

**YAŞAM SÜREÇLERİNE
ENTROPİ OPTİMİZASYON
YÖNTEMLERİNİN
UYGULAMALARI**

Sevda ÖZDEMİR
Yüksek Lisans Tezi

İstatistik Anabilim Dalı
Haziran - 2014

**Bu tez çalışması Anadolu Üniversitesi Bilimsel Araştırma Projeleri
Komisyonu Başkanlığı tarafından desteklenmiştir. Proje No: 1208F134**

JÜRİ VE ENSTİTÜ ONAYI

Sevda ÖZDEMİR'in "Yaşam Süreçlerine Entropi Optimizasyon Yöntemlerinin Uygulamaları" başlıklı İstatistik Anabilim Dalındaki, Yüksek Lisans Tezi 20.06.2014 tarihinde, aşağıdaki jüri tarafından Anadolu Üniversitesi Lisansüstü Eğitim-Öğretim ve Sınav Yönetmeliğinin ilgili maddeleri uyarınca değerlendirilerek kabul edilmiştir.

	<u>Adı Soyadı</u>	<u>İmza</u>
Üye (Tez Danışmanı) :	Prof. Dr. Aladdin ŞAMİLOV
Üye :	Prof. Dr. Memedağa MEMMEDLİ
Üye :	Prof. Dr. Yalçın KÜÇÜK

Anadolu Üniversitesi Fen Bilimleri Enstitüsü Yönetim Kurulu'nun
..... tarih ve sayılı kararıyla onaylanmıştır.

Enstitü Müdürü



ÖZET

Yüksek Lisans Tezi

YAŞAM SÜREÇLERİNE ENTROPİ OPTİMİZASYON YÖNTEMLERİNİN UYGULAMALARI

Sevda ÖZDEMİR

Anadolu Üniversitesi
Fen Bilimleri Enstitüsü
İstatistik Anabilim Dalı

Danışman: Prof. Dr. Aladdin ŞAMİLOV

2014, 81 sayfa

Bu tezde sağkalım verileri üzerine *MinMaxEnt*, *MaxMaxEnt* ve *MinMinxEnt*, *MaxMinxEnt* (GEOY) analizleri geliştirilmiştir. Ayrıca, bu analizler farklı anlamlarda birbirleri ile karşılaştırılmıştır. Böylece, ele alınmış yöntemlerin her birinin diğerine nazaran, belli anlamda avantaj sağlanmasına rağmen başka bir anlamda diğer yöntemin, bu yöntem üzerinde avantaj sağladığı gözlenmiştir. Geliştirilerin yöntemlerin performansı Ki-Kare, RMSE, Shannon entropi ölçümü ve Kullback-Leibler ölçümü kriterleri yardımıyla belirlenmiştir. Elde edilen $(MaxMaxEnt)_4$, $(MinMaxEnt)_4$, $(MinMinxEnt)_4$ ve $(MaxMinxEnt)_4$ dağılımlarına uygun olarak olasılık yoğunluk fonksiyonu $\hat{f}(t)$, birikimli dağılım fonksiyonu $\hat{F}(t)$, sağkalım fonksiyonu $\hat{S}(t)$ ve risk (hazard) fonksiyonu $\hat{h}(t)$ tahmin edilmiştir. Sonuçlar MATLAB 7.10.0 (R2010a) programı kullanılarak elde edilmiştir. Bu sonuçlar GEOY'nin sağkalım veri analizinde başarılı bir şekilde uygulanabilirliğini göstermiştir.

Anahtar Kelimeler: sağkalım fonksiyonu, sansürlü gözlem, tamamlayıcı olasılık, *MinMaxEnt*, *MaxMaxEnt*, *MinMinxEnt*, *MaxMinxEnt* dağılımları

ABSTRACT

Master of Science Thesis

GENERALIZED ENTROPY OPTIMIZATION METHODS ON LIFETIME DATA

Sevda ÖZDEMİR

Anadolu University

Graduate School of Sciences

Statistics Program

Supervisor: Prof. Dr. Aladdin ŞAMİLOV

2014, 81 pages

In this thesis we have developed *MinMaxEnt*, *MaxMaxEnt* and *MinMinxEnt*, *MaxMinxEnt* (GEOM) analysis on survival data. Moreover, mentioned methods are compared with each other in the difference senses. It is observed that each of considered methods possessing an advantage over the other method in the given sense may be give up one's place to the same method in the another sense. The performances of improved methods are established by Chi – Square criteria, Root Mean Square Error (RMSE) criteria, Shannon entropy measure and Kullback-Leibler measure. According to obtained distributions $(MaxMaxEnt)_4$, $(MinMaxEnt)_4$, $(MinMinxEnt)_4$ and $(MaxMinxEnt)_4$ estimator of Probability Density Function $\hat{f}(t)$, Cumulative Distribution Function $\hat{F}(t)$, Survival Function $\hat{S}(t)$ and Hazard Function $\hat{h}(t)$ are evaluated. The results are acquired by using statistical software MATLAB 7.10.0 (R2010a). The obtained results show that survival data analysis by GEOM acquires a new significance.

Key words: survival function, censored observation, supplementary probability, *MinMaxEnt*, *MaxMaxEnt*, *MinMinxEnt*, *MaxMinxEnt* distributions

TEŞEKKÜR

İstatistik Bölümüne “Ölçüm Teorisi ve Olasılık”, “İstatistiğin Matematiksel Yöntemleri”, “Entropi ve İnfomasyon Teorisi” isimli yüksek lisans ve doktora derslerini açarak bilgilerini ve deneyimlerini bizimle paylaşan, Uzmanlık Alan Dersi ve Seminer dersleri ile değerli vaktini bana ayıran, tez konusunun belirlenmesinde olduğu kadar, çalışmanın planlanması ve sonuçlanmasında da büyük emeği geçen, karşılaştığım güçlüklerle baş etmemi sağlayan, bana her zaman, her konuda destek olan çok sevdiğim, değerli hocam ve tez danışmanım Prof. Dr. Aladdin Şamilov’a yürekten teşekkür ederim.

Yüksek lisans aşamasında ders aldığım hocalarım ve Bölüm Başkanı Prof. Dr. Memmedağa Memmedli’ye, ayrı ayrı konularda bana yol gösteren, usanmadan her sorunuma çözüm arayan sevgili hocam Yard. Doç. Dr. Çiğdem Giriftinoğlu’na ve bana çeşitli konularda yardımcı olan Yard. Doç. Dr. Özer ÖZDEMİR’e teşekkürü borç bilirim.

Yüksek lisans tezime maddi destek sağlayan, güler yüzlerini esirgemeyen Anadolu Üniversitesi Proje Birimi çalışanlarına teşekkür ederim.

Her zaman yanımda olan, maddi ve manevi desteklerini benden hiçbir zaman esirgemeyen, beni yetiştiren, bu günlere emin adımlarla ulaşmamı sağlayan aileme bana emek verdikleri ve güvendikleri için tüm samimiyetimle teşekkür ederim.

Özellikle tezimi yazmamda bana destek olan kız kardeşim Nurbanu Özdemir, arkadaşım Veli Can Çalığışu ve Arş. Gör. Nihal Yılmaz’a en içten teşekkürlerimi sunarım.

Sevda ÖZDEMİR

Haziran – 2014

İÇİNDEKİLER

JÜRİ VE ENSTİTÜ ONAYI

ÖZET.....	i
ABSTRACT.....	ii
TEŞEKKÜR.....	iii
ŞEKİLLER DİZİNİ.....	vii
ÇİZELGELER DİZİNİ.....	ix

1. GİRİŞ.....	1
2. SAĞKALIM ANALİZİ.....	4
2.1. Sağkalım Süresi Değişkeninin Tanımı.....	4
2.2. Temel Kavramlar.....	4
2.2.1. Sağkalım Fonksiyonu $S(t)$	5
2.2.2. Olasılık Yoğunluk Fonksiyonu $f(t)$	5
2.2.3. Risk (Hazard) Fonksiyonu $h(t)$	6
2.2.4. $S(t)$, $f(t)$ ve $h(t)$ Fonksiyonları Arasındaki İlişkiler.....	7
2.3. Sansürlü (Durdurulmuş) Veriler.....	9
2.3.1. Planlanmış Sansür.....	9
2.3.1.1. 1. Tip Sansür (Time Censoring).....	10
2.3.1.2. 2. Tip Sansür (Order Censoring).....	11
2.3.2. Planlanmamış Sansür (Random Censoring).....	12
2.3.2.1. Sağdan Sansür.....	13
2.3.2.2. Soldan Sansür.....	14
3. GENELLEŞTİRİLMİŞ ENTROPİ OPTİMİZASYON YÖNTEMLERİ.....	15
3.1. Maksimum Entropi Yöntemi.....	16

3.1.1. Kesikli Rassal Değişkenler İçin Maksimum Entropi Yöntemi.....	16
3.1.2. Sürekli Rassal Değişkenler İçin Maksimum Entropi Yöntemi.....	21
3.2. Minimum Çapraz Entropi Yöntemi.....	23
3.2.1. Kesikli Rassal Değişkenler İçin Minimum Çapraz Entropi..... Yöntemi.....	23
3.2.2. Sürekli Rassal Değişkenler İçin Minimum Çapraz Entropi..... Yöntemi.....	25
4. SONLU SAYIDA KARAKTERİZE EDİCİ MOMENTLER YARDIMI İLE.. <i>MinMaxEnt</i> ve <i>MaxMaxEnt</i> DAĞILIMLARI	28
4.1. <i>MinMaxEnt</i> ve <i>MaxMaxEnt</i> Dağılımlarının Tanımı.....	28
4.2. <i>MinMaxEnt</i> ve <i>MaxMaxEnt</i> Yöntemleri ile Bağırsak Kanseri..... Verileri Analizi	29
4.3. <i>MinMaxEnt</i> ve <i>MaxMaxEnt</i> Yöntemleri ile Bağırsak..... Kanseri Verileri için Yaşam Süresi Rassal Değişkenlerinin Karakterize... Edici Fonksiyonları	38
4.4. <i>MinMinxEnt</i> ve <i>MaxMinxEnt</i> Yöntemleri ile Motor Arızası Verileri... Analizi.....	42
5. SONLU SAYIDA KARAKTERİZE EDİCİ MOMENTLER YARDIMI..... İLE <i>MinMinxEnt</i> ve <i>MaxMinxEnt</i> DAĞILIMLARI.....	50
5.1. <i>MinMinxEnt</i> ve <i>MaxMinxEnt</i> Dağılımlarının Tanımı	50
5.2. <i>MinMinxEnt</i> ve <i>MaxMinxEnt</i> Yöntemleri ile Bağırsak Kanseri..... Verileri Analizi	51
5.3. <i>MinMinxEnt</i> ve <i>MaxMinxEnt</i> Yöntemleri ile Bağırsak Kanseri..... Verileri için Yaşam Süresi Rassal Değişkenlerinin Karakterize..... Edici Fonksiyonları	61

6. <i>MinMaxEnt</i> , <i>MaxMaxEnt</i> VE <i>MinMinxEnt</i> , <i>MaxMinxEnt</i>	
YÖNTEMLERİNİN KARŞILAŞTIRILMASI	65
7. SONUÇLAR VE ÖNERİLER	67
KAYNAKLAR	70
EKLER.....	73
EK-1. Bağırsak Kanseri Verileri Üzerine Uygulanan ve Sansürlü Verileri.....	
Dikkate Alan <i>MinMaxEnt</i> Entropi Optimizasyon Probleminin.....	
Oluşturulması ve Çözümü için Hazırlanan Programın İki Kısıta Göre.....	
MATLAB Kodu	73
EK-2. Bağırsak Kanseri Verileri Üzerine Uygulanan ve Sansürlü Verileri.....	
Dikkate Alan <i>MaxMinxEnt</i> Entropi Optimizasyon Probleminin.....	
Oluşturulması ve Çözümü için Hazırlanan Programın İki Kısıta Göre.....	
MATLAB Kodu	77

ŞEKİLLER DİZİNİ

2.1. I. Tip Sansür'e Bir Örnek.....	11
2.2. II. Tip Sansür'e Bir Örnek	12
2.3. Sağdan Sansür'e Bir Örnek.....	13
2.4. Soldan Sansür'e Bir Örnek	14
4.1. Bağırsak kanseri verileri için (a) $(MinMaxEnt)_1$ Dağılımının Grafiği..... ve (b) $(MaxMaxEnt)_1$ Dağılımının Grafiği	31
4.2. Bağırsak kanseri verileri için (a) $(MinMaxEnt)_2$ Dağılımının grafiği..... ve (b) $(MaxMaxEnt)_2$ Dağılımının grafiği.....	32
4.3. Bağırsak kanseri verileri için (a) $(MinMaxEnt)_3$ Dağılımının grafiği..... ve (b) $(MaxMaxEnt)_3$ Dağılımının grafiği.....	34
4.4. Bağırsak kanseri verileri için (a) $(MinMaxEnt)_4$ Dağılımının grafiği..... ve (b) $(MaxMaxEnt)_4$ Dağılımının grafiği.....	35
4.5. Bağırsak kanseri verileri için $(MinMaxEnt)_4$ Dağılımına uygun..... (a) $\hat{f}(t)$ Fonksiyonunun Grafiği (b) $\hat{F}(t)$ Fonksiyonunun Grafiği,..... (b) $\hat{S}(t)$ Fonksiyonunun Grafiği ve (c) $\hat{h}(t)$ Fonksiyonunun Grafiği	40
4.6. Bağırsak kanseri verileri için $(MaxMaxEnt)_4$ Dağılımına uygun..... (a) $\hat{f}(t)$ Fonksiyonunun Grafiği (b) $\hat{F}(t)$ Fonksiyonunun Grafiği,..... (c) $\hat{S}(t)$ Fonksiyonunun Grafiği ve (d) $\hat{h}(t)$ Fonksiyonunun Grafiği	41
4.7. Motor arızası verileri için $(MinMaxEnt)_3$ Dağılımının (a) Olasılık..... Yoğunluk Fonksiyonu $\hat{f}(t)$, (b) Birikimli Dağılım Fonksiyonu $\hat{F}(t)$,..... (c) Sağkalım Fonksiyonu $\hat{S}(t)$ ve (d) Hazard Fonksiyonu $\hat{h}(t)$	47
4.8. Motor arızası verileri için $(MaxMaxEnt)_4$ Dağılımının (a) Olasılık..... Yoğunluk Fonksiyonu $\hat{f}(t)$, (b) Birikimli Dağılım Fonksiyonu $\hat{F}(t)$,..... (c) Sağkalım Fonksiyonu $\hat{S}(t)$ ve (d) Hazard Fonksiyonu $\hat{h}(t)$	49
5.1. Bağırsak kanseri verileri için (a) $(MinMinxEnt)_1$ Dağılımının grafiği..... ve (b) $(MaxMinxEnt)_1$ Dağılımının grafiği.....	53
5.2. Bağırsak kanseri verileri için (a) $(MinMinxEnt)_2$ Dağılımının grafiği..... ve (b) $(MaxMinxEnt)_2$ Dağılımının grafiği.....	55

5.3. Bağırsak kanseri verileri için (a) $(MinMinxEnt)_3$ Dağılımının grafiği.....	
ve (b) $(MaxMinxEnt)_3$ Dağılımının grafiği.....	56
5.4. Bağırsak kanseri verileri için (a) $(MinMinxEnt)_4$ Dağılımının grafiği.....	
ve (b) $(MaxMinxEnt)_4$ Dağılımının grafiği	58
5.5. Bağırsak kanseri verileri için $(MinMinxEnt)_4$ Dağılımına uygun.....	
(a) $\hat{f}(t)$ Fonksiyonunun Grafiği (b) $\hat{F}(t)$ Fonksiyonunun Grafiği,.....	
(c) $\hat{S}(t)$ Fonksiyonunun Grafiği ve (d) $\hat{h}(t)$ Fonksiyonunun Grafiği	63
5.6. Bağırsak kanseri verileri için $(MaxMinxEnt)_4$, Dağılımına uygun.....	
(a) $\hat{f}(t)$ Fonksiyonunun Grafiği (b) $\hat{F}(t)$ Fonksiyonunun Grafiği,.....	
(c) $\hat{S}(t)$ Fonksiyonunun Grafiği ve (d) $\hat{h}(t)$ Fonksiyonunun Grafiği	64

ÇİZELGELER DİZİNİ

4.1. 1935 – 1944 yılları arasında Connecticut şehrindeki erkek hastaların..... bağırsak kanseri verileri	29
4.2. 1935 – 1944 yılları arasında Connecticut şehrindeki erkek hastaların..... bağırsak kanseri verilerinin Gözlenen Olasılıkları ve Düzeltilmiş..... Olasılıkları.....	30
4.3. Bağırsak kanseri verileri için $g_0, g, g_0(x) = 1, g \in K_{0,1}$ vektör..... fonksiyonu yardımıyla elde edilen <i>MaxEnt</i> Dağılımları ve H_{max}	31
4.4. Bağırsak kanseri verileri için $g_0, g, g_0(x) = 1, g \in K_{0,2}$ vektör..... fonksiyonu yardımıyla elde edilen <i>MaxEnt</i> Dağılımları ve H_{max}	32
4.5. Bağırsak kanseri verileri için $g_0, g, g_0(x) = 1, g \in K_{0,3}$ vektör..... fonksiyonu yardımıyla elde edilen <i>MaxEnt</i> Dağılımları ve H_{max}	33
4.6. Bağırsak kanseri verileri için $g_0, g, g_0(x) = 1, g \in K_{0,4}$ vektör..... fonksiyonu yardımıyla elde edilen <i>MaxEnt</i> Dağılımları ve H_{max}	34
4.7. Bağırsak kanseri verileri için $(MinMaxEnt)_m, m = 1, \dots, 4$ Dağılımları ..	35
4.8. Bağırsak kanseri verileri için $(MaxMaxEnt)_m, m = 1, \dots, 4$ Dağılımları .	36
4.9. Bağırsak kanseri verileri için, $m = 1, 2, \dots, 4$ olmak üzere,..... $(MinMaxEnt)_m$ Dağılımları için Shannon entropi ölçüm değerleri H ,..... Kullback – Leibler ölçüm değerleri $D_{min}^{(m)}(p; q)$, Ki – Kare ve RMSE.....	37
4.10. Bağırsak kanseri verileri için, $m = 1, 2, \dots, 4$ olmak üzere,..... $(MinMaxEnt)_m$ Dağılımları için Shannon entropi ölçüm değeri H ,..... Kullback – Leibler ölçüm değeri $D_{max}^{(m)}(p; q)$, Ki – Kare ve RMSE.....	37
4.11. Bağırsak kanseri verileri için $(MinMaxEnt)_4$ Dağılımı ile yapılan..... Sağkalım Analizi.....	39

4.12. Bağırsak kanseri verileri için $(MaxMaxEnt)_4$ Dağılımı ile yapılan.....	
Sağkalım Analizi.....	40
4.13. Motor arızası verileri (1980)	42
4.14. Motor arızası verileri için Gözlenen ve Düzeltmiş Olasılıklar	43
4.15. Motor arızası verileri için $(MinMaxEnt)_m$, $m = 1, \dots, 4$ Dağılımları	43
4.16. Motor arızası verileri için $(MaxMaxEnt)_m$, $m = 1, \dots, 4$ Dağılımları....	44
4.17. Motor arızası verileri için $(MinMaxEnt)_m$, $m = 1, 2, \dots, 4$ için.....	
hesaplanan bazı kriter sonuçları.....	44
4.18. Motor arızası verileri için $(MaxMaxEnt)_m$, $m = 1, 2, \dots, 4$ için.....	
hesaplanan bazı kriter sonuçları.....	44
4.19. Motor arızası verileri için $(MinMaxEnt)_3$ Dağılımı ile Sağkalım.....	
Analizi.....	46
4.20. Motor arızası verileri için $(MaxMaxEnt)_4$ Dağılımı ile Sağkalım.....	
Analizi.....	48
5.1. Bağırsak kanseri verileri için $g_0, g, g_0(x) = 1, g \in K_{0,1}$ vektör.....	
fonksiyonu yardımıyla elde edilen $MinxEnt$ Dağılımları ve $D(p : q)$,.....	
D_{min}, D_{max} değerleri.....	52
5.2. Bağırsak kanseri verileri için $g_0, g, g_0(x) = 1, g \in K_{0,2}$ vektör.....	
fonksiyonu yardımıyla elde edilen $MinxEnt$ Dağılımları ve $D(p : q)$,.....	
D_{min}, D_{max} değerleri.....	53
5.3. Bağırsak kanseri verileri için $g_0, g, g_0(x) = 1, g \in K_{0,3}$ vektör.....	
fonksiyonu yardımıyla elde edilen $MinxEnt$ Dağılımları ve $D(p : q)$,.....	
D_{min}, D_{max} değerleri.....	55
5.4. Bağırsak kanseri verileri için $g_0, g, g_0(x) = 1, g \in K_{0,4}$ vektör.....	
fonksiyonu yardımıyla elde edilen $MinxEnt$ Dağılımları ve $D(p : q)$,.....	
D_{min}, D_{max} değerleri.....	57
5.5. Bağırsak kanseri verileri için $(MinMinxEnt)_m$, $m = 1, 2, 3, 4$	
Dağılımları	58
5.6. Bağırsak kanseri verileri için $(MaxMinxEnt)_m$, $m = 1, 2, 3, 4$	
Dağılımları	59
5.7. Bağırsak kanseri verileri için $(MinMinxEnt)_m$, $m = 1, 2, \dots, 4$	
Dağılımları için RMSE, Ki-Kare ve Kullback-Leibler ölçüm değerleri.....	59

5.8. Bağırsak kanseri verileri için $(MaxMinxEnt)_m$, $m = 1,2,\dots,4$	
Dağılımları için RMSE, Ki-Kare ve Kullback-Leibler ölçüm değerleri.....	59
5.9. Bağırsak kanseri verileri için $(MinMinxEnt)_4$ Dağılımı ile yapılan.....	
Sağkalım Analizi.....	62
5.10. Bağırsak kanseri verileri için $(MaxMinxEnt)_4$ Dağılımı ile yapılan.....	
Sağkalım Analizi.....	63

1. GİRİŞ

Sağkalım sürelerinin dağılımı nonparametrik ve bilinmiyorsa veya tüm nesnelerin sağkalım süreleri tam olarak bilinmiyorsa, bir başka ifadeyle, sağkalım süreleri sansürlü ise o zaman yeni bir yöntem geliştirmek gerekir. Bu yöntemlerden birisi Genelleştirilmiş Entropi Optimizasyon Yöntemleri (GEOY) olabilir. Çünkü tüm istatistiksel dağılımlar, uygun moment fonksiyonları seçilerek bulunan *MaxEnt* dağılımı olarak elde edilebilirler. Hatta Shannon entropi ölçümü veya Kullback – Leibler ölçümü ve bunlara ek karakterize edici moment fonksiyonları yardımıyla elde edilen Genelleştirilmiş Entropi Optimizasyon Dağılımları verilen istatistiksel veriyi daha tam olarak temsil etmektedirler. Bu sebeple, GEOD ile sağkalım veri analizi yeni bir anlam ve önem içermektedir.

Entropi Optimizasyon Yöntemleri (EOY) özellikle istatistikte, ekonomide fizikte, mühendislikte vb. birçok alanda önemli uygulamalara sahiptir.

Literatürde, istatistiksel verinin, bilinen istatistiksel dağılımlara uymadığı ancak entropi optimizasyon dağılımlarına iyi bir şekilde uyduğunu gösteren bir çok örnek vardır. Bunlardan birine de Şamilov'un uygulamasını da verdiği Entropi, İnforasyon ve Entropi Optimizasyon kitabında yer almaktadır. GEOY, informasyon teori anlamında, sırasıyla istatistiksel veriye en yakın ve istatistiksel veriden en uzak *MinMaxEnt* ve *MaxMaxEnt* dağılımlarının bulunmasını önermektedir (Shamilov, 2006, 2007). Bu sebeple GEOY, Sağkalım Veri Analizinde başarılı bir şekilde uygulanabilir.

Literatürde sağkalım veri analizinin farklı araştırmalarına ve uygulamalarına rastlanmaktadır. Özellikle Ebrahimi'nin 2000 yılında yayınladığı makalesinde, maksimum entropi prensibine dayandırılan risk (hazard) oran fonksiyonu tahmininin çeşitli problemleri araştırılmıştır. Bu araştırma, risk oran fonksiyonu üzerinde belli bilgi kısıtlarını tahmin eden farklı veri üreten dağılım modellerini uygulamak için yapılabilen maksimum entropi dağılımlarının çeşitli sınıflarını geliştirmeyi öngörür (Ebrahimi, 2000).

Despande ve Purohit ile Lee ve Wang yazdıkları kitaplarda sağkalım veri analizinin çeşitli uygulamalar içeren genel teorisini vermiş ve en yaygın kullanım yöntemlerini anlatmıştır, (Deshpande ve Purohit, 2005) ve (Lee ve Wang, 2003).

Bu tezde ise, *MinMaxEnt*, *MaxMaxEnt* ve *MinMinxEnt*, *MaxMinxEnt* yöntemleri, karakterize edici moment vektör fonksiyonlarının sonlu kümesi $K_{0,m}$ ve sırasıyla Shannon entropi ölçümü ve Kullback – Leibler ölçümü yardımıyla tanımlanan *MinMaxEnt*, *MaxMaxEnt* ve *MinMinxEnt*, *MaxMinxEnt* olasılık dağılımları formunda gerçekleştirilmiştir. Bu dağılımlar, *önsel* olasılık dağılımına en yakın veya *önsel* olasılık dağılımından en uzak dağılımlardır ve istatistiksel modellemede başarılı bir şekilde kullanılabilirler.

Bu araştırmanın sonuçlarını ifade etmek için ilk önce bazı yardımcı kavramlar ve olgular sunulmuştur. Sağkalım veri analizinde GEOY'nin önemli sonuçlara sahip olduğunu vurgulamak gerekir.

Bu tezde geçen bölümler, aşağıdaki şekilde özetlenebilir:

İkinci bölümde sağkalım analizinin genel teorisi, temel kavramları ve sağkalım fonksiyonlarının tanımları verilmiş, sansürlü veriler ve sansür çeşitlerine değinilmiştir.

Üçüncü bölümde Maksimum Entropi (*MaxEnt*) Dağılımlarının Kesikli ve Sürekli rassal değişkenler için elde edilme yöntemleri ve Minimum Çapraz Entropi (*MinxEnt*) Dağılımlarının Kesikli ve Sürekli rassal değişkenler için elde edilme yöntemleri, ispatları ile birlikte verilmiştir.

Dördüncü bölümde Genelleştirilmiş Entropi Optimizasyon Yöntemleri arasında *MinMaxEnt* ve *MaxMaxEnt* Yöntemleri anlatılmış ve bağırsak kanseri verileri ve motor arızası verileri için bu yöntemlerin uygulaması verilmiş ve elde edilen *MinMaxEnt* ve *MaxMaxEnt* dağılımlarının istatistiksel veriyi belli anlamlarda hangisinin diğerine göre daha iyi temsil ettiği çeşitli kriterler kullanılarak yorumlanmıştır. Daha sonra bu dağılımların tahmini sağkalım fonksiyonları bulunmuş ve bu fonksiyonların grafikleri çizilmiştir.

