hata kareler ortalaması (HKO) ve ortalama yaşam süreleri ($E(T)$) Tablo 5.5’de verilmiştir. Hesaplamalar Matlab programı yardımıyla yapılmıştır.

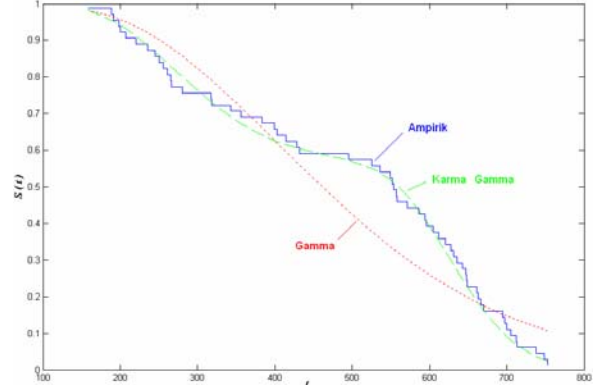
Tablo 5.5. Işınlanmış farelerin ölüm zamanları verisinin modellenmesinde kullanılan dağılımların parametre tahminleri, $\ln L$, D_{hesap} , HKO ve $E(T)$ Değerleri

Model	Tahminler		$\ln L$	D_{hesap} (p değeri)	HKO	$E(T)$
Gamma	$\hat{\alpha} = 5.6898$	$\hat{\beta} = 85.911$	-400.8227	0.189 (0.0275)	0.0083	488.8164
Karma Gamma	$\hat{\alpha}_1 = 92.975$	$\hat{\alpha}_2 = 10.421$	-381.7712	0.054 (0.9948)	0.0005	488.8226
	$\hat{\beta}_1 = 6.824$	$\hat{\beta}_2 = 28.609$				
	$\hat{\pi}_1 = 0.567$	$\hat{\pi}_2 = 0.433$				
Lognormal	$\hat{\mu} = 6.1015$	$\hat{\sigma} = 0.4562$	-403.6400	0.200 (0.0165)	0.0093	495.496
Karma Lognormal	$\hat{\mu}_1 = 5.650$	$\hat{\mu}_2 = 6.447$	-381.5347	0.047 (0.9994)	0.0005	488.7503
	$\hat{\sigma}_1 = 0.313$	$\hat{\sigma}_2 = 0.104$				
	$\hat{\pi}_1 = 0.434$	$\hat{\pi}_2 = 0.566$				
Weibull	$\hat{\alpha} = 3.0507$	$\hat{\beta} = 549.1703$	-397.0355	0.157 (0.1038)	0.0070	490.7649
Karma Weibull	$\hat{\alpha}_1 = 3.7614$	$\hat{\alpha}_2 = 9.7307$	-381.5347	0.061 (0.9789)	0.0004	488.9768
	$\hat{\beta}_1 = 313.520$	$\hat{\beta}_2 = 657.200$				
	$\hat{\pi}_1 = 0.397$	$\hat{\pi}_2 = 0.603$				

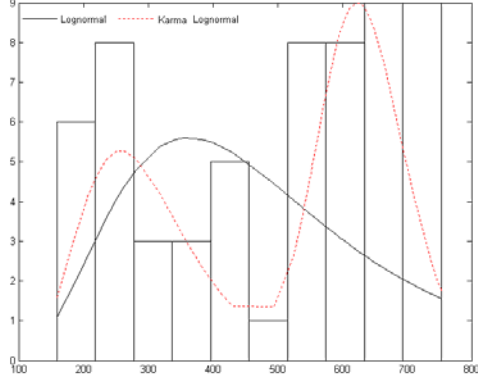
Tablo 5.5 incelendiğinde, tüm karşılaştırma kriterlerine göre ışınlamış farelerin ölüm zamanlarının modellenmesinde karma dağılım modellerinin klasik istatistik dağılımlarından daha iyi olduğu görülmektedir. Ortalama yaşam süreleri incelendiğinde; dağılımların karmalarının ortalama yaşam sürelerinin birbirlerine yaklaşık olarak benzer olduğu, klasik istatistik dağılımların ortalama yaşam sürelerinin birbirinde farklı olduğu görülmektedir. Işınlanmış farelerin ölüm zamanları için klasik istatistik dağılımların ve karma modellerin grafiksel karşılaştırılması Şekil 5.8’de verilmiştir.



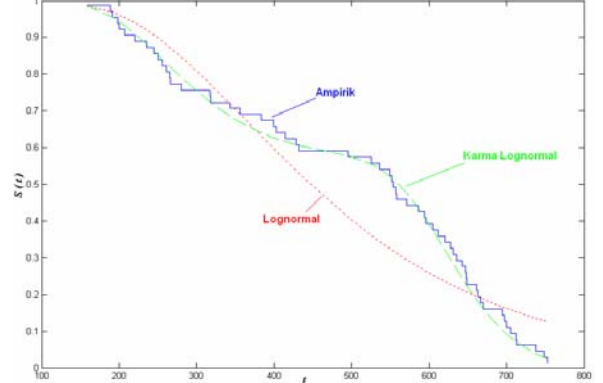
(a) Gamma ve karma Gamma dağılım modellerinin yoğunluk eğrileri



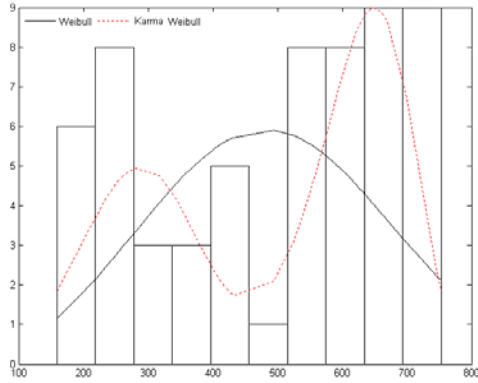
(b) Ampirik yaşam eğrisinin Gamma ve karma Gamma dağılım modellerinin yaşam eğrileri ile karşılaştırılması



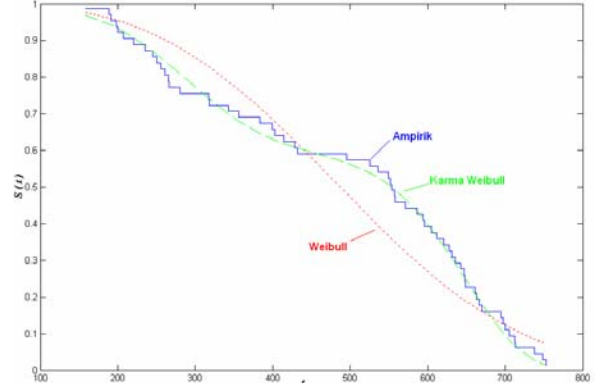
(c) Lognormal ve karma Lognormal dağılım modellerinin yoğunluk eğrileri



(d) Ampirik yaşam eğrisinin Lognormal ve karma Lognormal dağılım modellerinin yaşam eğrileri ile karşılaştırılması



(e) Weibull ve karma Weibull dağılım modellerinin yoğunluk eğrileri



(f) Ampirik yaşam eğrisinin Weibull ve karma Weibull dağılım modellerinin yaşam eğrileri ile karşılaştırılması

Şekil 5.8. Işınlanmış farelerinin ölüm zamanlarının modellenmesinde klasik istatistik dağılımları ile karma dağılım modellerinin grafiksel karşılaştırılması

Şekil 5.8’de yer alan yoğunluk eğrileri incelendiğinde heterojen yapıdaki ışınlanmış farelerin ölüm zamanlarının klasik istatistik dağılımlarıyla

modellenmesinde yapılacak hatanın, karma dağılım modelleri ile modellenmesinde yapılacak hatadan daha büyük olduğu görülmektedir. Buna göre heterojen yapıdaki bu verinin modellenmesinde karma dağılım modelleri klasik istatistik dağılım modellerine göre daha başarılıdır. Klasik istatistik dağılımları ve karma dağılım modellerinin yaşam eğrilerinin ampirik yaşam eğrisi ile karşılaştırılmasında da karma dağılım modellerinin veriyi modellemede daha başarılı olduğu açıkça görülmektedir.

5.3.2. Akciğer Kanseri Verisi

Dirican (2004) tarafından tanımlanan akciğer kanser verisi, akciğer kanseri tanısı alan 183 hastanın tanı konduktan sonraki yaşam sürelerinden (gün) oluşmaktadır. Akciğer kanseri tanısı alan hastaların tanı konulduktan sonraki yaşam süreleri Tablo 5.6’da verilmiştir.

Tablo 5.6. Akciğer kanseri tanısı alan hastaların tanı konduktan sonraki yaşam süreleri (Dirican 2004)

3	3	6	9	12	12	15	18	20	20
21	21	21	21	21	21	21	24	24	24
24	27	30	30	30	30	30	30	30	38
39	42	45	45	48	51	51	54	54	57
60	60	60	60	66	69	69	75	75	78
84	84	87	87	87	90	90	90	96	99
102	105	105	105	105	108	111	114	120	120
120	120	120	123	123	123	126	130	135	138
144	153	156	162	162	165	168	168	180	183
183	189	189	192	192	192	192	195	201	201
201	204	210	210	210	216	216	219	219	222
225	225	225	225	240	240	245	246	249	249
252	255	258	261	267	270	273	276	282	285
285	288	291	297	297	297	306	315	315	324
327	327	327	333	333	339	342	342	345	345
348	351	360	366	366	375	375	381	381	384
393	402	414	414	420	420	426	441	450	471
498	501	516	522	561	561	561	564	570	570
603	681	858							

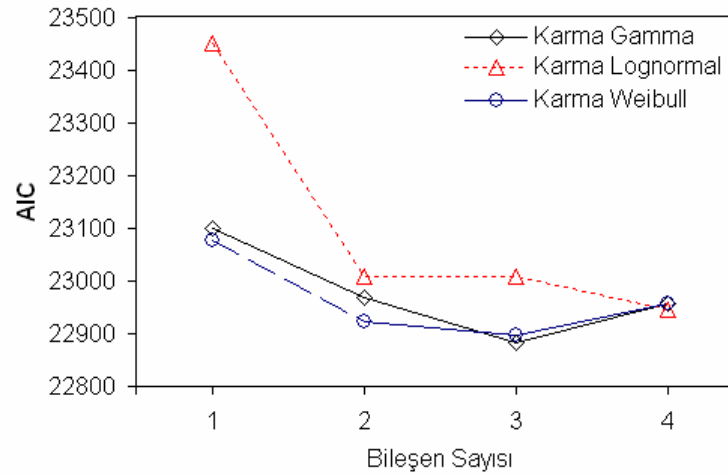
Akciğer kanseri tanısı konan hastaların yaşam sürelerinin modellenmesinde karma dağılım modellerinin kullanılması için öncelikle uygun bileşen sayısı belirlenir. Bu çalışmada uygun bileşen sayısı Akaike Bilgi Kriteri (AIC) ile

belirlenmiştir. İncelenen karma dağılım modellerinde farklı bileşen sayısına göre elde edilen $\ln L$ ve AIC değerleri Tablo 5.7’de verilmiştir.

Tablo 5.7. Karma dağılım modellerinde bileşen sayısına göre $\ln L$ ve AIC değerleri

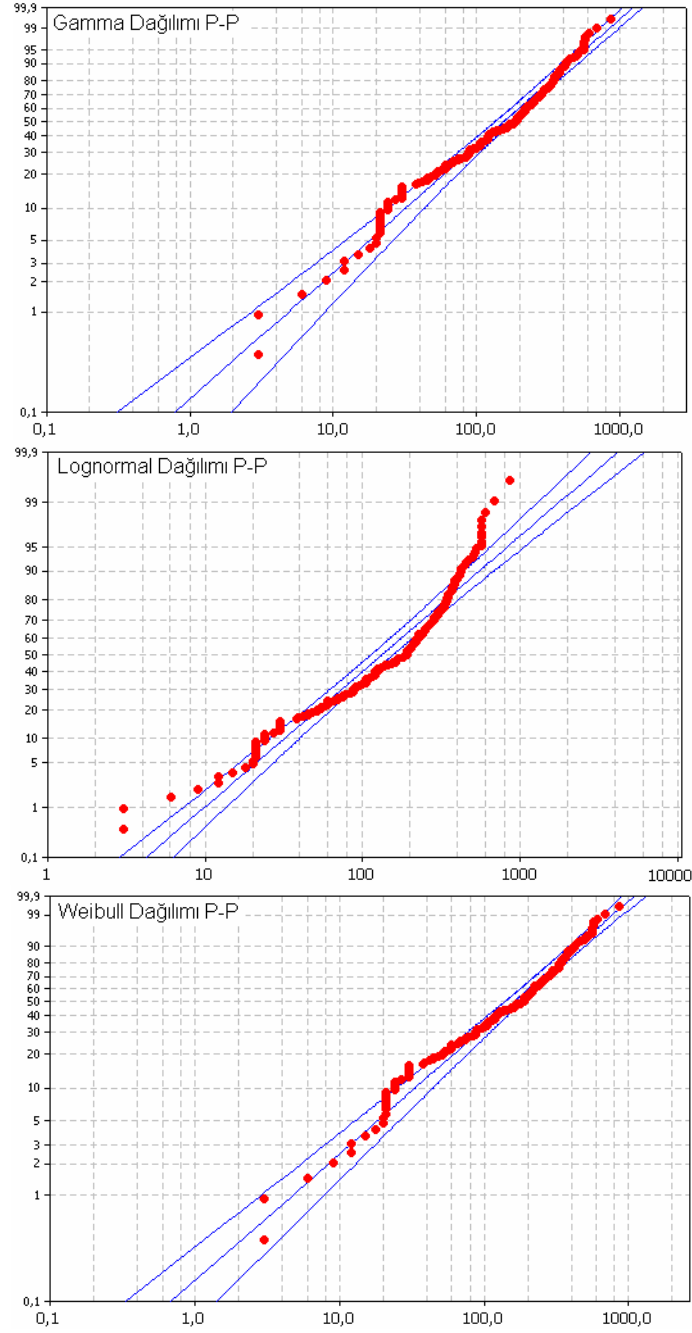
Bileşen Sayısı	Karma Gamma		Karma Lognormal		Karma Weibull	
	$\ln L$	AIC	$\ln L$	AIC	$\ln L$	AIC
1	-11548	23100	-11724	23452	-11536	23076
2	-11480	22970	-11500	23010	-11457	22924
3	-11434	22884	-11497	23010	-11440	22896
4	-11467	22956	-11462	22946	-11467	22956

Karma Gamma, karma Lognormal ve karma Weibull dağılım modellerinde bileşen sayısına göre Akaike bilgi kriterindeki değişimi gösteren çizgi grafiği Şekil 5.9’da verilmiştir.



Şekil 5.9. Karma Gamma, karma Lognormal ve karma Weibull dağılım modellerinde bileşen sayısına göre AIC değerlerinin değişimi

Şekil 5.9 ve Tablo 5.7’deki AIC değerlerine göre karma Gamma ve karma Weibull dağılım modellerinde uygun bileşen sayısı 3 iken, Karma Lognormal dağılım modelinde uygun bileşen sayısı 2’dir. Akciğer kanser veri seti için Gamma, Lognormal ve Weibull dağılımlarına ait P-P grafikleri Şekil 5.10’ da verilmiştir.



Şekil 5.10. Akciğer kanseri tanısı konan hastaların yaşam süreleri için P-P grafikleri

Şekil 5.10’da yer alan P-P grafikleri incelendiğinde akciğer kanseri tanısı alan hastaların yaşam sürelerinin modellenmesinde klasik istatistik dağılımlarının uygun modeller olmadığı görülmektedir.

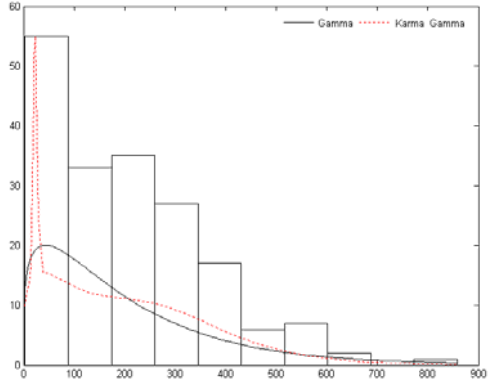
Akciğer kanseri tanısı alan hastaların yaşam sürelerinin modellenmesinde kullanılan modellere ait parametre tahminleri, $\ln L$ log olabilirlik değerleri, D_{hesap}

Kolmogorov-Smirnov test istatistiği, HKO hata kareler ortalaması ve $E(T)$ ortalama yaşam süreleri Tablo 5.8’de verilmiştir.

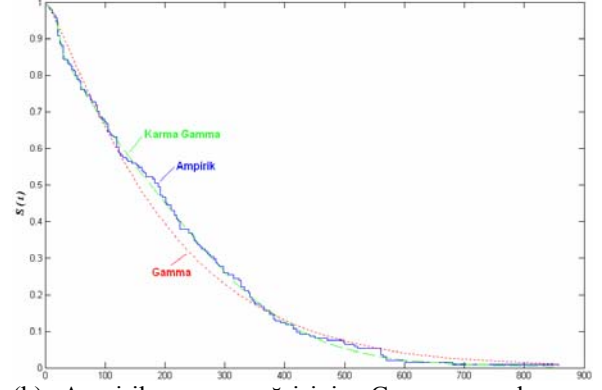
Tablo 5.8. Akciğer kanseri tanısı alan hastaların yaşam sürelerinin modellenmesinde kullanılan dağılımların parametre tahminleri, $\ln L$, D_{hesap} , HKO ve $E(T)$ Değerleri

Model	Parametre Tahminleri			$\ln L$	D_{hesap} (p değeri)	HKO	$E(T)$
Gamma	$\hat{\alpha} = 1.27$	$\hat{\beta} = 162.17$		-11548	0.078 (0.2155)	0.0007	205.80
Karma Gamma	$\hat{\alpha}_1 = 38.74$	$\hat{\alpha}_2 = 6.99$	$\hat{\alpha}_3 = 1.28$	-11434	0.027 (0.9993)	0.00005	205.80
	$\hat{\beta}_1 = 0.61$	$\hat{\beta}_2 = 47.39$	$\hat{\beta}_3 = 129.91$				
	$\hat{\pi}_1 = 0.07$	$\hat{\pi}_2 = 0.30$	$\hat{\pi}_3 = 0.63$				
Lognormal	$\hat{\mu} = 4.88$	$\hat{\sigma} = 01.11$		-11724	0.126 (0.0059)	0.0022	245.36
Karma Lognormal	$\hat{\mu}_1 = 5.69$	$\hat{\mu}_2 = 4.31$		-11500	0.037 (0.9636)	0.00011	211.85
	$\hat{\sigma}_1 = 0.38$	$\hat{\sigma}_2 = 1.10$					
	$\hat{\pi}_1 = 0.42$	$\hat{\pi}_2 = 0.58$					
Weibull	$\hat{\alpha} = 1.20$	$\hat{\beta} = 218.20$		-11536	0.068 (0.3659)	0.0006	205.29
Karma Weibull	$\hat{\alpha}_1 = 1.16$	$\hat{\alpha}_2 = 2.35$	$\hat{\alpha}_3 = 5.71$	-11440	0.032 (0.9920)	0.00007	205.80
	$\hat{\beta}_1 = 187.91$	$\hat{\beta}_2 = 337.25$	$\hat{\beta}_3 = 25.47$				
	$\hat{\pi}_1 = 0.59$	$\hat{\pi}_2 = 0.33$	$\hat{\pi}_3 = 0.08$				

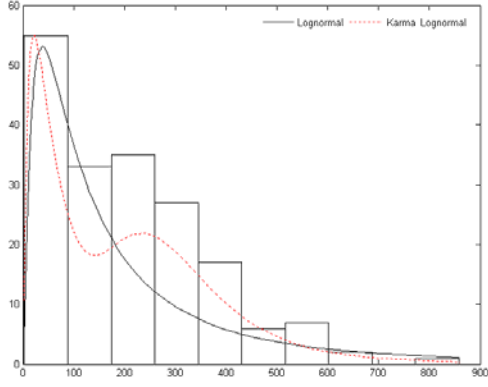
Tablo 5.8 incelendiğinde, tüm karşılaştırma kriterlerine göre akciğer kanseri tanısı alan kanser hastalarının yaşam sürelerinin modellenmesinde karma dağılım modellerinin klasik istatistik dağılımlarından daha uygun olduğu görülmektedir. Akciğer kanseri tanısı alan kanser hastalarının yaşam sürelerinin modellenmesinde, klasik istatistik dağılımları ile karma dağılım modellerinin grafiksel karşılaştırılması Şekil 5.11’de verilmiştir.



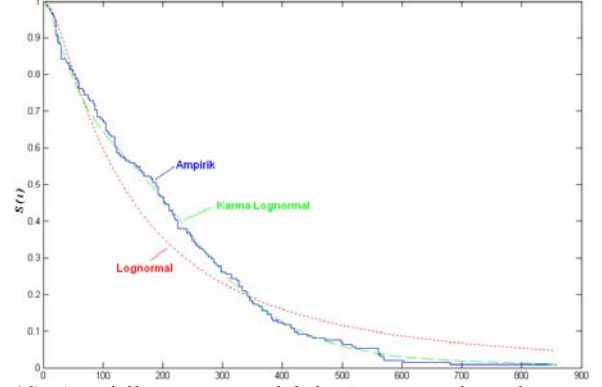
(a) Gamma ve karma Gamma dağılım modellerinin yoğunluk eğrileri



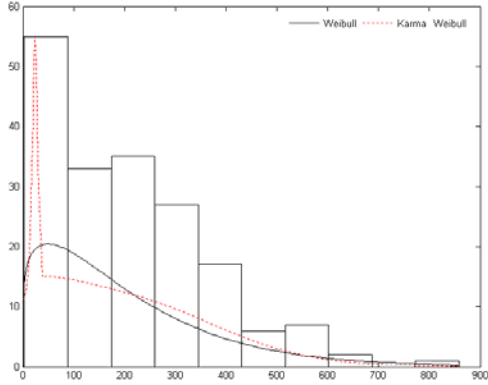
(b) Ampirik yaşam eğrisinin Gamma ve karma Gamma dağılım modellerinin yaşam eğrileri ile karşılaştırılması



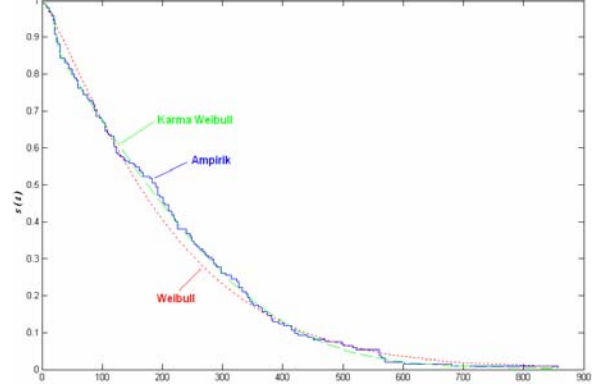
(c) Lognormal ve karma Lognormal dağılım modellerinin yoğunluk eğrileri



(d) Ampirik yaşam eğrisinin Lognormal ve karma Lognormal dağılım modellerinin yaşam eğrileri ile karşılaştırılması



(e) Weibull ve karma Weibull dağılım modellerinin yoğunluk eğrileri



(f) Ampirik yaşam eğrisinin Weibull ve karma Weibull dağılım modellerinin yaşam eğrileri ile karşılaştırılması

Şekil 5.11. Akciğer kanseri tanısı konan hastaların yaşam sürelerinin modellenmesinde klasik istatistik dağılımları ile karma dağılım modellerinin grafiksel karşılaştırılması

Şekil 5.11 incelendiğinde, akciğer kanseri tanısı alan kanser hastalarının yaşam sürelerinin modellenmesinde karma dağılım modellerinin, klasik istatistik dağılımlarına göre daha başarılı olduğu görülmektedir.

5.4. Sonuçlar

Bu bölümde heterojen yaşam süresi verilerin modellenmesinde karma dağılım modelleri incelenmiştir. Heterojen yaşam sürelerinin modellenmesinde karma Gamma, karma Lognormal ve karma Weibull dağılım modellerinde parametrelerin en çok olabilirlik tahminlerinin elde edilmesi için gerekli EM algoritmaları verilmiştir. Tanımlanan EM algoritmaların parametre tahminindeki etkinliği simülasyon çalışması ile ortaya konmuştur. Gerçek iki yaşam verisinde ele alınan karma dağılım modelleri klasik istatistik dağılımlarından Gamma, Lognormal ve Weibull dağılımları ile karşılaştırılmıştır. Karşılaştırmada log olabilirlik değeri, Kolmogorov-Smirnov test istatistiği ve hata kareler ortalaması kullanılmıştır. Işınlanmış farelerin ölüm zamanlarının ve akciğer kanseri tanısı konan hastaların yaşam sürelerinin modellenmesinde karma dağılım modelleri tüm karşılaştırma kriterlerine göre klasik istatistik dağılımlarından daha başarılı bulunmuştur.

6. YAŞAM ANALİZİNDE İKİ FARKLI DAĞILIMIN KARMASI

Heterojen yaşam sürelerinin modellenmesinde iki farklı dağılımın karmasından oluşan karma modeller de kullanılabilir (Erişoğlu ve ark 2011). Bu bölümde Üstel-Gamma karma dağılım modeli, Üstel-Weibull karma dağılım modeli ve Gamma-Weibull karma dağılım modeli oluşturulacaktır. İncelenen karma dağılım modellerinin parametrelerinin en çok olabilirlik tahminlerinde EM algoritması kullanılacaktır. EM algoritmasının parametre tahminindeki etkinliği simülasyon çalışması ile incelenecektir. Ayrıca iki gerçek heterojen yaşam süresi verisinin modellenmesinde iki farklı dağılımın karmasından elde edilen modeller standart olasılık dağılımları ile karşılaştırılacaktır.

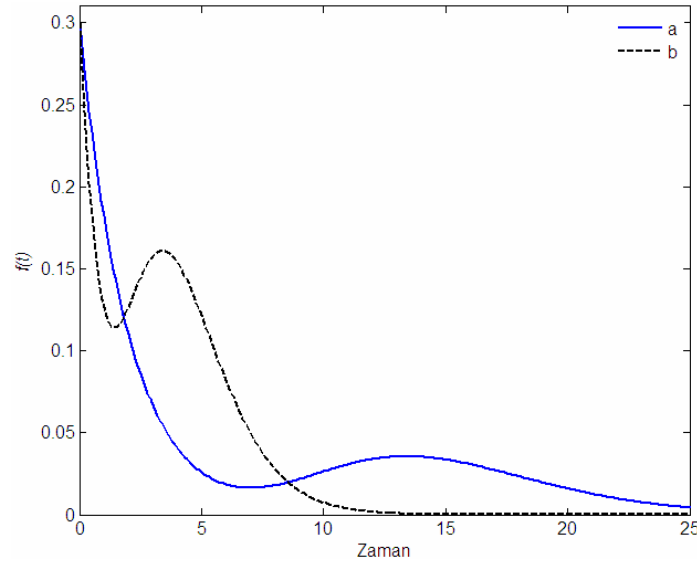
6.1. İki Farklı Dağılımın Karma Modelleri

6.1.1. Üstel-Gamma Karma Dağılımı

Üstel ve Gamma dağılımının karmasından oluşan Üstel-Gamma karma dağılım modelinin olasılık yoğunluk fonksiyonu

$$f(t) = \pi\lambda e^{-\lambda t} + (1-\pi)t^{\alpha-1} \frac{e^{-t/\beta}}{\beta^{\alpha}\Gamma(\alpha)} \quad (6.1)$$

eşitliği ile ifade edilir. Eşitlik (6.1)'de yer alan π , $\pi \in (0,1)$ olmak üzere Üstel dağılıma ait karma oranıdır. Üstel-Gamma karma dağılım modelinde farklı parametre değerleri için çizilen yoğunluk eğrileri Şekil 6.1'de verilmiştir.



Şekil 6.1. Üstel-Gamma karma dağılım modelinde farklı parametre değerleri için yoğunluk eğrileri a) $\pi = 0.6, \lambda = 0.5, \alpha = 10$ ve $\beta = 1.5$ (düz çizgi) b) $\pi = 0.3, \lambda = 1, \alpha = 5$ ve $\beta = 0.9$ (kesik çizgi)

Üstel-Gamma karma dağılım modelinde yaşam fonksiyonu, hazard fonksiyonu ve beklenen yaşam süresi sırasıyla

$$S(t | \psi) = \pi e^{-\lambda t} + (1 - \pi) \left(1 - \frac{\Gamma_x(\alpha)}{\Gamma(\alpha)} \right) \quad (6.2)$$

$$h(t) = \frac{\beta^\alpha \Gamma(\alpha) \pi \lambda e^{-\lambda t} + (1 - \pi) t^{\alpha-1} e^{-t\beta}}{\beta^\alpha \left\{ \Gamma(\alpha) \pi e^{-\lambda t} + (1 - \pi) (\Gamma(\alpha) - \Gamma_x(\alpha)) \right\}} \quad (6.3)$$

$$E(T) = \frac{\pi}{\lambda} + (1 - \pi) \alpha \beta \quad (6.4)$$

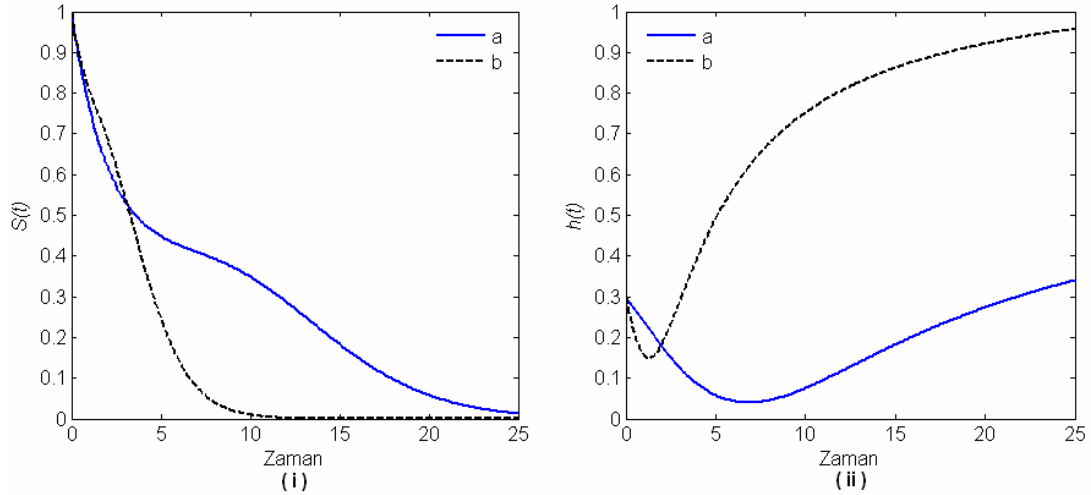
şeklinde tanımlanır. Eşitlik (6.2) ve (6.3)'de yer alan $\Gamma(\alpha)$, gamma fonksiyonudur

ve $\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx$ eşitliği ile tanımlanır. Eşitliklerde yer alan $\Gamma_x(\alpha)$ ise

tamamlanmamış gamma fonksiyonu olarak isimlendirilir ve $\Gamma_x(\alpha) = \int_0^x t^{\alpha-1} \exp(-t) dt$

eşitliği ile gösterilir. Üstel-Gamma karma dağılım modelinde farklı parametre

değerleri için çizilen yaşam fonksiyonu ve hazard fonksiyonu eğrileri Şekil 6.2’de verilmiştir.



Şekil 6.2. Üstel-Gamma karma dağılım modelinde farklı parametre değerleri için (i) yaşam fonksiyonu ve (ii) hazard fonksiyonu eğrileri a) $\pi = 0.6$, $\lambda = 0.5$, $\alpha = 10$ ve $\beta = 1.5$ (düz çizgi) b) $\pi = 0.3$, $\lambda = 1$, $\alpha = 5$ ve $\beta = 0.9$ (kesik çizgi)

Üstel-Gamma karma dağılım modelindeki parametrelerin en çok olabilirlik tahminlerini elde etmek için EM algoritmasının E adımında gözlemlerin Üstel bileşenine aitlik olasılıkları z_{1i} ,

$$\hat{z}_{1i} = \frac{\pi \lambda e^{-\lambda t}}{f(t)} \quad (6.5)$$

olarak tahmin edilir. Eşitlikte yer alan $f(t)$, Üstel-Gamma karma dağılım modelinin olasılık yoğunluk fonksiyonudur. Gözlemlerin Gamma bileşenine aitlik olasılıkları z_{2i} ,

$$\hat{z}_{2i} = 1 - \hat{z}_{1i} \quad (6.6)$$

eşitliği ile tahmin edilir. Böylece Üstel-Gamma karma dağılım modelinde tamamlanmış veri için log olabilirlik fonksiyonunun beklenen değeri

$$\begin{aligned}
 E(\ln L) &= \sum_{i=1}^n \left(\hat{z}_{1i} \ln \{ \pi \lambda e^{-\lambda t_i} \} + \hat{z}_{2i} \ln \{ (1-\pi) t_i^{\alpha-1} \frac{e^{-t_i/\beta}}{\beta^\alpha \Gamma(\alpha)} \} \right) \\
 E(\ln L) &= (\ln(\pi) + \ln(\lambda)) \sum_{i=1}^n \hat{z}_{1i} - \sum_{i=1}^n \hat{z}_{1i} t_i + (\ln(1-\pi) - \alpha \ln(\beta) - \ln(\Gamma(\alpha))) \sum_{i=1}^n \hat{z}_{2i} \\
 &\quad + \sum_{i=1}^n \hat{z}_{2i} (\alpha - 1) \ln(t_i) - \sum_{i=1}^n \frac{\hat{z}_{2i} t_i}{\beta}
 \end{aligned} \tag{6.7}$$

şeklinde elde edilir. EM algoritmasının M adımında, E adımında elde edilen $E(\ln L)$, $\pi \in (0,1)$ koşulu altında maksimize edilir. Üstel-Gamma karma dağılım modelinde Üstel bileşenin karma oranı

$$\hat{\pi} = \frac{\sum_{i=1}^n \hat{z}_{1i}}{n} \tag{6.8}$$

şeklinde tahmin edilir. Üstel-Gamma karma modelinde Üstel bileşenin λ parametresi

$$\hat{\lambda} = \sum_{i=1}^n \hat{z}_{1i} \left(\sum_{i=1}^n \hat{z}_{1i} t_i \right)^{-1} \tag{6.9}$$

eşitliği ile tahmin edilir.

Üstel-Gamma karma dağılımı modelinde Gamma bileşenine ait parametrelerinin tahmininde EM algoritması içerisinde iteratif yöntemlere ihtiyaç duyulmaktadır. $\psi(\cdot)$ digamma fonksiyonu, $\psi'(\cdot)$ trigamma fonksiyonu olmak üzere Newton-Raphson yönteminin $(r+1)$. iterasyonunda α parametresi

$$\hat{\alpha}_{r+1} = \hat{\alpha}_r - \frac{\ln(\hat{\alpha}_r) - \psi'(\hat{\alpha}_r) - \ln\left(\frac{\sum_{i=1}^n \hat{z}_{2i} t_i}{\sum_{i=1}^n \hat{z}_{2i}}\right) + \frac{\sum_{i=1}^n \hat{z}_{2i} \ln(t_i)}{\sum_{i=1}^n \hat{z}_{2i}}}{\frac{1}{\hat{\alpha}_r} - \psi'(\hat{\alpha}_r)} \quad (6.10)$$

şeklinde elde edilir. EM algoritmasında α parametresi tahmin edildikten sonra β parametresinin en çok olabilirlik tahmin edicisi

$$\hat{\beta} = \left(\hat{\alpha} \sum_{i=1}^n \hat{z}_{2i} \right)^{-1} \sum_{i=1}^n \hat{z}_{2i} t_i \quad (6.11)$$

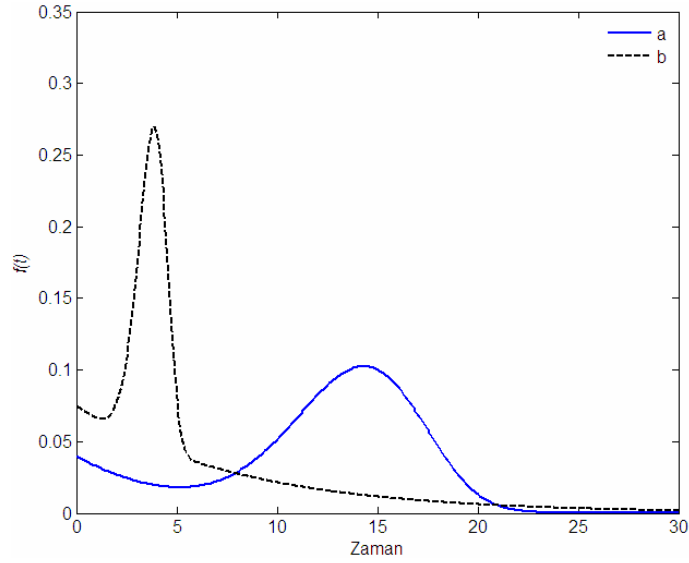
eşitliği ile elde edilir.

6.1.2. Üstel-Weibull Karma Dağılımı

Üstel ve Weibull dağılımının karmasından oluşan Üstel-Weibull karma dağılım modelinin olasılık yoğunluk fonksiyonu

$$f(t) = \pi \lambda e^{-\lambda t} + (1 - \pi) \frac{a}{b^a} t^{a-1} e^{-\left(\frac{t}{b}\right)^a} \quad (6.12)$$

eşitliği ile ifade edilir. Eşitlik (6.12)'de yer alan π , $\pi \in (0,1)$ olmak üzere Üstel dağılıma ait karma oranıdır. Üstel-Weibull karma dağılım modelinde farklı parametre değerleri için çizilen yoğunluk eğrileri Şekil 6.3'de verilmiştir.



Şekil 6.3. Üstel-Weibull karma dağılım modelinde farklı parametre değerleri için yoğunluk eğrileri a) $\pi = 0.2, \lambda = 0.2, a = 5$ ve $b = 15$ (düz çizgi) b) $\pi = 0.6, \lambda = 0.125, a = 6$ ve $b = 4$ (kesik çizgi)

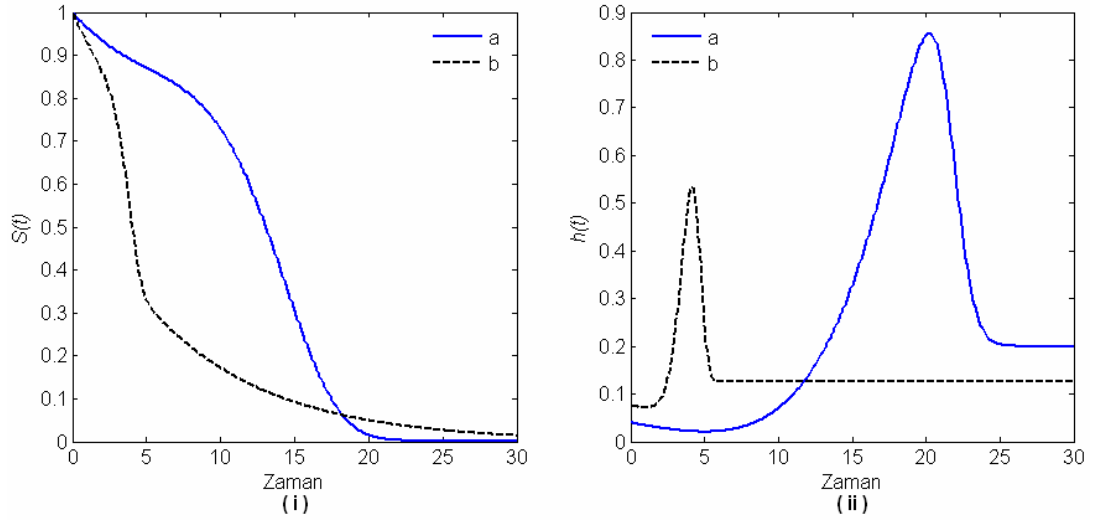
Üstel-Weibull karma dağılım modelinde yaşam fonksiyonu, hazard fonksiyonu ve beklenen yaşam süresi sırasıyla

$$S(t | \psi) = \pi e^{-\lambda t} + (1 - \pi) e^{-\left(\frac{t}{b}\right)^a} \quad (6.13)$$

$$h(t) = \frac{b^a \pi \lambda e^{-\lambda t} + (1 - \pi) a t^{a-1} e^{-\left(\frac{t}{b}\right)^a}}{b^a \pi e^{-\lambda t} + b^a e^{-\left(\frac{t}{b}\right)^a} - b^a \pi e^{-\left(\frac{t}{b}\right)^a}} \quad (6.14)$$

$$E(T) = \frac{\pi}{\lambda} + (1 - \pi) b \Gamma\left(1 + \frac{1}{a}\right) \quad (6.15)$$

şeklinde tanımlanır. Üstel-Weibull karma dağılım modelinde farklı parametre değerleri için çizilen yaşam fonksiyonu ve hazard fonksiyonu eğrileri Şekil 6.4'de verilmiştir.



Şekil 6.4. Üstel-Weibull karma dağılım modelinde farklı parametre değerleri için (i) yaşam fonksiyonu ve (ii) hazard fonksiyonu eğrileri a) $\pi = 0.2$, $\lambda = 0.2$, $a = 5$ ve $b = 15$ (düz çizgi) b) $\pi = 0.6$, $\lambda = 0.125$, $a = 6$ ve $b = 4$ (kesik çizgi)

Üstel-Weibull karma dağılım modelindeki parametrelerin en çok olabilirlik tahminlerini elde etmek için EM algoritmasının E adımında gözlemlerin Üstel bileşenine aitlik olasılıkları z_{1i} ,

$$\hat{z}_{1i} = \frac{\pi \lambda e^{-\lambda t}}{f(t)} \quad (6.16)$$

olarak tahmin edilir. Eşitlikte yer alan $f(t)$, Üstel-Weibull karma dağılım modelinin olasılık yoğunluk fonksiyonudur. Gözlemlerin Weibull bileşenine aitlik olasılıkları z_{2i} ,

$$\hat{z}_{2i} = 1 - \hat{z}_{1i} \quad (6.17)$$

eşitliği ile tahmin edilir. Böylece Üstel-Weibull karma dağılım modelinde tamamlanmış veri için log olabilirlik fonksiyonunun beklenen değeri

$$\begin{aligned}
 E(\ln L) &= \sum_{i=1}^n \left(\hat{z}_{1i} \ln \{ \pi \lambda e^{-\lambda t_i} \} + \hat{z}_{2i} \ln \left\{ (1-\pi) \frac{a}{b^a} t_i^{a-1} e^{-\left(\frac{t_i}{b}\right)^a} \right\} \right) \\
 E(\ln L) &= (\ln(\pi) + \ln(\lambda)) \sum_{i=1}^n \hat{z}_{1i} - \sum_{i=1}^n \hat{z}_{1i} t_i + (\ln(1-\pi) + \ln(a) - a \ln(b)) \sum_{i=1}^n \hat{z}_{2i} \\
 &\quad + (a-1) \sum_{i=1}^n \hat{z}_{2i} (a-1) \ln(t_i) - \frac{1}{b^a} \sum_{i=1}^n \hat{z}_{2i} t_i^a
 \end{aligned} \tag{6.18}$$

şeklinde elde edilir. EM algoritmasının M adımında, E adımında elde edilen $E(\ln L)$, $\pi \in (0,1)$ koşulu altında maksimize edilir. Üstel-Weibull karma dağılım modelinde Üstel bileşenin karma oranı

$$\hat{\pi} = \frac{\sum_{i=1}^n \hat{z}_{1i}}{n} \tag{6.19}$$

şeklinde tahmin edilir. Üstel-Weibull karma modelinde Üstel bileşenin λ parametresi

$$\hat{\lambda} = \sum_{i=1}^n \hat{z}_{1i} \left(\sum_{i=1}^n \hat{z}_{1i} t_i \right)^{-1} \tag{6.20}$$

eşitliği ile tahmin edilir.