Beşinci bölümde Genelleştirilmiş Entropi Optimizasyon Yöntemleri arasında *MinMinxEnt* ve *MaxMinxEnt* Yöntemleri anlatılmış ve bağırsak kanseri verileri için bu yöntemlerin uygulaması verilmiş ve elde edilen *MinMinxEnt* ve *MaxMinxEnt* dağılımlarının istatistiksel veriyi belli anlamlarda hangisinin diğerine göre daha iyi temsil ettiği çeşitli kriterler kullanılarak yorumlanmıştır. Daha sonra bu dağılımların tahmini sağkalım fonksiyonları bulunmuş ve bu fonksiyonların grafikleri çizilmiştir.

Altıncı bölümde elde edilen tüm *MinMaxEnt*, *MaxMaxEnt* ve *MinMinxEnt*, *MaxMinxEnt* dağılımları çeşitli kriterler kullanılarak birbiri ile karşılaştırılmış ve sonuçlar yorumlanmıştır.

Yedinci bölümde ise yapılan bu araştırmanın tüm sonuçları ve bu sonuçlardan yola çıkılarak, literatüre etkileri hakkında bazı önerilerde bulunulmuştur.

2. SAĞKALIM ANALİZİ

Sağkalım Analizi en eski istatistiksel tekniklerden biridir ve genellikle biyoistatistikte, endüstride, sigortacılar ve nüfus bilimciler tarafından geniş bir kullanım alanına sahiptir. Bu yöntem sadece yaşam sürelerinin (life time) değil ölçülebilir değerlerin analizinde de kullanılabilir. Yaşam süresi, tanımlanan herhangi bir olayın başladığı andan bitimine kadar geçen süredir. Örneğin bir hastalıkta kullanılan bir ilacın hastalığı iyileştirme süresi araştırıldığında veya başka bir çalışmada evli çiftlerin ne kadar evli kalacaklarının tespitinde ya da üretilen bir ürünün ne kadar zaman sonra bozulacağını veya arızalanacağını tahmin etmede sağkalım analizi kullanılır.

Araştırılan, tek bir gruptaki birimlerin yaşam süreleri ya da başarısızlık zamanlarının incelenmesi olabileceği gibi birden fazla gruptaki birimlerin başarısızlık zamanlarının (failure time) karşılaştırılması da olabilir. Bazı durumlarda ise yaşam süresini etkileyen açıklayıcı değişkenlerin analizdeki etkisi araştırılabilir.

2.1. Sağkalım Süresi Değişkeninin Tanımı

Sağkalım Analizinde sağ kalım süresi özetle, verilen bir olayın oluşması için gereken süre olarak tanımlanabilir. Bir başka ifade ile canlı veya cansız herhangi bir birimin yaşam süresi veya başarısızlık süresidir. Sağkalım süresi verileri belli bir etkinlik zamanını, bozulma, ölüm, tepki, depresme, bir hastalığın iyileşmesi ve kötüleşmesi gibi süreçleri ölçer. Böyle zamanlarda rassal değişkenlik vardır ve rassal değişken genellikle T ile gösterilir. T rassal değişkeni tanımı gereği her zaman sıfırdan büyüktür.

2.2. Temel Kavramlar

Her bir birimin başarısızlık zamanı olarak ifade edilen T rassal değişkeni, herhangi bir rassal değişken gibi bir dağılım oluşturur ve T ' nin olasılık dağılımı Sağkalım Analizi'nde genellikle dört fonksiyon tarafından karakterize edilir.

Bunlar; Birikimli Dağılım Fonksiyonu $F(t)$, Sağkalım Fonksiyonu $S(t)$, Olasılık Yoğunluk Fonksiyonu $f(t)$, ve Risk (Hazard) Fonksiyonu $h(t)$ 'dir.

2.2.1. Sağkalım Fonksiyonu $S(t)$

Sağkalım fonksiyonu bir bireyin t anından daha uzun süre hayatta kalma olasılığı olarak tanımlanır ve $S(t)$ ile gösterilir. O halde

$$S(t) = P\{T > t\} = \int_t^{\infty} f(u)du, \quad 0 < t < \infty$$

burada gerekli işlemler yapıldığında

$$S(t) = 1 - F(t) \quad (2.1)$$

olarak bulunur, burada $F(t)$, T 'nin birikimli dağılım fonksiyonu ve $S(t)$, t anının artmayan (monoton azalan) birikimli bir fonksiyonudur:

$$S(t) = \begin{cases} 1, & t = 0 \\ 0, & t = \infty \end{cases}$$

bir başka ifadeyle

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

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

dır.

2.2.2. Olasılık Yoğunluk Fonksiyonu $f(t)$

T başarısızlık süresi, her bir Δt başına $(t, t + \Delta t)$ küçük aralığında bir bireyin kötüleşmesi veya bir birimin başarısız olması olasılığının limiti olarak tanımlanan olasılık yoğunluk fonksiyonuna sahiptir. Bir başka ifadeyle $f(t)$,

$$f(t) = \frac{\lim_{\Delta t \rightarrow 0} P\{\text{herhangi bir bireyin } (t, t+\Delta t) \text{ aralığında kötüleşmesi}\}}{\Delta t}$$

$$f(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T \leq t + \Delta t)}{\Delta t} \quad (2.2)$$

şeklindedir ve t anında koşulsuz başarısızlığın yoğunluğu olarak yorumlanabilir. Dikkat etmek gerekir ki $S(t)$ ve $F(t)$, belli bir zaman aralığına ait olasılıklar olmalarına rağmen $f(t)$ bir zaman noktasına ait bir olasılıktır ve ani bir ölçümdür.

2.2.3. Risk (Hazard) Fonksiyonu $h(t)$

Risk fonksiyonu, başarısızlık hızı (failure rate), ani ölüm hızı (instantaneous death rate) veya ölümlülük gücü (force of mortality) olarak da isimlendirilebilir (Nelson 1972). Risk oranının tersi Mill's oranı (Mill's ratio) olarak bilinir.

Risk fonksiyonu, anın bir fonksiyonu olarak başarısızlık eğiliminin bir ölçüsüdür ve bu sebeple Sağlık Analizi'nde önemli bir yere sahiptir. O halde Sağlık süresi t ve risk fonksiyonu $h(t)$ olmak üzere;

$$h(t) = \frac{\lim_{\Delta t \rightarrow 0} P\left\{ \begin{array}{l} t \text{ anına kadar hayatta kaldığı bilinen herhangi bir bireyin} \\ (t, t+\Delta t) \text{ aralığında kötüleşmesi} \end{array} \right\}}{\Delta t}$$

şeklindedir, bir başka ifade ile

$$h(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T \leq t + \Delta t \mid T \geq t)}{\Delta t} \quad (2.3)$$

şeklindedir.

Bir bireyin t anında yaşamını sürdürdüğü biliniyorsa bu bireyin $(t, t+\Delta t)$ zaman aralığında yaşamını kaybetme olasılığı $h(t)\Delta t$ 'dir. Sürekli dağılımlar için Risk Fonksiyonu aşağıdaki özellikleri sağlar:

1. $h(t) > 0$
2. $\int_0^{\infty} h(t)dt = \infty$.

2.2.4. $S(t)$, $f(t)$ ve $h(t)$ Fonksiyonları Arasındaki İlişkiler

1. $h(t) = \frac{f(t)}{S(t)}$
2. $f(t) = -\frac{dS(t)}{dt}$
3. $h(t) = -\frac{S'(t)}{S(t)} = -\frac{d}{dt} \log S(t)$
4. $S(t) = \exp\left[-\int_0^t h(s)ds\right] = \exp[-H(t)]$

burada $H(t) = \int_0^t h(s)ds$ birikimli Risk Fonksiyonudur ve aşağıdaki özellikleri sağlar:

- 1) $H(t)$ artan bir fonksiyondur.
- 2) $\lim_{t \rightarrow \infty} H(t) = \infty$
- 3) $H(t)$ sağdan sürekli bir fonksiyondur ve $f(t) = h(t)\exp\left[-\int_0^{\infty} h(s)ds\right]$.

Şimdi, yukarıda bahsedilen ilişkilerin ispatı verilsin.

İspat.

$$\begin{aligned}
 1. \quad h(t) &= \lim_{\Delta t \rightarrow \infty} \frac{P\{t \leq T \leq t + \Delta t \mid T \geq t\}}{\Delta t} = \lim_{\Delta t \rightarrow \infty} \frac{1}{\Delta t} \frac{P\{t \leq T \leq t + \Delta t, T \geq t\}}{P\{T \geq t\}} \\
 &= \lim_{\Delta t \rightarrow \infty} \frac{1}{\Delta t} \frac{F(t + \Delta t) - F(t)}{F(t)} = \lim_{\Delta t \rightarrow \infty} \frac{\bar{F}(t) - \bar{F}(t + \Delta t)}{\Delta t \bar{F}(t)} = \frac{f(t)}{\bar{F}(t)} = \frac{f(t)}{S(t)}
 \end{aligned}$$

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

$$2. \quad S(t) = 1 - F(t) \Rightarrow F'(t) = -(-F'(t)) = -(S'(t)) = -\frac{dS(t)}{dt}$$

$$f(t) = -\frac{dS(t)}{dt} \quad \blacksquare$$

$$3. h(t) = -\frac{d}{dt} \log S(t)$$

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

burada her iki tarafın integrali alınırsa

$$-\int_0^t h(s) ds = \int_0^t \frac{d}{dt} \log S(t) ds$$

$$-\int_0^t h(s) ds = \log S(t)$$

$$S(t) = \exp \left[-\int_0^t h(s) ds \right]$$

$$S(t) = \exp[-H(t)] \quad \blacksquare$$

$$4. h(t) = \frac{f(t)}{S(t)}$$

$$f(t) = h(t)S(t)$$

$$f(t) = h(t) \exp \left[-\int_0^t h(s) ds \right] \quad \blacksquare$$

Daha farklı bir yorum getirilirse;

$$h(t) = \frac{f(t)}{S(t)} = \frac{f(t)}{1-F(t)} = \frac{f(t)}{1-\int_0^t f(s) ds}$$

$$h(t) = \frac{f(t)}{1-\int_0^t f(s) ds}$$

sadece yoğunluk fonksiyonuna bağlı bir fonksiyon olarak ifade edilir.

Risk Fonksiyonu her bir zaman noktasında var olan anlık başarısızlık oranıdır. Böylece yukarıdaki ilişkilerden de görülüyor ki Risk Fonksiyonu Sağkalım Analizinde yaşam sürdürme verilerinin modellenmesinde kullanılır ve modellemeye yarar sağlar.

2.3. Sansürlü (Durdurulmuş) Veriler

Sağkalım analizinde en büyük zorluk gözlenen verilerin başarısızlık zamanlarının çalışmanın herhangi bir yerinde, kişinin bilgisi dâhilinde olmadan gözlenememesidir. Böyle verilere sansürlü veriler (censoring data) denir. Başka bir tanım verilirse sansürleme, deney süresini bilinçli olarak kısaltma yöntemidir (Desphande ve Purohit, 2005). Örneğin; Tıpta bir klinik çalışmada bazı hastalar herhangi bir zamandan sonra tedaviye gelmeyebilir ya da çalışma bittiğinde hala yaşamını sürdürüyor olabilir veya endüstriyel çalışmalarda yapılan deneylerde bazı ürünler deney bitmiş olmasına rağmen bozulmamış olabilir yahut gözlem altındaki birey göç etmiş olabilir. Böyle durumlarda araştırmacı onların yaşam sürelerini gözlemleyemez. Böyle durumlarda, gözlenemeyen bu veriler, sansüre uğramış verilerdir.

İncelenen olayın türüne bağlı olarak farklı sansür tiplerinden söz edilebilir. Nelson sansür tiplerini Planlanmış Sansür ve Planlanmamış Sansür olmak üzere iki başlık altında toplamıştır. Aşağıda bu sansür tiplerinin Sağkalım Analizindeki yeri ve önemi anlatılmıştır (Başar, 1993).

2.3.1. Planlanmış Sansür

Planlanmış sansür, araştırmacı tarafından belirlenen kriterlere göre araştırmanın sona erdirilmesidir. Bir başka ifadeyle, araştırmacı kendi kararları doğrultusunda, planladığı kriterlere ulaştığı anda deneyi durdurmuş ise deney bittiği anda başarısızlığa ulaşmayan veriler gözlenemez ve bu veriler sansüre uğramış olurlar. Daha çok deneyin çok uzun süreceği durumlarda maddi açıdan veya zaman kısıtlılığından dolayı bilinçli olarak deneyin sonlandırılmasıdır. Kendi içinde 1. Tip Sansür (time censoring) ve 2. Tip Sansür (order (failure) censoring) olarak iki ana başlığa ayrılır.

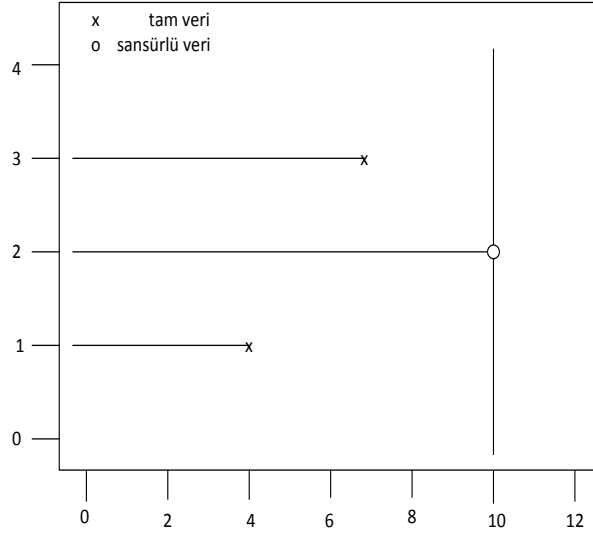
2.3.1.1 1.Tip Sansür (Time Censoring)

1. tip sansür (time censoring) özetle, önceden belirlenmiş bir zamanda çalışmanın durdurulmasıdır ve zamana bağlı sansür olarak da adlandırılabilir. Kendi içinde ise Tek Noktalı 1. Tip Sansür ve Çok Noktalı 1. Tip Sansür olarak ikiye ayrılır.

Tek Noktalı 1.Tip Sansür

X_1, X_2, \dots, X_n F dağılımından alınmış rassal bir örneklem ve t_0 sabitlenmiş sansür (durdurma) zamanı olsun. Tüm gözlemlerin araştırmaya alınmasıyla yapılan araştırma, araştırmacı tarafından belirlenen herhangi bir t_0 zamanında sona erdirilirse elde edilen bilgi, bazı gözlemler için başarısızlık zamanı, bazıları için ise sansür zamanı olacaktır. Başarısız olan gözlemlerin yaşam süreleri bilindiği gibi başarısız olmayanların ise yaşam sürelerinin sadece t_0 zamanından daha uzun olduğu söylenebilir. Böylece tek bir sansür zamanı olacaktır ve bu tür sansüre Tek Noktalı 1. Tip Sansür denir.

Örneğin; 3 Myeloma hastası için yapılan 10 aylık incelemede, 1'er aylık aralar ile yapılan kontroller sonucu her aralık boyunca hayatta kalan ve yaşamını yitiren bireylerin sayısı verilsin, 10. ayın sonunda ise çalışma durdurulsun. Bu süreçte, birinci ve üçüncü kişinin bilinen sebepler yüzünden hayatını kaybettiği fakat ikinci kişinin ise çalışma bitmesine rağmen hala hayatta olduğu bilinsin. İşte bu tip incelemelerde bu ikinci kişi 1. tip sansüre uğramış gözlem olarak nitelendirilir. Bu durumda ikinci kişinin başarısızlık zamanı ile ilgili sadece 10 aydan daha uzun olduğu söylenebilir ve bu durum Şekil 2.1'de gösterilmiştir.



Şekil 2.1. I. Tip Sansür'e Bir Örnek

Çok Noktalı 1. Tip Sansür

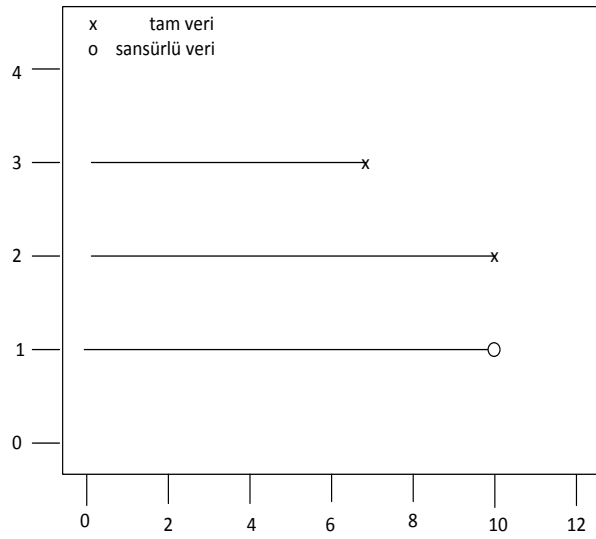
Çok Noktalı 1. Tip Sansürde ise Tek Noktalı 1. Tip Sansürden farklı olarak deney, birden fazla *durdurma zamanına* sahiptir. Bir başka ifadeyle, sadece bir tane durdurma zamanı t_0 anı yerine t_1, t_2, \dots, t_n gibi birkaç tane durdurma zamanı vardır.

2.3.1.2 2. Tip Sansür (Order Censoring)

2. tip sansür (order censoring) ise, bir çalışmada daha önceden belirlenen sayıda başarısızlığın meydana gelmesi sonucu, çalışmanın durdurulmasıdır. Bir başka ifadeyle, n birimlik rassal gözlemle başlanan bir çalışmada başarısız olan gözlemlerin sayısı k , $k < n$ ve $t_{(i)}$, i . gözlemin sağkalım süresi olmak üzere, incelenen çalışma k sayıda gözlemin başarısız olması ile durdurulursa geride kalan $n - k$ gözlemin başarısızlık zamanları incelenemez. Böylece $n - k$ gözlem, sansürlü gözlem olarak nitelendirilir, bu gözlemlerin sansür zamanı ise k . gözlemin sağkalım zamanı olan $t_{(k)}$ 'dir, burada k , n birimlik rassal gözlemin en küçük sıralı istatistiğidir, yani

$$t_{(1)} < t_{(2)} < t_{(3)} < \dots < t_{(k)} .$$

Örneğin 10 rassal seramik kondansatör ile sağkalım analizi yapılsın ve bu analiz, inceleme süresini azaltmak amacıyla, birinci kondansatörün başarısız olması ile durdurulsun. Bu çalışmaya göre, geride kalan iki kondansatörün başarısızlık süresi ile ilgili bilgi alınamaz ve bu iki gözlem, 2. Tip sansüre uğramış olur. Bu durum, Şekil 2.2’de gösterilmiştir.



Şekil 2.2. II. Tip Sansür'e Bir Örnek

Yukarıda bahsedilen *Planlanmış Sansür*, daha çok güvenilirlik çalışmalarında, başka ifadeyle, mühendislik çalışmalarında yaygın olarak kullanılır.

2.3.2. Planlanmamış Sansür (Random Censoring)

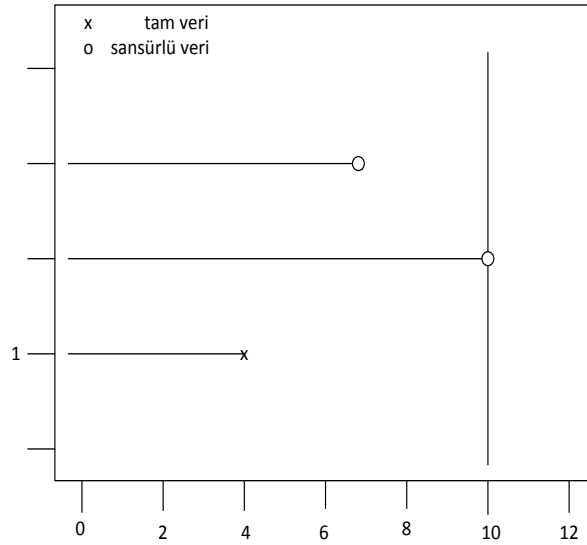
Planlanmamış sansür, özellikle biyomedikal verilerde daha yaygın olarak karşımıza çıkar. Yapılan çalışma sona ermeden, herhangi bir bireyin durdurma zamanı şansa bağlı nedenlerle gerçekleşmiş ise bu tür sansüre planlanmamış (rassal) sansür denir. Bir çalışmada her bir bireyin başarısızlık zamanı T ve

sansür zamanı L 'ye sahip olduğu varsayılın, burada temel varsayım; T ve L 'nin birbirinden bağımsız rassal değişkenler olduğudur. Planlanmamış sansür Sağdan Sansür (Right Censoring) ve Soldan Sansür (Left Censoring) olmak üzere ikiye ayrılır.

2.3.2.1 Sağdan Sansür

Bir gözlemin kesin başarısızlık zamanı bilinmiyor fakat sadece belli bir zamandan büyük ya da belli bir zamana eşit olduğu biliniyorsa bu tip sansüre sağdan sansür denir.

Örneğin üç akciğer kanseri on aylık deney süresince gözlemlenir. Üçüncü gözlemden altıncı aydan sonra haber alınamamış ise bu durumda bu gözlem Şekil 2.3'te görüldüğü gibi resmedilir.

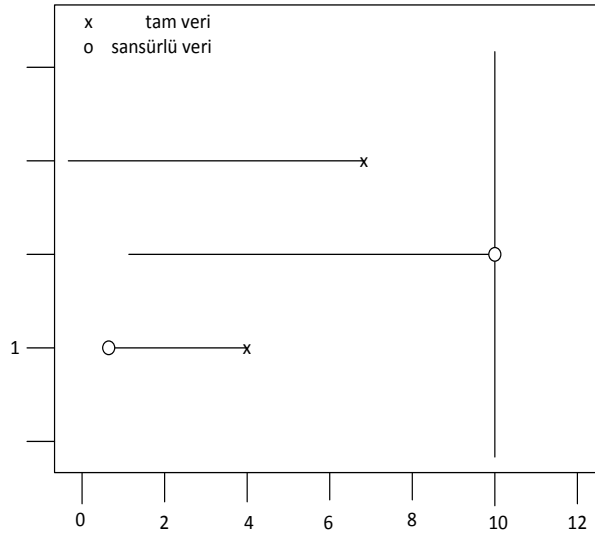


Şekil 2.3. Sağdan Sansür'e Bir Örnek

2.3.2.2 Soldan Sansür

Sağdan sansüre benzer şekilde, bir gözlemin kesin başarısızlık zamanının sadece belli bir zamandan küçük ya da belli bir zamana eşit olduğu biliniyorsa bu tip sansüre soldan sansür denir.

Örneğin, gene üç birey ele alınsın. Bunlardan birinci birey deneye ilk iki ay içinde dâhil edilsin ancak bu dâhil edilmiş tam zamanı bilinmiyor olsun. Bu durumda bu gözlemin grafiksel ifadesi Şekil 2.4'te görüldüğü gibidir.



Şekil 2.4. Soldan Sansür'e Bir Örnek

3. GENELLEŞTİRİLMİŞ ENTROPİ OPTİMİZASYON YÖNTEMLERİ

Genelleştirilmiş Entropi Optimizasyon Yöntemi (GEOY) denildiğinde istatistiksel veriyi Genelleştirilmiş Entropi Optimizasyon Problemi (GEOP)'un çözümü şeklinde modellemeyi öneren yöntem anlaşılmaktadır.

EOP: X rassal değişkeninin olasılık yoğunluk fonksiyonu $f^{(0)}(x)$, L entropi optimizasyon ölçümü, g ise m sayıda moment kısıtı üreten bir moment vektör fonksiyonu olsun. Verilen g moment fonksiyonuna uygun olan ve L 'ye ekstremum değer veren dağılımın bulunması problemi.

GEOP: X rassal değişkeninin olasılık yoğunluk fonksiyonu $f^{(0)}(x)$, L entropi optimizasyon ölçümü, K ise verilen moment vektör fonksiyonlar kümesi olsun. Verilen K kümesinden öyle $g^{(1)}, g^{(2)} \in K$ moment vektör fonksiyonlarının bulunması gerekir ki $g^{(1)}$ 'in belirlediği entropi optimizasyon yoğunluk fonksiyonu $f^{(1)}(x)$, verilmiş $\rho(x)$ yoğunluk fonksiyonuna L ölçümüne göre en yakın, $g^{(2)}$ 'nin belirlediği entropi optimizasyon yoğunluk fonksiyonu $f^{(2)}(x)$, verilmiş $\rho(x)$ yoğunluk fonksiyonuna L ölçümüne göre en uzak olsun. Burada L Shannon'un entropi ölçümü olarak seçilirse $f^{(1)}(x)$, *MinMaxEnt* Dağılımı; $f^{(2)}(x)$, *MaxMaxEnt* Dağılımı olarak; eğer L Kullback - Leibler ölçümü olarak seçilirse $f^{(1)}(x)$, *MinMinxEnt* Dağılımı; $f^{(2)}(x)$, *MaxMinxEnt* Dağılımı olarak adlandırılır (Shamilov, 2007).

GEOP'u çözmeye yöntemi Genelleştirilmiş Entropi Optimizasyon Yöntemleri (GEOY) olarak adlandırılır. Bir başka ifadeyle, Genelleştirilmiş Entropi Optimizasyon Dağılımları (GEOD)'un bulunması yöntemlerine GEOY denir.

Entropi Optimizasyon Yöntemleri (EOY) mühendislik, ekonomi ve çeşitli alanlarda, özellikle istatistikte önemli uygulamalara sahiptir. Literatürde, bilinen istatistiksel yöntemlerin istatistiksel veriye uymadığını ancak, entropi optimizasyon yöntemlerinin iyi bir şekilde uyduğunu gösteren bir çok örnek vardır. Tüm istatistiksel dağılımların *MaxEnt* dağılımı olarak elde edilebildiği

bilinmektedir ve hatta Genelleştirilmiş Entropi Optimizasyon Yöntemleri, verilen istatistiksel veriyi daha tam (eksiksiz) bir şekilde temsil etmektedir.

Bu bölümde, ilk olarak Maksimum entropi yöntemi ve bu yöntemin hem kesikli hem de sürekli rassal değişkenler için uygulanan Maksimum entropi yöntemleri daha sonra da Minimum çapraz entropi yöntemi ve bu yöntemin hem kesikli hem de sürekli rassal değişkenler için uygulanan Minimum çapraz entropi yöntemleri ele alınmıştır.

3.1. Maksimum Entropi Yöntemi

Bir fiziksel sistemin durumunu tanımlamak için yapılan ölçümlere dayandırılan olasılık dağılımını türetme yönteminin öncüsü Jaynes'tir. Bu yöntem, maksimum entropi yöntemi (*MaxEnt*), maksimum belirsizlik prensibi veya Jaynes'in formülü olarak bilinir (Kapur ve Kesevan, 1992).

Elde sadece bilgi olarak fiziksel sistemden elde edilmiş olasılık dağılımı ve karakterize edilmiş moment fonksiyonları varsa, bu bilgilere uygun sonsuz sayıda dağılım bulunabilir. Bu dağılımlardan Shannon entropi ölçümüne en büyük değeri veren dağılım bulma yöntemine *MaxEnt* yöntemi, bu yöntemden elde edilen dağılıma da *MaxEnt* dağılımı denir.

Şimdi sırasıyla kesikli rassal değişkenler için maksimum entropi yöntemi ve sürekli rassal değişkenler için maksimum entropi yöntemi ele alınacaktır.

3.1.1. Kesikli Rassal Değişkenler İçin Maksimum Entropi Yöntemi

n değere sahip X kesikli rassal değişkeninin veya n duruma sahip X kesikli fiziksel sisteminin entropisi

$$H(x) = - \sum_{i=1}^n p_i \ln p_i \quad (3.1)$$

fonksiyonu ile ifade edilmektedir. (3.1) fonksiyonunu,

$$\left. \begin{aligned} \sum_{i=1}^n p_i &= 1 \\ \sum_{i=1}^n p_i g_j(x_i) &= \mu_j, \quad j = 1, 2, \dots, m \end{aligned} \right\} \quad (3.2)$$

koşulları altında maksimum yapan (p_1, p_2, \dots, p_n) dağılımının bulunması bir optimizasyon problemidir. X rassal değişkeninin dağılımının bu optimizasyon probleminin çözümü olarak bulunmasını öneren yöntem *maksimum entropi yöntemi* denir. İlk defa Jaynes tarafından önerilen bu yöntem, Jaynes'in Maksimum Entropi Yöntemi olarak da bilinir.

(3.2) sisteminde $g_1(x), g_2(x), \dots, g_m(x)$ 'ler moment fonksiyonları olarak ifade edilir ve bu sistem n sayıda p_1, p_2, \dots, p_n bilinmeyenlerine sahiptir. Bu sistemin çözüme sahip olabilmesi için bazı koşulların sağlanması gerekmektedir. İlk olarak, $1, g_1(x), g_2(x), \dots, g_m(x)$ fonksiyonları lineer bağımsız olduğunda, bu fonksiyonların beklenen değerleri olan $1, \mu_1, \mu_2, \dots, \mu_m$ sayıları, (3.2) sistemi çözüme sahip olacak şekilde verilmelidir. İkinci olarak, $m + 1$ koşul sayısı ile n durum sayısı arasında $m + 1 < n$ bağıntısı bulunmalıdır. $m + 1 < n$ eşitsizliği sağlandığında, n bilinmeyenli $m + 1$ denklem oluşacağından, bu değişkenlerden $n - (m + 1)$ kadar bağımsız olacaktır. Bu durumda, p_1, p_2, \dots, p_n değişkenleri sonsuz sayıda değerlere sahip olur ve böylelikle, (3.1) fonksiyonunu maksimize eden (p_1, p_2, \dots, p_n) 'lerin seçilmesi anlamlı olacaktır. Eğer $m + 1 < n$ koşulu sağlanmazsa, bir başka ifade ile, $m + 1 = n$ veya $m + 1 > n$ olursa, birinci halde genellikle (3.1) fonksiyonunu maksimize etmek için bağımsız değişken kalmaz, ikinci halde ise (3.2) koşullarını sağlayan p_i 'ler olmayabilir.

(3.1) fonksiyonunun (3.2) koşulları altında maksimize edilmesi, kısaca (3.1), (3.2) problemi olarak ifade edilecektir. Burada (3.1) amaç fonksiyonu, (3.2) ise lineer koşullardır. Böylece, (3.1), (3.2) problemi koşullu ekstremum problemidir ve bilindiği gibi bu problem Lagrange çarpanları yöntemi ile çözülebilmektedir (Şamilov, 2009).

(3.1), (3.2) problemi Lagrange çarpanları yöntemi ile ele alındığında, (3.2) koşullarında ortaya çıkan μ_j sayıları keyfi verilemez. Pratik olarak, X rassal değişkenine ait istatistiksel frekans dağılımından $(\tilde{p}_1, \tilde{p}_2, \dots, \tilde{p}_n)$ ve verilmiş $g_j(x)$ moment fonksiyonlarından yola çıkarak μ_j 'ler hesaplandığında,

$$A = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ g_1(x_1) & g_1(x_2) & \cdots & g_1(x_n) \\ g_2(x_1) & g_2(x_2) & \cdots & g_2(x_n) \\ \vdots & \vdots & \vdots & \vdots \\ g_m(x_1) & g_m(x_2) & \cdots & g_m(x_n) \end{bmatrix}_{(m+1) \times n}$$

matrisinin rankı ve genişletilmiş

$$B = \begin{bmatrix} 1 & \cdots & 1 & 1 \\ g_1(x_1) & \cdots & g_1(x_n) & \mu_1 \\ g_2(x_1) & \cdots & g_2(x_n) & \mu_2 \\ \vdots & \vdots & \vdots & \vdots \\ g_m(x_1) & \cdots & g_m(x_n) & \mu_m \end{bmatrix}_{(m+1) \times (n+1)}$$

matrisinin rankı,

$$g_j(x_1)\tilde{p}_1 + \cdots + g_j(x_n)\tilde{p}_n = \mu_j, \quad j = 0, 1, \dots, m; \quad g_0(x) = 1, \quad \mu_j = 0$$

eşitlikleri sağlandığından, birbirine eşit olur. Bu nedenle de Kronecker-Capelli teoremine göre (3.2) koşullarını sağlayan p_1, p_2, \dots, p_n 'ler vardır. Başka bir deyişle, (3.2) eşitlikleri şeklinde verilmiş homojen olmayan lineer sistemin p_1, p_2, \dots, p_n 'ye göre çözümü vardır.

Bununla birlikte, (3.2) koşullarında bulunan $g_j(x_i)$ 'ler keyfi sayılar olabilir. Başka bir ifadeyle, $g_j(x_i)$ değerleri, a_{ji} şeklinde herhangi bir sayı ile ifade edilebilir.

(3.1) fonksiyonunu (3.2) koşulları altında maksimize etme problemine *maksimum entropi problemi* denir. Maksimum entropi problemi yardımıyla $(\tilde{p}_1, \tilde{p}_2, \dots, \tilde{p}_n)$ istatistiksel dağılımına entropi ölçümü açısından yakın olan (p_1, p_2, \dots, p_n) dağılımı bulunur. Burada, p_1, p_2, \dots, p_n 'ler (3.1) fonksiyonuna (3.2) koşulları altında maksimum değer veren olasılıklardır. (3.1) fonksiyonunu (3.2) koşulları altında maksimize etmek için $\lambda_0, \lambda_1, \dots, \lambda_m$ Langrange çarpanları yöntemiyle,

$$U = -\sum_{i=1}^n p_i \log p_i + \lambda_0 (\sum_{i=1}^n p_i - 1) + \sum_{j=1}^m \lambda_j (\sum_{i=1}^n p_i g_j(x_i) - \mu_j)$$

veya

$$U = -\sum_{i=1}^n p_i \ln p_i - \sum_{j=0}^m \sum_{i=1}^n (p_i g_j(x_i) - \mu_j), \quad g_0(x) = 1, \quad \mu_0 = 1 \quad (3.3)$$

yardımcı fonksiyonu kurulur ve bu fonksiyonun $p_1, p_2, \dots, p_n, \lambda_0, \lambda_1, \dots, \lambda_m$ değişkenlerine göre mutlak ekstremumu aranır.

Yardımcı (3.3) fonksiyonunun kendi değişkenlerine göre kısmi türevlerini alarak bu türevleri sıfıra eşitlemek gerekmektedir. Bu süreç yardımıyla, U fonksiyonuna ekstremum değer verebilecek kritik nokta bulunabilir. (3.3)'ten

$$\frac{\partial U}{\partial p_i} = -\ln p_i - 1 - \lambda_0 - \sum_{j=1}^m \lambda_j g_j(x_i) = 0, \quad i = 1, 2, \dots, n$$

$$\frac{\partial U}{\partial \lambda_0} = -(\sum_{i=1}^n p_i - 1) = 0$$

$$\frac{\partial U}{\partial \lambda_1} = -(\sum_{i=1}^n g_1(x_i) p_i - \mu_1) = 0$$

⋮

$$\frac{\partial U}{\partial \lambda_m} = -(\sum_{i=1}^n g_m(x_i) p_i - \mu_m) = 0.$$

Bu sisteme dikkat edilirse $\frac{\partial U}{\partial \lambda_0} = 0$ (3.2) koşullarının birincisini, $\frac{\partial U}{\partial \lambda_j} = 0$, $j = 1, 2, \dots, m$ (3.2) koşullarının j.sini ifade eder. Bu nedenle de $\frac{\partial U}{\partial p_i} = 0$, $i = 1, 2, \dots, n$ eşitliklerinden elde edilen p_i 'leri (3.2) koşullarında yerine yazarak Lagrange çarpanları bulunabilir.

$$\frac{\partial U}{\partial p_i} = -\ln p_i - 1 - \lambda_0 - \sum_{j=1}^m \lambda_j g_j(x_i) = 0$$

eşitliklerinden,

$$\ln p_i = -(1 + \lambda_0 + \sum_{j=1}^m \lambda_j g_j(x_i))$$

$$\begin{aligned}
p_i &= e^{-(1+\lambda_0)-\sum_{j=1}^m \lambda_j g_j(x_i)} \\
p_i &= e^{-\lambda^0-\sum_{j=1}^m \lambda_j g_j(x_i)}, \quad \lambda^0 = 1 + \lambda_0, \quad i = 1, 2, \dots, n
\end{aligned} \tag{3.4}$$

sonucuna varılır. (3.4) formülü ile p_1, p_2, \dots, p_n 'ler, $\lambda_0, \lambda_1, \dots, \lambda_m$ Lagrange çarpanlarına bağlı olarak ifade edilmiş olur.

Lagrange çarpanlarını bulmak için (3.4) ile ifade edilen p_1, p_2, \dots, p_n 'lerin (3.2) koşullarında yerine yazılmasıyla,

$$\left. \begin{aligned}
\sum_{i=1}^n e^{-\lambda^0-\sum_{j=1}^m \lambda_j g_j(x_i)} &= 1 \\
\sum_{i=1}^n g_j(x_i) e^{-\lambda^0-\sum_{j=1}^m \lambda_j g_j(x_i)} &= \mu_j
\end{aligned} \right\} \tag{3.5}$$

sistemi elde edilir.

Yukarıda da ifade edildiği gibi $\lambda^0, \lambda_1, \dots, \lambda_m$ Lagrange çarpanları (3.5) sisteminden bulunabilir. (3.5) sisteminin birinci denkleminde

$$\begin{aligned}
e^{\lambda^0} &= \sum_{i=1}^n e^{-\sum_{j=1}^m \lambda_j g_j(x_i)}, \\
\lambda^0 &= \ln \sum_{i=1}^n e^{-\sum_{j=1}^m \lambda_j g_j(x_i)}
\end{aligned} \tag{3.6}$$

bulunur. λ^0 'ın değeri, (3.5) sisteminin diğer denklemlerinde yerine yazıldığında,

$$f_i(\lambda_1, \dots, \lambda_m) \equiv \sum_{i=1}^n g_j(x_i) \frac{e^{-\sum_{j=1}^m \lambda_j g_j(x_i)}}{\sum_{i=1}^n e^{-\sum_{j=1}^m \lambda_j g_j(x_i)}} = \mu_j, \quad j = 1, 2, \dots, m \tag{3.7}$$

sistemi elde edilir (Şamilov, 2009). $j = 1, 2, \dots, m$ olmak üzere, (3.7) eşitliğinden $\lambda_1, \dots, \lambda_m$ değerlerinin μ_1, \dots, μ_m değerlerine bağlı olduğu görülmektedir.

3.1.2. Sürekli Rassal Değişkenler İçin Maksimum Entropi Yöntemi

Yoğunluk fonksiyonu $f(x)$ olan sürekli X rassal değişkeninin entropisi

$$H = - \int_a^b f(x) \ln f(x) dx \quad (3.8)$$

formülü ile verilir. İstatistiksel veriyi (3.8) fonksiyoneline

$$\int_a^b f(x) g_j(x) dx = \mu_j, \quad j = 0, 1, 2, \dots, m \quad (3.9)$$

koşulları altında maksimum değer veren dağılım ile modelleme yöntemine maksimum entropi yöntemi denir. (3.8) fonksiyoneli X rassal değişkeninin veya fiziksel sisteminin entropisini, (3.9) koşulları ise sistem hakkında ek bilgiyi ifade etmektedir. Böylece (3.8) fonksiyonelinin (3.9) koşulları altında maksimumunun bulunması, (3.9) koşulları yardımıyla sistemin sahip olduğu entropinin maksimum değerinin bulunması anlamına gelmektedir. Ayrıca ifade etmek gerekir ki (3.9) 'a ait olan her koşul sistem hakkında belli miktarda bilgi verdiği için, koşul sayısının artmasıyla entropi azalacaktır.

(3.8) fonksiyonelinin (3.9) koşulları altında maksimumu Lagrange çarpanları yöntemiyle bulunabilir. Bu durumda $\lambda_0, \lambda_1, \dots, \lambda_m$ 'ler Lagrange çarpanları olmak üzere, U Lagrange fonksiyoneli,

$$U \equiv - \int_a^b f(x) \ln f(x) dx - \sum_{j=0}^m \lambda_j \left(\int_a^b f(x) g_j(x) dx - \mu_j \right) \quad (3.10)$$

şeklindedir. (3.10) fonksiyoneli

$$U \equiv \int_a^b \left\{ -f(x) \ln f(x) - \sum_{j=0}^m \lambda_j g_j(x) f(x) \right\} dx - \sum_{j=0}^m \lambda_j \mu_j$$

veya

$$U \equiv \int_a^b F(f, \lambda) dx - \sum_{j=0}^m \lambda_j \mu_j \quad (3.11)$$

olarak yazılabilir, burada $F(f, \lambda) = -f(x) \ln f(x) - \sum_{j=0}^m \lambda_j g_j(x) f(x)$ 'dir.

(3.11) fonksiyonelinin ekstremumunun varlığı için gerek koşul, bu fonksiyonele ekstremum değer veren f fonksiyonunun

$$F_f - \frac{d}{dx} F_{f'} = 0 \quad (3.12)$$

Euler denklemini sağlamasıdır. (3.12) denkleminde $F_{f'} = 0$ olduğundan, bu denklem $F_f = 0$ denkleminde dönüşür. F 'in ifadesinden yola çıkarak Euler denkleminin

$$F_f = -\ln f(x) - 1 - \sum_{j=0}^m \lambda_j g_j(x) = 0$$

olduğu sonucuna varılır. Buradan,

$$\begin{aligned} \ln f(x) &= -1 - \sum_{j=0}^m \lambda_j g_j(x) \\ f(x) &= e^{-1 - \lambda_0 - \sum_{j=1}^m \lambda_j g_j(x)} \\ f(x) &= e^{-\tilde{\lambda}_0 - \sum_{j=1}^m \lambda_j g_j(x)}, \quad \tilde{\lambda}_0 = 1 + \lambda_0 \end{aligned} \quad (3.13)$$

elde edilir. (3.13) fonksiyonu (3.9) koşullarında yerine yazılarak λ 'lar bulunabilir:

$$\int_a^b e^{-\tilde{\lambda}_0 - \sum_{j=1}^m \lambda_j g_j(x)} g_k(x) dx = \mu_k, \quad k = 0, 1, 2, \dots, m \quad (3.14)$$

formülünde $k = 0$ iken $g_0(x) = 1$ ve $\mu_0 = 1$ olduğundan

$$\begin{aligned} \int_a^b e^{-\tilde{\lambda}_0 - \sum_{j=1}^m \lambda_j g_j(x)} dx &= 1 \\ e^{-\tilde{\lambda}_0} &= \frac{1}{\int_a^b e^{-\sum_{j=1}^m \lambda_j g_j(x)} dx} \\ \tilde{\lambda}_0 &= \ln \left\{ \int_a^b e^{-\sum_{j=1}^m \lambda_j g_j(x)} dx \right\} \end{aligned} \quad (3.15)$$

elde edilir ve $\tilde{\lambda}_0$ 'ın bu değeri (3.15)'in diğer eşitliklerinde yerine yazılıp k . eşitliğin sol tarafı F_k ile gösterildiğinde,

$$F_k(\lambda) = \frac{\int_a^b e^{-\sum_{j=1}^m \lambda_j g_j(x)} g_k(x) dx}{\int_a^b e^{-\sum_{j=1}^m \lambda_j g_j(x)} dx} = \mu_k, \quad k = 0, 1, 2, \dots, m \quad (3.16)$$

sistemine ulaşılır, burada $\lambda = (\lambda_1, \lambda_2, \dots, \lambda_m)$.

3.2. Minimum Çapraz Entropi Yöntemi

Ünlü bilim adamı Kullback'in öncülüğünü yaptığı Minimum çapraz entropi yöntemi, *MaxEnt* yöntemi gibi önemli bir entropi optimizasyon yöntemidir. Uygulama alanının geniş olması ve teorik önemi sebebiyle *MaxEnt* yöntemi ile rekabet eder (Kapur ve Kevesan, 1992). Burada, Minimum çapraz entropi yöntemi kısaca *MinxEnt* yöntemi olarak ifade edilecektir.

Elde sadece bilgi olarak fiziksel sistemden elde edilmiş önsel bir olasılık dağılımı ve önceden verilmiş karakterize edilmiş moment fonksiyonları varsa, önsel dağılıma en yakın dağılımın belirlenmesi için Kullback – Leibler ölçümüne en küçük değeri veren dağılımı bulma yöntemine *MinxEnt* yöntemi, bu yöntemden elde edilen dağılıma da *MinxEnt* dağılımı denir (Kullback 1959, Kullback – Leibler, 1951).

Şimdi sırasıyla kesikli rassal değişkenler için maksimum entropi yöntemi ve sürekli rassal değişkenler için maksimum entropi yöntemi ele alınacaktır.

3.2.1. Kesikli Rassal Değişkenler İçin Minimum Çapraz Entropi Yöntemi

X kesikli rassal değişkeni, p_1, p_2, \dots, p_n olasılıkları ile x_1, x_2, \dots, x_n değerlerine sahip olsun. X rassal değişkeninin, $p = (p_1, p_2, \dots, p_n)$ olasılık dağılımı hakkında önsel bir bilgi, $q = (q_1, q_2, \dots, q_n)$ dağılımı şeklinde verildiğinde p ile q arasındaki “yakınlığı” ifade eden Kullback-Leibler ölçümü

$$D(p : q) = \sum_{i=1}^n p_i \ln \frac{p_i}{q_i} \quad (3.17)$$

olarak yazılır, burada $\sum_{i=1}^n p_i = 1$, $\sum_{i=1}^n q_i \geq 1$, $p_i \geq 0$, $q_i \geq 0$ olur. (3.17) fonksiyonunu

$$\left. \begin{aligned} \sum_{i=1}^n p_i &= 1 \\ \sum_{i=1}^n p_i g_j(x_i) &= \mu_j, \quad j = 1, 2, \dots, m \end{aligned} \right\} \quad (3.18)$$

koşulları altında minimize etme problemi ele alınsın. $m + 1 < n$ eşitsizliği (3.18) koşullarını sağlayan p_1, p_2, \dots, p_n 'lerin $n - (m + 1)$ tanesinin serbest kalmasını gerektirir. Böylece sonsuz sayıdaki çözümler arasından Kullback-Leibler ölçümünde minimum değer veren p_i 'ler seçilir. MaxEnt yönteminde olduğu gibi burada da Lagrange çarpanları yöntemi uygulanarak Taylor Formülü yardımıyla (3.17) fonksiyonunun minimum değere sahip olduğu gösterilebilir (Şamilov, 2009).

Söz konusu yöntemin uygulanmasıyla oluşturulan yardımcı fonksiyon,

$$U \equiv \sum_{i=1}^n p_i \ln \frac{p_i}{q_i} - \lambda^0 (\sum_{i=1}^n p_i - 1) - \sum_{j=1}^m \lambda_j \sum_{i=1}^n (p_i g_j(x_i) - \mu_j) \quad (3.19)$$

şeklinindedir. (3.19) yardımcı fonksiyonunun p_i ve λ_j 'lere göre kısmi türevleri hesaplanarak sıfıra eşitlendiğinde p_i 'ler,

$$p_i = q_i e^{-\sum_{j=0}^m \lambda_j g_j(x_i)}, \quad i = 1, 2, \dots, n \quad (3.20)$$

olarak elde edilir, burada $\lambda_0 = 1 + \lambda^0$, $g_0(x) = 1$ 'dir.

(3.20) 'deki p_i 'ler (3.18) koşullarında yerine yazıldığında λ_j 'lere göre denklemler sistemi ortaya çıkar. Bu sistemden λ_j Lagrange çarpanları bulunabilir. (3.20), (3.18)'de yerine yazıldığında,

$$\left. \begin{aligned} \sum_{i=1}^n q_i e^{-\lambda_0 - \sum_{j=1}^m \lambda_j g_j(x_i)} &= 1 \\ \sum_{i=1}^n g_k(x_i) q_i e^{-\lambda_0 - \sum_{j=1}^m \lambda_j g_j(x_i)} &= \mu_k, \quad k = 1, 2, \dots, m \end{aligned} \right\} \quad (3.21)$$

sistemine ulaşılır. (3.21)'in ilk denkleminde,

$$e^{\lambda_0} = \sum_{i=1}^n q_i e^{-\sum_{j=1}^m \lambda_j g_j(x_i)}$$

ve buradan da,

$$\lambda_0 = \ln \sum_{i=1}^n q_i e^{-\sum_{j=1}^m \lambda_j g_j(x_i)} \quad (3.22)$$

elde edilir. Böylece, λ_0 , (3.22) formülü yardımı ile $\lambda_1, \lambda_2, \dots, \lambda_m$ 'lere bağlı ifade edilmiş olur. Bu durum dikkate alındığında (3.21) sistemi,

$$f_k(\lambda_1, \dots, \lambda_m) = \frac{\sum_{i=1}^n g_k(x_i) q_i e^{-\sum_{j=1}^m \lambda_j g_j(x_i)}}{e^{\lambda_0}} - \mu_k = 0, \quad k = 0, 1, \dots, m \quad (3.23)$$

denklemler sistemine dönüştürülmüş olur, burada λ_0 , (3.22) formülü ile ifade edilir (Şamilov, 2009). $j = 1, 2, \dots, m$ olmak üzere, (3.23) eşitliğinden $\lambda_1, \dots, \lambda_m$ değerlerinin μ_1, \dots, μ_m değerlerine bağlı olduğu görülmektedir.

3.2.2. Sürekli Rassal Değişkenler İçin Minimum Çapraz Entropi Yöntemi

Yoğunluk fonksiyonu $f(x)$ olan X rassal değişkeni için önsel dağılım $q(x)$ olsun. Bu durumda $f(x)$ ile $q(x)$ arasındaki Kullback-Leibler ölçümü

$$D(f : q) = \int_a^b f(x) \ln \frac{f(x)}{q(x)} dx \quad (3.24)$$

olarak tanımlanır.

MinxEnt yöntemi, (3.24) Kullback - Leibler ölçümünü

$$\int_a^b f(x) g_j(x) dx = \mu_j, \quad j = 1, 2, \dots, m \quad (3.25)$$

koşulları altında minimize eden dağılımın bulunması yöntemidir. Söz konusu minimizasyon problemi, koşullu ekstremum problemi olup *MaxEnt* yönteminde uygulanan Lagrange çarpanları yöntemiyle çözülebilir. (3.24), (3.25) koşullu ekstremum problemi,

$$U \equiv - \int_a^b f(x) \ln f(x) dx + \sum_{j=0}^m \lambda_j \left(\int_a^b f(x) g_j(x) dx - \mu_j \right) \quad (3.26)$$

yardımcı fonksiyonelinin mutlak ekstremum problemine dönüşür. (3.26) fonksiyonelinin ekstremumunun varlığı için gerek koşulu ifade eden Euler denklemi

$$F_f - \frac{d}{dx} F_{f'} = 0 \quad (3.27)$$

şeklindedir, burada

$$F = f(x) \ln \frac{f(x)}{q(x)} + \sum_{j=0}^m \lambda_j g_j(x) f(x) \quad (3.28)$$

ile ifade edilir ve bu fonksiyon yardımı ile (3.26) fonksiyoneli

$$U \equiv - \int_a^b (F(f, \lambda) + \sum_{j=0}^m \lambda_j \mu_j) dx - \sum_{j=0}^m \lambda_j \mu_j$$

olarak ifade edilir. F 'in ifadesinden görüldüğü gibi $F_{f'} = 0$ olur. Bu nedenle de $F_f = 0$ 'dır ve buradan

$$\begin{aligned} \ln \frac{f(x)}{q(x)} &= -1 - \sum_{j=0}^m \lambda_j g_j(x) \\ \frac{f(x)}{q(x)} &= e^{-1-\lambda_0-\sum_{j=1}^m \lambda_j g_j(x)} \\ f(x) &= q(x) e^{-\tilde{\lambda}_0-\sum_{j=1}^m \lambda_j g_j(x)}, \quad \tilde{\lambda}_0 = 1 + \lambda_0 \end{aligned} \quad (3.29)$$

sonucuna varılır.

(3.29) fonksiyonu (3.25) koşullarında yerine yazılarak $\tilde{\lambda}_0, \lambda_1, \dots, \lambda_m$ 'ler,

$$\int_a^b q(x) e^{-\tilde{\lambda}_0 - \sum_{j=1}^m \lambda_j g_j(x)} g_j(x) dx = \mu_j, \quad j = 0, 1, 2, \dots, m \quad (3.30)$$

lineer olmayan denklemler sisteminden bulunabilir. (3.30) 'da $j = 0$ olduğunda

$$\begin{aligned} \int_a^b q(x) e^{-\tilde{\lambda}_0 - \sum_{j=1}^m \lambda_j g_j(x)} g_j(x) dx &= 1 \\ e^{-\tilde{\lambda}_0} &= \frac{1}{\int_a^b q(x) e^{-\sum_{j=1}^m \lambda_j g_j(x)} dx} \\ \tilde{\lambda}_0 &= \ln \left\{ q(x) \int_a^b e^{-\sum_{j=1}^m \lambda_j g_j(x)} dx \right\} \end{aligned} \quad (3.31)$$

olarak elde edilir. $\tilde{\lambda}_0$ 'ın bu değeri (3.30) 'un $j = 0, 1, 2, \dots, m$ değerlerine uygun denklemlerinde yerine yazılarak, $\lambda = (\lambda_1, \lambda_2, \dots, \lambda_m)$ şeklinde gösterildiğinde,

$$F_k(\lambda) = \frac{\int_a^b q(x) e^{-\sum_{j=1}^m \lambda_j g_j(x)} g_k(x) dx}{\int_a^b q(x) e^{-\sum_{j=1}^m \lambda_j g_j(x)} dx} = \mu_k, \quad k = 1, 2, \dots, m \quad (3.32)$$

olur. (3.32) sistemi $\lambda_1, \lambda_2, \dots, \lambda_m$ bilinmeyenlerine göre lineer olmayan denklemler sistemidir ve görüldüğü gibi $\lambda_1, \lambda_2, \dots, \lambda_m$ 'ler integral işareti altındadır.