Üstel-Weibull karma dağılımı modelinde Weibull bileşenine ait parametrelerinin tahmininde EM algoritması içerisinde iteratif yöntemlere ihtiyaç duyulmaktadır. Newton-Raphson yönteminde $(r+1)$. iterasyonunda a parametresinin tahmini

$$\hat{a}_{r+1} = \hat{a}_r + \frac{A_r + (1/\hat{a}_r) - (C_r/B_r)}{(1/\hat{a}_r^2) + (B_r D_r - C_r^2)/B_r^2} \tag{6.21}$$

şeklinde elde edilir. Burada $A_r = \left(\sum_{i=1}^n \hat{z}_{2i} \right)^{-1} \sum_{i=1}^n \hat{z}_{2i} \ln t_i$, $B_r = \sum_{i=1}^n \hat{z}_{2i} t_i^{\hat{a}_r}$, $C_r = \sum_{i=1}^n \hat{z}_{2i} t_i^{\hat{a}_r} \ln t_i$

ve $D_r = \sum_{i=1}^n \hat{z}_{2i} t_i^{\hat{a}_r} (\ln t_i)^2$ şeklinde tanımlanır. EM algoritmasında a parametresi tahmin edildikten sonra b parametresinin en çok olabilirlik tahmin edicisi

$$\hat{b} = \left\{ \left(\sum_{i=1}^n \hat{z}_{2i} \right)^{-1} \left(\sum_{i=1}^n \hat{z}_{2i} t_i^a \right) \right\}^{1/a} \quad (6.22)$$

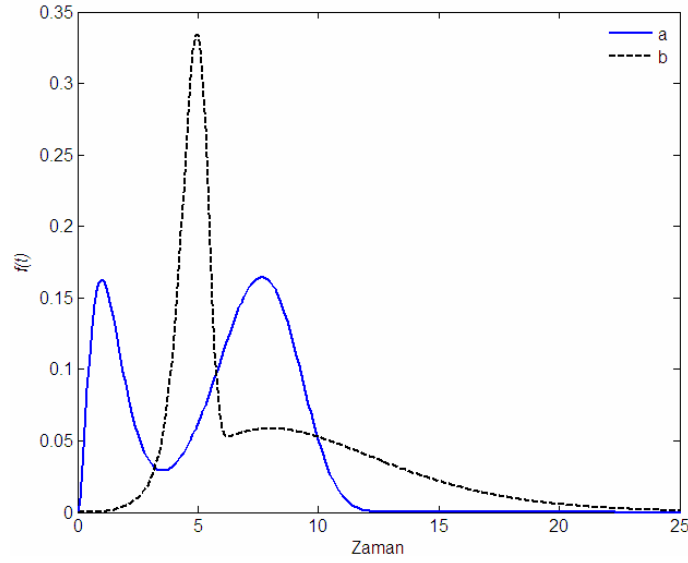
eşitliği ile elde edilir.

6.1.3. Gamma-Weibull Karma Dağılımı

Gamma ve Weibull dağılımının karmasından oluşan Gamma-Weibull karma dağılım modelinin olasılık yoğunluk fonksiyonu

$$f(t) = \pi t^{\alpha-1} \frac{e^{-t/\beta}}{\beta^\alpha \Gamma(\alpha)} + (1-\pi) \frac{a}{b^a} t^{a-1} e^{-\left(\frac{t}{b}\right)^a} \quad (6.23)$$

eşitliği ile ifade edilir. Eşitlik (6.23)'de yer alan π , $\pi \in (0,1)$ olmak üzere Gamma dağılıma sahip bileşene ait karma oranıdır. Gamma-Weibull karma dağılım modelinde farklı parametre değerleri için çizilen yoğunluk eğrileri Şekil 6.5'de verilmiştir.



Şekil 6.5. Gamma-Weibull karma dağılım modelinde farklı parametre değerleri için yoğunluk eğrileri a) $\pi = 0.3, \alpha = 3, \beta = 0.5, a = 5$ ve $b = 8$ (düz çizgi) b) $\pi = 0.6, \alpha = 5, \beta = 2, a = 10$ ve $b = 5$ (kesik çizgi)

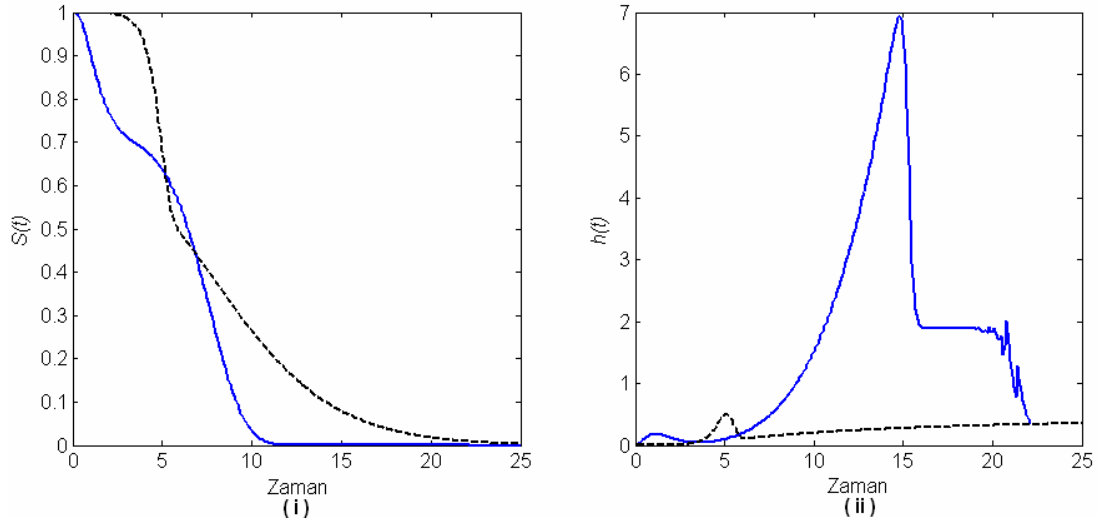
Gamma-Weibull karma dağılım modelinde yaşam fonksiyonu, hazard fonksiyonu ve beklenen yaşam süresi sırasıyla

$$S(t | \psi) = \pi \left(1 - \frac{\Gamma_x(\alpha)}{\Gamma(\alpha)} \right) + (1 - \pi) e^{-\left(\frac{t}{b}\right)^\alpha} \quad (6.24)$$

$$h(t) = \frac{\Gamma(a) \left\{ b^a t^{a-1} e^{-\frac{t}{b}} + \beta^\alpha \Gamma(\alpha) (1 - \pi) a t^{a-1} e^{-\left(\frac{t}{b}\right)^\alpha} \right\}}{b^a \beta^\alpha \Gamma(\alpha) \left\{ \pi (\Gamma(a) - \Gamma_x(a)) + \Gamma(a) (1 - \pi) e^{-\left(\frac{t}{b}\right)^\alpha} \right\}} \quad (6.25)$$

$$E(T) = \pi \alpha \beta + (1 - \pi) b \Gamma\left(1 + \frac{1}{a}\right) \quad (6.26)$$

şeklinde tanımlanır. Gamma-Weibull karma dağılım modelinde farklı parametre değerleri için çizilen yaşam fonksiyonu ve hazard fonksiyonu eğrileri Şekil 6.6'de verilmiştir.



Şekil 6.6. Gamma-Weibull karma dağılım modelinde farklı parametre değerleri için (i) yaşam fonksiyonu ve (ii) hazard fonksiyonu eğrileri a) $\pi = 0.3$, $\alpha = 3$, $\beta = 0.5$, $a = 5$ ve $b = 8$ (düz çizgi) b) $\pi = 0.6$, $\alpha = 5$, $\beta = 2$, $a = 10$ ve $b = 5$ (kesik çizgi)

Gamma-Weibull karma dağılım modelindeki parametrelerin en çok olabilirlik tahminlerini elde etmek için EM algoritmasının E adımında gözlemlerin Gamma bileşenine aitlik olasılıkları z_{1i} ,

$$\hat{z}_{1i} = \pi t^{\alpha-1} \frac{e^{-t/\beta}}{\beta^\alpha \Gamma(\alpha) f(t)} \quad (6.27)$$

olarak tahmin edilir. Eşitlikte yer alan $f(t)$, Gamma-Weibull karma dağılım modelinin olasılık yoğunluk fonksiyonudur. Gözlemlerin Weibull bileşenine aitlik olasılıkları z_{2i} ,

$$\hat{z}_{2i} = 1 - \hat{z}_{1i} \quad (6.28)$$

eşitliği ile tahmin edilir. Böylece Gamma-Weibull karma dağılım modelinde tamamlanmış veri için log olabilirlik fonksiyonunun beklenen değeri

$$\begin{aligned}
 E(\ln L) &= \sum_{i=1}^n \left(\hat{z}_{1i} \ln \left\{ \pi t^{\alpha-1} \frac{e^{-t/\beta}}{\beta^\alpha \Gamma(\alpha)} \right\} + \hat{z}_{2i} \ln \left\{ (1-\pi) \frac{a}{b^a} t_i^{a-1} e^{-\left(\frac{t_i}{b}\right)^a} \right\} \right) \\
 E(\ln L) &= (\ln(\pi) - \alpha \ln(\beta) - \ln(\Gamma(\alpha))) \sum_{i=1}^n \hat{z}_{1i} + \sum_{i=1}^n \hat{z}_{1i} (\alpha - 1) \ln(t_i) - \sum_{i=1}^n \frac{\hat{z}_{1i} t_i}{\beta} \\
 &+ (\ln(1-\pi) + \ln(a) - a \ln(b)) \sum_{i=1}^n \hat{z}_{2i} + (a-1) \sum_{i=1}^n \hat{z}_{2i} \ln(t_i) - \frac{1}{b^a} \sum_{i=1}^n \hat{z}_{2i} t_i^a \quad (6.29)
 \end{aligned}$$

şeklinde elde edilir. EM algoritmasının M adımında, E adımında elde edilen $E(\ln L)$, $\pi \in (0,1)$ koşulu altında maksimize edilir. Gamma-Weibull karma dağılım modelinde Gamma bileşenin karma oranı

$$\hat{\pi} = \frac{\sum_{i=1}^n \hat{z}_{1i}}{n} \quad (6.30)$$

şeklinde tahmin edilir. Gamma-Weibull karma modelinde Gamma bileşenine ait parametrelerinin tahmininde EM algoritması içerisinde iteratif yöntemlere ihtiyaç duyulmaktadır. $\psi(\cdot)$ digamma fonksiyonu, $\psi'(\cdot)$ trigamma fonksiyonu olmak üzere Newton-Raphson yönteminin $(r+1)$. iterasyonunda α parametresi

$$\hat{\alpha}_{r+1} = \hat{\alpha}_r - \frac{\ln(\hat{\alpha}_r) - \psi(\hat{\alpha}_r) - \ln \left(\frac{\sum_{i=1}^n \hat{z}_{1i} t_i}{\sum_{i=1}^n \hat{z}_{1i}} \right) + \frac{\sum_{i=1}^n \hat{z}_{1i} \ln(t_i)}{\sum_{i=1}^n \hat{z}_{1i}}}{\frac{1}{\hat{\alpha}_r} - \psi'(\hat{\alpha}_r)} \quad (6.31)$$

şeklinde elde edilir. EM algoritmasında α parametresi tahmin edildikten sonra β parametresinin en çok olabilirlik tahmin edicisi

$$\hat{\beta} = \left(\hat{\alpha} \sum_{i=1}^n \hat{z}_{1i} \right)^{-1} \sum_{i=1}^n \hat{z}_{1i} t_i \quad (6.32)$$

eşitliği ile elde edilir.

Gamma-Weibull karma dağılımı modelinde Weibull bileşenine ait parametrelerinin tahmininde de EM algoritması içerisinde iteratif yöntemlere ihtiyaç duyulmaktadır. Newton-Raphson yönteminde $(r+1)$. iterasyonunda a parametresinin tahmini

$$\hat{a}_{r+1} = \hat{a}_r + \frac{A_r + (1/\hat{a}_r) - (C_r/B_r)}{(1/\hat{a}_r^2) + (B_r D_r - C_r^2)/B_r^2} \quad (6.33)$$

şeklinde elde edilir. Burada $A_r = \left(\sum_{i=1}^n \hat{z}_{2i} \right)^{-1} \sum_{i=1}^n \hat{z}_{2i} \ln t_i$, $B_r = \sum_{i=1}^n \hat{z}_{2i} t_i^{\hat{a}_r}$,

$C_r = \sum_{i=1}^n \hat{z}_{2i} t_i^{\hat{a}_r} \ln t_i$ ve $D_r = \sum_{i=1}^n \hat{z}_{2i} t_i^{\hat{a}_r} (\ln t_i)^2$ şeklinde tanımlanır. EM algoritmasında a

parametresi tahmin edildikten sonra b parametresinin en çok olabilirlik tahmin edicisi

$$\hat{b} = \left\{ \left(\sum_{i=1}^n \hat{z}_{2i} \right)^{-1} \left(\sum_{i=1}^n \hat{z}_{2i} t_i^a \right) \right\}^{1/a} \quad (6.34)$$

eşitliği ile elde edilir.

6.2. İki Farklı Dağılımın Karma Modelinde EM Algoritmasının Parametre Tahminindeki Etkinliği

İki farklı dağılımın karma modellerindeki parametrelerin en çok olabilirlik tahminlerinin elde edilmesi için tanımlanan EM algoritmalarının parametre tahminindeki etkinliği göstermek için örneklem genişliği 100 birim, tekrar sayısı 1000 olmak üzere simülasyonlar gerçekleştirilmiştir. Simülasyon çalışmasında yoğunluk eğrileri sırasıyla Şekil 6.1, Şekil 6.3 ve Şekil 6.5’de verilen Üstel-Gamma, Üstel-Weibull ve Gamma-Weibull karma dağılımına sahip kitleler için parametre

tahminleri gerçekleştirilmiştir. Tahminler arasındaki fark 10^{-4} 'den daha az olduğu zaman yakınsamanın olduğu varsayılmıştır.

Üstel-Gamma dağılım modeline sahip kitleler için gerçekleştirilen simülasyon sonrası EM algoritması ile elde edilen parametrelerin en çok olabilirlik tahminlerinin ortalamaları ve standart hataları Tablo 6.1'de verilmiştir.

Tablo 6.1.Üstel-Gamma karma dağılım modelinde EM algoritması ile elde edilen parametrelerin en çok olabilirlik tahminlerinin ortalamaları ve standart hataları

Üstel-Gamma		$\hat{\pi}$	$\hat{\lambda}$	$\hat{\alpha}$	$\hat{\beta}$
$\pi = 0.6, \lambda = 0.5$	Ortalama	0.5933	0.5178	10.1686	1.5039
$\alpha = 10, \beta = 1.5$	S.Hata	0.0017	0.0029	0.0492	0.0071
$\pi = 0.3, \lambda = 1$	Ortalama	0.3004	1.0242	4.9991	0.9050
$\alpha = 5, \beta = 0.9$	S.Hata	0.0027	0.0155	0.0221	0.0036

Üstel-Weibull dağılım modeline sahip kitleler için gerçekleştirilen simülasyon sonrası EM algoritması ile elde edilen parametrelerin en çok olabilirlik tahminlerinin ortalamaları ve standart hataları Tablo 6.2'de verilmiştir.

Tablo 6.2.Üstel-Weibull karma dağılım modelinde EM algoritması ile elde edilen parametrelerin en çok olabilirlik tahminlerinin ortalamaları ve standart hataları

Üstel-Weibull		$\hat{\pi}$	$\hat{\lambda}$	$\hat{a}$	$\hat{b}$
$\pi = 0.2, \lambda = 0.2$	Ortalama	0.1999	0.2054	5.0447	14.9697
$a = 5, b = 15$	S.Hata	0.0016	0.0014	0.0154	0.0121
$\pi = 0.6, \lambda = 0.125$	Ortalama	0.5944	0.1277	6.0599	4.1560
$a = 6, b = 4$	S.Hata	0.0021	0.0005	0.0173	0.0306

Gamma-Weibull dağılım modeline sahip kitleler için gerçekleştirilen simülasyon sonrası EM algoritması ile elde edilen parametrelerin en çok olabilirlik tahminlerinin ortalamaları ve standart hataları Tablo 6.3'de verilmiştir.

Tablo 6.3. Gamma-Weibull karma dağılım modelinde EM algoritması ile elde edilen parametrelerin en çok olabilirlik tahminlerinin ortalamaları ve standart hataları

Gamma-Weibull		$\hat{\pi}$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{a}$	$\hat{b}$
$\pi = 0.3, \alpha = 3, \beta = 0.5$	Ortalama	0.3074	3.1260	0.5027	5.1347	8.0102
$a = 5, b = 8$	S.Hata	0.0016	0.0203	0.0044	0.0167	0.0070
$\pi = 0.6, \alpha = 5, \beta = 2$	Ortalama	0.5958	5.1335	2.0139	10.1805	5.0044
$a = 10, b = 5$	S.Hata	0.0019	0.0261	0.0107	0.0451	0.0035

Yakınsama tüm durumlarda başarılıdır. Bu EM algoritmasının sayısal olarak tutarlı olduğunu vurgular. Ortalama tahmin ve bunların standart hata değerleri de EM tahminlerinin tutarlı olduğunu gösterir. Simülasyon sonuçlarına göre EM yaklaşımının farklı karma oranlarında da iyi çalıştığı görülmektedir.

6.3. Uygulama

Heterojen yapı gösteren iki gerçek yaşam süresi verisi, standart olasılık dağılımları Üstel, Gamma ve Weibull dağılım modelleri ile iki farklı dağılımın karmasından oluşan Üstel-Gamma, Üstel-Weibull ve Gamma-Weibull karma dağılım modelleri kullanılarak modellenecektir. Standart olasılık dağılımları Üstel, Gamma ve Weibull dağılımlarının olasılık yoğunluk fonksiyonları

$$f_{\text{Üstel}}(t; \lambda) = \lambda e^{-\lambda t} \quad (6.35)$$

$$f_{\text{Gamma}}(t; \alpha, \beta) = t^{\alpha-1} \frac{e^{-t/\beta}}{\beta^{\alpha} \Gamma(\alpha)} \quad (6.36)$$

$$f_{\text{Weibull}}(t; a, b) = \frac{a}{b^a} t^{a-1} e^{-\left(\frac{t}{b}\right)^a} \quad (6.37)$$

şeklinde tanımlanmıştır. Üstel-Gamma, Üstel-Weibull ve Gamma-Weibull karma dağılım modellerinin standart olasılık dağılımlarından Üstel, Gamma ve Weibull dağılım modelleri ile karşılaştırılmasında Kolmogorov-Smirnov test istatistiği (D_{hesap}) ve hata kareler ortalaması (HKO) değerleri kullanılacaktır.

6.3.1. Dış Sulama Aygıtlarının Yaşam Süreleri

Çalışmada kullanılacak ilk veri seti Jiang ve Murthy (1997) makalelerinde çalıştığı dış sulama aygıtlarının yaşam süreleridir (Colvert ve Boardman 1976). Dış sulama aygıtlarının yaşam süreleri Tablo 6.4’de verilmiştir.