4. SONLU SAYIDA KARAKTERİZE EDİCİ MOMENTLER YARDIMI İLE *MinMaxEnt* ve *MaxMaxEnt* DAĞILIMLARI

4.1. *MinMaxEnt* ve *MaxMaxEnt* Dağılımlarının Tanımı

$K_0 = \{g_1, \dots, g_r\}$, r elemanlı karakterize edici moment fonksiyonlar kümesi olsun. K_0 'ın tüm m elemanlı kombinasyonlarının kümesi $K_{0,m}$ ile gösterilsin, burada $K_{0,m}$ 'in her bir elemanı m kısıtlı bir g vektörüdür.

MinMaxEnt ve *MaxMaxEnt* problemlerini çözmek için $(g_0(x), g^{(1)}(x))$, $(g_0(x), g^{(2)}(x))$ vektör fonksiyonlarını bulmak gerekir, burada $g_0(x) = 1$, $g^{(1)} \in K_{0,m}$ ve $g^{(2)} \in K_{0,m}$ sırasıyla, $U(g)$ 'ye Sahannon Entropi ölçümüne göre, minimum ve maksimum değerleri veren karakterize edici moment vektörleridir. $U(g)$ 'nin $g_0(x)$ ve tüm m -boyutlu $g(x)$, $g \in K_{0,m}$ vektör fonksiyonları tarafından üretilen kısıtlara göre minimum veya maksimum değerlere ulaştığına dikkat edilmelidir. Diğer yandan, $U(g)$ 'nin minimum değeri $g(x)$, $g \in K_{0,m}$ 'e uygun olan H_{max} değerinin en küçük, $U(g)$ 'nin maksimum değeri $g(x)$, $g \in K_{0,m}$ 'e uygun olan H_{max} değerinin en büyük değeridir. Eğer $(g_0, g^{(1)})$, $U(g)$ 'ye en küçük değeri veriyorsa o zaman $p^{(1)} = (p_1^{(1)}, \dots, p_n^{(1)})$ dağılımı $(g_0, g^{(1)})$ vektör fonksiyonuna göre veriye uygun olan dağılımdır ve *MinMaxEnt* Dağılımı olarak adlandırılır. Eğer $(g_0, g^{(2)}(x))$ de $U(g)$ 'ye en büyük değeri veriyorsa o zaman $p^{(2)} = (p_1^{(2)}, \dots, p_n^{(2)})$ dağılımı $(g_0, g^{(2)}(x))$ vektör fonksiyonuna göre veriye uygun olan dağılımdır ve *MaxMaxEnt* Dağılımı olarak adlandırılır.

Bu tezde, *MaxEnt* yöntemi için de K_0 'ın kısıtları $g_1(x) = x$, $g_2(x) = x^2$, $g_3(x) = \ln x$, $g_4(x) = (\ln x)^2$, $g_5(x) = \ln(1 + x^2)$ karakterize edici moment vektör fonksiyonları olarak seçilmiştir. Bir başka ifadeyle, $K_0 = \{g_1, \dots, g_5\}$ olarak ifade edilir.

4.2. *MinMaxEnt* ve *MaxMaxEnt* Yöntemleri ile Bağırsak Kanseri Verileri Analizi

Bu bölümde, sağkalım veri analizi, Genelleştirilmiş Entropi Optimizasyon Yöntemleri uygulanarak gerçekleştirilecektir. Bunun için 1935 – 1944 yılları arasında Connecticut şehrindeki bağırsak kanserine yakalanmış erkek hastaların verileri göz önünde bulundurulmuş (Deshpande ve Purohit, 2005) ve veriler Çizelge 4.1’de gösterilmiştir. *MinMaxEnt* ve *MaxMaxEnt* yöntemleri, sonlu sayıda karakterize edici moment fonksiyonları kümesi, Shannon entropi ölçümü ve Kullback-Leibler entropi ölçümü yardımıyla üretilen istatistiksel dağılımlar ile elde edilmiş ve bu yöntemlerin sağkalım veri analizi üzerine uygulanışı gösterilmiştir. Sonuçlar MATLAB 7.10.0 (R2010a) istatistiksel programları kullanılarak elde edilmiştir. Elde edilen Genelleştirilmiş Entropi Optimizasyon Dağılımlarının performans değerlendirmeleri ise Ki – Kare ve RMSE kriterleri ile Shannon Entropi ölçümü ve Kullback – Leibler ölçümü yardımıyla karşılaştırılmış ve sonuçlar yorumlanmıştır. Böylece, uygulanan yöntemler birbiri ile karşılaştırıldığında her birinin diğeri üzerinde bir anlamda avantajı olduğu halde başka bir anlamda dezavantajı olabildiği gözlenmiştir. Aşağıda, *MaxEnt* yönteminin yukarıda bahsedilen veriye uygulaması ifade edilmektedir.

Çizelge 4.1. 1935 – 1944 yılları arasında Connecticut şehrindeki erkek hastaların bağırsak kanseri verileri

Sağkalım süreleri (yıl)	Aralık başlangıcında sağkalan hasta sayısı	Aralıkta ölen hasta sayısı	Aralıkta sansüre uğrayan hasta sayısı
t	n_i	d_i	c_i
1	388	167	2
2	219	45	1
3	173	45	1
4	127	19	0
5	108	17	0
6	91	11	1
7	79	8	0
8	71	5	0
9	66	6	1
10	59	7	0

Burada zamansal sansüre uğrayan 52 hasta çalışmaya dâhil edilmeyerek *gözlenen ölüm olasılıkları* hesaplanmış ve bu olasılıkların toplamı 0.9821 olarak bulunmuştur. Bu değer 1'den küçük olması sebebiyle $1 - 0.9821 = 0.0179$ farkı, belirtilen aralıktaki her bir rassal sansürlü gözleme eşit olarak paylaştırılmış ve sansürlü gözlemlerin ait oldukları aralıktaki verilerin olasılıklarına eklenerek yeni olasılıklar elde edilmiştir. Elde edilen bu yeni olasılıklara ise “*düzeltilmiş olasılıklar*” adı verilmiştir ve sonuçlar Çizelge 4.2’de gösterilmiştir. Bu düzeltme yönteminin kabul edilebilirliğinin temeli önemli bir anlam içerir.

Çizelge 4.2. 1935 – 1944 yılları arasında Connecticut şehrindeki erkek hastaların bağırsak kanseri verilerinin Gözlenen Olasılıkları ve Düzeltilmiş Olasılıkları

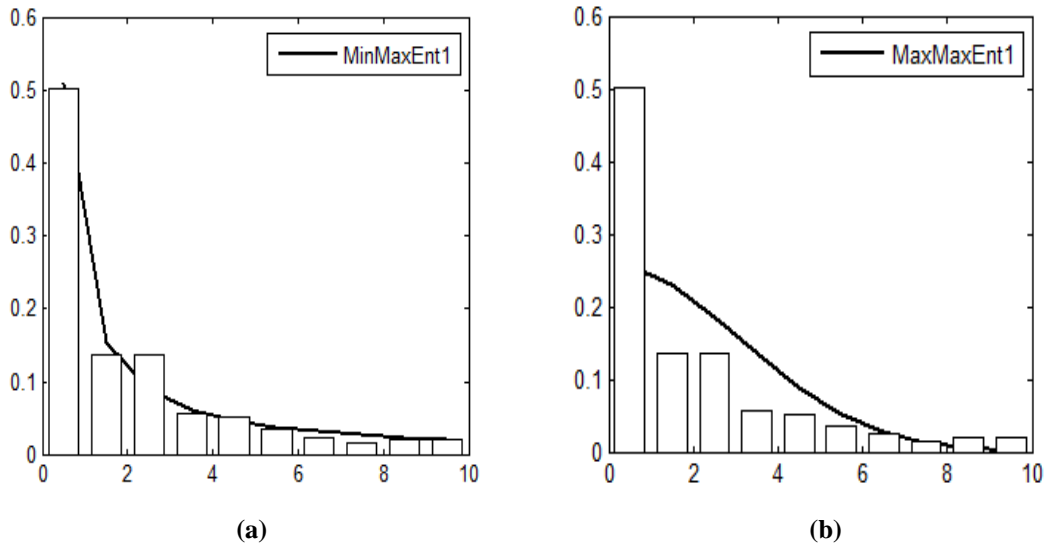
t	n_i	d_i	c_i	Gözlenen olasılıklar p_i	Düzeltilmiş Olasılıklar p_i^*
1	388	167	2	0.4970	0.5030
2	219	45	1	0.1339	0.1369
3	173	45	1	0.1339	0.1369
4	127	19	0	0.0565	0.0565
5	108	17	0	0.0506	0.0506
6	91	11	1	0.0327	0.0357
7	79	8	0	0.0238	0.0238
8	71	5	0	0.0149	0.0149
9	66	6	1	0.0179	0.0208
10	59	7	0	0.0208	0.0208

$m = 1, 2, \dots, 4$ ve $g_0(x) = 1$ olmak üzere (g_0, g) , $g \in K_{0,m}$, vektör fonksiyonlarına uygun tüm *MaxEnt* dağılımları ve H_{max} değerleri Çizelge 4.3 – 4.6 arasında verilmiştir. Bu çizelgelerde koyu yazılarak belirtilmiş $(MinMaxEnt)_m$, $(MaxMaxEnt)_m$, dağılımları da elde edilmiş ve elde edilen bu dağılımlar Şekilde 4.1 – 4.4 arasında gösterilmiştir.

Çizelge 4.3. Bağırsak kanseri verileri için (g_0, g) , $g_0(x) = 1$, $g \in K_{0,1}$ vektör fonksiyonu yardımıyla elde edilen *MaxEnt* Dağılımları ve H_{max} değerleri

(g_0, g)	(g_0, g_1)	(g_0, g_2)	(g_0, g_3)	(g_0, g_4)	(g_0, g_5)
<i>MaxEnt</i> Dağılımları	0.3805	0.2566	0.5115	0.2366	0.4594
	0.2370	0.2308	0.1541	0.2965	0.2104
	0.1476	0.1867	0.0882	0.1830	0.1092
	0.0919	0.1359	0.0611	0.1087	0.0667
	0.0573	0.0889	0.0464	0.0662	0.0453
	0.0357	0.0523	0.0373	0.0418	0.0331
	0.0222	0.0277	0.0311	0.0273	0.0254
	0.0138	0.0132	0.0266	0.0183	0.0202
	0.0086	0.0056	0.0232	0.0126	0.0165
0.0054	0.0022	0.0205	0.0089	0.0138	
H_{max}	2.4613	2.6554	2.3840	2.6471	2.3941

Çizelge 4.3'e bakıldığında *MinMaxEnt* ve *MaxMaxEnt* dağılımları sırasıyla $(g_0, g_3) = (1, \ln x)$ ve $(g_0, g_2) = (1, x^2)$ karakterize edici vektör fonksiyonları yardımıyla elde edilmiştir. Bu dağılımların Shannon entropi ölçüm değerleri ise $H((MinMaxEnt)_1) = 2.3840$ ve $H((MaxMaxEnt)_1) = 2.6554$ şeklindedir. Bu dağılımların grafikleri Şekil 4.1 'de gösterilmiştir.

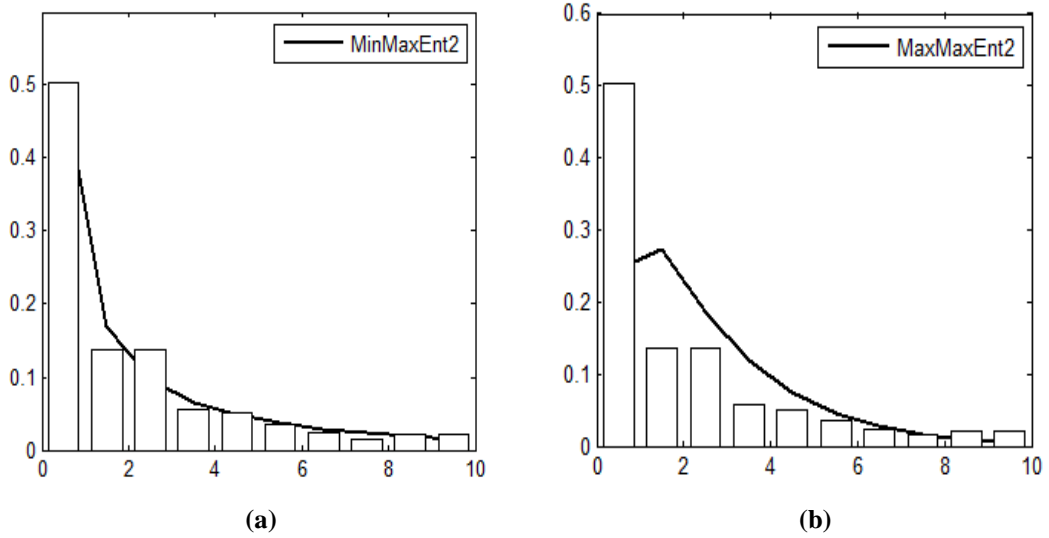


Şekil 4.1. Bağırsak kanseri verileri için (a) $(MinMaxEnt)_1$ Dağılımının Grafiği ve (b) $(MaxMaxEnt)_1$ Dağılımının Grafiği

Çizelge 4.4. Bağırsak kanseri verileri için (g_0, g) , $g_0(x) = 1$, $g \in K_{0,2}$ vektör fonksiyonu yardımıyla elde edilen $MaxEnt$ Dağılımları ve H_{max} değerleri

(g_0, g)	(g_0, g_1, g_2)	(g_0, g_1, g_3)	(g_0, g_1, g_4)	(g_0, g_1, g_5)	(g_0, g_2, g_3)
<i>MaxEnt</i> Dağılımları	0.4772	0.4948	0.5019	0.4861	0.5020
	0.2113	0.1762	0.1444	0.2012	0.1664
	0.1054	0.1013	0.1052	0.0981	0.0972
	0.0592	0.0672	0.0806	0.0592	0.0665
	0.0374	0.0478	0.0597	0.0410	0.0488
	0.0267	0.0354	0.0422	0.0312	0.0370
	0.0214	0.0269	0.0286	0.0253	0.0286
	0.0193	0.0209	0.0186	0.0216	0.0223
	0.0196	0.0165	0.0117	0.0190	0.0174
0.0225	0.0131	0.0072	0.0173	0.0136	
H_{max}	2.3937	2.3799	2.3908	2.3903	2.3802

(g_0, g)	(g_0, g_2, g_4)	(g_0, g_2, g_5)	(g_0, g_3, g_4)	(g_0, g_3, g_5)	(g_0, g_4, g_5)
<i>MaxEnt</i> Dağılımları	0.2490	0.4864	0.4973	0.4860	0.5022
	0.2811	0.2017	0.1767	0.1943	0.1818
	0.1864	0.0982	0.0982	0.1018	0.0936
	0.1160	0.0586	0.0643	0.0635	0.0594
	0.0709	0.0402	0.0459	0.0441	0.0427
	0.0428	0.0305	0.0347	0.0327	0.0331
	0.0255	0.0249	0.0273	0.0255	0.0270
	0.0149	0.0215	0.0221	0.0206	0.0228
	0.0086	0.0195	0.0183	0.0170	0.0198
0.0048	0.0185	0.0154	0.0144	0.0176	
H_{max}	2.6416	2.3904	2.3806	2.3815	2.3846



Şekil 4.2. Bağırsak kanseri verileri için (a) $(MinMaxEnt)_2$ Dağılımının grafiği ve (b) $(MaxMaxEnt)_2$ Dağılımının grafiği

Çizelge 4.4'e bakıldığında *MinMaxEnt* ve *MaxMaxEnt* dağılımları sırasıyla $(g_0, g_1, g_3) = (1, x, \ln x)$ ve $(g_0, g_2, g_4) = (1, x^2, (\ln x)^2)$ karakterize edici vektör fonksiyonları yardımıyla elde edilmiştir. Bu dağılımların Shannon entropi ölçüm değerleri ise $H((MinMaxEnt)_1) = 2.3799$ ve $H((MaxMaxEnt)_1) = 2.6416$ şeklindedir. Bu dağılımların grafikleri Şekil 4.2'de gösterilmiştir.

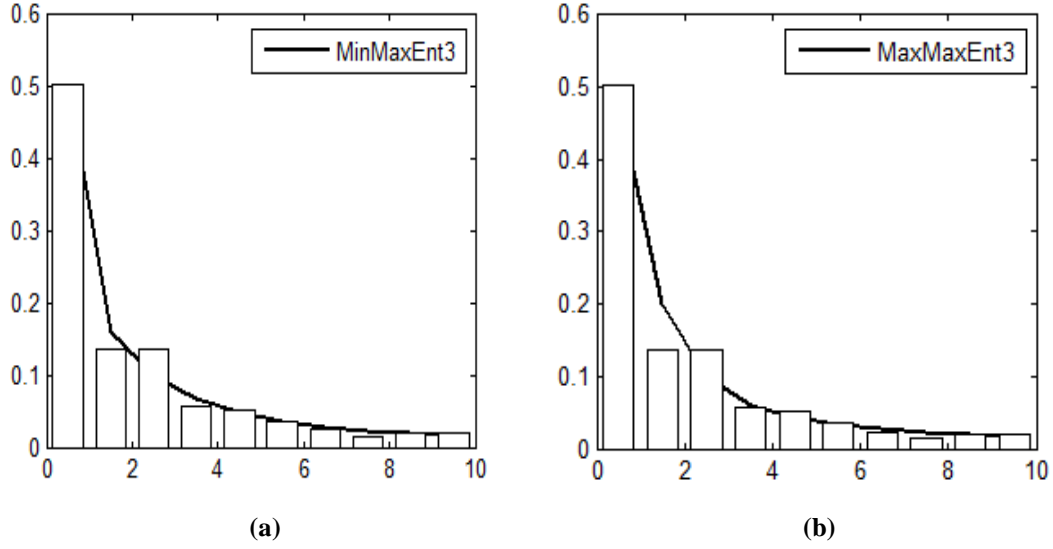
Çizelge 4.5. Bağırsak kanseri verileri için (g_0, g) , $g_0(x) = 1$, $g \in K_{0,3}$ vektör fonksiyonu yardımıyla elde edilen *MaxEnt* Dağılımları ve H_{max} değerleri

(g_0, g)	(g_0, g_1, g_2, g_3)	(g_0, g_1, g_2, g_4)	(g_0, g_1, g_2, g_5)	(g_0, g_1, g_3, g_4)	(g_0, g_1, g_3, g_5)
	0.4963	0.5011	0.4760	0.4997	0.5007
	0.1707	0.1612	0.2004	0.1620	0.1592
	0.0976	0.0996	0.0992	0.0976	0.0980
<i>MaxEnt</i>	0.0652	0.0679	0.0607	0.0678	0.0687
Dağılımları	0.0472	0.0485	0.0427	0.0500	0.0507
	0.0360	0.0360	0.0328	0.0380	0.0383
	0.0284	0.0278	0.0268	0.0294	0.0294
	0.0231	0.0224	0.0229	0.0230	0.0229
	0.0192	0.0189	0.0202	0.0182	0.0179
	0.0163	0.0167	0.0182	0.0144	0.0141
H_{max}	2.3798	2.3761	2.3903	2.3794	2.3790

(g_0, g)	(g_0, g_2, g_3, g_5)	(g_0, g_2, g_3, g_5)	(g_0, g_2, g_3, g_5)	(g_0, g_2, g_3, g_5)	(g_0, g_2, g_3, g_5)
	0.4994	0.4986	0.5001	0.4972	0.5012
	0.1620	0.1664	0.1631	0.1687	0.1562
	0.0973	0.0965	0.0962	0.0952	0.0998
<i>MaxEnt</i>	0.0679	0.0659	0.0666	0.0651	0.0705
Dağılımları	0.0504	0.0485	0.0494	0.0486	0.0514
	0.0384	0.0373	0.0380	0.0379	0.0383
	0.0297	0.0293	0.0297	0.0300	0.0291
	0.0231	0.0234	0.0235	0.0238	0.0224
	0.0179	0.0189	0.0186	0.0188	0.0174
	0.0140	0.0153	0.0148	0.0147	0.0138
H_{max}	2.3801	2.3802	2.3802	2.3822	2.3779

Çizelge 4.5'e bakıldığında *MinMaxEnt* ve *MaxMaxEnt* dağılımları $(g_0, g_1, g_2, g_4) = (1, x, x^2, (\ln x)^2)$ ve $(g_0, g_1, g_2, g_5) = (1, x, x^2, \ln(1 + x^2))$ karakterize edici vektör fonksiyonları yardımıyla elde edilmiştir. Bu dağılımların Shannon entropi ölçüm değerleri $H((MinMaxEnt)_1) = 2.3761$ ve

$H((MaxMaxEnt)_1) = 2.3903$ şeklindedir. Bu dağılımların grafikleri Şekil 4.3'te resmedilmiştir.



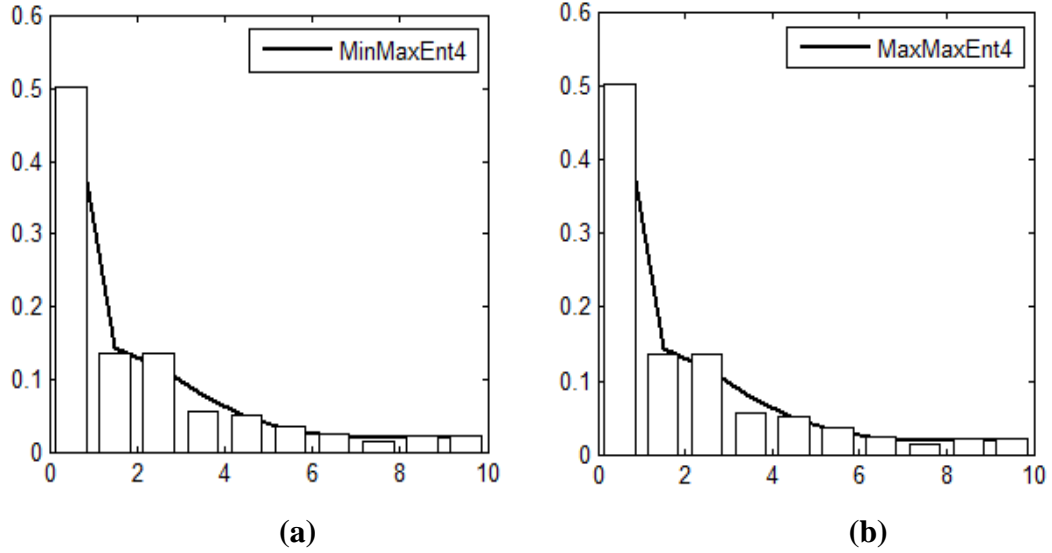
Şekil 4.3. Bağırsak kanseri verileri için (a) $(MinMaxEnt)_3$ Dağılımının grafiği ve (b) $(MaxMaxEnt)_3$ Dağılımının grafiği

Çizelge 4.6. Bağırsak kanseri verileri için (g_0, g) , $g_0(x) = 1$, $g \in K_{0,4}$ vektör fonksiyonu yardımıyla elde edilen $MaxEnt$ Dağılımları ve H_{max} değerleri

(g_0, g)	$(g_0, g_1, g_2, g_3, g_4)$	$(g_0, g_1, g_2, g_3, g_5)$	$(g_0, g_1, g_2, g_4, g_5)$	$(g_0, g_1, g_3, g_4, g_5)$	$(g_0, g_2, g_3, g_4, g_5)$
	0.5024	0.5025	0.5024	0.5026	0.5025
	0.1442	0.1438	0.1439	0.1428	0.1433
	0.1151	0.1157	0.1155	0.1176	0.1166
$MaxEnt$	0.0780	0.0781	0.0781	0.0780	0.0781
Dağılımları	0.0488	0.0485	0.0486	0.0473	0.0480
	0.0312	0.0310	0.0310	0.0304	0.0307
	0.0220	0.0220	0.0220	0.0221	0.0220
	0.0180	0.0182	0.0181	0.0188	0.0184
	0.0180	0.0181	0.0181	0.0187	0.0184
	0.0224	0.0222	0.0223	0.0217	0.0220
H_{max}	2.3687	2.3686	2.3686	2.3682	2.3684

Çizelge 4.6'ya bakıldığında $MinMaxEnt$ ve $MaxMaxEnt$ dağılımları $(g_0, g_1, g_3, g_4, g_5) = (1, x, \ln x, (\ln x)^2, \ln(1 + x^2))$ ve $(g_0, g_1, g_2, g_3, g_4) = (1, x, x^2, \ln x, (\ln x)^2)$ karakterize edici vektör fonksiyonları yardımıyla elde

edilmiştir. Bu dağılımların Shannon entropi ölçüm değerleri $H((MinMaxEnt)_1) = 2.3682$ ve $H((MaxMaxEnt)_1) = 2.3687$ şeklindedir. Bu dağılımların grafikleri Şekil 4.4'te gösterilmiştir.



Şekil 4.4. Bağırsak kanseri verileri için (a) $(MinMaxEnt)_4$ Dağılımının grafiği ve (b) $(MaxMaxEnt)_4$ Dağılımının grafiği

Elde edilen $MinMaxEnt$ Dağılımları Çizelge 4.7'de ve $MaxMaxEnt$ Dağılımları Çizelge 4.8'de gösterilmiştir.

Çizelge 4.7. Bağırsak kanseri verileri için $(MinMaxEnt)_m$, $m = 1, \dots, 4$ Dağılımları

t	d_i	c_i	p_i^*	$(MinMaxEnt)_1$	$(MinMaxEnt)_2$	$(MinMaxEnt)_3$	$(MinMaxEnt)_4$
1	167	2	0.5030	0.5115	0.4948	0.5011	0.5026
2	45	1	0.1369	0.1541	0.1762	0.1612	0.1428
3	45	1	0.1369	0.0882	0.1013	0.0996	0.1176
4	19	0	0.0565	0.0611	0.0672	0.0679	0.0780
5	17	0	0.0506	0.0464	0.0478	0.0485	0.0473
6	11	1	0.0357	0.0373	0.0354	0.0360	0.0304
7	8	0	0.0238	0.0311	0.0269	0.0278	0.0221
8	5	0	0.0149	0.0266	0.0209	0.0224	0.0188
9	6	1	0.0208	0.0232	0.0165	0.0189	0.0187
10	7	0	0.0208	0.0205	0.0131	0.0167	0.0217

Çizelge 4.8. Bağırsak kanseri verileri için $(MaxMaxEnt)_m$, $m = 1, \dots, 4$ Dağılımları

t	d_i	c_i	p_i^*	$(MaxMaxEnt)_1$	$(MaxMaxEnt)_2$	$(MaxMaxEnt)_3$	$(MaxMaxEnt)_4$
1	167	2	0.5030	0.2566	0.2490	0.4760	0.5024
2	45	1	0.1369	0.2308	0.2811	0.2004	0.1442
3	45	1	0.1369	0.1867	0.1864	0.0992	0.1151
4	19	0	0.0565	0.1359	0.1160	0.0607	0.0780
5	17	0	0.0506	0.0889	0.0709	0.0427	0.0488
6	11	1	0.0357	0.0523	0.0428	0.0328	0.0312
7	8	0	0.0238	0.0277	0.0255	0.0268	0.0220
8	5	0	0.0149	0.0132	0.0149	0.0229	0.0180
9	6	1	0.0208	0.0056	0.0086	0.0202	0.0180
10	7	0	0.0208	0.0022	0.0048	0.0182	0.0224

Bu sonuçlardan yola çıkarak $MaxEnt$ Dağılımlarının veriye uygunluklarını araştırmak için uyum testi yapılmıştır. Bu bağlamda, bu dağılımların veriye uyumunu elde etmek için Ki – Kare ve RMSE kriterleri ile Shannon entropi ölçümü ve Kullback-Leibler ölçümü gibi çeşitli kriterler kullanılmıştır. Burada, Kullback - Leibler ölçümü

$$D(p; q) = \sum_{i=1}^{10} p_i \ln \frac{p_i}{q_i}$$

şeklindedir. Bu ölçüm, $m = 1, \dots, 4$ olmak üzere, $q = (p_1, p_2, \dots, p_{10})$ gözlenen olasılıkları ile elde edilen tüm $(MinMaxEnt)_m$ ve $(MaxMaxEnt)_m$ dağılımları arasındaki Kullback – Leibler mesafesi, sırasıyla $D(p_{min}^{(m)}; q)$ ve $D(p_{max}^{(m)}; q)$ olarak hesaplanmış ve sonuçlar Çizelge 4.9 ve Çizelge 4.10’de gösterilmiştir. Burada Kullback – Leibler ölçümünü kullanmadaki amaç, $MaxEnt$ yöntemi ile elde edilmiş dağılımların, veriden elde edilen gözlenen olasılıklara ne kadar yakın olduklarını ifade etmektir. Hatta böylece, sansürlü verinin, $MaxEnt$ yöntemindeki önemi dikkate alınmış olur.

Çizelge 4.9. Bağırsak kanseri verileri için, $m = 1, 2, \dots, 4$ olmak üzere, $(MinMaxEnt)_m$ Dağılımları için Shannon entropi ölçüm değerleri H , Kullback – Leibler ölçüm değerleri $D(p_{min}^{(m)}; q)$, Ki – Kare ve RMSE değerleri

<i>MinMaxEnt</i> Dağılımları	H	$D(p_{min}^{(m)}; q)$	Ki – Kare hesap değeri	Ki – Kare tablo değeri	RMSE
$(MinMaxEnt)_1$	2.3840	0.0500	0.7740	$\chi_{8, \alpha}^2 = 15.51$	0.1680
$(MinMaxEnt)_2$	2.3799	0.0471	0.6512	$\chi_{7, \alpha}^2 = 14.07$	0.1841
$(MinMaxEnt)_3$	2.3761	0.0422	0.5025	$\chi_{6, \alpha}^2 = 12.59$	0.1374
$(MinMaxEnt)_4$	2.3682	0.0343	0.2458	$\chi_{5, \alpha}^2 = 11.07$	0.0984

Not: Burada $p_{min}^{(m)} = (MinMaxEnt)_m$, $m = 1, \dots, 4$ ifadesidir.

Çizelge 4.10. Bağırsak kanseri verileri için, $m = 1, 2, \dots, 4$ olmak üzere, $(MaxMaxEnt)_m$ Dağılımları için Shannon entropi ölçüm değeri H , Kullback – Leibler ölçüm değeri $D(p_{max}^{(m)}; q)$, Ki – Kare ve RMSE değerleri

<i>MaxMaxEnt</i> Dağılımları	H	$D(p_{max}^{(m)}; q)$	Ki – Kare hesap değeri	Ki – Kare tablo değeri	RMSE
$(MaxMaxEnt)_1$	2.6554	0.2931	11.6661	$\chi_{8, \alpha}^2 = 15.51$	1.9600
$(MaxMaxEnt)_2$	2.6416	0.2962	9.5441	$\chi_{7, \alpha}^2 = 14.07$	0.8416
$(MaxMaxEnt)_3$	2.3903	0.0588	0.8717	$\chi_{6, \alpha}^2 = 12.59$	0.1465
$(MaxMaxEnt)_4$	2.3687	0.0347	0.2598	$\chi_{5, \alpha}^2 = 11.07$	0.0981

Not: Burada $p_{max}^{(m)} = (MaxMaxEnt)_m$, $m = 1, \dots, 4$ ifadesidir.

Çizelge 4.9 ve Çizelge 4.10 dikkate alındığına RMSE kriterine göre istatistiksel veri için her bir $(MinMaxEnt)_m$ ($m = 1, 2, 3$) dağılımı $(MaxMaxEnt)_m$ ($m = 1, 2, 3$) dağılımından daha iyidir ancak $(MaxMaxEnt)_4$ dağılımı veriye, $(MinMaxEnt)_4$ dağılımından daha uygundur.

Aynı çizelgelerde tüm $(MinMaxEnt)_m$ ve $(MaxMaxEnt)_m$ dağılımları Shannon entropi ölçümü ve Kullback – Leibler ölçümüne göre incelendiğinde kısıt sayısı arttıkça H ve $D(p; q)$ değerlerinin gitgide azaldığı ve $(MinMaxEnt)_m$ dağılımlarının $(MaxMaxEnt)_m$ dağılımlarına gittikçe

yaklaştığı hatta, $(MinMaxEnt)_4$ dağılımının $(MaxMaxEnt)_4$ dağılımına neredeyse eşit olduğu söylenir. Böylece, Shannon entropi ölçümü için aşağıdaki *Sonuç 4.1* verilebilir:

Sonuç 4.1. $m = 1, \dots, 4$ olmak üzere $(MaxMaxEnt)_m$ ve $(MinMaxEnt)_m$ sırasıyla, farklı $g \in K_{0,m}$ olmakla $g(x)$ moment fonksiyonu tarafından üretilen m moment kısıtlı $MaxMaxEnt$ ve $MinMaxEnt$ dağılımlarını gösterebilir, $m_1 < m_2$ olduğu zaman

$$H((MaxMaxEnt)_{m_1}) > H((MaxMaxEnt)_{m_2})$$

$$H((MinMaxEnt)_{m_1}) > H((MinMaxEnt)_{m_2})$$

eşitsizlikleri elde edilir. Diğer yandan, moment kısıt sayısı arttıkça hem $MaxMaxEnt$ hem de $MinMaxEnt$ dağılımlarının entropi değerleri azalır. Dahası herhangi bir m için

$$H((MaxMaxEnt)_m) > H((MinMaxEnt)_m)$$

eşitsizliği elde edilir.

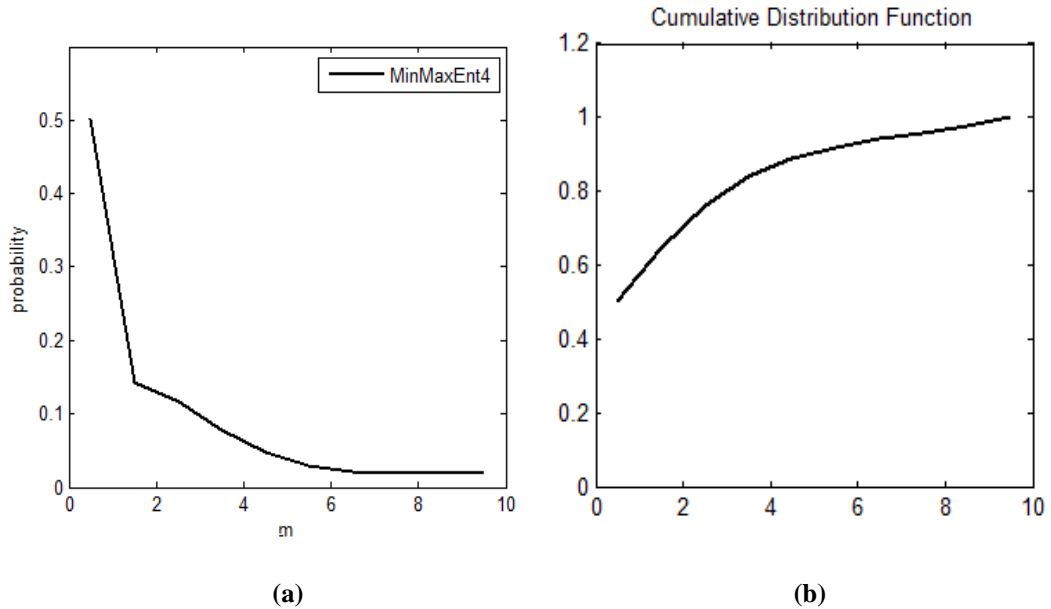
4.3. $MinMaxEnt$ ve $MaxMaxEnt$ Yöntemleri Yöntemleri ile Bağlı Kanser Verileri için Yaşam Süresi Rassal Değişkenlerinin Karakterize Edici Fonksiyonları

Bu bölümde sağkalım veri analizi, $m = 1, 2, \dots, 4$ olmak üzere, $(MinMaxEnt)_m$ ve $(MaxMaxEnt)_m$ dağılımları arasından istatistiksel veriyi en iyi biçimde ifade eden $(MinMaxEnt)_4$ ve $(MaxMaxEnt)_4$ dağılımları ile yapılmıştır. Elde edilen Genelleştirilmiş Entropi Optimizasyon Dağılımlarına göre Sağkalım fonksiyonları olan olasılık yoğunluk fonksiyonu $\hat{f}(t)$, birikimli dağılım fonksiyonu $\hat{F}(t)$, Sağkalım fonksiyonu $\hat{S}(t)$ ve hazard (risk) fonksiyonu $\hat{h}(t)$ fonksiyonları tahmin edilmiş ve grafikleri çizilmiştir.

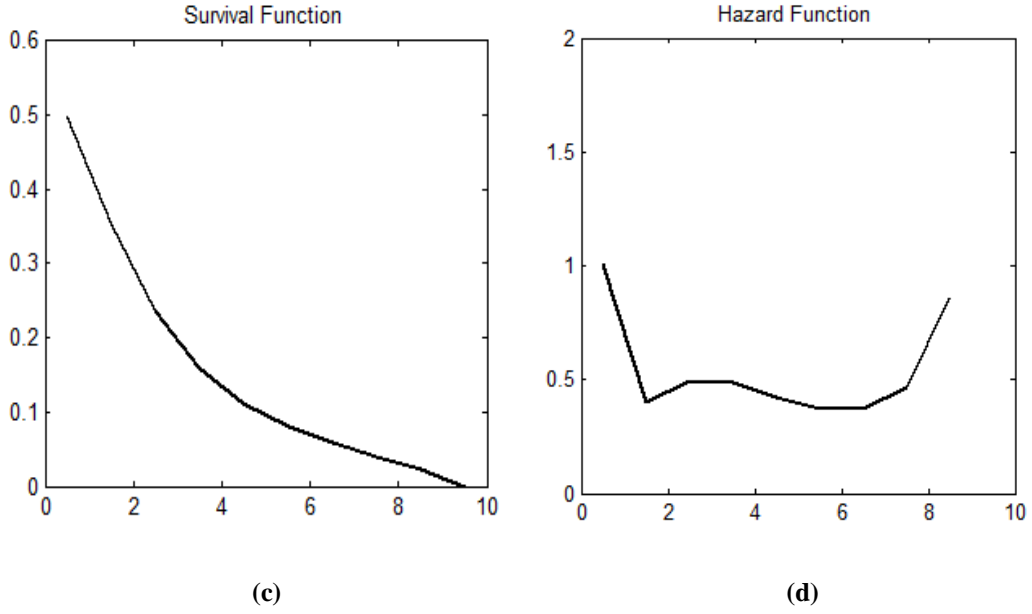
Çizelge 4.11’de $(MinMaxEnt)_4$ dağılımı için tahmini Olasılık Yoğunluk Fonksiyonu $\hat{f}(t)$, Birikimli Dağılım Fonksiyonu $\hat{F}(t)$, Sağkalım Fonksiyonu $\hat{S}(t)$ ve Hazard (Risk) Fonksiyonu $\hat{h}(t)$ değerleri verilmiştir. Bu dağılımlara göre çizilen $\hat{F}(t)$, $\hat{S}(t)$ ve $\hat{h}(t)$ ‘nin grafikleri Şekil 4.5’te gösterilmiştir.

Çizelge 4.11. Bağırsak kanseri verileri için $(MinMaxEnt)_4$ Dağılımı ile yapılan Sağkalım Analizi

t	n_i	d_i	c_i	$\hat{f}(t) = (MinMaxEnt)_4$	$\hat{F}(t)$	$\hat{S}(t)$	$\hat{h}(t) = \hat{f}(t)/\hat{S}(t)$
1	388	167	2	0.5026	0.5026	0.4974	1.0105
2	219	45	1	0.1428	0.6474	0.3526	0.4050
3	173	45	1	0.1176	0.7650	0.2350	0.5004
4	127	19	0	0.0780	0.8430	0.1570	0.4968
5	108	17	0	0.0473	0.8903	0.1097	0.4311
6	91	11	1	0.0304	0.9207	0.0793	0.3833
7	79	8	0	0.0221	0.9428	0.0572	0.3863
8	71	5	0	0.0188	0.9616	0.0384	0.4896
9	66	6	1	0.0187	0.9803	0.0197	0.9492
10	59	7	0	0.0217	1.0000	0.0000	--



Şekil 4.5. Bağırsak kanseri verileri için $(MinMaxEnt)_4$ Dağılımına uygun (a) $\hat{f}(t)$ Fonksiyonunun Grafiği (b) $\hat{F}(t)$ Fonksiyonunun Grafiği,

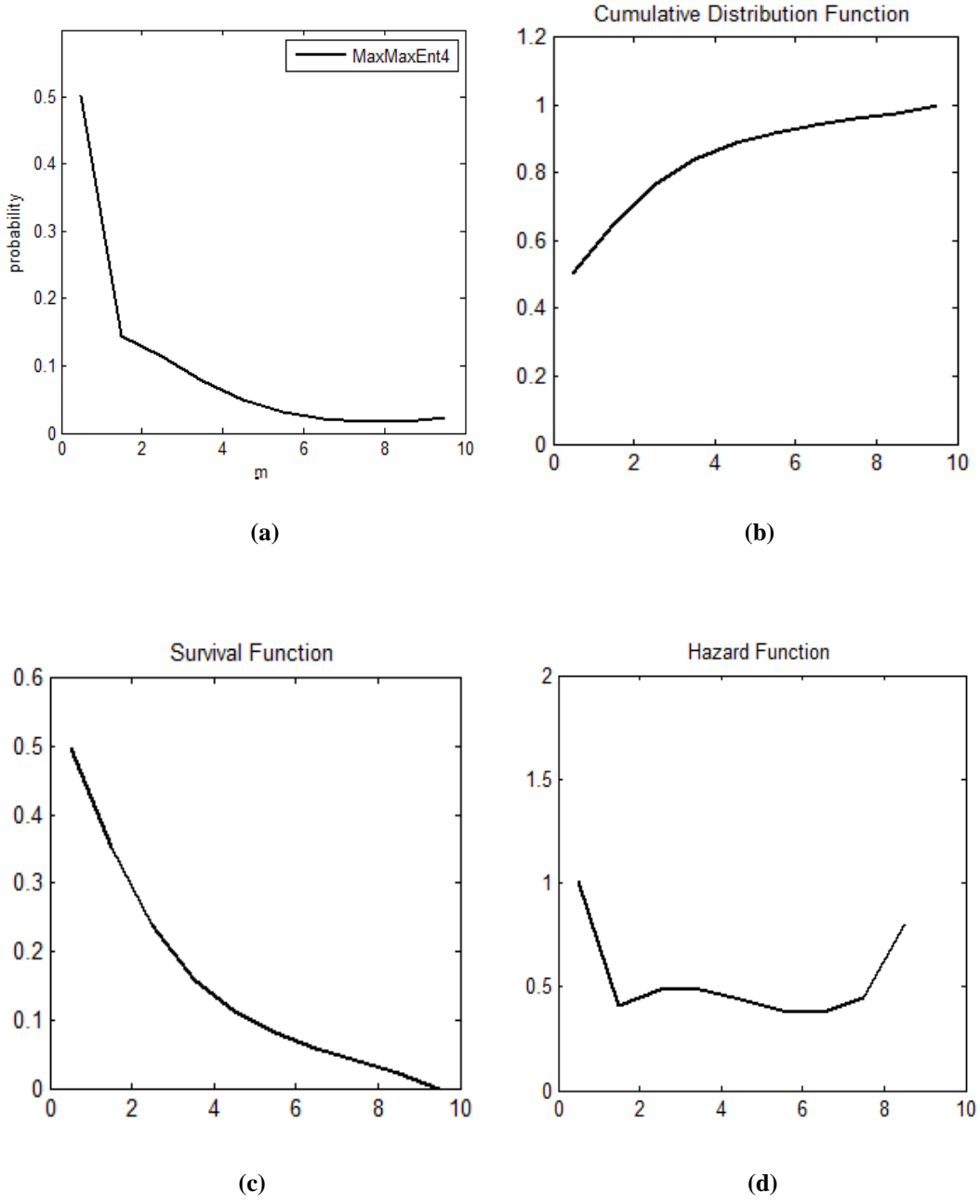


Şekil 4.6 (devam). Bağırsak kanseri verileri için $(MinMaxEnt)_4$ Dağılımına uygun (c) $\hat{S}(t)$ Fonksiyonunun Grafiği ve (d) $\hat{h}(t)$ Fonksiyonunun Grafiği

Çizelge 4.12’de $(MaxMaxEnt)_4$ Dağılımı için $\hat{f}(t)$, $\hat{F}(t)$, $\hat{S}(t)$ ve $\hat{h}(t)$ değerleri verilmiş ve bu dağılımlara göre çizilen $\hat{F}(t)$, $\hat{S}(t)$ ve $\hat{h}(t)$ ‘nin grafikleri Şekil 4.6’da gösterilmiştir.

Çizelge 4.12. Bağırsak kanseri verileri için $(MaxMaxEnt)_4$ Dağılımı ile yapılan Sağkalım Analizi

t	n_i	d_i	c_i	$\hat{f}(t) = (MaxMaxEnt)_4$	$\hat{F}(t)$	$\hat{S}(t)$	$\hat{h}(t) = \hat{f}(t)/\hat{S}(t)$
1	388	167	2	0.5024	0.5026	0.4974	1.0105
2	219	45	1	0.1442	0.6474	0.3526	0.4050
3	173	45	1	0.1151	0.7650	0.2350	0.5004
4	127	19	0	0.0780	0.8430	0.1570	0.4968
5	108	17	0	0.0488	0.8903	0.1097	0.4311
6	91	11	1	0.0312	0.9207	0.0793	0.3833
7	79	8	0	0.0220	0.9428	0.0572	0.3863
8	71	5	0	0.0180	0.9616	0.0384	0.4896
9	66	6	1	0.0180	0.9803	0.0197	0.9492
10	59	7	0	0.0224	1.0000	0.0000	--



Şekil 4.7. Bağırsak kanseri verileri için $(MaxMaxEnt)_4$ Dağılımına uygun (a) $\hat{f}(t)$ Fonksiyonunun Grafiği (b) $\hat{F}(t)$ Fonksiyonunun Grafiği, (c) $\hat{S}(t)$ Fonksiyonunun Grafiği ve (d) $\hat{h}(t)$ Fonksiyonunun Grafiği

4.4. *MinMinxEnt* ve *MaxMinxEnt* Yöntemleri ile Motor Arızası Verileri Analizi

Başka bir sağkalım veri seti için *MinMaxEnt* ve *MaxMaxEnt* yöntemlerinin uygulamaları Bölüm 4.3'te olduğu gibi benzer hesaplamalar yapılarak incelenmiştir. Bu incelemede ise veri seti olarak 1980 yılına ait motor arızası verileri alınmış ve bu sağkalım verileri Çizelge 4.13'te gösterilmiştir.

Burada, başlangıçta 200 sağkalım verisi olarak deneye başlanmış ancak deney sırasında 97 gözlem zamansal sansüre uğramış, geride kalan verilerin gözlenen q_i olasılıkları hesaplanmış ve gözlenen olasılıklar toplamı 0.8155 olarak bulunmuştur. Bu toplamın 1'den küçük olması sebebiyle $1 - 0.8155 = 0.1845$ farkı, her bir rassal sansürlü gözleme eşit olarak dağıtılmış ve böylece gözlenen olasılıklar 1'e tamamlanarak *düzeltilmiş* p_i^* olasılıkları hesaplanmıştır. Gözlenen olasılıklar ve düzeltilmiş olasılıklar Çizelge 4.14'te verilmiştir.

Çizelge 4.13. Motor arızası verileri (1980)

Sağkalım Zamanı (yıl)	Aralık Başında Çalışan Motor Sayısı	Aralık Boyunca Bozulan Motor Sayısı	Aralık Boyunca Sansüre Uğrayan Motor Sayısı
t	n_i	d_i	c_i
0-1	200	5	0
1-2	195	10	1
2-3	184	12	5
3-4	167	8	2
4-5	157	10	0
5-6	147	15	6
6-7	126	9	3
7-8	114	8	1
8-9	105	4	0
9-10	101	3	1

Çizelge 4.14. Motor arızası verileri için Gözlenen ve Düzeltilmiş Olasılıklar

t	n_i	d_i	c_i	Gözlenen Olasılıklar q_i	Düzeltilmiş Olasılıklar p_i^*
1	388	167	2	0.0485	0.0485
2	219	45	1	0.0971	0.1068
3	173	45	1	0.1165	0.1650
4	127	19	0	0.0777	0.0971
5	108	17	0	0.0971	0.0971
6	91	11	1	0.1456	0.2039
7	79	8	0	0.0874	0.1165
8	71	5	0	0.0777	0.0874
9	66	6	1	0.0388	0.0388
10	59	7	0	0.0291	0.0388

Bu veri seti için de *MaxEnt* yöntemini uygulamada K_0 'ın kısıtları olan karakterize edici moment vektör fonksiyonları olarak $g_1(x) = x$, $g_2(x) = x^2$, $g_3(x) = \ln x$, $g_4(x) = (\ln x)^2$, $g_5(x) = \ln(1 + x^2)$ seçilmiştir. Böylece $K_0 = \{g_1, \dots, g_5\}$. $g_0(x) = 1$ ve $g \in K_{0,m}$, $m = 1, 2, \dots, 4$ olmak üzere (g_0, g) 'ye karşılık gelen tüm $(MinMaxEnt)_m$, $(MaxMaxEnt)_m$, $m = 1, 2, \dots, 4$ dağılımları bir önceki bölümde olduğu gibi hesaplanmış ve bu dağılımlar Çizelge 4.15 ve Çizelge 4.16'da gösterilmiştir.

Çizelge 4.15. Motor arızası verileri için $(MinMaxEnt)_m$, $m = 1, \dots, 4$ Dağılımları

t	d_i	c_i	p_i^*	$(MinMaxEnt)_1$	$(MinMaxEnt)_2$	$(MinMaxEnt)_3$	$(MinMaxEnt)_4$
1	5	0	0.0485	0.1269	0.0483	0.0489	0.0475
2	10	1	0.1068	0.1249	0.1036	0.1034	0.1217
3	12	5	0.1650	0.1210	0.1363	0.1314	0.1192
4	8	2	0.0971	0.1153	0.1493	0.1458	0.1292
5	10	0	0.0971	0.1081	0.1456	0.1466	0.1422
6	15	6	0.2039	0.0997	0.1296	0.1342	0.1421
7	9	3	0.1165	0.0905	0.1068	0.1117	0.1225
8	8	1	0.0874	0.0809	0.0820	0.0844	0.0898
9	4	0	0.0388	0.0711	0.0589	0.0578	0.0560
10	3	1	0.0388	0.0615	0.0397	0.0358	0.0298

Çizelge 4.16. Motor arızası verileri için $(MaxMaxEnt)_m$, $m = 1, \dots, 4$ Dağılımları

t	d_i	c_i	p_i^*	$(MaxMaxEnt)_1$	$(MaxMaxEnt)_2$	$(MaxMaxEnt)_3$	$(MaxMaxEnt)_4$
1	5	0	0.0485	0.0989	0.0863	0.0501	0.0477
2	10	1	0.1068	0.0995	0.1449	0.0986	0.1198
3	12	5	0.1650	0.0998	0.1299	0.1416	0.1221
4	8	2	0.0971	0.1000	0.1126	0.1543	0.1306
5	10	0	0.0971	0.1001	0.0999	0.1436	0.1397
6	15	6	0.2039	0.1002	0.0914	0.1227	0.1395
7	9	3	0.1165	0.1003	0.0861	0.0999	0.1230
8	8	1	0.0874	0.1004	0.0833	0.0792	0.0921
9	4	0	0.0388	0.1004	0.0824	0.0619	0.0570
10	3	1	0.0388	0.1005	0.0832	0.0480	0.0284

Elde edilen dağılımların performansını incelemek için Ki - Kare, RMSE, Shannon entropi ölçümü ve Kullback - Leibler ölçümü gibi çeşitli kriterler kullanılmış ve sonuçlar Çizelge 4.17 ve Çizelge 4.18'de gösterilmiştir.

Çizelge 4.17. Motor arızası verileri için $(MinMaxEnt)_m$, $m = 1, 2, \dots, 4$ için hesaplanan bazı kriter sonuçları

$MinMaxEnt$ Dağılımları	H	$D(p_{min}^{(m)}; q)$	Ki - Kare Hesap Değeri	Ki - Kare Tablo Değeri	RMSE
$(MinMaxEnt)_1$	3.2854	0.3938	4.4310	$\chi_{8, \alpha}^2 = 15.51$	0.3158
$(MinMaxEnt)_2$	3.2041	0.3348	0.6512	$\chi_{7, \alpha}^2 = 14.07$	0.1873
$(MinMaxEnt)_3$	3.2000	0.3300	1.7787	$\chi_{6, \alpha}^2 = 12.59$	0.1799
$(MinMaxEnt)_4$	3.1932	0.3193	1.6161	$\chi_{5, \alpha}^2 = 11.07$	0.1830

Çizelge 4.18. Motor arızası verileri için $(MaxMaxEnt)_m$, $m = 1, 2, \dots, 4$ için hesaplanan bazı kriter sonuçları

$MaxMaxEnt$ Dağılımları	H	$D(p_{max}^{(m)}; q)$	Ki - Kare Hesap Değeri	Ki - Kare Tablo Değeri	RMSE
$(MaxMaxEnt)_1$	3.3219	0.4441	5.3820	$\chi_{8, \alpha}^2 = 15.51$	0.3492
$(MaxMaxEnt)_2$	3.2921	0.4009	4.9233	$\chi_{7, \alpha}^2 = 14.07$	0.3492
$(MaxMaxEnt)_3$	3.2155	0.3457	2.2804	$\chi_{6, \alpha}^2 = 12.59$	0.2104
$(MaxMaxEnt)_4$	3.1937	0.3198	1.6383	$\chi_{5, \alpha}^2 = 11.07$	0.1888

Çizelge 4.17 ve Çizelge 4.18'den görülüyor ki tüm $(MinMaxEnt)_m$ ve $(MaxMaxEnt)_m$, $m = 1,2,3,4$ dağılımları verilen sağkalım verileri için Ki-Kare kriteri anlamında uygundur.