Tablo 6.4. Dış sulama aygıtlarının yaşam süreleri (Colvert ve Boardman 1976)

1.175	7.02	7.58	9.76	15.02	15.57
17.39	19.55	22.47	23.24	23.96	25.05
32.44	36.87	42.76	43.14	43.81	46.95
56.33	58.99	59.08	60.37	61.01	77.86
122.61	129.31	130.42	137.57	142.27	142.98
148.29	150.79	151.21	155.62	157.93	160.27
160.72	169.79	186.26	197.6	224.83	233.64
242.07	256.86	260.77	261.68	277.99	283.95
288.94	295.48	314.76	317.06	332.07	339.46
362.61	369.47	370.74	401.06	403.39	414.78
426.71	450.62	455.84	457.94	466.61	468.64
469.09	476.42	481.41	481.82	488.15	490.06
493.67	494.38	503.72	508.93	509.01	518.32
532.29	534.62	545.23	547.41	558.41	571.1
585.52	589.11	592.93	607.15	623.1	647.91

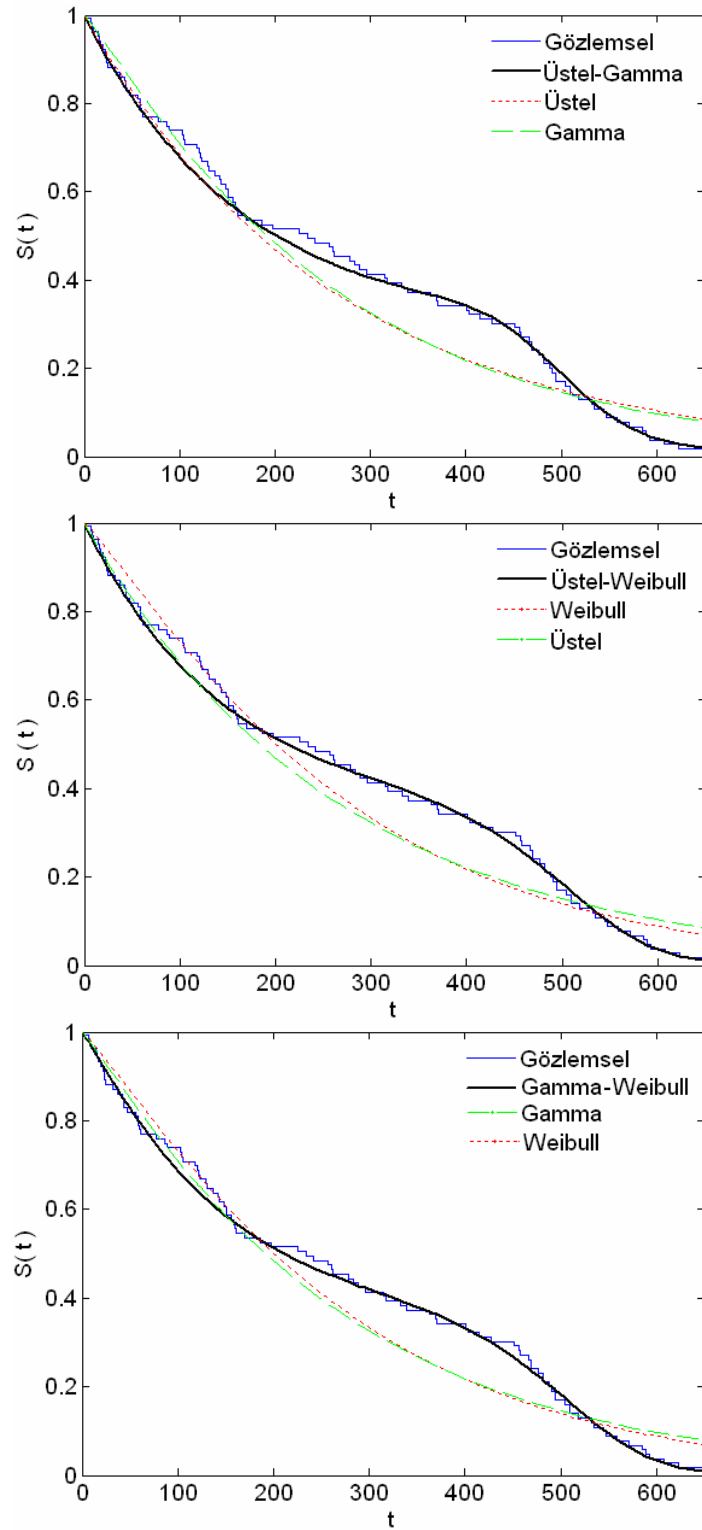
Dış sulama aygıtlarının yaşam süreleri standart olasılık dağılımlarından Üstel, Gamma ve Weibull dağılımları ile Üstel-Gamma, Üstel-Weibull ve Gamma-Weibull karma dağılım modelleri ile temsil edilecektir.

Üstel, Gamma ve Weibull dağılımları ile Üstel-Gamma, Üstel-Weibull ve Gamma-Weibull karma dağılım modellerinin dış sulama aygıtlarının yaşam süreleri verisine uyumunu incelemek için parametre tahminleri, Kolmogorov-Smirnov test istatistiği (D_{hesap}) ve hata kareler ortalaması (HKO) değerleri Tablo 6.5’de verilmiştir.

Tablo 6.5. Diş sulama aygıtlarının yaşam süreleri veri seti için Üstel, Gamma ve Weibull dağılım modelleri ve Üstel-Gamma, Üstel-Weibull ve Gamma-Weibull karma dağılım modellerinin parametre tahminleri, D_{hesap} ve HKO değerleri

	Parametre Tahminleri	D_{hesap} (p değeri)	HKO
Üstel	$\hat{\lambda} = 0.0038$	0.1127 (0.2031)	0.0034
Gamma	$\hat{\alpha} = 1.1219, \hat{\beta} = 235.8211$	0.1151 (0.1841)	0.0032
Weibull	$\hat{a} = 1.1464, \hat{b} = 276.3050$	0.1182 (0.1617)	0.0033
	$\hat{\pi} = 0.7109$		
Üstel-Gamma	$\hat{\lambda} = 0.006$ $\hat{\alpha} = 65.5052, \hat{\beta} = 7.7484$	0.0591 (0.9112)	0.00047
	$\hat{\pi} = 0.6561$		
Üstel-Weibull	$\hat{\lambda} = 0.0068$ $\hat{a} = 6.8394, \hat{b} = 520.9826$	0.0575 (0.9273)	0.00041
	$\hat{\pi} = 0.6507$		
Gamma-Weibull	$\hat{\alpha}_1 = 1.1412, \hat{\beta}_1 = 127.1518$ $\hat{a} = 6.8419, \hat{b} = 521.5024$	0.0500 (0.9780)	0.00034

Tablo 6.5 incelendiğinde, hem Kolmogorov-Smirnov test istatistiği hem de HKO değerlerine göre diş sulama aygıtlarının yaşam sürelerinin modellenmesinde Üstel-Gamma, Üstel-Weibull ve Gamma-Weibull karma dağılım modellerinin Üstel, Gamma ve Weibull dağılım modellerine göre daha başarılı olduğu görülmektedir. İki farklı karma dağılım modelleri içerisinde ise iki kritere göre de en başarılı model Gamma-Weibull karma dağılım modelidir.



Şekil 6.7. Diş sulama aygıtlarının yaşam sürelerine ait yaşam fonksiyonları

Diş sulama aygıtlarının yaşam sürelerinin modellenmesinde; gözlemsel yaşam fonksiyonu ile Üstel, Gamma ve Weibull dağılım modelleri ve Üstel-Gamma,

Üstel-Weibull ve Gamma-Weibull karma dağılım modellerinin yaşam fonksiyonlarının grafiksel karşılaştırmaları Şekil 6.7’de verilmiştir. Şekil 6.7’e göre Üstel-Gamma, Üstel-Weibull ve Gamma-Weibull karma modellerinin, Üstel, Gamma ve Weibull dağılımlarına göre diş sulama aygıtlarının yaşam sürelerinin modellenmesinde daha başarılı olduğu açık olarak görülmektedir.

6.3.2. Kanser Verisi

Çalışmada kullanılacak diğer veri seti ise Gusmao ve ark (2008) makalelerinde çalıştığı kanser tanısı alan ve radyoterapi sonrası ölen 42 hastanın tanı konduktan sonraki yaşam süreleridir. Kanser tanısı alan 42 hastanın tanı konduktan sonraki yaşam süreleri Tablo 6.6’da verilmiştir.

Tablo 6.6. Kanser tanısı alan 42 hastanın tanı konduktan sonraki yaşam süreleri (Gusmao ve ark 2008)

7	34	42	63	64	83
84	91	108	112	129	133
133	139	140	140	146	149
154	157	160	160	165	173
176	218	225	241	248	273
277	297	405	417	420	440
523	583	594	1101	1146	1417

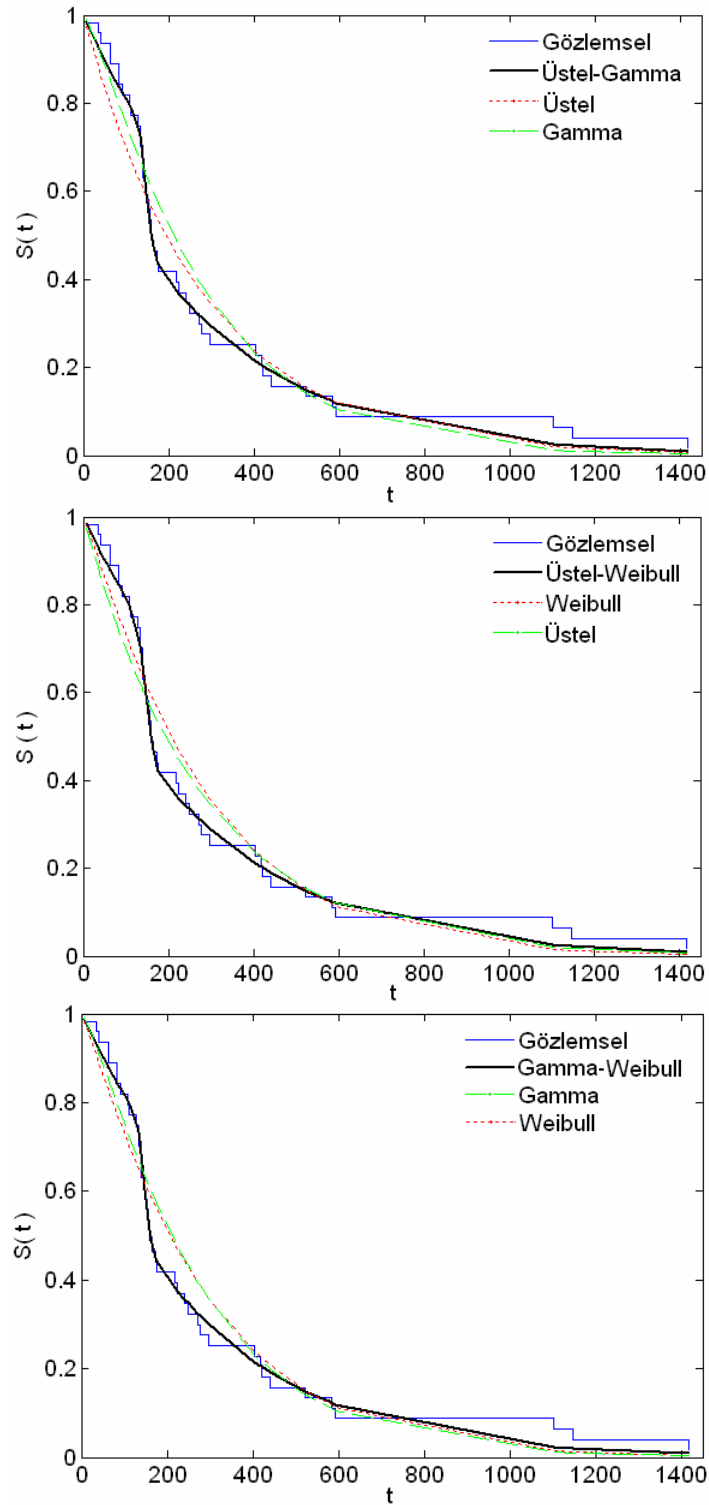
Kanser tanısı alan hastaların tanı konduktan sonraki yaşam süreleri Üstel, Gamma ve Weibull dağılımları ve Üstel-Gamma, Üstel-Weibull ve Gamma-Weibull karma dağılımları ile temsil edilecektir. Veri setinin modellenmesinde kullanılan dağılımlara ait parametre tahminleri, D_{hesap} ve HKO değerleri Tablo 6.7’de verilmiştir.

Tablo 6.7.Kanser tanısı alan hastaların tanı konduktan sonraki yaşam sürelerinin modellenmesi için Üstel, Gamma ve Weibull dağılım modelleri ve Üstel-Gamma, Üstel-Weibull ve Gamma-Weibull karma dağılım modelleri ile modellenmesinde elde edilen parametre tahminleri, D_{hesap} ve HKO değerleri

	Parametre Tahminleri	D_{hesap} (p değeri)	HKO
Üstel	$\hat{\lambda} = 0.0036$	0.1220 (0.5594)	0.0055
Gamma	$\hat{\alpha}_1 = 1.3026,$ $\hat{\beta}_1 = 215.0764$	0.1571 (0.2511)	0.0054
Weibull	$\hat{\alpha}_2 = 290.9438, \hat{\beta}_2 = 1.0916$	0.1437 (0.3510)	0.0055
Üstel-Gamma	$\hat{\pi} = 0.7262$ $\hat{\lambda} = 0.0030$ $\hat{\alpha}_1 = 116.7280, \hat{\beta}_1 = 1.2765$	0.0471 (0.9999)	0.00049
Üstel-Weibull	$\hat{\pi} = 0.6996$ $\hat{\lambda} = 0.00296$ $\hat{a} = 9.8770, \hat{b} = 153.7472$	0.0380 (0.9999)	0.00038
Gamma-Weibull	$\hat{\pi} = 0.2719$ $\hat{\alpha} = 118.2485, \hat{\beta} = 1.2602$ $\hat{a} = 1.0337, \hat{b} = 333.8992$	0.0558 (0.9995)	0.000622

Tablo 6.7 incelendiğinde, hem Kolmogorov-Smirnov test istatistiği hem de HKO değerlerine göre kanser tanısı alan hastaların tanı konduktan sonraki yaşam sürelerinin modellenmesinde Üstel-Gamma, Üstel-Weibull ve Gamma-Weibull karma dağılım modellerinin Üstel, Gamma ve Weibull dağılım modellerine göre daha başarılı olduğu görülmektedir. İki farklı karma dağılım modelleri içerisinde ise iki kritere göre de en başarılı model Üstel-Weibull karma dağılım modelidir.

Kanser hastalarının yaşam sürelerinin modellenmesinde; gözlemsel yaşam fonksiyonu ile Üstel, Gamma ve Weibull dağılım modelleri ve Üstel-Gamma, Üstel-Weibull ve Gamma-Weibull karma dağılım modellerinin yaşam fonksiyonlarının grafiksel karşılaştırmaları Şekil 6.8’de verilmiştir. Şekil 6.8 incelendiğinde, kanser hastalarının yaşam sürelerinin modellenmesinde Üstel-Gamma, Üstel-Weibull ve Gamma-Weibull karma modellerinin, Üstel, Gamma ve Weibull dağılımlarına göre daha başarılı olduğu açık olarak görülmektedir.



Şekil 6.8. Kanser tanısı alan hastaların tanı konduktan sonraki yaşam sürelerine ait yaşam fonksiyonları

6.4. Sonuçlar

Bu çalışmada heterojen yaşam verilerin modellenmesinde Üstel-Gamma, Üstel-Weibull ve Gamma-Weibull gibi iki farklı dağılımın karması önerilmiştir. Önerilen karma modellerin parametrelerinin en çok olabilirlik tahminleri EM algoritması ile elde edilmiştir. Üstel-Gamma, Üstel-Weibull ve Gamma-Weibull karma dağılımları için önerilen algoritmaların parametreleri tahmin etmedeki etkinliği simülasyon çalışması ile ortaya konmuştur. Simülasyon çalışmasına göre, EM algoritması karma modellerin parametre tahminlerinde başarılı olduğu görülmüştür.

Heterojen yaşam verilerinin modellenmesinde önerilen iki farklı dağılımın karmasından oluşan modellerin standart olasılık dağılımlara göre etkinliği iki gerçek veri seti üzerinde gösterilmiştir. Yaşam sürelerinin modellenmesinde yaygın olarak kullanılan Üstel, Gamma ve Weibull dağılımları önerilen iki farklı dağılımın karması ile iki gerçek veri setinin modellenmesinde karşılaştırılmıştır. Karşılaştırmada Kolmogorov-Smirnov test istatistiği ve Hata Kareler Ortalaması değerleri olmak üzere iki farklı kriter kullanılmıştır. Kullanılan iki kritere göre de heterojen yaşam sürelerinin modellenmesinde önerilen iki farklı dağılımın karmasından oluşan modeller standart olasılık dağılımlarına göre daha başarılı bulunmuştur.

Diş sulama aygıtlarının yaşam sürelerinin modellenmesinde Gamma ve Weibull dağılımlarının karmasından oluşan Gamma-Weibull karma dağılımı en başarılı model olarak belirlenmiştir. Kanseri tanısı alan hastaların tanı konduktan sonraki yaşam sürelerinin modellenmesinde ise Üstel ile Weibull dağılımlarının karmasından oluşan Üstel-Weibull karma dağılımı en başarılı model olarak belirlenmiştir.

7. KARIŞTIRILMIŞ DAĞILIMLARIN KARMA MODELİ

Yaşam süresi verilerinin modellenmesinde son dönemde, birçok yeni dağılım önerilmiştir (Adamidis ve Loukas 1998, Adamidis ve ark 2005, Kuş 2007, Tahmasbi ve Rezaei 2008). Bunun yanında özellikle heterojen yaşam sürelerinin modellenmesinde karma dağılım modelleri yaygın olarak kullanılmaktadır (Aalen 1992, Jiang ve Murthy 1997, Ling ve Pan 1998, Angelis ve ark 1999, Al-Hussaini ve ark 2000, Marin ve ark 2005).

Bu bölümde heterojen yaşam verilerinin modellenmesi için Adamidis ve ark (2005) tarafından önerilen genelleştirilmiş Üstel-Geometrik (EEG) dağılımının karma dağılım modeli oluşturulacaktır. Çalışmada öncelikle dönüştürülmüş uç değer dağılımında α parametresinin Üstel dağılıma sahip bir rassal değişken olarak tanımlanması ile elde edilen EEG dağılımı incelenecektir. EEG dağılımının heterojen yaşam sürelerinin modellenmesinde kullanılabilmesi için karma EEG dağılım modeli tanımlanarak, parametrelerinin en çok olabilirlik tahminleri için gerekli olan EM algoritması verilecektir. Tanımlanan EM algoritmasının parametre tahminindeki etkinliği simülasyon çalışması ile gösterilecektir. Ayrıca, iki bileşenli karma EEG dağılım modeli ile heterojen yapı gösteren iki gerçek veri seti modellenecektir. Gerçek veri setlerinin modellenmesinde karma EEG dağılım modeli, tek bileşenli Weibull, tek bileşenli EEG ve karma Weibull dağılım modeli ile karşılaştırılacaktır.

7.1. Genelleştirilmiş Üstel-Geometrik Dağılım

Yaşam verileri karmaşık bir yapıya sahip olduğundan dolayı standart olasılık dağılım modelleri heterojen yaşam verilerinin modellenmesinde direk olarak kullanılmaz. Bundan dolayı karıştırılmış dağılım heterojen yaşam verisi gibi karmaşık durumlarda kullanılabilir. Adamidis ve ark (2005), kitle birimlerinin farklı ölçekteki artan başarısızlık oranları ile fiziksel olarak ayırt edilemez olduğu için kitle heterojenliğinin modellenmesinde EEG dağılımını önermişlerdir.

T rassal değişkeni α ve β parametreleri ile dönüştürülmüş uç değer dağılımına sahip olsun. Dağılımın olasılık yoğunluk fonksiyonu

$$f(t | \alpha; \beta) = \alpha \beta e^{\beta t + \alpha(1 - e^{\beta t})}, \quad \alpha, \beta \in R_+ \quad (7.1)$$

şeklinde tanımlanmıştır. Dönüştürülmüş uç değer dağılımının yoğunluk fonksiyonundaki α parametresinin aşağıdaki gibi Üstel dağılıma sahip olduğu varsayılın.

$$g(\alpha; \gamma) = \gamma e^{-\gamma \alpha}, \quad \alpha, \gamma \in R_+ \quad (7.2)$$

Bu durumda T 'in koşulsuz yaşam dağılımı $f(t | \alpha; \beta)$ 'nin $g(\alpha; \gamma)$ tarafından karıştırılmasıyla elde edilebilir. Böylece karıştırılmış dağılımın olasılık yoğunluk fonksiyonu

$$f(t; \beta, \gamma) = \int_0^{\infty} f(t | \alpha; \beta) g(\alpha; \gamma) d\alpha \quad (7.3)$$

$$= \int_0^{\infty} \alpha \beta e^{\beta t + \alpha(1 - e^{\beta t})} \gamma e^{-\gamma \alpha} d\alpha$$

$$= \gamma \beta e^{\beta t} \int_0^{\infty} \alpha e^{\alpha(1 - e^{\beta t} - \gamma)} d\alpha$$

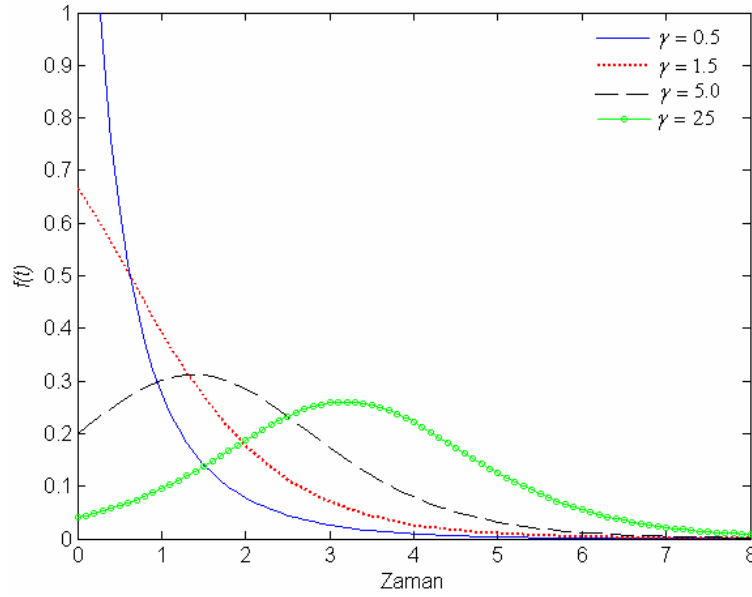
$$\Gamma(z) = k^z \int_0^{\infty} t^{z-1} e^{-kt} dt \text{ dönüşümü ile}$$

$$= \frac{\gamma \beta e^{\beta t} \Gamma(2)}{(e^{\beta t} + \gamma - 1)^2}$$

elde edilir. Elde edilen eşitlikte $\Gamma(2) = 1$ yerine yazılır ve düzenleme yapılırsa EEG dağılımının olasılık yoğunluk fonksiyonu

$$f(t; \beta, \gamma) = \beta \gamma e^{-\beta t} \{1 - (1 - \gamma)e^{-\beta t}\}^{-2} \quad (7.4)$$

elde edilmiş olur. Eşitlik (7.4) deki olasılık yoğunluk fonksiyonu Adamidis ve ark (2005) tarafından T 'in koşulsuz yaşam zamanı dağılımı için istatistiki bir dağılım modeli olarak kullanılmış ve EEG dağılımı olarak adlandırılmıştır. EEG dağılımında $\beta=1$ ve farklı γ parametre değerlerine göre elde edilen yoğunluk eğrileri Şekil 7.1'de verilmiştir.