RMSE kriterine göre verilen sağkalım veri setini, $m = 1,2,3,4$ olmak üzere, her bir $(MinMaxEnt)_m$ dağılımı, kendi kısıt sayısı ile eşit kısıta sahip olan $(MaxMaxEnt)_m$ dağılımından daha iyi temsil etmektedir. Buna ek olarak, istatistiksel veri setine $(MinMaxEnt)_1$ dağılımı $(MaxMaxEnt)_1$ ve $(MaxMaxEnt)_2$ dağılımlarından daha yakın bir dağılım iken $m = 2,3,4$ olmak üzere, her bir $(MinMaxEnt)_m$ dağılımı tüm $(MaxMaxEnt)_m$ ($m = 1, \dots, 4$) dağılımlarından daha iyidir. Bu sonuçlar gösteriyor ki, $m = 1, \dots, 4$ olmak üzere, $(MinMaxEnt)_m$ dağılımları arasından $(MinMaxEnt)_3$ dağılımı en uygun, $(MaxMaxEnt)_m$ dağılımları arasından da $(MaxMaxEnt)_4$ dağılımı en elverişli dağılımdır.

Shannon entropi ölçümü ve Kullback – Leibler anlamında sağkalım verilerine GEOD'un kullanılabilirliğini belirlemek amacıyla GEOD ölçüm değerlerini dikkate almak gereklidir. Bu sebeple, tüm GEOD için Shannon entropi ölçüm ve Kullback – Leibler ölçüm değerleri karşılaştırıldığında tüm $MaxEnt$ dağılımları arasından $(MinMaxEnt)_4$ dağılımının en iyi olduğu sonucuna varılır.

Bu sonuçlara göre, $(MinMaxEnt)_3$ ve $(MaxMaxEnt)_4$ dağılımları için Sağkalım Analizi Fonksiyonları olan Olasılık Yoğunluk Fonksiyonu $\hat{f}(t)$, Birikimli Dağılım Fonksiyonu $\hat{F}(t)$, Sağkalım Fonksiyonu $\hat{S}(t)$ ve Hazard Fonksiyonu $\hat{h}(t)$ belirlenmiş ve sonuçlar Çizelge 4.19 ve Çizelge 4.20'de verilmiştir. Ayrıca, bu fonksiyonların grafikleri Şekil 4.7 ve Şekil 4.8'de gösterilmiştir.

Sonuç 4.1 bu bölüm için de geçerlidir, buna ilave olarak, Kullback – Leibler ölçümü *Sonuç 4.2* şeklinde özetlenebilir.

Sonuç 4.2. Gözlenen olasılıklar $q = (q_1, q_2, \dots, q_{10})$ ve $m = 1, \dots, 4$ olmak üzere $g \in K_{0,m}$ olmakla $g(x)$ moment fonksiyonları tarafından üretilen m moment kısıtlı $p_{min}^{(m)} = (MinMaxEnt)_m$ ve $p_{max}^{(m)} = (MaxMaxEnt)_m$

dağılımları sırasıyla *MinMaxEnt* ve *MaxMaxEnt* dağılımlarını gösterirse $m_1 < m_2$ olduğu zaman

$$D(p_{min}^{(m_1)}; q) > D(p_{min}^{(m_2)}; q)$$

$$D(p_{max}^{(m_1)}; q) > D(p_{max}^{(m_2)}; q)$$

eşitsizlikleri elde edilir. Diğer yandan, moment kısıt sayısı arttıkça hem *MaxMaxEnt* hem de *MinMaxEnt* dağılımlarının entropi değerleri azalır. Dahası herhangi bir m için

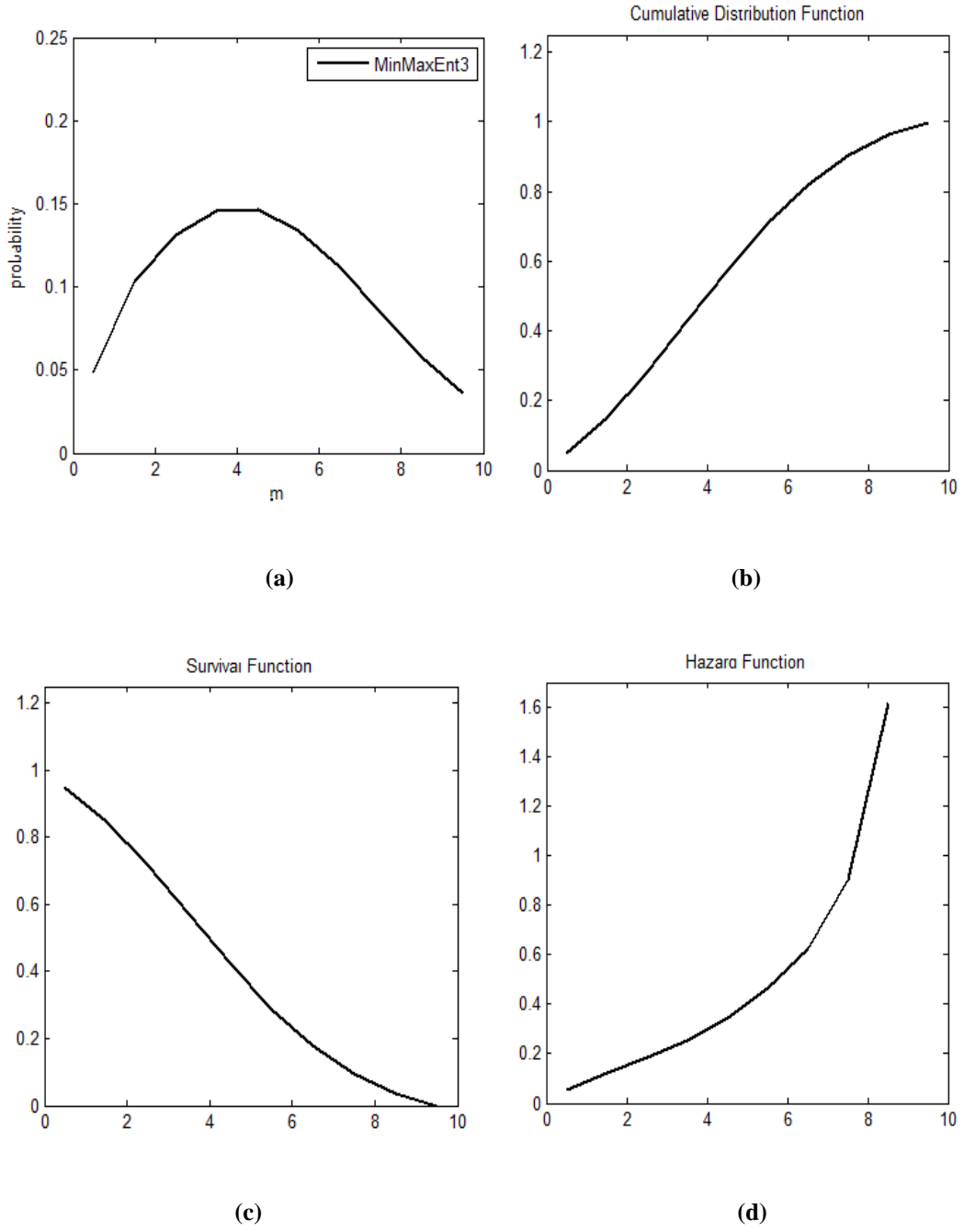
$$D(p_{max}^{(m)}; q) > D(p_{min}^{(m)}; q)$$

eşitsizliği elde edilir.

Bir önceki bölümde kullanılan yöntemlere benzer yöntemler uygulanarak diğer *MaxEnt* dağılımlarının da Sağkalım Analizi Fonksiyonları belirlenebilir ve grafikleri çizilebilir.

Çizelge 4.19. Motor arızası verileri için $(MinMaxEnt)_3$ Dağılımı ile Sağkalım Analizi

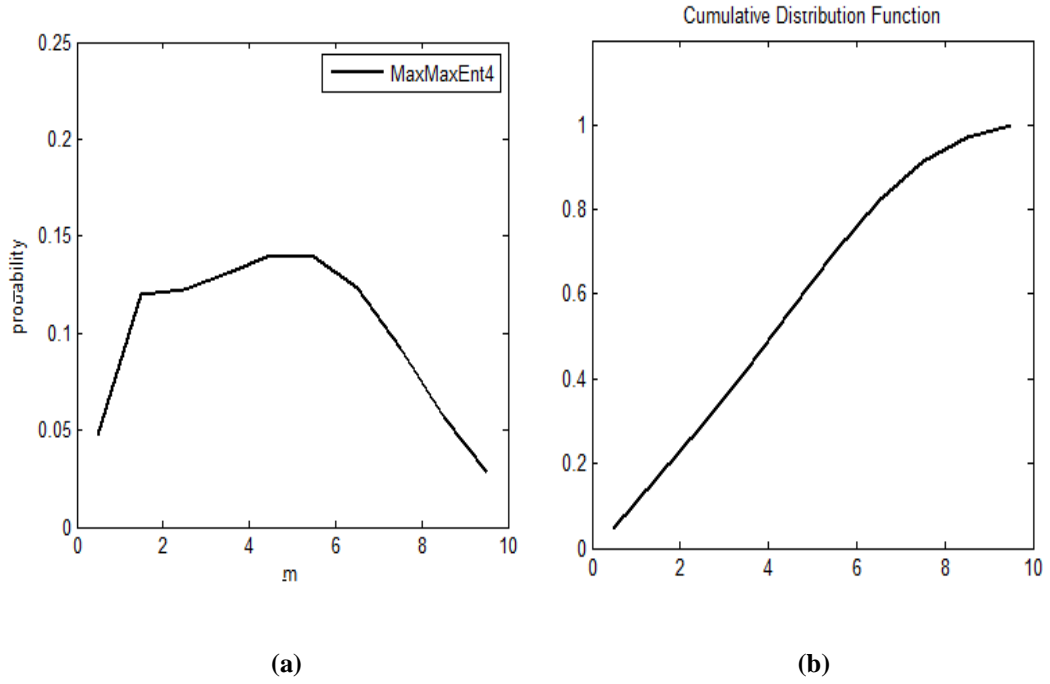
t	n_i	d_i	c_i	$\hat{f}(t) = (MinMaxEnt)_3$	$\hat{F}(t)$	$\hat{S}(t)$	$\hat{h}(t) = \hat{f}(t)/\hat{S}(t)$
1	200	5	0	0.0489	0.0489	0.9511	0.0514
2	195	10	1	0.1034	0.1523	0.8477	0.1220
3	184	12	5	0.1314	0.2837	0.7163	0.1834
4	167	8	2	0.1458	0.4295	0.5705	0.2556
5	157	10	0	0.1466	0.5761	0.4239	0.3458
6	147	15	6	0.1342	0.7103	0.2897	0.4632
7	126	9	3	0.1117	0.8220	0.1780	0.6275
8	114	8	1	0.0844	0.9064	0.0936	0.9017
9	105	4	0	0.0578	0.9642	0.0358	1.6145
10	101	3	1	0.0358	1.0000	0.0000	--



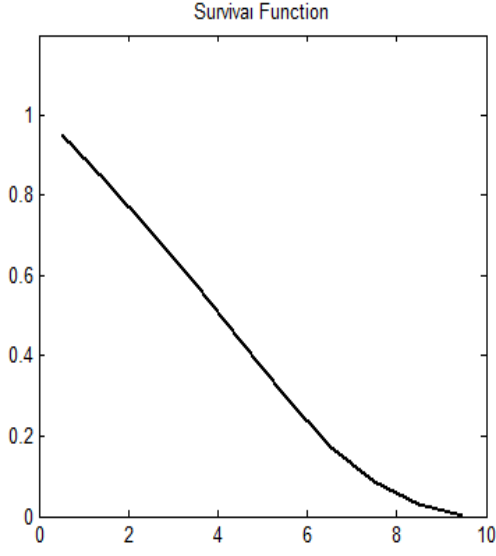
Şekil 4.8. Motor arızası verileri için $(MinMaxEnt)_3$ Dağılımına uygun (a) Olasılık Yoğunluk Fonksiyonu $\hat{f}(t)$, (b) Birikimli Dağılım Fonksiyonu $\hat{F}(t)$, (c) Sağkalım Fonksiyonu $\hat{S}(t)$ ve (d) Hazard Fonksiyonu $\hat{h}(t)$

Çizelge 4.20. Motor arızası verileri için $(MaxMaxEnt)_4$ Dağılımı ile Sağkalım Analizi

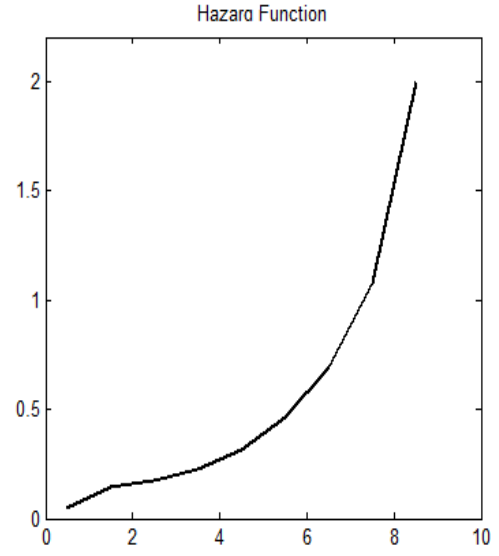
t	n_i	d_i	c_i	$\hat{f}(t) = (MaxMaxEnt)_4$	$\hat{F}(t)$	$\hat{S}(t)$	$\hat{h}(t) = \hat{f}(t)/\hat{S}(t)$
1	200	5	0	0.0477	0.0477	0.9523	0.0501
2	195	10	1	0.1198	0.1675	0.8325	0.1439
3	184	12	5	0.1221	0.2896	0.7104	0.1719
4	167	8	2	0.1306	0.4202	0.5798	0.2253
5	157	10	0	0.1397	0.5599	0.4401	0.3174
6	147	15	6	0.1395	0.6994	0.3006	0.4641
7	126	9	3	0.1230	0.8224	0.1776	0.6926
8	114	8	1	0.0921	0.9145	0.0855	1.0772
9	105	4	0	0.0570	0.9715	0.0285	2.0000
10	101	3	1	0.0284	0.9999	0.0001	--



Şekil 4.9. Motor arızası verileri için $(MaxMaxEnt)_4$ Dağılımına uygun (a) Olasılık Yoğunluk Fonksiyonu $\hat{f}(t)$, (b) Birikimli Dağılım Fonksiyonu $\hat{F}(t)$,



(c)



(d)

Şekil 4.10 (devam). Motor arızası verileri için $(MaxMaxEnt)_4$ Dağılımına uygun (c) Sağkalım Fonksiyonu $\hat{S}(t)$ ve (d) Hazard Fonksiyonu $\hat{h}(t)$

5. SONLU SAYIDA KARAKTERİZE EDİCİ MOMENTLER YARDIMI İLE *MinMinxEnt* ve *MaxMinxEnt* DAĞILIMLARI

5.1. *MinMinxEnt* ve *MaxMinxEnt* Dağılımlarının Tanımı

MinMinxEnt ve *MaxMinxEnt* problemlerini çözmek için $(g_0(x), g^{(1)}(x))$, $(g_0(x), g^{(2)}(x))$ vektör fonksiyonlarını bulmak gerekir, burada $g_0(x) = 1$, $g^{(1)} \in K_{0,m}$ ve $g^{(2)} \in K_{0,m}$ sırasıyla, $U(g)$ 'ye Kullback - Leibler ölçümüne göre, minimum ve maksimum değerleri veren karakterize edici moment vektörleridir. $U(g)$ 'nin $g_0(x)$ ve tüm m -boyutlu $g(x)$, $g \in K_{0,m}$ vektör fonksiyonları tarafından üretilen kısıtlara göre minimum veya maksimum değerlere ulaştığına dikkat edilmelidir. Diğer yandan, $U(g)$ 'nin minimum değeri $g(x)$, $g \in K_{0,m}$ 'e uygun olan D_{min} değerinin en küçük, $U(g)$ 'nin maksimum değeri $g(x)$, $g \in K_{0,m}$ 'e uygun olan D_{min} değerinin en büyük değeridir. Eğer $(g_0, g^{(1)})$, $U(g)$ 'ye en küçük değeri veriyorsa o zaman $p^{(1)} = (p_1^{(1)}, \dots, p_n^{(1)})$ dağılımı $(g_0, g^{(1)})$ vektör fonksiyonuna göre veriye uygun olan dağılımdır ve *MinMinxEnt* Dağılımı olarak adlandırılır. Eğer $(g_0, g^{(2)}(x))$ de $U(g)$ 'ye en büyük değeri veriyorsa o zaman $p^{(2)} = (p_1^{(2)}, \dots, p_n^{(2)})$ dağılımı $(g_0, g^{(2)}(x))$ vektör fonksiyonuna göre veriye uygun olan dağılımdır ve *MaxMinxEnt* Dağılımı olarak adlandırılır.

Bu tezde, *MinxEnt* yöntemi için de K_0 'ın kısıtları $g_1(x) = x$, $g_2(x) = x^2$, $g_3(x) = \ln x$, $g_4(x) = (\ln x)^2$, $g_5(x) = \ln(1 + x^2)$ karakterize edici moment vektör fonksiyonları olarak seçilmiştir. Bir başka ifadeyle, $K_0 = \{g_1, \dots, g_5\}$ olarak ifade edilir.

5.2. *MinMinxEnt* ve *MaxMinxEnt* Yöntemleri ile Bağırsak Kanseri Verileri Analizi

MinMinxEnt ve *MaxMinxEnt* yöntemlerinin sağkalım veri analizine uygulanabilir şekilde ifade edilmesi için gereken olgular ele alınsın. Bu amaçla önce, söz konusu dağılımların tanımlanabilir olması için sonlu sayıda karakterize edici moment vektör fonksiyonlarından yola çıkarak iki, üç, dört ve beş kısıtlarını doğuran tüm mümkün vektör fonksiyonlar ele alınacaktır. Bu kısıtların doğurduğu *MinMinxEnt* ve *MaxMinxEnt* dağılımları tanımlanacaktır.

MaxMaxEnt ve *MinMaxEnt* dağılımları belirlenirken iki, üç, vs. koşullu dağılımlar arasından sırası ile entropisi en büyük olan ve entropisi en küçük olan dağılımlar seçilmiştir. Bu ise sonuçta, bu dağılımlar içerisinde maksimum bilgi ve minimum bilgi içeren dağılımları vermiştir. *MaxMinxEnt* ve *MinMinxEnt* dağılımları ise seçilmiş önsel dağılıma Kullback – Leibler ölçümü açısından en uzak ve en yakın dağılımların bulunması problemi. Burada da *MaxMaxEnt* ve *MinMaxEnt* dağılımlarında olduğu gibi *MinMinxEnt* ve *MaxMinxEnt* dağılımları iki, üç, vs. koşullu dağılımlar arasından sırası ile Kullback – Leibler ölçümüne göre en uzak ve en yakın dağılımların bulunması problemi.

Bu bölümde, Bölüm 4.3'te hesaplanan *düzeltilmiş olasılıklar* Shannon entropi ölçümü yerine Kullback – Leibler ölçümü açısından yorumlanmıştır. Böylece, olasılık toplamları 1'den küçük olan dağılıma en yakın olan dağılımın bulunması, belli anlamda *sansürlü* veriyi dikkate alarak problemi çözme imkanı sağlamaktadır. Bu durum göz önünde bulundurularak, *MinMinxEnt* Yöntemi Çizelge 4.1'de verilen veriye uygulanmış ve $(MinMinxEnt)_m$, $m = 1, 2, \dots, 4$ dağılımları elde edilmiştir.

Yukarıda belirtilen bu problem, EOD arasından Kullback – Leibler ölçümü ile ilgili olan, gözlenen olasılık dağılımına en yakın veya gözlenen olasılık dağılımından en uzak dağılımın seçilmesi ile *MinMinxEnt* ve *MaxMinxEnt* yöntemleri uygulanarak çözülmüştür.

(g_0, g) , $g_0(x) = 1$, $g \in K_{0,m}$, $m = 1, 2, \dots, 4$ vektör fonksiyonlarına uygun tüm *MinxEnt* dağılımları ve $D(p : q)$ değerleri Çizelge 5.1'de

gösterilmiştir. Bu tablolar yardımıyla $D_{min} = D((MinMinxEnt)_m: q)$, $D_{max} = D((MaxMinxEnt)_m: q)$ $m = 1, 2, \dots, 4$ dağılımları da elde edilebilir ve elde edilen bu dağılımlar Çizelge 5.1’de gösterilmiştir.

Çizelge 5.1. Bağırsak kanseri verileri için (g_0, g) , $g_0(x) = 1$, $g \in K_{0,1}$ vektör fonksiyonu yardımıyla elde edilen *MinxEnt* Dağılımları ve $D(p : q)$, D_{min} , D_{max} değerleri

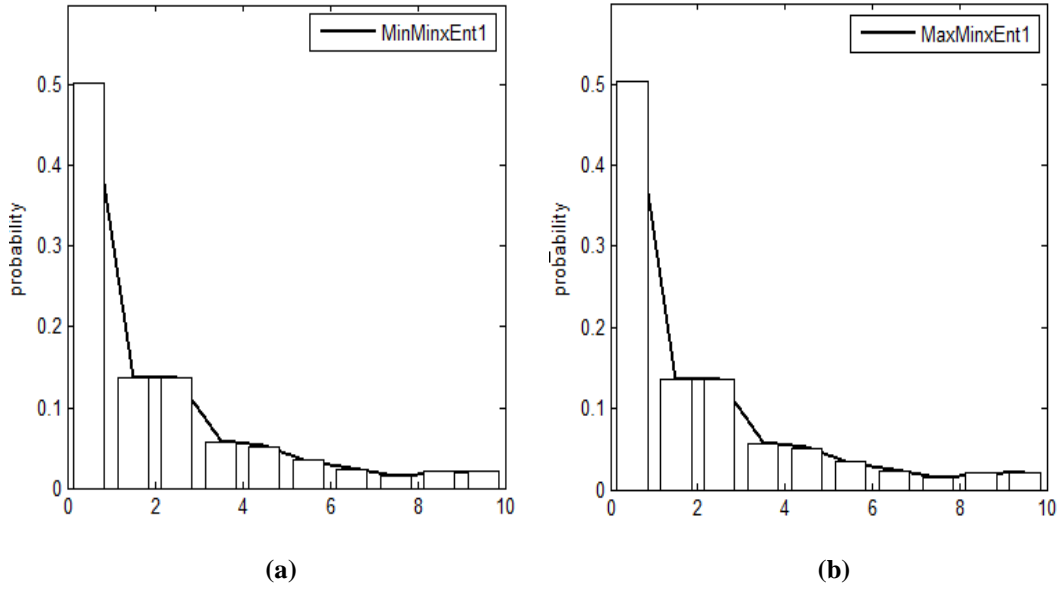
(g_0, g)	(g_0, g_1)	(g_0, g_2)	(g_0, g_3)	(g_0, g_4)	(g_0, g_5)
	0.5029	0.5039	0.5025	0.5040	0.5027
	0.1361	0.1359	0.1366	0.1355	0.1363
	0.1366	0.1362	0.1371	0.1362	0.1369
<i>MinxEnt</i>	0.0579	0.0577	0.0581	0.0578	0.0580
Dağılımları	0.0520	0.0518	0.0520	0.0520	0.0521
	0.0338	0.0337	0.0337	0.0338	0.0338
	0.0247	0.0246	0.0246	0.0247	0.0246
	0.0155	0.0155	0.0154	0.0155	0.0154
	0.0187	0.0187	0.0185	0.0187	0.0185
	0.0219	0.0221	0.0216	0.0219	0.0216
$D(p : q)$	0.026053347	0.026054655	0.026038793	0.026047815	0.026043573

Çizelge 5.1’e bakıldığında *MinMinxEnt* ve *MaxMinxEnt* dağılımları sırasıyla $(g_0, g_3) = (1, \ln x)$ ve $(g_0, g_2) = (1, x^2)$ karakterize edici vektör fonksiyonları yardımıyla elde edilmiştir. Bu dağılımların Kullback – Leibler ölçüm değerleri

$$D_{min} = D((MinMaxEnt)_1: q) = 0.026038793,$$

$$D_{max} = D((MaxMaxEnt)_1: q) = 0.026054655$$

şeklindedir. Bu dağılımların grafikleri Şekil 5.1’de gösterilmiştir.