Şekil 7.1. EEG dağılımında $\beta=1$ ve farklı γ parametre değerlerine göre elde edilen yoğunluk eğrileri

EEG dağılımının olasılık fonksiyonu $0 < \gamma \leq 2$ aralığı için $t=0$ durumunda modu $\beta\gamma^{-1}$ 'dir ve monoton azalan bir fonksiyondur. EEG dağılımının olasılık fonksiyonu $\beta\gamma^{-1}$ 'den $t = \beta^{-1} \ln(\gamma-1)$ 'deki modu $\beta\gamma\{4(\gamma-1)\}^{-1}$ 'e kadar artar, buradan $\gamma > 2$ için 0 değerine asimptotik olarak azalarak yaklaşır.

EEG dağılımında birikimli dağılım fonksiyonu $F(t)$,

$$F(t) = (1 - e^{-\beta t}) \{1 - (1 - \gamma)e^{-\beta t}\}^{-1} \quad (7.5)$$

yaşam fonksiyonu $S(t)$,

$$S(t) = \gamma e^{-\beta t} \{1 - (1 - \gamma)e^{-\beta t}\}^{-1} \quad (7.6)$$

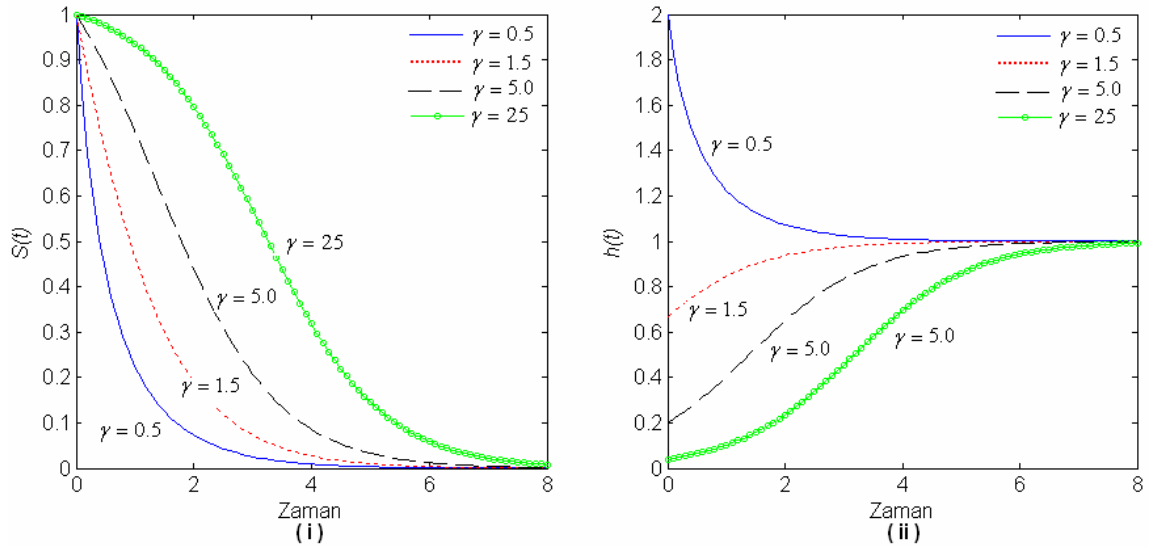
beklenen değer $E(T)$,

$$E(T) = \frac{-\gamma_k \ln \gamma_k}{\beta_k (1 - \gamma_k)} \quad (7.7)$$

ve hazard fonksiyonu $h(t)$,

$$h(t) = \beta \{1 - (1 - \gamma)e^{-\beta t}\}^{-1} \quad (7.8)$$

şeklinde. EEG dağılımında $\beta = 1$ ve farklı γ parametre değerlerine göre elde edilen yaşam fonksiyonu ve hazard fonksiyonu eğrileri Şekil 7.2’de verilmiştir.



Şekil 7.2. EEG dağılımında $\beta = 1$ ve farklı γ parametre değerlerine göre elde edilen (i) yaşam fonksiyonu ve (ii) hazard fonksiyonu eğrileri

7.2. EEG Dağılımının Karması

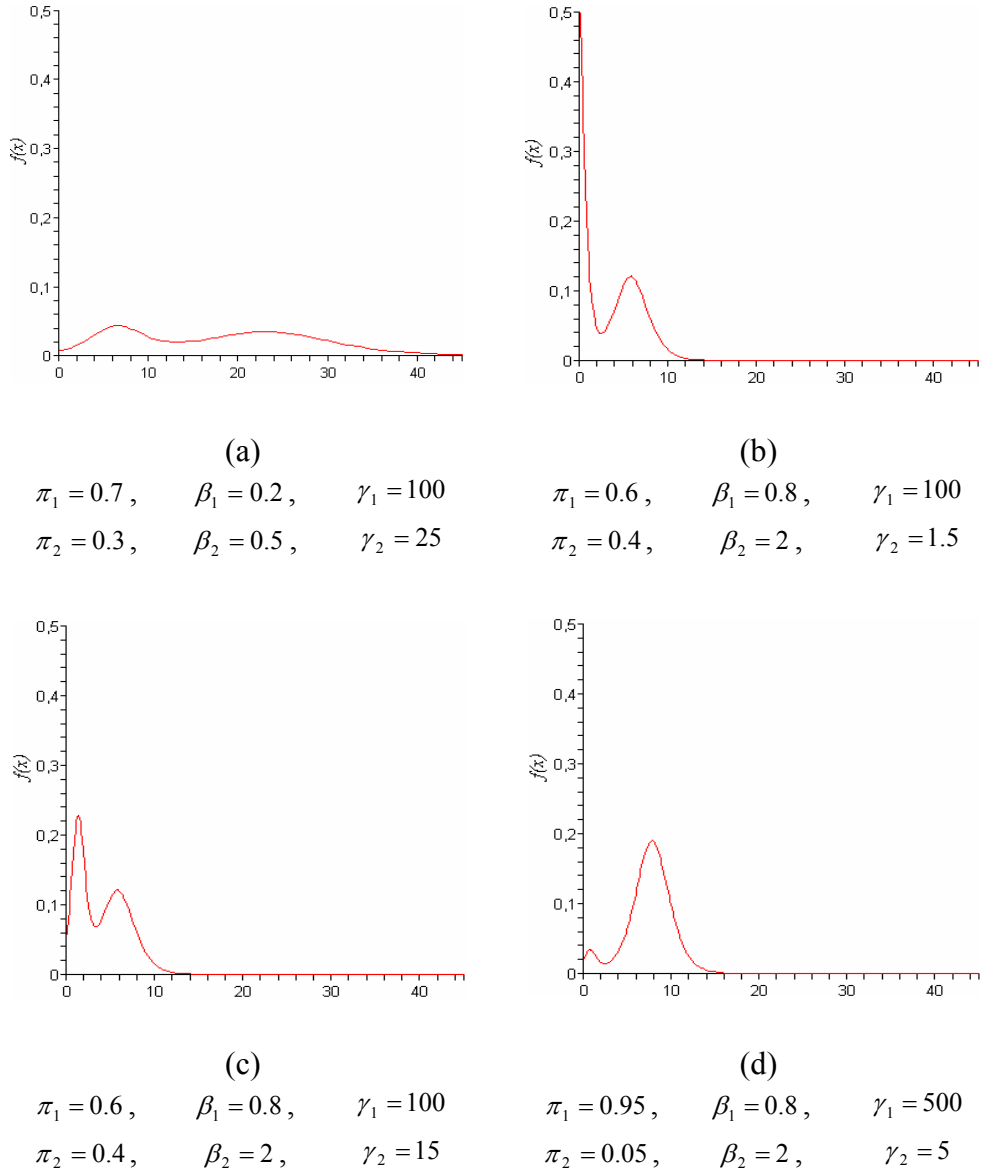
Bazı durumlarda karıştırılmış dağılım modelleri heterojen yaşam verilerinin modellenmesi için yeterli olmayabilir. Çünkü yaşam verisi oldukça karmaşık bir yapıya sahip olabilir. Bu yüzden heterojen yaşam verisinin modellenmesi için karıştırılmış dağılımın sonlu karmasına ihtiyaç duyarız. Karıştırılmış dağılımların sonlu karmaları karmaşık yapıdaki heterojen yaşam verisi için en iyi istatistik modelleridir (Erişoğlu ve Erol 2010).

g bileşenli EEG dağılımının karmasının olasılık yoğunluk fonksiyonu

$$f(t; \psi) = \sum_{k=1}^g \pi_k \left(\beta_k \gamma_k e^{-\beta_k t} \left\{ 1 - (1 - \gamma_k) e^{-\beta_k t} \right\}^{-2} \right) \quad (7.9)$$

şeklindedir.

Farklı parametreler değerleri için iki bileşenli EEG dağılımının ($g = 2$) karmasının grafikleri Şekil 7.3’de gösterilmiştir.



Şekil 7.3. Farklı parametre değerleri için iki bileşenli EEG dağılımının karmasının olasılık yoğunluk fonksiyonları

Karma EEG dağılım modelinin birikimli dağılım fonksiyonu

$$F(t) = \sum_{k=1}^g \pi_k \left((1 - e^{-\beta_k t}) \left\{ 1 - (1 - \gamma_k) e^{-\beta_k t} \right\}^{-1} \right) \quad (7.10)$$

şeklindedir. EEG dağılımlarının karmasında T rassal değişkeninin beklenen değeri

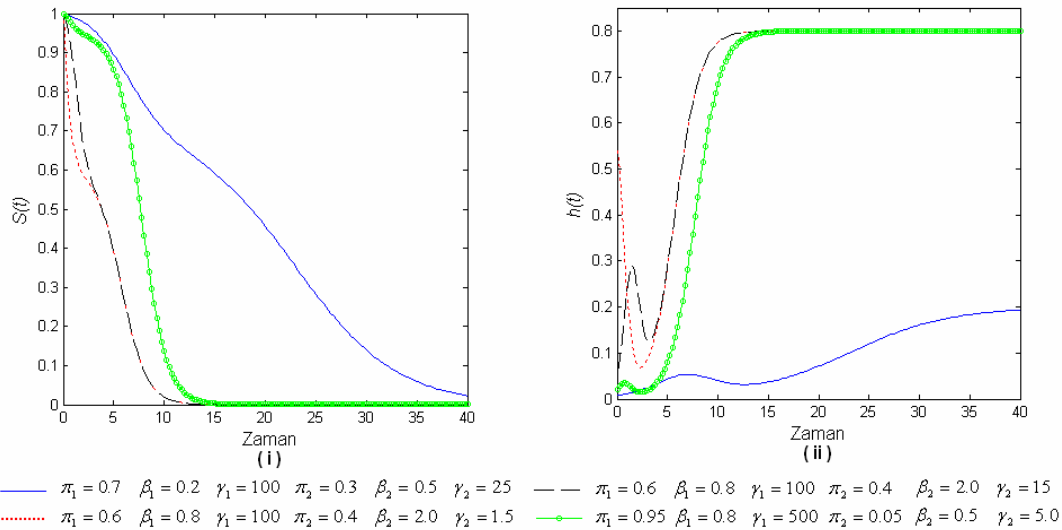
$$E(T) = \sum_{k=1}^g \pi_k \left(\frac{-\gamma_k \ln \gamma_k}{\beta_k (1 - \gamma_k)} \right) \quad (7.11)$$

şeklindedir. Karma EEG dağılım modelinin yaşam fonksiyonu $S(t; \psi)$ ve hazard fonksiyonu $h(t; \psi)$ sırasıyla

$$S(t; \psi) = \sum_{k=1}^g \pi_k \left(\gamma_k e^{-\beta_k t} \left\{ 1 - (1 - \gamma_k) e^{-\beta_k t} \right\}^{-1} \right) \quad (7.12)$$

$$h(t; \psi) = \frac{\sum_{k=1}^g \pi_k \left(\beta_k \gamma_k e^{-\beta_k t} \left\{ 1 - (1 - \gamma_k) e^{-\beta_k t} \right\}^{-2} \right)}{\sum_{k=1}^g \pi_k \left(\gamma_k e^{-\beta_k t} \left\{ 1 - (1 - \gamma_k) e^{-\beta_k t} \right\}^{-1} \right)} \quad (7.13)$$

şeklinde elde edilir. Karma EEG dağılımının modelinde farklı parametre değerlerine göre elde edilen yaşam fonksiyonu ve hazard fonksiyonu eğrileri Şekil 7.4'de verilmiştir.



Şekil 7.4. Karma EEG dağılımının modelinde farklı parametre değerlerine göre elde edilen (i) yaşam fonksiyonu ve (ii) hazard fonksiyonu eğrileri

7.3. EM Algoritması İle Karma EEG Dağılımının Parametrelerinin En Çok Olabilirlik Tahminleri

Karma EEG dağılım modelindeki parametrelerin en çok olabilirlik tahminlerini elde etmek için algoritmanın E adımında kayıp gözlem vektörünün elemanları

$$\hat{z}_{ki} = \frac{\pi_k f(t_i; \theta_k)}{\sum_{s=1}^g \pi_s f_s(t_i; \theta_s)} \quad (7.14)$$

şeklinde tahmin edilir. Eşitlikte yer alan $f(t_i; \theta_k)$, k . EEG bileşenine ait olasılık yoğunluk fonksiyonunu göstermektedir. Tamamlanmış veri için olabilirlik fonksiyonunun doğal logaritmasının beklenen değeri,

$$E(\ln L) = \sum_{i=1}^n \sum_{k=1}^g \hat{z}_{ki} \ln \left(\pi_k \left(\beta_k \gamma_k e^{-\beta_k t_i} \left\{ 1 - (1 - \gamma_k) e^{-\beta_k t_i} \right\}^{-2} \right) \right) \quad (7.15)$$

şeklindedir. EM algoritmasının M adımında, E adımında elde edilen $E(\ln L)$, $\sum_{k=1}^g \pi_k = 1$ koşulu altında maksimize edilir. Bunun için $\ln \tilde{L}$ Langrance yöntemi ile $E(\ln L)$ maksimize edilmeye çalışılır.

$$\ln \tilde{L} = \sum_{i=1}^n \sum_{k=1}^g \hat{z}_{ki} \ln(\pi_k) + \hat{z}_{ki} \left\{ \ln(\beta_k \gamma_k) - \beta_k t_i - 2 \ln(1 - (1 - \gamma_k) e^{-\beta_k t_i}) \right\} - \lambda \left(\sum_{k=1}^g \pi_k - 1 \right) \quad (7.16)$$

Eşitlik (7.16)'da verilen ifadeyi maksimize eden π_k , β_k ve γ_k parametrelerini tahmin etmek için, sırasıyla tüm parametrelere göre ayrı ayrı $\ln \tilde{L}$ 'nin ilk türevleri alınarak sıfıra eşitlenir. Böylece $\ln \tilde{L}$ yi maksimize etmek için gerekli koşullar aşağıdaki eşitliklerden çıkarılabilir.

$$\hat{\pi}_k = \frac{1}{n} \sum_{i=1}^n \hat{z}_{ki}, \quad k=1, \dots, g \quad (7.17)$$

$$\frac{\partial \ln \tilde{L}}{\partial \beta_k} = \frac{\sum_{i=1}^n \hat{z}_{ki}}{\beta_k} + \sum_{i=1}^n \hat{z}_{ki} t_i - 2 \sum_{i=1}^n \frac{\hat{z}_{ki} t_i}{(1 - (1 - \gamma_k) e^{-\beta_k t_i})} = 0, \quad k=1, \dots, g \quad (7.18)$$

$$\frac{\partial \ln \tilde{L}}{\partial \gamma_k} = \frac{(1 + \gamma_k) \sum_{i=1}^n \hat{z}_{ki}}{\gamma_k} - 2 \sum_{i=1}^n \hat{z}_{ki} \left\{ \frac{1}{(1 - (1 - \gamma_k) e^{-\beta_k t_i})} \right\} = 0, \quad k=1, \dots, g \quad (7.19)$$

Görüldüğü gibi, π_k için tahmin $\hat{z}_{ki}$ etiket değerlerine bağlı olarak (7.17)'deki gibi elde edilebilir. Bu gözlemlerimizin karma yapıdaki heterojen yaşam verilerinin hangi sınıfından ya da grubundan geldiğini öğrenmek demektir. Böylece karma yapıdaki heterojen yaşam verilerinin gözlemleri etiketlenmiş olur. (7.18) ve (7.19) denklemlerinde verilen β_k ve γ_k için lineer olmayan denklemlerin çözümü nümerik bir metotla elde edilmelidir. Burada EM algoritması tekrar uygulanır.

Kısaca, EM algoritması T gözlem değerlerine A gözlenemeyen değişkenin ilave edilmesiyle uygulanır. Çünkü T karıştırılmış değişken yapıdadır. T 'nin olasılık yoğunluk fonksiyonu $f_k(t | \alpha_k; \beta_k)$ olarak tanımlanır. $f_k(t | \alpha_k; \beta_k)$ 'nin α_k parametresi $g_k(\alpha_k; \gamma_k)$ olasılık yoğunluk fonksiyonu ile rasgele bir değişken olarak rol almaktadır. Böylece A_k , α_k ile ilişkilidir. Dolayısıyla A_k karıştırılmış yapıdaki iki dağılımın karıştırıcı değişkenidir. Böylece tam verinin yoğunluk fonksiyonu

$$f(t, \alpha_k; \theta_k) = \alpha_k \beta_k \gamma_k e^{\beta_k t + \alpha_k (1 - \gamma_k - e^{-\beta_k t})}, \quad \alpha_k, \beta_k, \gamma_k \in R_+ \quad (7.20)$$

şeklinde hesaplanır. $l_{kc} = l(\theta_k; Y_{kc})$ tam veri için log olabilirlik fonksiyonu olsun. $Y_{kc} = (T_i, A_{ki}; i=1, \dots, n)$ sonraki yoğunluğa karşılık gelir. $Y_{kc} = (T_i, A_{ki}; i=1, \dots, n)$ karıştırılmış değişken gözlemleri ile tanımlanır ve karıştırılmış dağılımın karıştıran değişkeninin değeri ile ilgilidir. EM döngüsünün E adımı verilerin $\theta_k^{(r)} = (\beta_k^{(r)}, \gamma_k^{(r)})$ değerleri için

$$E(l_{kc}) \propto \ln(\beta_k \gamma_k) \sum_{i=1}^n \hat{z}_{ki} + \beta_k \sum_{i=1}^n \hat{z}_{ki} t_i + (1 - \gamma_k) \sum_{i=1}^n \hat{z}_{ki} E(A_{ki}) - \sum_{i=1}^n \hat{z}_{ki} e^{\beta_k t_i} E(A_{ki}) \quad (7.21)$$

hesaplanır. Burada beklenenler y_{obs} ve $\theta_k^{(r)}$ bağlıdır. $A_k | T$ 'nın koşullu olasılık yoğunluk fonksiyonu $f(\alpha_k | t; \theta_k) = \alpha_k \{1 - (1 - \gamma_k) e^{-\beta_k t}\}^2 e^{\alpha_k (1 - \gamma_k - e^{-\beta_k t}) + 2\beta_k t}$, $t, \alpha_k, \beta_k, \gamma_k \in R_+$ ile tanımlanır. Böylece (7.21) nolu eşitlikte yer alan beklenen değer $E(A_k | T; \theta_k) = 2e^{-\beta_k t} \{1 - (1 - \gamma_k) e^{-\beta_k t}\}^{-1}$ şeklinde hesaplanır. M adımında, E adımında verilen $\theta_k^{(r)} = (\beta_k^{(r)}, \gamma_k^{(r)})$ değerleri için oluşturulan (7.21) nolu eşitlik maksimum edilmeye çalışılır. Bunun için parametrelere göre kısmi türevler alınıp sıfıra eşitlenmesi ile,

$$\frac{\sum_{i=1}^n \hat{z}_{ki}}{\beta_k^{(r+1)}} + \sum_{i=1}^n \hat{z}_{ki} t_i - \sum_{i=1}^n \hat{z}_{ki} t_i e^{\beta_k^{(r+1)} t_i} E(A_{ki}) = 0, \quad k = 1, \dots, g \quad (7.22)$$

$$\frac{\sum_{i=1}^n \hat{z}_{ki}}{\gamma_k^{(r+1)}} - \sum_{i=1}^n \hat{z}_{ki} E(A_{ki}) = 0, \quad k = 1, \dots, g \quad (7.23)$$

eşitlikleri elde edilir. (7.22) nolu eşitliğin çözümünde Newton-Raphson metodu kullanılabilir. $\beta_k^{(r+1)}$ 'ın Newton-Raphson metodu ile elde edilen en çok olabilirlik tahmin edicisi,

$$\hat{\beta}_k^{(r+1)} = \frac{\beta_k^{(r)} \left(2 \left(\sum_{i=1}^n \hat{z}_{ki} \right) + \left(\sum_{i=1}^n \hat{z}_{ki} t_i^2 e^{\beta_k^{(r)} t_i} E(A_{ki}) \right) (\beta_k^{(r)})^2 + \left(\sum_{i=1}^n \hat{z}_{ki} t_i \right) \beta_k^{(r)} - \left(\sum_{i=1}^n \hat{z}_{ki} t_i e^{\beta_k^{(r)} t_i} E(A_{ki}) \right) \beta_k^{(r)} \right)}{\left(\sum_{i=1}^n \hat{z}_{ki} \right) + \left(\sum_{i=1}^n \hat{z}_{ki} t_i^2 e^{\beta_k^{(r)} t_i} E(A_{ki}) \right) (\beta_k^{(r)})^2} \quad (7.24)$$

şeklindedir. $\gamma_k^{(r+1)}$ 'ın en çok olabilirlik tahmin edicisi ise, (7.23) nolu eşitlik kullanılarak

$$\hat{\gamma}_k^{(r+1)} = \frac{\sum_{i=1}^n \hat{z}_{ki}}{2 \sum_{i=1}^n \hat{z}_{ki} e^{-\beta_k^{(r)} t_i} \left\{ 1 - (1 - \gamma_k^{(r)}) e^{-\beta_k^{(r)} t_i} \right\}^{-1}} \quad (7.25)$$

doğrudan elde edilir. İki bileşenli EEG dağılımının karma modelindeki $\hat{\pi}_k$, $\hat{\beta}_k^{(r+1)}$ ve $\hat{\gamma}_k^{(r+1)}$ parametrelerinin tahmini için iterasyonda sırasıyla (7.17), (7.24) ve (7.25)

eşitlikleri kullanılır. EM algoritmasında başlangıç değerleri için $\beta_k^{r=0} = \frac{\sum_{i=1}^n \hat{z}_{ki}}{\sum_{i=1}^n \hat{z}_{ki} t_i}$

şeklinde hesaplanırken $\gamma_k^{r=0}$ rasgele olarak belirlenmektedir.