Şekil 5.1. Bağırsak kanseri verileri için (a) $(MinMinxEnt)_1$ Dağılımının grafiği ve (b) $(MaxMinxEnt)_1$ Dağılımının grafiği

Çizelge 5.2. Bağırsak kanseri verileri için (g_0, g) , $g_0(x) = 1$, $g \in K_{0,2}$ vektör fonksiyonu yardımıyla elde edilen $MinxEnt$ Dağılımları ve $D(p : q)$, D_{min} , D_{max} değerleri

(g_0, g)	(g_0, g_1, g_2)	(g_0, g_1, g_3)	(g_0, g_1, g_4)	(g_0, g_1, g_5)	(g_0, g_2, g_3)
$MinxEnt$ Dağılımları	0.5034	0.5034	0.5028	0.5036	0.5033
	0.1360	0.1359	0.1361	0.1359	0.1360
	0.1363	0.1364	0.1367	0.1362	0.1364
	0.0578	0.0578	0.0579	0.0577	0.0578
	0.0519	0.0520	0.0520	0.0519	0.0519
	0.0337	0.0338	0.0338	0.0338	0.0337
	0.0247	0.0247	0.0247	0.0247	0.0247
	0.0155	0.0155	0.0155	0.0155	0.0155
	0.0187	0.0187	0.0187	0.0187	0.0187
	0.0220	0.0219	0.0218	0.0220	0.0220
$D(p : q)$	0.02605563	0.02605452	0.02605340	0.02605656	0.02605594

Çizelge 5.2 (devam). Bağırsak kanseri verileri için (g_0, g) , $g_0(x) = 1$, $g \in K_{0,2}$ vektör fonksiyonu yardımıyla elde edilen *MinxEnt* Dağılımları ve $D(p : q)$, D_{min} , D_{max} değerleri

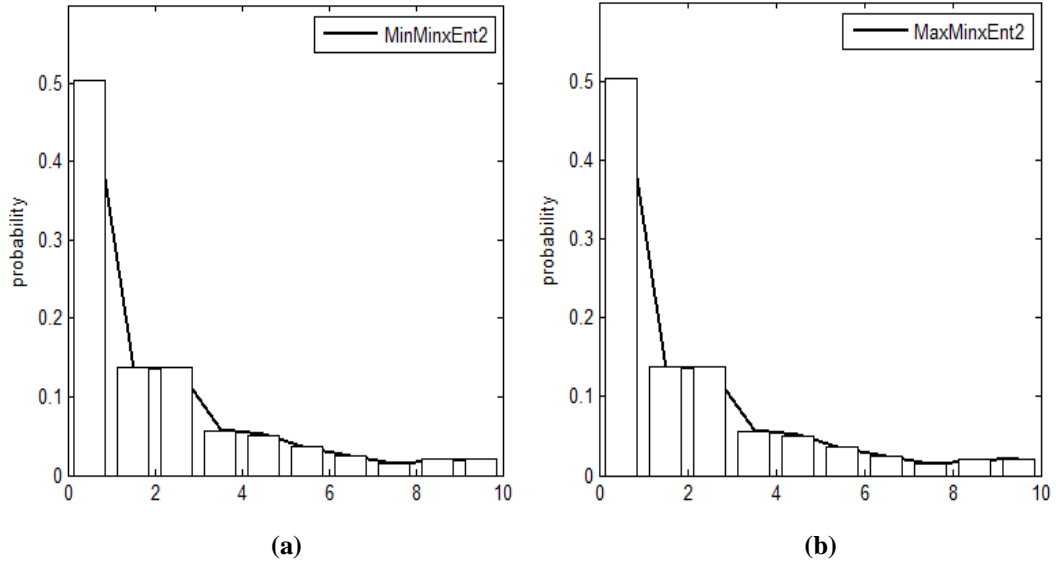
(g_0, g)	(g_0, g_2, g_4)	(g_0, g_2, g_5)	(g_0, g_3, g_4)	(g_0, g_3, g_5)	(g_0, g_4, g_5)
<i>MinxEnt</i> Dağılımları	0.5039	0.5035	0.5033	0.5032	0.5034
	0.1361	0.1360	0.1358	0.1359	0.1357
	0.1362	0.1363	0.1365	0.1367	0.1364
	0.0576	0.0578	0.0579	0.0580	0.0579
	0.0517	0.0519	0.0521	0.0521	0.0520
	0.0336	0.0337	0.0338	0.0338	0.0338
	0.0246	0.0247	0.0247	0.0247	0.0247
	0.0155	0.0155	0.0155	0.0155	0.0155
	0.0188	0.0187	0.0186	0.0186	0.0186
	0.0221	0.0220	0.0218	0.0217	0.0218
$D(p : q)$	0.02605531	0.02605536	0.02604991	0.02604576	0.0260492

Çizelge 5.2'ye bakıldığında *MinMinxEnt* ve *MaxMinxEnt* dağılımları sırasıyla $(g_0, g_3, g_5) = (1, \ln x, \ln(1 + x^2))$ ve $(g_0, g_1, g_5) = (1, x, \ln(1 + x^2))$ karakterize edici vektör fonksiyonları yardımıyla elde edilmiştir. Bu dağılımların Kullback – Leibler ölçüm değerleri

$$D_{min} = D((MinMinxEnt)_{2:q}) = 0.026045760 ,$$

$$D_{max} = D((MaxMinxEnt)_{2:q}) = 0.026056568$$

şeklindedir. Bu dağılımların grafikleri Şekil 5.2'de gösterilmiştir.



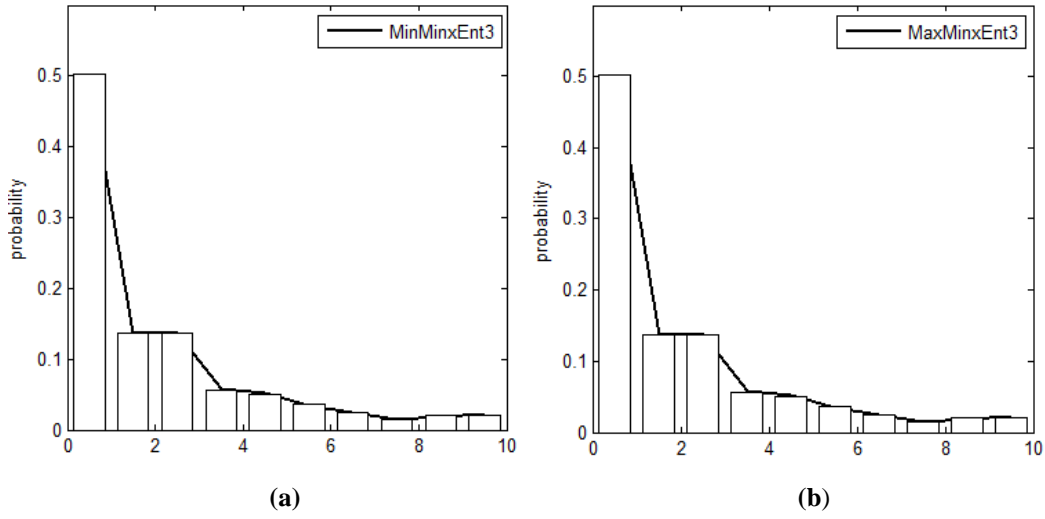
Şekil 5.2. Bağırsak kanseri verileri için (a) $(MinMinxEnt)_2$ Dağılımının grafiği ve (b) $(MaxMinxEnt)_2$ Dağılımının grafiği

Çizelge 5.3. Bağırsak kanseri verileri için (g_0, g) , $g_0(x) = 1$, $g \in K_{0,3}$ vektör fonksiyonu yardımıyla elde edilen $MinxEnt$ Dağılımları ve $D(p : q)$, D_{min} , D_{max} değerleri

(g_0, g)	(g_0, g_1, g_2, g_3)	(g_0, g_1, g_2, g_4)	(g_0, g_1, g_2, g_5)	(g_0, g_1, g_3, g_4)	(g_0, g_1, g_3, g_5)
<i>MinxEnt</i> Dağılımları	0.5032	0.5030	0.5037	0.5031	0.5030
	0.1362	0.1365	0.1357	0.1366	0.1368
	0.1364	0.1365	0.1360	0.1364	0.1363
	0.0578	0.0577	0.0577	0.0576	0.0576
	0.0518	0.0518	0.0520	0.0517	0.0517
	0.0337	0.0336	0.0339	0.0336	0.0336
	0.0246	0.0246	0.0248	0.0246	0.0246
	0.0155	0.0155	0.0156	0.0155	0.0155
	0.0187	0.0187	0.0187	0.0188	0.0188
0.0220	0.0221	0.0219	0.0221	0.0221	
$D(p : q)$	0.026056161	0.026058095	0.026058114	0.026061714	0.026063488

Çizelge 5.3 (devam). Bağırsak kanseri verileri için (g_0, g) , $g_0(x) = 1$, $g \in K_{0,3}$ vektör fonksiyonu yardımıyla elde edilen *MinxEnt* Dağılımları ve $D(p : q)$, D_{min} , D_{max} değerleri

(g_0, g)	(g_0, g_2, g_3, g_5)	(g_0, g_2, g_3, g_5)	(g_0, g_2, g_3, g_5)	(g_0, g_2, g_3, g_5)	(g_0, g_2, g_3, g_5)
<i>MinxEnt</i> Dağılımları	0.5031	0.5031	0.5030	0.5032	0.5030
	0.1367	0.1364	0.1366	0.1364	0.1369
	0.1364	0.1365	0.1365	0.1365	0.1362
	0.0576	0.0577	0.0577	0.0577	0.0575
	0.0517	0.0518	0.0517	0.0518	0.0516
	0.0336	0.0336	0.0336	0.0336	0.0336
	0.0246	0.0246	0.0246	0.0246	0.0246
	0.0155	0.0155	0.0155	0.0155	0.0155
	0.0188	0.0187	0.0187	0.0187	0.0188
0.0221	0.0221	0.0221	0.0221	0.0221	
$D(p : q)$	0.026062419	0.026057436	0.026058723	0.026057195	0.026066724



Şekil 5.3. Bağırsak kanseri verileri için (a) $(MinMinxEnt)_3$ Dağılımının grafiği ve (b) $(MaxMinxEnt)_3$ Dağılımının grafiği

Çizelge 5.3'e bakıldığında *MinMinxEnt* ve *MaxMinxEnt* dağılımları sırasıyla $(g_0, g_1, g_2, g_3) = (1, x, x^2, \ln x)$ ve $(g_0, g_2, g_3, g_5) = (1, x^2, \ln x, \ln(1 + x^2))$ karakterize edici vektör fonksiyonları yardımıyla elde edilmiştir. Bu dağılımların Kullback – Leibler ölçüm değerleri

$$D_{min} = D((MinMinxEnt)_3: q) = 0.026056161,$$

$$D_{max} = D((MaxMinxEnt)_3: q) = 0.026066724$$

şeklindedir. Bu dağılımların grafikleri Şekil 5.3'te gösterilmiştir.

Çizelge 5.4. Bağırsak kanseri verileri için (g_0, g) , $g_0(x) = 1$, $g \in K_{0,4}$ vektör fonksiyonu yardımıyla elde edilen *MinxEnt* Dağılımları ve $D(p : q)$, D_{min} , D_{max} değerleri

(g_0, g)	$(g_0, g_1, g_2, g_3, g_4)$	$(g_0, g_1, g_2, g_3, g_5)$	$(g_0, g_1, g_2, g_4, g_5)$	$(g_0, g_1, g_3, g_4, g_5)$	$(g_0, g_2, g_3, g_4, g_5)$
<i>MinxEnt</i> Dağı- lımları	0.5029	0.5029	0.5029	0.5029	0.5029
	0.1375	0.1375	0.1375	0.1374	0.1375
	0.1354	0.1354	0.1354	0.1355	0.1355
	0.0572	0.0573	0.0573	0.0573	0.0573
	0.0517	0.0518	0.0518	0.0518	0.0518
	0.0339	0.0339	0.0339	0.0339	0.0339
	0.0250	0.0249	0.0249	0.0249	0.0249
	0.0157	0.0157	0.0157	0.0156	0.0157
	0.0188	0.0188	0.0188	0.0188	0.0188
	0.0217	0.0218	0.0218	0.0218	0.0218
$D(p : q)$	0.026084381	0.02608291	0.02608330	0.02607863	0.02608078

Çizelge 5.4'e bakıldığında *MinMinxEnt* ve *MaxMinxEnt* dağılımları sırasıyla

$$(g_0, g_1, g_3, g_4, g_5) = (1, x, \ln x, (\ln x)^2, \ln(1 + x^2)) \quad \text{ve}$$

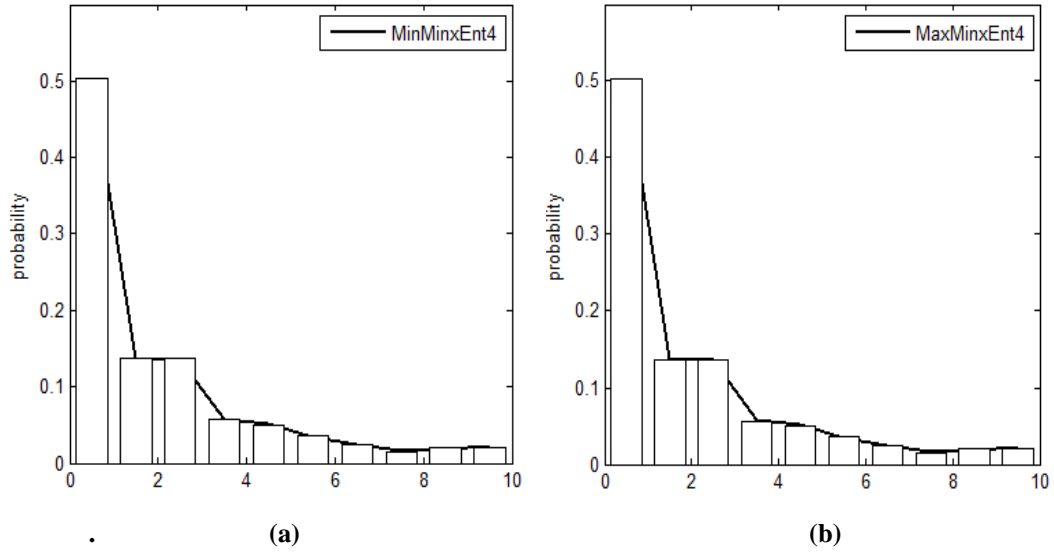
$$(g_0, g_1, g_2, g_3, g_4) = (1, x, x^2, \ln x, (\ln x)^2)$$

karakterize edici vektör fonksiyonları yardımıyla elde edilmiştir. Bu dağılımların Kullback – Leibler ölçüm değerleri

$$D_{min} = D((MinMinxEnt)_4: q) = 0.026078635$$

$$D_{max} = D((MaxMinxEnt)_4: q) = 0.026084381$$

şeklindedir. Bu dağılımların grafikleri Şekil 5.4'te gösterilmiştir.



Şekil 5.4. Bağırsak kanseri verileri için (a) $(MinMinxEnt)_4$ Dağılımının grafiği ve (b) $(MaxMinxEnt)_4$ Dağılımının grafiği

$m = 1, \dots, 4$ olmak üzere, elde edilen tüm $(MinMinxEnt)_m$ ve $(MaxMinxEnt)_m$ dağılımları sırasıyla Çizelge 5.5 ve Çizelge 5.6'da gösterilmiştir.

Çizelge 5.5. Bağırsak kanseri verileri için $(MinMinxEnt)_m$, $m = 1, 2, 3, 4$ Dağılımları

t	d_i	c_i	p_i^*	$(MinMinxEnt)_1$	$(MinMinxEnt)_2$	$(MinMinxEnt)_3$	$(MinMinxEnt)_4$
1	167	2	0.5030	0.5025	0.5032	0.5032	0.5029
2	45	1	0.1369	0.1366	0.1359	0.1362	0.1374
3	45	1	0.1369	0.1371	0.1367	0.1364	0.1355
4	19	0	0.0565	0.0581	0.0580	0.0578	0.0573
5	17	0	0.0506	0.0520	0.0521	0.0518	0.0518
6	11	1	0.0357	0.0337	0.0338	0.0337	0.0339
7	8	0	0.0238	0.0246	0.0247	0.0246	0.0249
8	5	0	0.0149	0.0154	0.0155	0.0155	0.0156
9	6	1	0.0208	0.0185	0.0186	0.0187	0.0188
10	7	0	0.0208	0.0216	0.0217	0.0220	0.0218

Çizelge 5.6. Bağırsak kanseri verileri için $(MaxMinxEnt)_m$, $m = 1,2,3,4$ Dağılımları

t	d_i	c_i	p_i^*	$(MaxMinxEnt)_1$	$(MaxMinxEnt)_2$	$(MaxMinxEnt)_3$	$(MaxMinxEnt)_4$
1	167	2	0.5030	0.5039	0.5036	0.5030	0.5029
2	45	1	0.1369	0.1359	0.1359	0.1369	0.1375
3	45	1	0.1369	0.1362	0.1362	0.1362	0.1354
4	19	0	0.0565	0.0577	0.0577	0.0575	0.0572
5	17	0	0.0506	0.0518	0.0519	0.0516	0.0517
6	11	1	0.0357	0.0337	0.0338	0.0336	0.0339
7	8	0	0.0238	0.0246	0.0247	0.0246	0.0250
8	5	0	0.0149	0.0155	0.0155	0.0155	0.0157
9	6	1	0.0208	0.0187	0.0187	0.0188	0.0188
10	7	0	0.0208	0.0221	0.0220	0.0221	0.0217

Bu sonuçlardan yola çıkarak $MinxEnt$ Dağılımlarının veriye uygunluklarını elde etmek için RMSE, Ki-Kare ve dağılımların Kullback-Leibler ölçümü gibi çeşitli kriterler kullanılmıştır. Elde edilen sonuçlar Çizelge 5.7 ve Çizelge 5.8’de gösterilmiştir.

Çizelge 5.7. Bağırsak kanseri verileri için $(MinMinxEnt)_m$, $m = 1,2,\dots,4$ Dağılımları için RMSE, Ki – Kare ve Kullback – Leibler ölçüm değerleri

$MinMinxEnt$ Dağılımları	$D(p : q)$	Ki – Kare Hesap Değeri	Ki – Kare Tablo Değeri	RMSE
$(MinMinxEnt)_1$	0.026038793	0.0118	$\chi^2_{8,\alpha} = 15.51$	0.0338
$(MinMinxEnt)_2$	0.026045760	0.0116	$\chi^2_{7,\alpha} = 14.07$	0.0333
$(MinMinxEnt)_3$	0.026056161	0.0112	$\chi^2_{6,\alpha} = 12.59$	0.0329
$(MinMinxEnt)_4$	0.026078635	0.0103	$\chi^2_{5,\alpha} = 11.07$	0.0317

Çizelge 5.8. Bağırsak kanseri verileri için $(MaxMinxEnt)_m$, $m = 1,2,\dots,4$ Dağılımları için RMSE, Ki – Kare ve Kullback – Leibler ölçüm değerleri

$MaxMinxEnt$ Dağılımları	$D(p : q)$	Ki – Kare Hesap Değeri	Ki – Kare Tablo Değeri	RMSE
$(MaxMinxEnt)_1$	0.026054655	0.0115	$\chi^2_{8,\alpha} = 15.51$	0.0332
$(MaxMinxEnt)_2$	0.026056568	0.0112	$\chi^2_{7,\alpha} = 14.07$	0.0329
$(MaxMinxEnt)_3$	0.026066724	0.0107	$\chi^2_{6,\alpha} = 12.59$	0.0323
$(MaxMinxEnt)_4$	0.026084381	0.0104	$\chi^2_{5,\alpha} = 11.07$	0.0320

Çizelge 5.7 ve Çizelge 5.8'e bakıldığında χ^2 istatistiksel kriterine göre uyum testi yapıldığında Bağırsak kanseri hastalarının yaşam sürelerini modellemede tüm *MinMinxEnt* ve *MaxMinxEnt* Dağılımlarının veri setini açıkladığı görülmektedir.

RMSE kriterine göre her bir $(MaxMinxEnt)_m$ ($m = 1,2,3$) dağılımı $(MaxMinxEnt)_m$ ($m = 1,2,3$) dağılımına karşılık veri için daha iyidir ancak, $(MinMinxEnt)_4$ dağılımı $(MaxMinxEnt)_4$ dağılımına göre istatistiksel veri için daha uygun bir dağılımdır. Diğer yandan, bu uygulama için aşağıdaki *Sonuç 5.1* elde edilir.

Sonuç 5.1. $R((MinMinxEnt)_{m_1})$ ve $R((MaxMinxEnt)_{m_2})$ sırasıyla $(MinMinxEnt)_{m_1}$ ve $(MaxMinxEnt)_{m_2}$ dağılımlarının RMSE değerleri olsun. Bu durumda, eğer $m_1 < m_2$ ise o zaman aşağıdaki eşitsizlikler sağlanır.

$$R((MinMinxEnt)_{m_1}) > R((MinMinxEnt)_{m_2})$$

$$R((MaxMinxEnt)_{m_1}) > (R((MaxMinxEnt)_{m_2})).$$

Ayrıca, $m \leq 3$ ise

$$R((MinMinxEnt)_m) > R((MaxMinxEnt)_m),$$

eşitsizliği ve $m = 4$ ise

$$R((MinMinxEnt)_4) < R((MaxMinxEnt)_4)$$

eşitsizliği elde edilir.

Kullback - Leibler ölçümüne göre ise $(MinMinxEnt)_4$ dağılımı, veri için $(MaxMinxEnt)_4$ dağılımdan daha iyidir. Sonuçlar Çizelge 5.7 ve Çizelge 5.8'de gösterilmiştir.

$K_{0,m}$ kümesindeki bilinen karakterize edici moment vektör fonksiyonları ve Kullback-Leibler ölçümü kullanılarak bulunan *MinMinxEnt* ve *MaxMinxEnt* dağılımlarının bazı özellikleri *Sonuç 5.2* de eşitsizlikler halinde özetlenmiştir.

Sonuç 5.2. $g(x)$, $g \in K_{0,m}$ moment fonksiyonları ve $q = (q_1, \dots, q_n)$ tarafından üretilen m moment kısıtlı $(MinMinxEnt)_m$ ve $(MaxMinxEnt)_m$ sırasıyla $MinMinxEnt$ ve $MaxMinxEnt$ dağılımlarını gösterebilir. Bu durumda, $m_1 < m_2$ olduğunda aşağıdaki eşitsizlikler sağlanır:

$$D((MinMinxEnt)_{m_1} : q) < D((MinMinxEnt)_{m_2} : q)$$

$$D((MaxMinxEnt)_{m_1} : q) < D((MaxMinxEnt)_{m_2} : q).$$

Başka bir ifadeyle, $MinMinxEnt$ dağılımlarının Kullback-Leibler ölçümü, artan moment kısıtı sayısına bağlı olarak artar. Ayrıca, herhangi bir m sayısı için eşitsizlik

$$D((MaxMinxEnt)_m : q) > D((MinMinxEnt)_m : q)$$

şeklindedir.

Bu sonuçlar, her iki dağılımın da sağkalım veri analizinde çeşitli problemleri çözmede uygulanabildiğini gösterir.

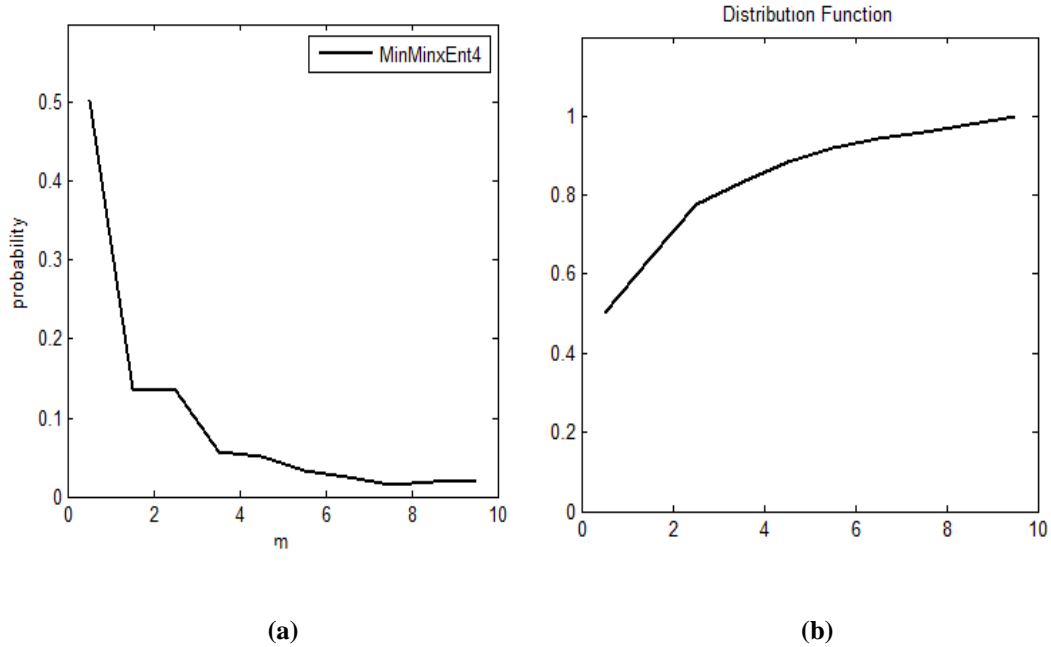
5.3. $MinMinxEnt$ ve $MaxMinxEnt$ Yöntemleri ile Bağırsak Kanseri Verileri için Yaşam Süresi Rassal Değişkenlerinin Karakterize Edici Fonksiyonları

Bu bölümde, $MinMinxEnt$ ve $MaxMinxEnt$ yöntemleri ile bağırsak kanseri verileri için yaşam süresi rassal değişkenlerinin karakterize edici fonksiyonları olan Olasılık Yoğunluk Fonksiyonu $\hat{f}(t)$, Birikimli Dağılım Fonksiyonu $\hat{F}(t)$, Sağkalım Fonksiyonu $\hat{S}(t)$ ve Hazard (Risk) Fonksiyonu $\hat{h}(t)$ tahmin edilmiştir. Buradan yola çıkarak, bağırsak kanseri veri analizi, RMSE kriterine göre $(MinMinxEnt)_m$ ve $(MaxMinxEnt)_m$, $m = 1, 2, \dots, 4$ dağılımları arasında istatistiksel veriyi en iyi biçimde ifade eden $(MinMinxEnt)_4$ ve $(MaxMinxEnt)_4$ dağılımları ile yapılmıştır.

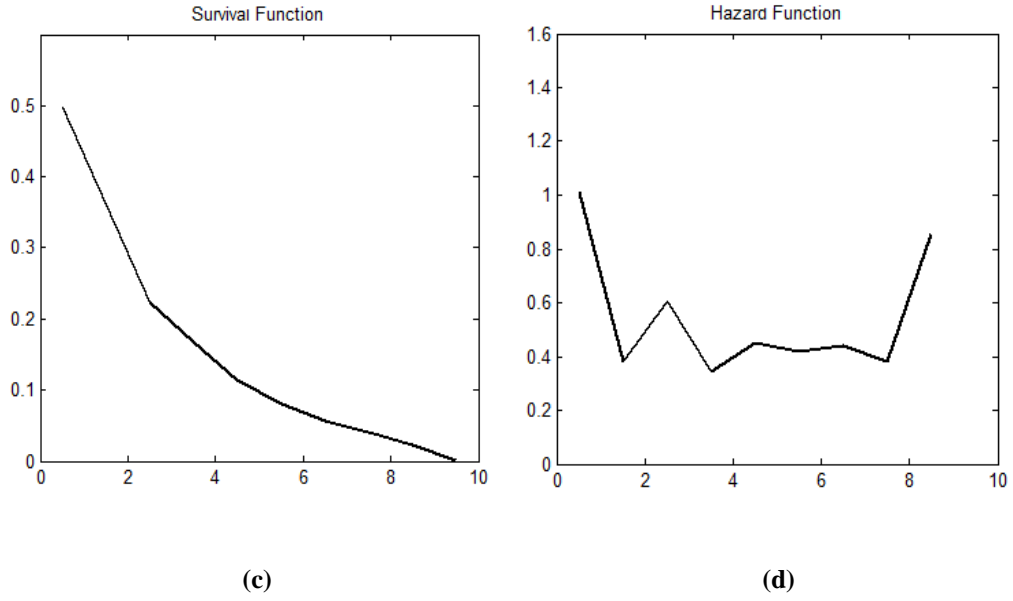
Çizelge 5.9’da $(MinMinxEnt)_4$ Dağılımı için tahmini Olasılık Yoğunluk Fonksiyonu $\hat{f}(t)$, Birikimli Dağılım Fonksiyonu $\hat{F}(t)$, Sağkalım Fonksiyonu $\hat{S}(t)$ ve Hazard (Risk) Fonksiyonu $\hat{h}(t)$ değerleri verilmiştir. Bu dağılımlara göre çizilen $\hat{f}(t)$, $\hat{F}(t)$, $\hat{S}(t)$ ve $\hat{h}(t)$ ‘nin grafikleri Şekil 5.5’te gösterilmiştir.