Simülasyonlarda EM algoritmasının yakınsaması incelenmiştir. ψ 'nin 4 farklı değeri için her biri rasgele olarak iki bileşenli EEG dağılımının karmasından 500 boyutlu 300 örneklem üretilmiştir. Tahminler arasındaki fark 10^{-5} 'den daha az olduğu zaman yakınsamanın olduğu varsayılmıştır. Simülasyondan bulunan 1200 veri setinden elde edilen sonuçlar Tablo 7.1'de görüldüğü gibidir. Tabloda 300 en çok olabilirlik tahminlerinde elde edilen ortalama tahmin $av(\hat{\pi}, \hat{\theta})$ ve bunların standart hatası $se(\hat{\pi}, \hat{\theta})$ verilmektedir.

Tablo 7.1. İki bileşenli EEG dağılımının karmasından 500 boyutlu 300 örneklemde elde edilen EM tahminlerinin ortalamaları ve standart hataları

Kitle		$\hat{\pi}$	$\hat{\beta}_1$	$\hat{\gamma}_1$	$\hat{\beta}_2$	$\hat{\gamma}_2$
$\pi = 0.7, \beta_1 = 2.5, \gamma_1 = 20$ $\beta_2 = 10, \gamma_2 = 5$	$av(\hat{\pi}, \hat{\theta})$	0.6963	2.5205	16.7508	10.3671	6.1557
	$se(\hat{\pi}, \hat{\theta})$	0.0028	0.0120	0.4015	0.1374	0.2490
$\pi = 0.65, \beta_1 = 1.5, \gamma_1 = 6$ $\beta_2 = 20, \gamma_2 = 20$	$av(\hat{\pi}, \hat{\theta})$	0.6513	1.5041	6.2688	20.4742	23.5769
	$se(\hat{\pi}, \hat{\theta})$	0.0019	0.0070	0.0951	0.1384	0.6181
$\pi = 0.85, \beta_1 = 1, \gamma_1 = 30$ $\beta_2 = 20, \gamma_2 = 10$	$av(\hat{\pi}, \hat{\theta})$	0.8498	1.0031	30.9838	20.1648	11.3017
	$se(\hat{\pi}, \hat{\theta})$	0.0010	0.0031	0.3907	0.1863	0.3087
$\pi = 0.8, \beta_1 = 0.5, \gamma_1 = 50$ $\beta_2 = 5, \gamma_2 = 10$	$av(\hat{\pi}, \hat{\theta})$	0.7997	0.5015	52.0956	5.0646	11.4430
	$se(\hat{\pi}, \hat{\theta})$	0.0011	0.0015	0.7540	0.0434	0.3129

Simülasyon sonucunda elde edilen ortalama parametre tahminleri incelendiğinde EM algoritmasının karma dağılım modellerindeki farklı parametre değerlerinin tahmininde oldukça başarılı olduğu görülmektedir. $av(\hat{\pi}, \hat{\theta})$ ve $se(\hat{\pi}, \hat{\theta})$ değerleri de EM tahminlerinin tutarlı olduğunu gösterir. Simülasyon sonuçlarına göre EM yaklaşımının farklı karma oranlarında da iyi çalıştığı görülmektedir.

7.4. Uygulamalar

Heterojen yaşam sürelerinin modellenmesi için önerilen karıştırılmış EEG dağılımının karma modeli tek bileşenli EEG ve Weibull dağılımı ile iki bileşenli Weibull karma modeliyle karşılaştırılacaktır. Literatürde iki farklı alt sınıftan oluşan iki gerçek heterojen yaşam verisi modellenmeye çalışılacaktır. Karma EEG dağılım modelinin diğer dağılım modelleri ile karşılaştırılmasında Akaike bilgi kriteri (AIC), Kolmogorov-Smirnov test istatistiği (D_{hesap}) ve hata kereler ortalaması (HKO) değerleri kullanılacaktır.

7.4.1. Yaşam Yorgunluğu Veri Seti

Yaşam yorgunluğu veri seti iki farklı tipte 25 numuneyi içerir ve artan olarak listelenmiştir (Ling ve Pan 1998). İki farklı türde 25 birimin yaşam yorgunluğu verisi Tablo 7.2’de verilmiştir.

Tablo 7.2. İki farklı tipteki 25 numunenin yaşam yorgunluğu (Ling ve Pan 1998)

5.2426	6.3014	7.4485	8.1348	8.4369	9.0949	9.9299
10.2580	10.9970	11.3670	11.6950	12.6310	13.7140	15.5040
19.8910	23.7850	26.4670	28.3450	31.0980	34.8500	35.6890
37.2250	41.0590	44.8920	49.8390			

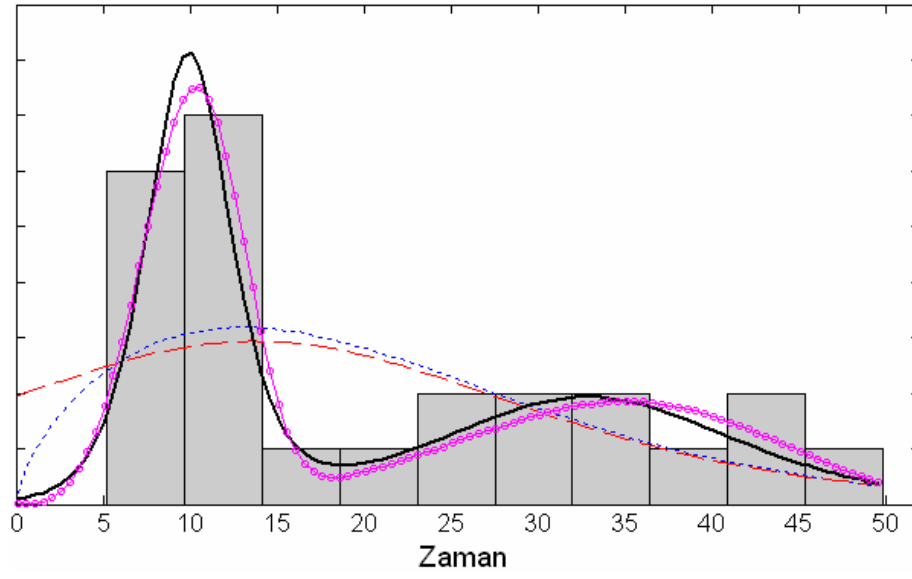
Tablo 7.2’de verilen yaşam yorgunluğu verisine Weibull dağılım modelinin olasılık yoğunluk fonksiyonu, EEG dağılım modelinin olasılık yoğunluk fonksiyonu, iki bileşenli Weibull dağılımın karma modelinin olasılık yoğunluk fonksiyonu ve iki bileşenli EEG dağılımın karma modelinin olasılık yoğunluk fonksiyonu uygulanmıştır.

Tablo 7.3. Yaşam yorgunluğu veri seti için ilgilenilen dağılım modellerinde parametre tahminleri, log olabilirlik fonksiyonu ($\ln L$), AIC , D_{hesap} ve HKO değerleri

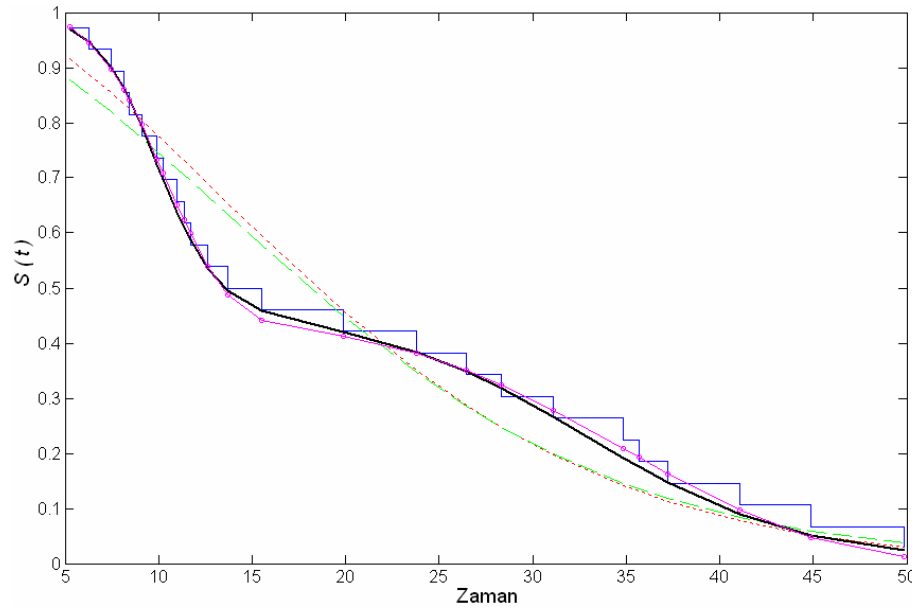
Dağılımlar	Parametre tahminleri	$\ln L$	AIC	D_{hesap} (p değeri)	HKO
Weibull Dağılımı	$\hat{\beta} = 1.6360$ $\hat{\alpha} = 23.1352$	-96.5679	197.1358	0.1537 (0.5961)	0.006269374
EEG Dağılımı	$\hat{\beta} = 0.0968$ $\hat{\gamma} = 4.7884$	-97.8598	199.8196	0.133381 (0.7653)	0.005283
İki bileşenli Weibull Dağılımın Karması	$\hat{\pi} = 0.56$ $\hat{\beta}_1 = 4.0386$ $\hat{\beta}_2 = 4.3134$ $\hat{\alpha}_1 = 11.0932$ $\hat{\alpha}_2 = 37.2930$	-90.2449	190.4880	0.0265 (0.9999)	0.0002295
İki bileşenli EEG Dağılımın Karması	$\hat{\pi} = 0.4619$ $\hat{\beta}_1 = 0.1736$ $\hat{\beta}_2 = 0.6185$ $\hat{\gamma}_1 = 301.5254$ $\hat{\gamma}_2 = 466.8431$	-90.6066	191.2132	0.0322206 (0.9999)	0.00020634

Tablo 7.3 incelendiğinde log olabilirlik değerleri, AIC değerleri, K-S test istatistiği ve HKO değerlerine göre yaşam yorgunluğu veri setinin modellenmesinde iki bileşenli EEG dağılımının karması Weibull dağılımına ve EEG dağılımına göre daha uygun olduğu görülmektedir. Veri modellemesinde sık sık kullanılan iki bileşenli Weibull dağılımının karması ile iki bileşenli EEG dağılımının karması karşılaştırıldığında benzer sonuçlar vermektedir. HKO değerlerine göre iki bileşenli EEG dağılımının karması iki bileşenli Weibull dağılımının karmasından ayrıca daha iyi olduğu görülmektedir.

Yaşam yorgunluğu verisinin Weibull dağılım modelinin olasılık yoğunluk fonksiyonuna, EEG dağılım modelinin olasılık yoğunluk fonksiyonuna, iki bileşenli Weibull dağılımın karma modelinin olasılık yoğunluk fonksiyonuna ve iki bileşenli EEG dağılımın karma modelinin olasılık yoğunluk fonksiyonuna uyumunun grafiksel karşılaştırılması Şekil 7.5.(a)'da verilmiştir. Ampirik yaşam fonksiyonu ile Weibull dağılım modelinin yaşam fonksiyonunun, EEG dağılım modelinin yaşam fonksiyonunun, iki bileşenli Weibull dağılımın karma modelinin yaşam fonksiyonunun ve iki bileşenli EEG dağılımın karma modelinin yaşam fonksiyonunun grafiksel karşılaştırmaları Şekil 7.5.(b)'de verilmiştir.



(a)



(b)

Şekil 7.5. (a) Yaşam yorgunluğu verisinin Weibull dağılım modelinin yoğunluk eğrisi (nokta çizgi), EEG dağılım modelinin yoğunluk eğrisi (ince kırık çizgi), iki bileşenli Weibull dağılımın karma modelinin yoğunluk eğrisi (dairesel çizgi) ve iki bileşenli EEG dağılımın karma modelinin yoğunluk eğrisi (kalın çizgi) uyumunun grafiksel karşılaştırılması (b) Ampirik yaşam fonksiyonu (çokgen çizgi) ile Weibull dağılım modelinin yaşam fonksiyonunun (nokta çizgi), EEG dağılım modelinin yaşam fonksiyonunun (ince kırık çizgi), iki bileşenli Weibull dağılımın karma modelinin yaşam fonksiyonunun (dairesel çizgi) ve iki bileşenli EEG dağılımın karma modelinin yaşam fonksiyonunun (kalın çizgi) grafiksel karşılaştırmaları

Yaşam yorgunluğu verisi için iki bileşenli EEG dağılımının karmasının Weibull dağılımından, EEG dağılımından ve iki bileşenli Weibull dağılımının karmasından daha uygun olduğu Şekil 7.5.(a)'da görülmektedir. Şekil 7.5.(a)' da iki bileşenli EEG dağılımının karmasında ilk ve ikinci bileşenlerin yoğunluk fonksiyonları için $\gamma > 2$ olduğu görülmektedir. Böylece yaşam yorgunluğu veri seti için bileşen yoğunlukları önce asimptotik olarak sıfır değerinde mod değerine artar ve sonra asimptotik olarak sıfır değerine yaklaşır. Weibull dağılımının yaşam fonksiyonu, EEG dağılımının yaşam fonksiyonu, iki bileşenli Weibull dağılımının karmasının yaşam fonksiyonu ve iki bileşenli EEG dağılımının karmasının yaşam fonksiyonu ile uygun ampirik yaşam fonksiyonu Şekil 7.5.(b)' de görüldüğü gibi grafiksel olarak karşılaştırıldığında yaşam yorgunluğu veri seti için iki bileşenli EEG dağılımının karmasının yaşam fonksiyonunun daha uygun olduğu görülmektedir.

7.4.2. Diş Sulama Aygıtlarının Yaşam Süreleri

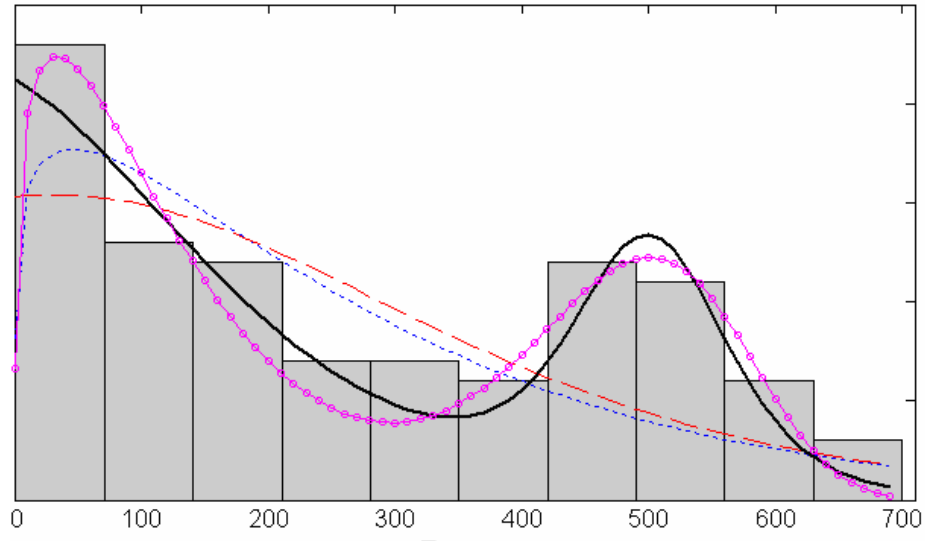
Çalışmada kullanılacak diğer veri seti ise Jiang and Murthy (1997) makalelerinde çalıştığı diş sulama aygıtlarının yaşam süreleridir (Colvert and Boardman 1976). Bu veri setinde yer alan 98 bozulma süresi Weibull, EEG, karma Weibull ve karma EEG dağılım modelleri ile temsil edilecektir. Diş sulama aygıtlarının yaşam süreleri veri seti için ilgilenilen dağılım modellerinde parametre tahminleri, log olabilirlik fonksiyonu değerleri, AIC değerleri, K-S test istatistiği ve HKO değerleri Tablo 7.4'de verilmiştir.

Tablo 7.4. Diş sulama aygıtlarının yaşam süreleri için ilgilenilen dağılım modellerinde parametre tahminleri, log olabilirlik fonksiyonu ($\ln L$), AIC , D_{hesap} ve HKO değerleri

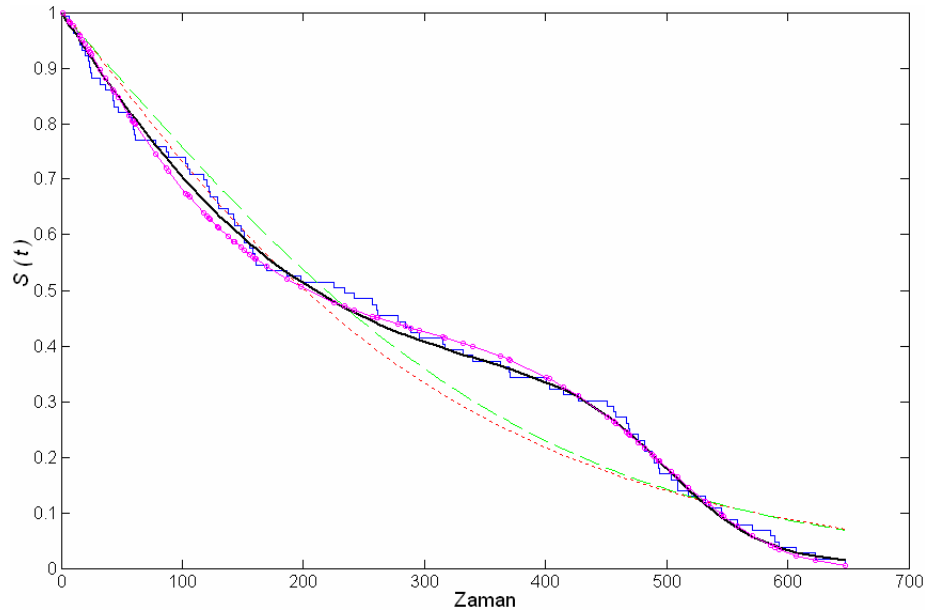
Dağılımlar	Parametre tahminleri	$\ln L$	AIC	D_{hesap} (p değeri)	HKO
Weibull Dağılımı	$\hat{\beta} = 1.1464$ $\hat{\alpha} = 276.3046$	-643.4262	1290.852	0.1182 (0,1293)	0.312031077
EEG Dağılımı	$\hat{\beta} = 0.0053$ $\hat{\gamma} = 2.1816$	-642.306	1288.612	0.111017 (0,17849)	0.002990401
İki bileşenli Weibull Dağılımın Karması	$\hat{\pi} = 0.5918$ $\hat{\beta}_1 = 1.2428$ $\hat{\beta}_2 = 6.4971$ $\hat{\alpha}_1 = 123.7294$ $\hat{\alpha}_2 = 515.3055$	-626.3399	1262.68	0.0582 (0,8942)	0.0005325
İki bileşenli EEG Dağılımın Karması	$\hat{\pi} = 0.6825$ $\hat{\beta}_1 = 0.0082$ $\hat{\beta}_2 = 0.0249$ $\hat{\gamma}_1 = 1.6609$ $\hat{\gamma}_2 = 261380$	-625.548	1261.096	0.038889 (0,99842)	0.000285392

Tablo 7.4 incelendiğinde log olabilirlik değerleri, AIC değerleri, K-S test istatistiği ve HKO değerlerine göre diş sulama aygıtlarının yaşam süreleri veri setinin modellenmesine iki bileşenli EEG dağılımın karmasının Weibull dağılımına ve EEG dağılımına göre daha uygun olduğu görülmektedir. Tablo 7.4'deki sonuçlara göre, diş sulama aygıtlarının yaşam süreleri veri setinin modellenmesinde tüm karşılaştırma kriterlerine göre iki bileşenli EEG dağılımının karmasının iki bileşenli Weibull dağılımının karmasından daha iyi olduğu görülmektedir.

Diş sulama aygıtlarının yaşam sürelerinin modellenmesinde kullanılan dağılım modellerinin olasılık yoğunluk fonksiyonu eğrileri Şekil.7.6.(a)'da verilmiştir. Ampirik yaşam fonksiyonu ile Weibull dağılım modelinin yaşam fonksiyonunun, EEG dağılım modelinin yaşam fonksiyonunun, iki bileşenli Weibull dağılımın karma modelinin yaşam fonksiyonunun ve iki bileşenli EEG dağılımın karma modelinin yaşam fonksiyonunun grafiksel karşılaştırmaları Şekil 7.6.(b)'de verilmiştir.



(a)



(b)

Şekil 7.6. (a) Dış sulama aygıtlarının yaşam süreleri Weibull dağılım modelinin yoğunluk eğrisi(nokta çizgi), EEG dağılım modelinin yoğunluk eğrisi(ince kırık çizgi), iki bileşenli Weibull dağılımın karma modelinin yoğunluk eğrisi(daireli çizgi) ve iki bileşenli EEG dağılımın karma modelinin yoğunluk eğrisi(kalın çizgi) uyumunun grafiksel karşılaştırılması (b) Ampirik yaşam fonksiyonu (çokgen çizgi) ile Weibull dağılım modelinin yaşam fonksiyonunun (nokta çizgi), EEG dağılım modelinin yaşam fonksiyonunun (ince kırık çizgi), iki bileşenli Weibull dağılımın karma modelinin yaşam fonksiyonunun (daireli çizgi) ve iki bileşenli EEG dağılımın karma modelinin yaşam fonksiyonunun (kalın çizgi) grafiksel karşılaştırmaları

Diş sulama aygıtlarının yaşam süreleri için iki bileşenli EEG dağılımının karmasının Weibull dağılımından, EEG dağılımından ve iki bileşenli Weibull dağılımının karmasından daha uygun olduğu Şekil 7.6.(a)'da görülmektedir. Şekil 7.6.(a)'da iki bileşenli EEG dağılımının karmasının yoğunluk fonksiyonunda ilk bileşen için $0 < \gamma \leq 2$ aralığında ve ikinci bileşen için $\gamma > 2$ aralığında olduğu görülmektedir. Böylece, ilk bileşenin yoğunluk fonksiyonu $t = 0$ 'da mod değeri ile monoton artan ve ikinci bileşenin yoğunluk fonksiyonu önce mod değerinde artan iken sonra asimptotik olarak sıfır değerine azalır. Weibull dağılımının yaşam fonksiyonu, EEG dağılımının yaşam fonksiyonu, iki bileşenli Weibull dağılımının karmasının yaşam fonksiyonu ve iki bileşenli EEG dağılımının karmasının yaşam fonksiyonu ile uygun ampirik yaşam fonksiyonu Şekil 7.6.(b)'de görüldüğü gibi grafiksel olarak karşılaştırıldığında diş sulama aygıtlarının yaşam süreleri veri seti için iki bileşenli EEG dağılımının karmasının yaşam fonksiyonunun daha uygun olduğu görülmektedir. Bunun sebebi iki bileşenli EEG dağılımının karmasının uzun kuyruklu veri setlerine uygun olmasıdır.

7.5. Sonuçlar

Bu çalışmada heterojen yaşam veri setlerine EEG dağılımının karmaşı önerildi. Karma EEG dağılımının parametrelerinin en çok olabilirlik tahminleri EM algoritması ile elde edildi. İki bileşenli EEG dağılımının karmaşı iki gerçek heterojen yaşam veri setlerinin modellenmesinde başarıyla uygulandı.

Grafikler incelendiğinde Weibull dağılımı ve EEG dağılımı gibi tek bileşenli dağılım modellerinin yaşam yorgunluğu ve diş sulama aygıtlarının yaşam süreleri veri setlerinin modellenmesi için uygun değildir. Fakat Tablo 7.3 ve Tablo 7.4'deki K-S test istatistiğine göre Weibull dağılımı ve EEG dağılımı gibi tek yoğunluk modellerini ret edilmemektedir. Bu sonuçlara göre, uygun model seçiminde sadece Kolmogorov-Smirnov test istatistiğinin kullanılmasının yetersiz olduğu görülmektedir.

Yaşam yorgunluğu veri seti ve diş sulama aygıtlarının yaşam süreleri veri setine uyumu için önerilen iki bileşenli EEG dağılımının karma modeli iki bileşenli

Weibull dađılımının karma modeline bir alternatiftir ve belli uç durumlarda iki bileőenli EEG dađılımının karma modeli iki bileőenli Weibull dađılımının karma modelinden daha elveriőli olabilir.

İki dađılımının karmasının avantajı farklı karakteristik özelliklere sahip dađılımların beraber kullanımıdır. Bu çalışmada önerilen dađılım yaklaşımı iki karıőtırılmıő dađılımın karması olarak da adlandırılabilir.

8.SONUÇLAR VE ÖNERİLER

Yaşam analizi, tanımlanan bir olayın meydana gelmesi için geçen sürelerin analizinde kullanılan en eski analizlerden biridir. Yaşam analizine konu olan yaşam süresi rassal değişkeni tanımlı gereği pozitif ve sürekli bir rassal değişkendir. Yaşama sürelerinin analizinde kullanılan teknikler genellikle parametrik ve parametrik olmayan yaklaşımlar olmak üzere iki gruba ayrılmaktadır. Parametrik olmayan yaklaşımlarda gözlemlere ait ampirik dağılım fonksiyonları kullanılmaktadır. Bu konuda yapılan çalışmalar özellikle son yüzyılın ikinci çeyreğinde kendini göstermektedir. 1958 yılında, yaşam fonksiyonlarının tahmini için çağrım-limit yöntemi Kaplan ve Meier (1958) tarafından geliştirilmiştir. Cox (1972) Oransal Hazard Modelini geliştirmiştir ve Miller (1981) tarafından yaşam analizi için parametrik olmayan yöntemler uygulanmıştır. Yaşam analizi çalışmalarında kullanılan parametrik yaklaşımlarda en önemli aşama, yaşam süresi verisini en uygun dağılım ile modellemektir. Yaşam süresi verilerinin modellenmesinde Üstel, Gamma, Lognormal, Weibull dağılımı gibi özellik gösteren standart olasılık dağılımları yaygın olarak kullanılmaktadır.

Yaşam analizinde ele alınan yaşam süresi verisinin heterojen veya çok modlu olması durumunda standart olasılık dağılımları veriyi modellemede yetersiz kalmaktadır. Son dönemlerde bu tür verilerin modellenmesinde karma dağılım modelleri yaygın olarak kullanılmaya başlanmıştır. Karma dağılım modelleri farklı karakteristik özellikler taşıyan dağılımların bir arada kullanılmasından dolayı veri modellemede standart olasılık dağılımlarına göre daha esnek ve daha kullanışlıdır.

Son dönemde yaşam sürelerinin analizi için tanımlanan çok sayıda yeni karıştırılmış dağılım bulunmaktadır. Dağılımların birleştirilmesi yöntemlerinden biri olan karıştırma işleminde, genelde basit bir forma sahip olan bir dağılımın parametrelerinden bir veya birkaçı rassal değişken olarak düşünülerek farklı bir dağılıma sahip olduğu varsayılmaktadır. Bu yöntem ile basit bir forma sahip olan dağılımlardan daha kullanışlı ve daha esnek dağılımlar elde edilebilir.

Heterojen özellik taşıyan yaşam süresi verilerinin modellenmesi için karma dağılım modellerinin incelendiği bu çalışmada aynı dağılımın farklı parametrelere

sahip bileşenleri ile oluşturulan karma dağılım modelleri yanında farklı iki dağılımın karmasından oluşan karma dağılım modelleri ve karıştırılmış dağılımların sonlu karma modelleri tanımlanmıştır.

Bölüm 3’de yaşam analizi ile ilgili temel tanım ve gösterimlere yer verilmiştir. Yaşam analizinde önemli olan üç fonksiyon olasılık yoğunluk fonksiyonu, yaşam fonksiyonu ve hazard fonksiyonu tanımlanarak bu fonksiyonlar arasındaki matematiksel ilişkiler gösterilmiştir. Ayrıca yaşam sürelerinin modellenmesinde yaygın bir kullanıma sahip olan Üstel, Weibull, Gamma ve Lognormal dağılımlarına ait karakteristik özellikler incelenmiştir.

Bölüm 4’de karma dağılım modellerinin oluşturulması, karma dağılım modellerinin karakteristik özellikleri, karma dağılım modellerinde uygun bileşen sayısının seçimi, karma dağılım modelinde tamamlanmamış veri yapısı ve parametre tahmini incelenmiştir. Bölüm 4’de ayrıca karıştırılmış dağılımlarla ilgili temel tanım ve gösterimler verildikten sonra bazı önemli karıştırılmış dağılımların elde edilmesi gösterilmiştir.

Bölüm 5’de heterojen yaşam sürelerinin modellenmesi için karma Gamma dağılım modeli, karma Lognormal dağılım modeli ve karma Weibull dağılım modelleri incelenmiştir. İlgilenilen karma dağılım modellerinde parametrelerin en çok olabilirlik tahmin edicilerinin elde edilmesi için gerekli EM algoritmaları verilmiştir. Algoritmaların parametre tahminindeki etkinliği simülasyon çalışması ile ortaya konmuştur. Simülasyon çalışmasında farklı karma oranlarında ve farklı karakteristik özellikteki karma dağılım modellerindeki parametrelerin en çok olabilirlik tahminlerinin elde edilmesinde EM algoritmasının oldukça başarılı olduğu görülmüştür. Parametre tahminlerinin standart hatalarının oldukça küçük olması EM algoritmasının sayısal olarak tutarlı olduğu göstermiştir.

Bölüm 5’de önerilen karma dağılım modelleri ile Jiang ve Murthy (1995) tarafından tanımlanan ışınlanmış farelerinin ölüm zamanları verisi ve Dirican (2004) tarafından tanımlanan akciğer kanseri tanısı alan hastalarının tanı konduktan sonraki yaşam sürelerinden oluşan akciğer kanser verisi başarılı bir şekilde modellenmiştir. Bu iki veri setinden ışınlanmış farelerin ölüm zamanları verisindeki bileşen sayısının 2 olduğu bilinmekte iken akciğer kanseri verisinde uygun bileşen sayısı Akaike bilgi

kriteri (AIC) ile belirlenmiştir. AIC değerlerine göre karma Gamma ve karma Weibull dağılım modellerinde uygun bileşen sayısı 3 iken, karma Lognormal dağılım modelinde uygun bileşen sayısı 2 olarak belirlenmiştir. İki gerçek heterojen yaşam süresinin modellenmesinde kullanılan karma dağılım modelleri log olabilirlik, Kolmogorov-Smirnov test istatistiği ve hata kareler ortalaması (HKO) bakımından standart olasılık dağılımları Gamma, Weibull ve Lognormal dağılımları ile karşılaştırılmıştır. Tüm karşılaştırma kriterlerine göre iki yaşam süresi verisinin modellenmesinde karma dağılım modelleri standart olasılık dağılım modellerinden daha başarılı bulunmuştur.

Bölüm 6'da heterojen yaşam sürelerinin modellenmesi için farklı iki dağılımın karmasından oluşan Üstel-Gamma, Üstel-Weibull ve Gamma-Weibull karma dağılım modelleri önerilmiştir. Tanımlanan Üstel-Gamma, Üstel-Weibull ve Gamma-Weibull karma dağılım modellerinin karakteristik özellikleri incelenerek, model parametrelerinin en çok olabilirlik tahminleri için gerekli EM algoritmaları verilmiştir. Bu bölümde de simülasyon çalışması ile farklı iki dağılımın karmasından oluşan karma dağılım modellerinde EM algoritmasının parametre tahminindeki etkinliği gösterilmiştir. Ayrıca Colvert ve Boardman (1976) tarafından tanımlanan diş sulama aygıtlarının yaşam süreleri ve Gusmao ve ark (2008) tarafından tanımlanan kanser verisinin modellenmesinde Üstel-Gamma, Üstel-Weibull ve Gamma-Weibull karma dağılım modelleri, standart olasılık dağılımları Üstel, Gamma ve Weibull dağılımları ile karşılaştırılmıştır. Karşılaştırmada Kolmogorov-Smirnov test istatistiği ve HKO değerleri olmak üzere iki farklı kriter kullanılmıştır. Kullanılan iki kritere göre de heterojen yaşam sürelerinin modellenmesinde önerilen iki farklı dağılımın karmasından oluşan modeller standart olasılık dağılımlarına göre daha başarılı bulunmuştur. Diş sulama aygıtlarının yaşam sürelerinin modellenmesinde Gamma ve Weibull dağılımlarının karmasından oluşan Gamma-Weibull karma dağılımı en başarılı model olarak belirlenmiştir. Kanser tanısı alan hastaların tanı konduktan sonraki yaşam sürelerinin modellenmesinde ise Üstel ile Weibull dağılımlarının karmasından oluşan Üstel-Weibull karma dağılımı en başarılı model olarak belirlenmiştir.

Bölüm 7’de Adamidis ve ark (2005) tarafından önerilen genelleştirilmiş Üstel-Geometrik (EEG) dağılımının karma dağılım modeli oluşturulmuştur. Dönüştürülmüş uç değer dağılımının, üstel dağılımla karıştırılması ile elde edilen genelleştirilmiş Üstel-Geometrik dağılımının karakteristik özellikleri incelendikten sonra genelleştirilmiş Üstel-Geometrik (EEG) dağılımının karma dağılım modeli oluşturulmuştur. Bu yeni karma dağılım modeli karıştırılmış dağılımların karma modeli olarak tanımlanmaktadır. Karma EEG dağılımının parametrelerinin en çok olabilirlik tahminleri için EM algoritmasının verildiği çalışmada, algoritmanın etkinliği simülasyon çalışması ile gösterilmiştir. Gerçek heterojen yaşam süresi verilerinin modellenmesinde karma EEG dağılım modelinin, tek bileşenli EEG ve Weibull dağılımları ile karma Weibull dağılım modeli karşısındaki performansı Ling ve Pan (1998) tarafından tanımlanan yaşam yorgunluğu verisi ve Colvert ve Boardman (1976) tarafından tanımlanan diş sulama aygıtlarının yaşam sürelerinin modellenmesinde gösterilmiştir. Karşılaştırmada Akaike bilgi kriteri, Kolmogorov-Smirnov test istatistiği ve hata kereler ortalaması değerleri kullanılmıştır. Karşılaştırma sonrası karma EEG dağılım modelinin, tek bileşenli EEG ve Weibull dağılımlarına göre veri modellemede daha başarılı olduğu görülmüştür. Diş sulama aygıtlarının yaşam sürelerinin modellenmesinde tüm karşılaştırma kriterlerine göre karma EEG dağılım modelinin, karma Weibull dağılım modelinden daha iyi olduğu görülmüştür. Yaşam yorgunluğu verisinin modellenmesinde karma EEG dağılım modeli ile karma Weibull dağılım modeli benzer sonuçlar vermiştir. Hatta HKO kriterine göre karma EEG dağılım modelinin karma Weibull dağılım modelinden daha iyi olduğu görülmüştür. Sonuç olarak önerilen karma EEG dağılım modelinin, heterojen yaşam verilerinin modellenmesinde yaygın bir kullanıma sahip olan karma Weibull dağılımına güçlü bir alternatif olduğu görülmüştür.

Standart olasılık dağılımlarına göre veri modellemede daha esnek ve daha kullanışlı olan karma dağılım modellerinin yaşam sürelerinin modellenmesindeki kullanımının detaylı olarak gösterildiği çalışma sonucunda heterojen yaşam süresi verilerinin modellenmesinde karma dağılım modellerinin üstünlükleri gösterilmiştir. Karıştırılmış dağılımların karma modellerinin de heterojen yaşam süresi verilerinin

modellenmesinde kullanılabileceğinin gösterildiği çalışmada tam veri durumu ile ilgilenilmiştir.

KAYNAKLAR

- AALEN, O., 1988. Heterogeneity in Survival Analysis, *Statistics in Medicine*, 7:1121-1137.
- AALEN, O., 1992. Modelling Heterogeneity in Survival Analysis by the Compound Poisson Distribution. *the Annals of Applied Probability*, 2(4):951-972.
- ADAMIDIS, K., DIMITRAKOPOULOU, T., and LOUKAS, S., 2005. On an Extension of the Exponential Geometric Distribution. *Statistics & Probability Letters*, 73:259-269.
- ADAMIDIS, K., and LOUKAS, S., 1998. A Lifetime Distribution with Decreasing Failure Rate. *Statist. Probab. Lett.*, 39:35-42.
- AITCHISON, J., and BROWN J.A.C., 1957. *The Lognormal Distribution*. Cambridge University Press, Cambridge, 176p.
- AKAIKE, H., 1974. A New Look at the Statistical Model Identification. *IEEE Transactions on Automatic Control*, 19:716-723.
- AL-HUSSAINI, E.K., AL-DAYIAN, G.R., and ADHAM, S.A., 2000. On Finite Mixture of Two-Component Gompertz Lifetime Model. *J. Statist. Comput. Simul.*, 67:1-20.
- AMBAGASPITIYA, R.S., 1995. Manipulating Lagrangian Distributions and Associated Compound Distributions with Maple. *Actuarial Research Clearing House*, 1:357-370.
- AMBAGASPITIYA, R.S., 1998. Compound Bivariate Lagrangian Poisson Distributions. *Insurance: Mathematics and Economics*, 23:21-31.
- ANGELIS, DE, R., CAPOCACCIA, R., HAKULINEN, T., SODERMAN, B., and VERDECCHIA, A., 1999. Mixture Models for Cancer Survival Analysis: Application to Population-Based Data with Covariates. *Statistics in Medicine*, 18:441-454.
- ANTZOULAKOS, D.L., and CHADJICONSTANTINIDIS, S., 2004. On Mixed and Compound Mixed Poisson Distributions. *Scandinavian Actuarial Journal*, 3:161-188.

- BALAKRISHNAN, N., GUPTA, R.C., KUNDU, D., LEIVA, V., and SANHUEZA, A., 2011. On Some Mixture Models Based on the Birnbaum–Saunders Distribution and Associated Inference. *Journal of Statistical Planning and Inference*, 141:2175-2190.
- BARGER, K.J.A., 2006. Mixtures of Exponential Distributions to Describe the Distribution of Poisson Means in Estimating the Number of Unobserved Classes. Unpublished Masters Thesis, the Faculty of the Graduate School of Cornell University, 96p.
- BAŞAR, E., 1993. Yaşam Tabloları Analizinde Kullanılan Bazı İstatistiksel Tekniklerin Böbrek Nakli Verilerine Uygulaması. Hacettepe Üniversitesi Fen Bilimleri Enstitüsü Yayınlanmamış Doktora Tezi, Ankara, 132s.
- BEBBINGTON, M., LAI, C.D., and ZITIKIS, R., 2007. Modeling Human Mortality using Mixtures of Bathtub Shaped Failure Distributions. *Journal of Theoretical Biology*, 245:528-538.
- BERKSON, J., 1942. The Calculation of Survival Rates, in *Carcinoma and Other Malignant Lesions of the Stomach*. edited by W. Walters, H. K. Gray, and J. T. Priestley. W.B. Saunders, Philadelphia, 467-484.
- BERRETONI, J.N., 1964. Practical Applications of the Weibull Distribution. *Industrial Quality Control*, 21:71-79.
- BHATTACHARYA, C.T., 1967. A Simple Method of Resolution of a Distribution into Gaussian Components. *Biometrics*, 23:115-137.
- BIRNBAUM, Z.W., and SAUNDERS, S.C., 1958. A Statistical Model for Life-Length of Materials. *J. Amer. Stat. Assoc.*, 53(281):151-160.
- BLISCHKE, W.R., 1964. Estimating the Parameters of Mixtures of Binomial Distributions. *J. Amer. Statist. Assoc.*, 59:510-528.
- BOAG, J.W., 1949. Maximum Likelihood Estimates of Proportion of Patients Cured by Cancer Therapy. *Journal of the Royal Statistical Society Series B*, 11:15-44.

- BOGARDUS, C., LILLIOJA, S., NYOMBA, B.L., ZURLO, F., SWINBURN, B., PUENTE, A.E., KNOWLER, W.C., RAVUSSIN, E., MOTT, D.M., and BENNET, P.H., 1989. Distribution of in Vivo Insulin Action in Pima Indians as Mixture of Three Normal Distributions. *Diabetes*, 38(11):1423-1432.
- BOROVKOV, A.A., 1984. *Asymptotic Methods in Queueing Theory*. Wiley, NewYork, 292p.
- BROWN, G.W., and FLOOD, M.M., 1947. Tumbler Mortality. *Journal of the American Statistical Association*, 42:562-574.
- CASSIE, R.M., 1954. Some Uses of Probability Paper in the Analysis of Size of Frequency Distributions. *Austral. J. Mar. Fish.. Freshw. Res.*, 5:513-522.
- CEKANAVICIUS, V., and ROOS, B., 2004. Two-Parametric Compound Binomial Approximations. *Lithuanian Mathematical Journal*, 44:354-373.
- CELEUX, G., and SOROMENHO, G., 1996. An Entrophy Criterion for Assessing the Number of Cluster in A Mixture Model. *Classification Journal* 13:195-212.
- CHEN, W.C., HILL, B.M., GREENHOUSE, J.B., and FAYOS, J.V., 1985. Bayesian Analysis of Survival Curves for Cancer Patients Following Treatment. *Bayesian Statistics*, 2:299-328.
- CHOI, C., and WETTE, R., 1969. Maximum Likelihood Estimation of the Parameters of the Gamma Distribution and Their Bias. *Technometrics*, 11(4):683-690.
- CLARK, V. A., CHAPMAN, J. M., COULSON, A. H., and HASSELBLAD, V. 1968. Dividing the Blood Pressure from the Los Angeles Heart Study into Two Normal Distributions. *Johns Hopkins. Med. J.*, 122:77-83.
- COHEN, A.C., 1963. Progressively Censored Sample in Life Testing. *Technometrics*, 5:327-339.
- COHEN, A.C., 1965. Maximum Likelihood Estimation in the Weibull Distribution based on Complete on Censored Samples. *Technometrics*, 7:579-588.
- COHEN, A.C., 1967. Estimation in Mixtures of Two Normal Distributions. *Technometrics*, 9:15-28.

- COLVERT, R.E., and BOARDMAN, T.J., 1976. Estimation in the Piece-Wise Constant Hazard Rate Model. *Commun. Statist.-Theor. Math.*, 11:1013-1029.
- COX, D.R., 1972. Regression Models and Life Tables. *Journal of the Royal Statistical Society*, 34:187-220.
- COX, D.R., and OAKES, D., 1984. *Analysis of Survival Data*. Chapman and Hall, London, 201p.
- CRAIGMILE, P.F., and TITTERINGTON, D.M., 1997. Parameter Estimation for Finite Mixtures of Uniform Distributions. *Communications in Statistics-Theory and Methods*, 26(8):1981-1995.
- CRAN, G.W., 1976. Graphical Estimation Methods for Weibull Distributions. *Microelectronics & Reliability*, 15:47-52.
- ÇALIŞ, N., 2005. Karma Dağılım Modellerinde Bileşen Sayısını Tahmin Etmek İçin Yöntemler. Çukurova Üniversitesi Fen Bilimleri Enstitüsü Yayınlanmamış Yüksek Lisans Tezi, Adana, 107s.
- DAVIS, D.J., 1952. An Analysis of Some Failure Data. *J.Amer. Stat. Assoc.*, 47:113-150.
- DAY, N.E., 1969. Estimating the Components of a Mixture of Normal Distributions. *Biometrika*, 56:463-474.
- DEMPSTER, A.P., LAIRD, N.M., and RUBIN, D.B., 1977. Maximum-Likelihood from Incomplete Data via the EM Algorithm. *J. Roy. Statist. Soc. B*, 39:1-38.
- DİRİCAN, A., 2004. Kliniğimizde akciğer kanseri tanısı alan hastaların prospektif olarak değerlendirilmesi ve sağkalıma etki eden faktörlerin belirlenmesi, Ondokuz Mayıs Üniversitesi Tıpta Uzmanlık Tezi, Samsun, 122s.
- DOUGLAS, J.B., 1980. *Analysis with Standard Contagious Distribution*. Statistical Distributions in Scientific Work Vol.4, International Co-operative Publishing House, Fairland, Maryland, USA, 520p.
- EPSTEIN, B., 1958. Exponential Distribution and its Role in Life Testing. *Indian Quality Control*, 15:4-6.
- EPSTEIN, B., and SOBEL, M., 1953. Life Testing. *J. Amer. Stat. Assoc.*, 48(263):486-502.

- ERİŞOĞLU, Ü., ERİŞOĞLU, M., and EROL, H., 2011. A Mixture Model of Two Different Distributions Approach to the Analysis of Heterogeneous Survival Data. *International Journal of Computational and Mathematical Sciences*, 5(2):75-79.
- ERİŞOĞLU, Ü., and EROL, H., 2010. Modeling Heterogeneous Survival Data Using Mixture of Extended Exponential-Geometric Distributions. *Communications in Statistics-Simulation and Computation*, 39(10):1939-1952.
- EVERITT, B.S., and HAND, D.J., 1981. *Finite Mixture Distributions*. Chapman and Hall, London, 143p.
- FALLS, L.W., 1970. Estimation of Parameters in Compound Weibull Distributions. *Technometrics*, 12(2):399-407.
- FAREWELL, V.T., and SPROTT, D.A., 1988. The Use of a Mixture Model in the Analysis of Count. *Biometrics*, 44(4):1191-1194.
- FELLER, W., 1943. On a General Class of "Contagious" Distributions. *The Annals of Mathematical Statistics*, 14(4):389-400.
- FISHER, R.A., 1912. *Messenger of Mathematics*. 41:155-160.
- FISHER, R.A., and TIPPETT, L.H.C., 1928. Limiting Forms of the Frequency Distribution of the Largest or Smallest Member of a Sample. *Proceedings of the Cambridge Philosophical Society*, 24:180-190.
- FOWLKES, E.B., 1979. Some Methods for Studying the Mixture of Two Normal(Lognormal) distributions. *J. Amer. Statist. Assoc.* 74:561-575.
- FRALEY, C., and RAFTERY, A.E., 1998. How Many Clusters? Which Clustering Methods? Answers via Model-Based Cluster Analysis. *The Computer Journal*, 41(8):578-588.
- GADDUM, J.H., 1945. *Log Normal Distributions*. Nature, London, 156p.
- GORDON, N.H., 1990. Application of the Theory of Finite Mixtures for the Estimation of Cure Rates of Treated Cancer Patients. *Statistics in Medicine*, 9:397-407.

- GREENWOOD, M., and YULE, G.U., 1920. An Inquiry into the Nature of Frequency Distribution Representative of Multiple Happenings with Particular Reference to the Occurrence of Multiple Attacks of Disease or of Repeated Accidents. *J. Roy. Stat. Soc.*, 83:255-279.
- GRUBBSTRÖM, R.W., and TANG, O., 2006. The Moments and Central Moments of a Compound Distribution. *European Journal of Operational Research*, 170:106-119.
- GUENTHER, W.C., 1975. The Inverse Hypergeometric a Useful Model. *Statistica Neerlandica*, 29(4):124-144.
- GUPTA, A.K., and MIYAWAKI, T., 1978. On a Uniform Mixture Model, *Biom. J.*, 20:631-637.
- GURLAND, J., 1957. Some Interrelations among Compound and Generalized Distributions. *Biometrika*, 44:263-268.
- GUSMÃO, F.R.S., JÚNIOR, E.L.S., SANTOS, E.S., and ORTEGA, E.M.M., 2008. Distribuição Weibull Inversa: Uma Aplicação a Dados Médicos com Censuras, 8. Encontro Regional de Matemática Aplicada e Computacional, 1-4.
- HARDING, J.P., 1949. The Use of Probability Paper for the Graphical Analysis of Polymodel Frequency Distributions. *J. Marine Biol. Assoc.*, 28:141-153.
- HASSELBLAD, V., 1966. Estimation of Parameters of a Mixture of Normal Distributions. *Technometrics*, 8:431-444.
- HASSELBLAD, V., 1969. Estimation of Finite Mixture of Distributions from the Exponential Family. *J. Amer. Statist. Assoc.*, 64:1459-1471.
- HOUGAARD, P., 1984. Life Table Methods for Heterogeneous Populations: Distributions Describing the Heterogeneity. *Biometrika*, 71:75-83.
- <http://apps.isiknowledge.com> (Erişim Tarihi: 09 Eylül 2010)
- JENSEN, F., and PETERSEN, N.E., 1982. Bum-In, An Engineering Approach to the Design and Analysis of Bum-In Procedures. John Wiley & Sons, 506p.
- JIANG, R., ZUO, M.J., and LI, H.-X., 1999. Weibull and Inverse Weibull Mixture Models Allowing Negative Weights. *Reliability Engineering and System Safety*, 66:227-234.

- JIANG, S., and KECECIOGLU, D., 1992. Graphical Representation of Two Mixed-Weibull Distributions. *IEEE Transactions on Reliability*, 41(2):241-247.
- JIANG, R., and MURTHY, D.N.P., 1995. Modeling Failure-Data by Mixture of 2 Weibull Distributions: A Graphical Approach. *IEEE Transactions on Reliability*, 44:477-488.
- JIANG, R., and MURTHY, D.N.P., 1997. Two Sectional Models Involving Three Weibull Distributions. *Quality and Reliability Engineering International*, 13:83-96.
- JOHN, S., 1970a. On Identifying the Population of Origin of Each Observation in a Mixture of Observations from Two Normal Populations. *Technometrics*, 12:553-563.
- JOHN, S., 1970b. On Identifying the Population of Origin of Each Observation in a Mixture of Observations from Two Gamma Populations. *Technometrics*, 12:565-568.
- JOHNSON, N.L., KOTZ, S., and KEMP, A.W., 1992. *Univariate Discrete Distributions*. John Wiley & Sons, New York, 646p.
- JOHNSON, N.L., KOTZ, S., and BALAKRISHNAN, N., 1995. *Continuous Univariate Distributions Volume 1*, John Wiley & Sons, New York, 756p.
- JOHNSON, N.L., KOTZ, S., and BALAKRISHNAN, N., 1995. *Continuous Univariate Distributions Volume 2*, John Wiley & Sons, New York, 717p.
- KABIR, A.B.M.L., 1968. Estimation of Parametrers of a Finite Mixture of Distributions. *J.R. Statist. Soc. B.*, 30:472-482.
- KAYLAN, A.R., and HARRIS, C.M., 1981. Efficient Algorithms to Derive Maximum Likelihood Estimates for Finite Exponential and Weibull Mixtures. *Computers and Operations Research*, 8(2): 97-104.
- KAO, J.H.K., 1958. Computer Methods for Estimating Weibull Parameters in Reliability Studies. *IRE Trans. Rel. and Quality Ctrl.*, PGRQC-13:15-22.
- KAO, J.H.K., 1959. A Graphical Estimation of Mixed Weibull Parameters in Life Testing of Electron Tubes. *Technometrics*, 1:389-407.
- KAPLAN, E.L., and MEIER, P., 1958. Nonparametric Estimation from in Complete Observations. *Journal of the American Statistical Association*, 53:457-481.

- KARLIS, D., and XEKALAKI, E., 2005. Mixed Poisson Distributions. *International Statistical Review*, 73(1):35-58.
- KECECIOGLU, D.B., and WANG, W., 1998. Parameter Estimation for Mixed Weibull Distribution. *Proc. 44th Annual Reliability and Maintainability Symposium*, Anaheim, 247-252.
- KLEIN, J.P., and MOESCHBERGER, M.L., 1997. *Survival Analysis: Techniques for Censored and Truncated Data*, Springer, New York, 536p.
- KLEINBANM, D.G., and KLEIN, M., 2005. *Survival Analysis: A Self-Learning Text. Second Edition*, Springer, USA, 590p.
- KUMAR, K.D., NICKLIN, E.H., and PAULSON, A.S., 1979. Comment on a Paper by Quandt and Ramsey. *J. Amer. Statist. Assoc.*, 74:52-55.
- KUŞ, C., 2007. A New Lifetime Distribution. *Computational Statistics and Data Analysis* 51:4497-4509.
- LAWLESS, J.F., 2003. *Statistics Models and Methods for Lifetime Data. Second Edition*, John Wiley & Sons, New Jersey, 630p.
- LEE, E.T., and WANG, J.W., 2003. *Statistical Methods For Survival Data Analysis. Third Edition*, John Wiley & Sons, New York, 513p.
- LEEMIS, M.L., 1986. Lifetime Distribution Identites. *IEEE Transactions on Reliability*, 35:170-174.
- LI, Y., 2003. *Some Aging Properties of Mixture Distributions*. University of Pittsburg, Unpublished Phd. Thesis, 154p.
- LIEBLEIN, J., and ZELEN, M., 1956. Statistical Investigation of the Fatigue Life of Deep-Groove Ball Bearings. *Journal of Research, National Bureau of Standards*, 57:273-316.
- LING, J., and PAN, J., 1998. A New Method for Selection of Population Distribution and Parameter Estimation. *Reliability Engineering and System Safety*, 60:247-255.
- LIU, N., 2003. *Statistical Properties of Maximum Likelihood Estimators for Complete and Censored Data from Mixed Exponential and Weibull Distributions*. Graduate School of Arts and Sciences of Washington University, Unpublished Phd. Thesis, 104p.

- LIU, D., WEN, S., and WANG, L., 2002. Poisson-Gumbel Mixed Compound Distribution and its Application. Chinese Science Bulletin, 47:1901-1906.
- LIU, D., WANG, L., and PANG, L., 2006. Theory of Multivariate Compound Extreme Value Distribution and its Application to Extreme Sea State Prediction. Chinese Science Bulletin, 5:2926-2930.
- LONDON, D., 1988. Survival Models and Their Estimation. Actex Publication, Winsted, 326p.
- MACHIN, D., CHEUNG, Y.B., and PARMAR, M.K., 2006. Survival Analysis: A Practical Approach. Second Edition, John Wiley & Sons, England, 266p.
- MARIN, J.M., RODRIGUEZ-BERNAL, M.T., and WIPER, M.P., 2005. Using Weibull Mixture Distributions to Model Heterogeneous Survival Data. Communications in Statistics-Simulation and Computation, 34:673-684.
- MARKS, R.G., and RAO, P.V., 1979. An Estimation Procedure for Data Containing Outliers with a one-Directional Shift in the Mean. J. Amer. Statist. Assoc., 74:614-620.
- MASUYAMA, M., 1977. A Mixture of Two Gamma Distributions Applied To Rheumatoid Arthritis. Rep. Stat. Appl. Res., JUSE, 24(2):28/82-31/85.
- MCALISTER, D., 1879. The Law of the Geometric Mean. Proceedings of the Royal Society, 29:367-376.
- MCLACHLAN, G.J, and KRISHNAN, T., 1997. The EM Algorithm and Extensions. Wiley, New York, 274p.
- MCLACHLAN, G.J., and PEEL, D., 2001. Finite Mixture Model, Wiley, NewYork, 419p.
- MEEKER, W.Q., and LUVALLE, M.J., 1995. An Accelerated Life Test Model Based on Reliability Kinetics. Technometrics, 37(2):123-127.
- MILLER, R.G., 1981. Survival Analysis. John Wiley & Sons, New York, 238p.
- MURAT, M., and SZYNAL, D., 2000. On Computational Formulas for Densities and Moments of Compound Distributions. Journal Mathematical Sciences, 99:1286-1299.
- MURTHY, D.N.P., XIE, M., and JIANG, R., 2004. Weibull Models. John Wiley & Sons, New Jersey, 383p.

- NELSON, W., 1982. *Applied Life Data Analysis*. John Wiley & Sons, Canada, 634p.
- ORMONEIT, D., and TRESP, V., 1998. Averaging, Maximum Penalized Likelihood and Bayesian Estimation for Improving Gaussian Mixture Probability Density Estimates. *IEEE Transactions on Neural Networks*, 9(4):639-650.
- PATRA, K., 2000. *Innovative Approaches to Reliability and Survival Analysis*. University of Connecticut, Unpublished Phd. Thesis, 115p.
- PAVLIC, M., BRAND, R.J., and CUMMINGS, S.R., 2001. Estimating Probability of Nonresponse to Treatment Using Mixture Distributions. *Statistics in Medicine*, 20:1739-1753.
- PEARSON, K., 1894. Contribution to the Mathematical Theory of Evolution. *Phil. Trans. Roy. Soc. A.*, 185:71-110.
- PLAIT, A., 1962. The Weibull Distribution. *Ind Qual. Contr.*, 19:17-27.
- POLLARD, H.S., 1934. On the Relative Stability of the Median and the Arithmetic Mean, with Particular Reference to Certain Frequency Distributions which can be Dissected into Normal Distributions. *Ann. Math. Stat.*, 5:227-262.
- QUANDT, R.E., and RAMSEY, J.B., 1978. Estimating Mixture of Normal Distributions and Switching Regressions. *J. Amer. Statist. Assoc.*, 73:730-738.
- QIAN, J., 1994. *A Bayesian Weibull Survival Model*. Institute of Statistical and Decision Sciences, Duke University: North Carolina, Unpublished Ph.D. Thesis, 119p.
- RAO, C.R., 1948. The Utilization of Multiple Measurements in Problems of Biological Classification. *J. R. Statist. Soc. B.*, 10:159-203.
- RAPER, L.P., BALKAU, B., TAYLOR, R., MILNER, B., COLLINS, V., and ZIMMET, P., 1983. Plasma Glucose Distributions in Two Pacific Populations: The Bimodality Phenomenon. *Tohok. J. Exp. Med.*, 141:199-206.
- RAZALI, A.M., and SALIH, A.A., 2009. Combining Two Weibull Distributions Using a Mixing Parameter. *European Journal of Scientific Research*, 31(2):296-305.

- REDNER, R.A., and WALKER, H.F., 1984. Mixture Densities, Maximum Likelihood, and the EM Algorithm. *SIAM Rev.*, 26:195-239.
- RIDER, P.R., 1961a. The Method of Moments Applied to a Mixture of Two Exponential Distributions. *Ann. Math. Statist.*, 32:143-147.
- RIDER, P.R., 1961b. Estimating the Parameters of Mixed Poisson, Binomial, and Weibull Distributions by the Method of Moments. *Bull. Inst. Internat. Stat.* (Part 2), 33:225-232.
- RIDOLFI, A., and IDIER, J., 2002. Penalized Maximum Likelihood Estimation for Normal Mixture Distributions. Technical Report, 200285, 1-19.
- ROEDER, K., 1990. Density Estimation with Confidence Sets Exemplified by Superclusters and Voids in the Galaxies. *Journal of the American Statistical Association*, 85:617-624.
- ROSEN, P., and RAMMLER, B., 1933. The Laws Governing the Fineness of Powdered Coal. *Journal of the Institute of Fuels*, 6:246-249.
- RUFO, M.J., PEREZ, C.J., and MARTIN, J., 2009. Local Parametric Sensitivity for Mixture Models of Lifetime Distributions. *Reliability Engineering and System Safety*, 94:1238-1244.
- SATTERTHWAITE, F.E., 1942. Generalized Poisson Distribution. *The Annals of Mathematical Statistics*, 13(4):410-417.
- SAVANUR, S.J., CHADHA, V.K., and JAGANNATHA, P.S., 2002. Mixture Model for Analysis of Tuberculin Surveys. *Ind. J. Tub.*, 49:147-152.
- SERVİ, T., 2009. Çok Değişkenli Karma Dağılım Modeline Dayalı Kümeleme Analizi. Yayınlanmamış Doktora Tezi. Çukurova Üniversitesi Fen Bilimleri Enstitüsü İstatistik Anabilim Dalı, Adana. 239s.
- SHAWKY, A.I., and BAKOBAN, R.A., 2009. On Finite Mixture of two-Component Exponentiated Gamma Distribution. *Journal of Applied Sciences Research*, 5(10):1351-1369.
- SHEPHERD, B.K., SCHAALE, G.B., SMITH, M.J., MRRAY, B.K., and O'NEILL, K.L. 2000. Statistical Analysis of Commet Assay Using a Mixture of Gamma Distributions. *Quantitative Microbiology*, 2:69-79.

- SOROMENHO, G., 1993. Comparing Approaches for Testing the Number of Components in a Finite Mixture Model. *Computational Statistics*, 9:65-78.
- STEPHEN, T.S., 1993. Estimation Methods for Finite Mixture Distributions. Carleton University, Unpublished Masters Thesis, 171p.
- SUCHINDRAN, C.M., and LACHENBRUCH, P.A., 1974. Estimates of Parameters in a Probability Model for First Livebirth Interval. *J. Amer. Statist. Assoc.*, 69:507-513.
- SULTAN, K.S., ISMAIL, M.A., and AL-MOISHEER, A.S., 2007. Mixture of Two Inverse Weibull Distributions: Properties and Estimation. *Computational Statistics & Data Analysis*, 51:5377-5387.
- TAHMASBI, R., and REZAEI, S., 2008. A Two-Parameter Lifetime Distribution with Decreasing Failure Rate. *Computational Statistics and Data Analysis*, 52:3889-3901.
- TALLIS, G.M., and LIGHT, R., 1968. The Use of Fractional Moments for Estimating the Parameters of a Mixed Exponential Distributions. *Technometrics*, 10:161-175.
- THOMAS, E.A.C., 1966. Mathematical Models for the Clustered Firing of Single Cortical Neurones. *Brit. J. Math. Stat. Psychol.*, 19:151-162.
- THOMPSON, J.R., 1968. Some Shrinkage Techniques for Estimating The mean. *Journal of the American Statistical Association*, 63:113-123.
- TITTERINGTON, D.M., SMITH, A.F.M., and MAKOV, U.E., 1985. Statistical Analysis of Finite Mixture Distributions. John Wiley and Sons, Chichester, 243p.
- WONDMAGEGNEHU, E.T., 2002. Mixture of Distributions with Increasing Failure Rates. University of Pittsburgh, Unpublished Pht Thesis, 125p.
- WEIBULL, W., 1951. A Statistical Distribution Function of Wide Applicability. *J. Applied Mech.*, 18:293-297.
- WELDON, W.F.R., 1892. Certain Correlated Variations in Crangon Vulgaris. *Proceedings of the Royal Society of London*, 51:2-21.
- WELDON, W.F.R., 1893. On Certain Correlated Variations in Carcinus Moenas. *Proceedings of the Royal Society of London*, 54:318-329.

- YE, Q., 2006. Survival Analysis Estimation and Testing Assuming a Two Component Exponential Mixture. Stony Brook University, Unpublished Phd. Thesis, 69p.
- ZHANG, Y., 2008. Parametric Mixture Models in Survival Analysis with Applications. The Graduate Board of Temple University, Unpublished Phd Thesis, 85p.

ÖZGEÇMİŞ

1980 yılında Ankara’da doğdu. İlkokulu Van’da, ortaokul ve lise öğrenimini İstanbul’da tamamladı. 1998 yılında Selçuk Üniversitesi Fen Edebiyat Fakültesi İstatistik Bölümünü kazandı. Lisans eğitimimi 2002 yılında tamamlayarak aynı yıl Selçuk Üniversitesi Fen Bilimleri Enstitüsü İstatistik Anabilim Dalında yüksek lisans eğitimine başladı. 2002 yılının sonunda Selçuk Üniversitesi Fen Edebiyat Fakültesi İstatistik Bölümünde araştırma görevlisi olarak çalışmaya başladı. Yüksek lisans eğitimini 2005 yılında bitirdi. 2007 yılında doktora eğitimi için 2547 Sayılı Yükseköğretim Kanunu’nun 35. Maddesi gereği Çukurova Üniversitesi Fen Bilimleri Enstitüsü’ne araştırma görevlisi olarak görevlendirilerek İstatistik Anabilim Dalı’nda doktora eğitimine başladı.