Çizelge 5.9. Bağırsak kanseri verileri için $(MinMinxEnt)_4$ Dağılımı ile yapılan Sağkalım Analizi

t	n_i	d_i	c_i	$\hat{f}(t) = (MinMinxEnt)_4$	$\hat{F}(t)$	$\hat{S}(t)$	$\hat{h}(t) = \hat{f}(t)/\hat{S}(t)$
1	388	167	2	0.5029	0.5029	0.4971	1.0117
2	219	45	1	0.1374	0.6403	0.3597	0.3820
3	173	45	1	0.1355	0.7758	0.2242	0.6044
4	127	19	0	0.0573	0.8331	0.1669	0.3433
5	108	17	0	0.0518	0.8849	0.1151	0.4500
6	91	11	1	0.0339	0.9188	0.0812	0.4175
7	79	8	0	0.0249	0.9437	0.0563	0.4423
8	71	5	0	0.0156	0.9593	0.0407	0.3833
9	66	6	1	0.0188	0.9781	0.0219	0.8584
10	59	7	0	0.0218	0.9999	0.0001	--



Şekil 5.5. Bağırsak kanseri verileri için $(MinMinxEnt)_4$ Dağılımına uygun (a) $\hat{f}(t)$ Fonksiyonunun Grafiği (b) $\hat{F}(t)$ Fonksiyonunun Grafiği,

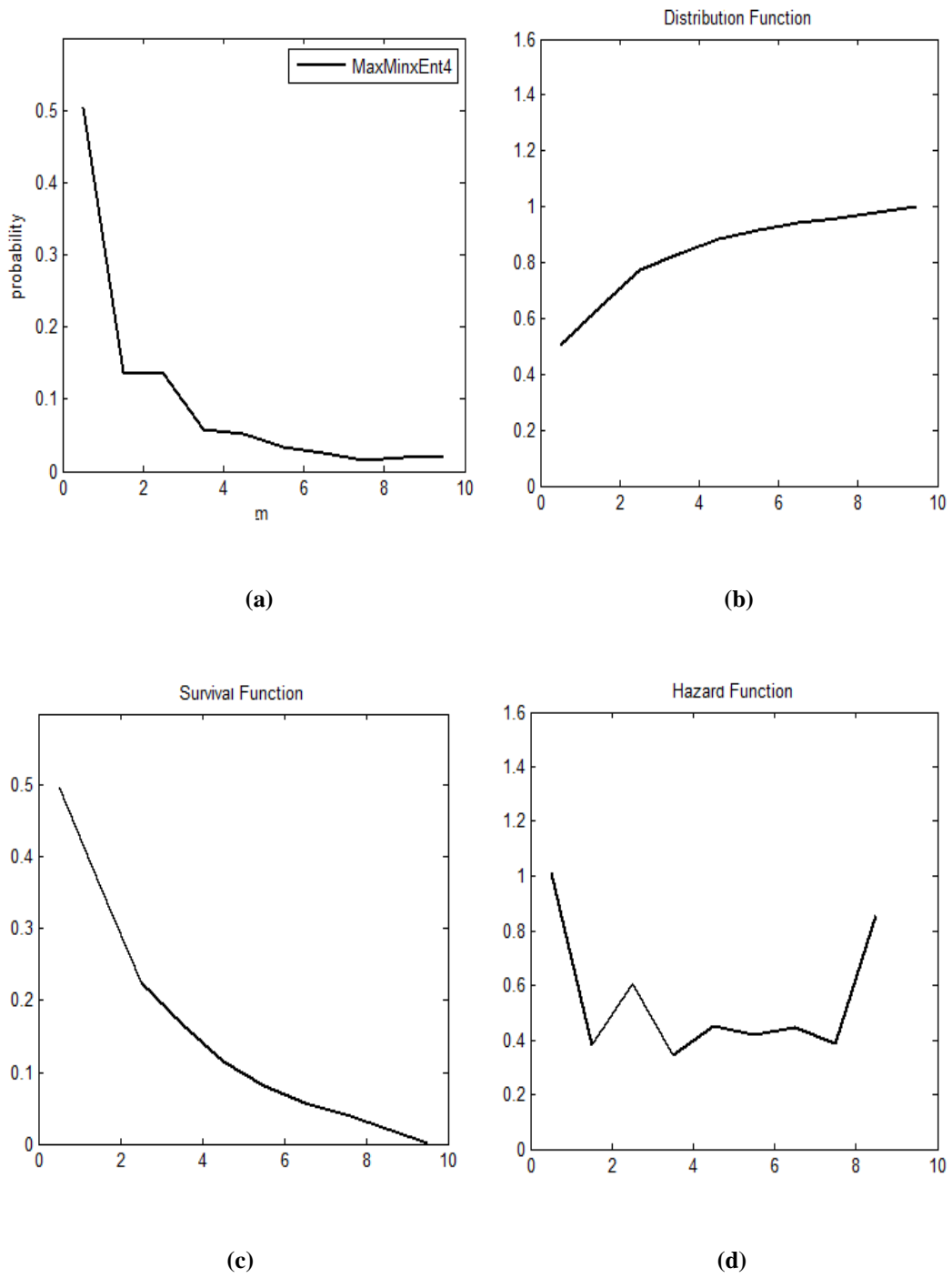


Şekil 5.6 (devam). Bağırsak kanseri verileri için $(MinMinxEnt)_4$ Dağılımına uygun (c) $\hat{S}(t)$ Fonksiyonunun Grafiği ve (d) $\hat{h}(t)$ Fonksiyonunun Grafiği

Çizelge 5.10'de $(MaxMinxEnt)_4$ Dağılımı için tahmini Olasılık Yoğunluk Fonksiyonu $\hat{f}(t)$, Birikimli Dağılım Fonksiyonu $\hat{F}(t)$, Sağkalım Fonksiyonu $\hat{S}(t)$ ve Hazard (Risk) Fonksiyonu $\hat{h}(t)$ değerleri verilmiştir. Bu dağılımlara göre çizilen $\hat{f}(t)$, $\hat{F}(t)$, $\hat{S}(t)$ ve $\hat{h}(t)$ 'nin grafikleri Şekil 5.6'te gösterilmiştir.

Çizelge 5.10. Bağırsak kanseri verileri için $(MaxMinxEnt)_4$ Dağılımı ile yapılan Sağkalım Analizi

t	n_i	d_i	c_i	$\hat{f}(t) = (MaxMinxEnt)_4$	$\hat{F}(t)$	$\hat{S}(t)$	$\hat{h}(t) = \hat{f}(t)/\hat{S}(t)$
1	388	167	2	0.5029	0.5029	0.4971	1.0117
2	219	45	1	0.1375	0.6404	0.3596	0.3824
3	173	45	1	0.1354	0.7758	0.2242	0.6039
4	127	19	0	0.0572	0.8330	0.1670	0.3425
5	108	17	0	0.0517	0.8847	0.1153	0.4484
6	91	11	1	0.0339	0.9186	0.0814	0.4165
7	79	8	0	0.0250	0.9436	0.0564	0.4433
8	71	5	0	0.0157	0.9593	0.0407	0.3857
9	66	6	1	0.0188	0.9781	0.0219	0.8584
10	59	7	0	0.0217	0.9998	0.0002	--



Şekil 5.7. Bağırsak kanseri verileri için $(MaxMinxEnt)_4$, Dağılımına uygun (a) $f(t)$ Fonksiyonunun Grafiği (b) $F(t)$ Fonksiyonunun Grafiği, (c) $S(t)$ Fonksiyonunun Grafiği ve (d) $h(t)$ Fonksiyonunun Grafiği

6. *MinMaxEnt*, *MaxMaxEnt* VE *MinMinxEnt*, *MaxMinxEnt* YÖNTEMLERİNİN KARŞILAŞTIRILMASI

Bölüm 4.2 ve Bölüm 5.2’de 1935 – 1944 yılları arasında Connecticut şehrindeki bağırsak kanserine yakalanmış erkek hastaların sağkalım verilerine sırasıyla Shannon entropi ölçümü, Kullback – Leibler ölçümü ve sonlu sayıda karakterize edici moment fonksiyonu yardımıyla üretilmiş istatistiksel dağılımları temsil eden *MinMaxEnt*, *MaxMaxEnt* ve *MinMinxEnt*, *MaxMinxEnt* yöntemleri ayrı ayrı uygulanmıştır. Geliştirilen yöntemlerin performansı Ki – Kare kriteri yardımıyla incelenmiştir. Bu incelemenin sonucu olarak, sırasıyla elde edilen tüm $(MinMaxEnt)_m$, $(MaxMaxEnt)_m$ ve $(MinMinxEnt)_m$, $(MaxMinxEnt)_m$, $m = 1, \dots, 4$ dağılımlarının Ki – Kare kriterine göre yukarıda belirtilen sağkalım verileri için uygun olduğu ifade edilmiştir. Buna ek olarak *MinMaxEnt*, *MaxMaxEnt* ve *MinMinxEnt*, *MaxMinxEnt* yöntemleri RMSE, Shannon entropi ölçümü ve Kullback – Leibler ölçümü gibi farklı anlamlarda birbiri ile karşılaştırılmıştır. Böylece, gözlenmiştir ki ele alınmış yöntemlerin her biri, diğerine nazaran, belli anlamda avantaj sağlanmasına rağmen başka bir anlamda diğer yöntem, bu yöntem üzerinde avantaj sağlamaktadır.

Örneğin, RMSE kriteri anlamında, istatistiksel veri için $(MaxMaxEnt)_4$ dağılımının $(MinMaxEnt)_4$ dağılımından ve $(MinMinxEnt)_4$ dağılımının da $(MaxMinxEnt)_4$ dağılımından daha uygun olduğu sonucuna varılmıştır. Shannon entropi ölçümüne göre ise sağkalım verileri için $(MinMaxEnt)_4$ dağılımının $(MaxMaxEnt)_4$ dağılımından daha uygun olduğu, Kullback – Leibler ölçümüne göre de sağkalım verilerini $(MinMinxEnt)_4$ dağılımının $(MaxMinxEnt)_4$ dağılımından daha iyi temsil ettiği görülmüştür.

Bu bölümde ise, adı geçen tüm Genelleştirilmiş Entropi Optimizasyon Dağılımları Ki – Kare, RMSE, Shannon entropi ölçümü ve Kullback – Leibler ölçümü olmak üzere farklı kriterler kullanılarak *birbirleri* ile karşılaştırılmıştır. Kullanılan kritere göre hangisinin daha iyi olduğu vurgulanmıştır.

Çizelge 4.9 ve Çizelge 5.7’ye bakıldığında $p^{(1)} = (MinMaxEnt)_4$ ve $\tilde{p}^{(1)} = (MinMinxEnt)_4$ dağılımlarının, gözlenen dağılım q ’ya yakınlıklarının Kullback – Leibler ölçümlerinin sırasıyla $D(p^{(1)}; q) = 0.0343$ ve

$D(\tilde{p}^{(2)}; q) = 0.026078635$ olduğu görülmektedir. Buradan anlaşılıyor ki, q 'ya $(MinMinxEnt)_4$ dağılımı $(MinMaxEnt)_4$ dağılımından daha yakın bir dağılımdır. Böylece Kullback – Leibler ölçümü anlamında, bu veriyi $(MinMinxEnt)_4$ dağılımının $(MinMaxEnt)_4$ dağılımından daha iyi temsil ettiği sonucu elde edilir.

Elde edilen bu dağılımlar RMSE kriterine göre ele alındığında, Çizelge 5.7'den $(MinMinxEnt)_4$ 'ün RMSE değerinin 0.317 ve Çizelge 4.10'dan $(MaxMaxEnt)_4$ 'ün RMSE değerinin ise 0.0984 olduğu görülmektedir. Bu kriterlere göre, istatistiksel veriye, $(MaxMaxEnt)_4$ dağılımı $(MinMinxEnt)_4$ dağılımından daha uygundur.

Bu sonuçlar dikkate alındığında, GEOY yardımıyla daha önceden gözlenen veriye uyan bilinmeyen dağılımı elde etmek için veri ile bu dağılım arasındaki yakınlık derecesinin belirlenmesi çok önemlidir. Sonuç olarak, bu problemin çözümü için gereken dağılımın bulunması var olan yöntemler arasından belli koşullara ve belli anlamda yakınlık derecesine göre seçilmelidir.

$K_{0,m}$ kümesine ait karakterize edici moment vektör fonksiyonlarına uygun olarak Shannon entropi ölçümü, Kullback-Leibler ölçümü ve RMSE yardımıyla belirlenen $MaxMaxEnt$, $MinMaxEnt$ ve $MaxMinxEnt$, $MinMinxEnt$ dağılımlarının bazı özellikleri Sonuç 4.1, Sonuç 4.2, Sonuç 5.1 ve Sonuç 5.2'de özetlenmiştir.

Bu sonuçlara teorik anlamda GEOD'un karşılaştırılması gibi bakılabilir.

Bu sonuçlar $MaxMaxEnt$, $MinMaxEnt$ ve $MaxMinxEnt$, $MinMinxEnt$ dağılımlarının sağkalım veri analizinde çeşitli problemlerin çözümünde uygulanabilir olduğunu göstermektedir.

7. SONUÇLAR VE ÖNERİLER

Bu tezde, elde edilen sonuçlar hakkında özetle, aşağıdaki yorumlar yapılmıştır.

İlk olarak, Bağırsak kanseri verileri ve Motor arızaları verileri olmak üzere farklı iki sağkalım verilerinden elde edilen önsel dağılım için *MaxEnt* yöntemi yardımıyla, seçilen moment kısıtlarına bağlı özel fonksiyoneller ile bu fonksiyonellere sırasıyla minimum ve maksimum değer veren ve informasyonu maksimum yapan *MinMaxEnt* ve *MaxMaxEnt* dağılımlarından bahsedilmiştir. Böylece, seçilen moment kısıtları ve Shannon entropi ölçümü yardımıyla *MaxEnt* dağılımları elde edilmiş, bu dağılımlar arasından Shannon entropi ölçümü en küçük *MinMaxEnt* ve Shannon entropi ölçümü en büyük *MaxMaxEnt* dağılımları belirlenmiştir. Bu dağılımların nümerik yöntemlerle bulunması, entropi optimizasyon problemleri şeklinde ortaya konmuş ve bu problemlerin çözümleri için MATLAB 7.10.0 (R2010a)'da özel program yazılmış, ulaşılan sonuçlar yorumlanmıştır.

İkinci olarak, Bağırsak kanseri verilerinden elde edilen önsel dağılım için *MinxEnt* yöntemi yardımıyla, seçilen moment kısıtlarına bağlı özel fonksiyoneller ile bu fonksiyonellere sırasıyla minimum ve maksimum değer veren *MinMinxEnt* ve *MaxMinxEnt* dağılımlarından bahsedilmiştir. Böylece, seçilen moment kısıtları ve Kullback – Leibler ölçümü yardımıyla *MinxEnt* dağılımları elde edilmiş, bu dağılımlar arasından önsel dağılıma en yakın *MinMinxEnt* ve önsel dağılımdan en uzak *MaxMinxEnt* dağılımları belirlenmiştir. Bu dağılımların nümerik yöntemlerle bulunması da, entropi optimizasyon problemleri şeklinde ortaya konmuş ve bu problemlerin çözümleri için de MATLAB 7.10.0 (R2010a)'da özel program yazılmış, ulaşılan sonuçlar yorumlanmıştır.

Bu sonuçlara göre, Sağkalım Analizinde, sağkalım verilerinden elde edilen önsel dağılıma uygun dağılımların belirlenmesinde, bilinen istatistiksel yöntemlere alternatif olarak GEOY sunulmaktadır. Çünkü elde edilen her bir GEOD'un performansı Ki – Kare, RMSE, Shannon entropi ölçümü ve Kullback –

Leibler ölçümü gibi kriterler kullanılarak araştırılmış ve bu dağılımların belli anlamlarda sağkalım veri setini açıkladığı gösterilmiştir.

Bağırsak kanseri verileri için GEOD'ı arasından $m = 1, \dots, 4$ olmak üzere, $(MinMaxEnt)_m$ ve $(MaxMaxEnt)_m$ dağılımları ele alınırsa, RMSE anlamında $(MaxMaxEnt)_4$ dağılımının veri seti için en uygun dağılım olduğu sonucuna varılmıştır.

Gene bağırsak kanseri verileri için GEOD'ı arasından $m = 1, \dots, 4$ olmak üzere, $(MinMaxEnt)_m$ ve $(MaxMaxEnt)_m$ dağılımları ele alınırsa, RMSE anlamında $(MinxMinxEnt)_4$ dağılımının veri seti için en iyi dağılım olduğu ifade edilmiştir.

Ayrıca, moment kısıtları sayısı arttıkça, $MinMaxEnt$ ve $MinMinxEnt$ dağılımlarının da diğer dağılımlara göre veriyi modellemede etkilerinin arttığı gözlenmiştir.

Motor arızası verileri için ise $MaxEnt$ dağılımları arasından $m = 1, \dots, 4$ olmak üzere, $(MinMaxEnt)_m$ ve $(MaxMaxEnt)_m$ dağılımları ele alınırsa, RMSE anlamında $(MinMaxEnt)_3$ dağılımının veri seti için en uygun dağılım olduğuna, Shannon entropi ölçümü açısından incelendiğinde ise $(MinMaxEnt)_4$ dağılımının veri setini açıklayan en iyi dağılım olduğu sonucuna varılmıştır.

Tüm GEOD dikkate alındığında ise Kullback – Leibler ölçümü anlamında, bağırsak kanseri verilerini $(MinMinxEnt)_4$ dağılımının $(MinMaxEnt)_4$ dağılımından daha iyi temsil ettiği, RMSE kriterine göre de istatistiksel veriye, $(MaxMaxEnt)_4$ dağılımının $(MinMinxEnt)_4$ dağılımından daha uygun olduğu sonucuna ulaşılır.

Bu araştırmalar ışığında verilecek öneriler aşağıdaki şekilde özetlenebilir:

Bu sonuçlar $MaxMaxEnt$, $MinMaxEnt$ ve $MaxMinxEnt$, $MinMinxEnt$ dağılımlarının sağkalım veri analizinde çeşitli problemlerin çözümünde uygulanabilir olduğunu göstermektedir. Bu sebeple, klasik istatistiksel yöntemlerin yanı sıra, özellikle bilinen bir dağılıma uymayan sağkalım verileri için GEOY güvenle uygulanabilir.

Saękalım verilerinden elde edilen önsel dağılımın olasılıklar toplamının 1'den küçük olması sansürlü verilerden kaynaklanmaktadır. Böyle durumlarda, olasılık toplamları 1' den küçük olan dağılıma en yakın olan dağılımın bulunması, belli anlamda *sansürlü* veriyi dikkate alarak problemi çözme imkanı sağlamaktadır. Bu durum göz önünde bulundurularak, *MinMinxEnt* ve *MaxMaxEnt* Yöntemleri Saękalım Veri Analizi için mükemmel bir yöntem ve alternatif bir çözüm yolu üretebilir. Bu sebeple GEOY, Saękalım Veri Analizinde başarılı bir şekilde uygulanabilir.

KAYNAKLAR

- Anonim (2006), *Anadolu Üniversitesi Fen Bilimleri Enstitüsü Tez Yazım Yönergesi*.
http://www.fbe.anadolu.edu.tr/duyurular/enstituozel/tr/yonetm/tez_yazim_ynrg.pdf
- Başer E., (1993), *Yaşam Tabloları Analizinde Kullanılan Bazı İstatistiksel Tekniklerin Böbrek Nakli Verilerine Uygulanması*, Doktora Tezi, Hacettepe Üniversitesi, Fen Bilimleri Enstitüsü, Ankara.
- Cramer H. (1966), *Mathematical Methods of Statistics*, Princeton University Press.
- Cover T. M. (1991), *Elements of Information Theory*, Jhon Wiley, New York.
- Deborah Kaminski ve Cheryl Geisler (2012), “Survival Analysis of Faculty Retention in Science and Engineering by Gender”, *Science* 335:864-866.
- Deshpande J. V. ve Purohit S.G. (2005), *Life Time Data: Statistical Models and Methods*, Series on Quality, Reliability and Engineering Statistics Vol. 11, India.
- Eyal M. Reingold, Erik D. Reichle, Mackenzie G. Glaholt, Heather Sheridan (2012), “Direct Lexical Control of Eye Movements in Reading: Evidence from a Survival Analysis of Fixation Durations”, *Cognitive Psychology*, Vol. 65, Issue 2:177–206.
- Huan Wang, Hongsheng Dai (2012), “Accelerated Failure Time Models for Censored Survival Data under Referral Bias”, *Biostatistics*, Vol. 14, Issue 2:313-326.
- Jaynes E. T. (1957), “Information Theory and Statistical Mechanics”, *Phys. Rev.*, 106, 620-630.
- Kapur J. N., Kesavan (1992), *Entropy Optimization Principles with Applications*, Academic Press.
- Kullback S. ve Leibler R. A. (1951), “on Information and Sufficiency”, *Annals of Mathematical Statistics*, 22, 79-96.
- Kullback S., (1959), *Information Theory and Statistics*, Wiley.

- Lee E. T. ve Wang J.W. (2003), *Statistical Methods for Survival Data Analysis*, Wiley-Interscience, Oklahoma.
- Nader Ebrahimi (2000), “The Maximum Entropy Method for Lifetime Distributions”, *Sankhyā: The Indian Journal of Statistics, Series A* (1961-2002) Vol. 62, No. 2:236-243.
- Papoulis A., Pillai S. U. (2002), *Probablitiy, Random Variables and Stochastic Processes*, McGraw-Hill Book.
- Patricia Guyot, Ae Ades, Mario Jnm Ouwens, Nicky J Welton (2012), “Enhanced Secondary Analysis of Survival Data: Reconstructing The Data from Published Kaplan-Meier Survival Curves”, *Bmc Medical Research Methodology* 12:9.
- Pierre Joly, Thomas A. Gerds, Vibeke Qvist, Daniel Commenges, Niels Keiding (2012), “Estimating Survival of Dental Fillings on The Basis of Interval-Censored Data and Multi-State Models”. *Statistics In Medicine*, Vol 31 Issue 11-12.
- Şamilov A. (2009), *Entropi, İnfomasyon ve Entropi Optimizasyon*, T.C. Anadolu Universitesi Yayınları, Eskisehir.
- Shamilov A. (2006), “A Development of Entropy Optimization Methods”, *Wseas Transactions on Mathematics*, Vol. 5, Issue 5:568-575.
- Shamilov A. (2007), “Generalized Entropy Optimization Problems and The Existence of Their Solutions”, *Physica A: Statistical Mechanics and Its Applications*, 382(2):465-472
- Shamilov A. (2010), “Generalized Entropy Optimization Problems with Finite Moment Functions Sets”, *Journal of Statistics and Management Systems*, Vol. 13, Issue 3:595-603.
- Shamilov A., Giriftinoglu C., Usta I., Mert Kantar Y. (2008), “A New Concept of Relative Suitability of Moment Function Sets”, *Applied Mathematics and Computation*, 206:521-529.

Shamilov A., Giriftinoglu C., Ozdemir S. (2013), “Survival Data Analysis by Generalized Entropy Optimization Methods”, *Proceedings of the 13th International Conference on Computational and Mathematical Methods in Science and Engineering*, CMMSE 2013 24-27, Prague Czech Republic.

Shamilov A., Giriftinoglu C., Ozdemir S. (2013), “MinMinxEnt ve MaxMinxEnt Yöntemleri ile Sağkalım Veri Analizi”, *Uluslararası 8. İstatistik Kongresi Bildiri Kitabı*, Kemer, ANTALYA

Shamilov A., Giriftinoglu C., Ozdemir S. (2014), “An Application of Generalized Entropy Optimization Methods in Survival Data Analysis”, *On Actual Problems of Mathematics and Mechanics Proceedings of the International Conference Devoted to the 55-th Anniversary of the Institute Mathematics and Mechanics*, Baku, Azerbaijan.

Shamilov A., Ozdemir S., Yilmaz N. (2014), “Generalized Entropy Optimization Methods for Survival Data”, *ALT2014: 5th International Conference on Accelerated Life Testing and Degradation Models*, Pau, France

Usta İ., (2006), *MaxEnt ve MinxEnt Optimizasyon Prensiplerine Bağlı Nümerik İncelemeler ve İstatistiksel Uygulamalar*, Yüksek Lisans Tezi, Anadolu Üniversitesi, Fen Bilimleri Enstitüsü, Eskişehir.

EKLER

EK-1. Bağırsak Kanseri Verileri Üzerine Uygulanan ve Sansürlü Verileri Dikkate Alan *MinMaxEnt* Entropi Optimizasyon Probleminin Oluşturulması ve Çözümü için Hazırlanan Programın İki Kısıta Göre MATLAB Kodu

1. clear all
2. g1=inline('x');
3. g2=inline('x.^2');
4. g3=inline('log(x)');
5. g4=inline('(log(x).^2)');
6. g5=inline('log(1+x.^2)');
7. Frk=[167 45 45 19 17 11 8 5 6 7];
8. sansr=[2 1 1 0 0 1 0 0 1 0];
9. veri=[0.5 1.5 2.5 3.5 4.5 5.5 6.5 7.5 8.5 9.5];
10. syms w1 w2 w3 w4 w5 ;
11. H=[w1,w2,w3,w4,w5];
12. syms T1 T2 T3 T;
13. Topfrk=sum(Frk)+sum(sansr);
14. v=length(veri);
15. s=length(sansr);
16. b(1:v)=1;
17. G=[b;g1(veri);g2(veri);g3(veri);g4(veri);g5(veri)];
18. [str,stn]=size(G);
19. for p=1:stn
20. olasilik(p)=Frk(p)/Topfrk
21. end
22. F=1-sum(olasilik)
23. S=F/6
24. y=sansr.*S
25. sonP=olasilik+y
26. moment(1:str)=0;

```
27. moment(1)=1;
28. for j=2:str
29. for i=1:stn
30. moment(j)=moment(j)+sonP(i)*G(j,i);
31. end
32. end
33. kombinasyon=1 ;
34. combos=combtnts(1:(str-1),kombinasyon);
35. [c,e]=size(combos);
36. for m=1:c
37. T1=0;
38. T2=0;
39. T3=0;
40. for i=1:stn
41. for j=1:e
42. T3=T3-(H(j)*G(combos(m,j)+1,i)) ;
43. end
44. for j=1:e
45. T1=T1+G(combos(m,j)+1,i)*exp(T3) ;
46. end
47. T2=T2+exp(T3);
48. T3=0 ;
49. end
50. for j=1:e
51. sonuc1(j)=(T1/T2)-moment(combos(m,j)+1) ;
52. end
53. for i=1:e
54. x0(i,1)=0;
55. end
56. fid=fopen('sevda.m', 'w');
57. fprintf(fid,'function fu=sevda(y)\n');
58. if kombinasyon==1
```

```
59. fprintf(fid,'w1=y(1)\n');
60. end
61. if kombinasyon==2
62. fprintf(fid,'w1=y(1)\n');
63. fprintf(fid,'w2=y(2)\n');
64. end
65. if kombinasyon==3
66. fprintf(fid,'w1=y(1)\n');
67. fprintf(fid,'w2=y(2)\n');
68. fprintf(fid,'w3=y(3)\n');
69. end
70. if kombinasyon==4
71. fprintf(fid,'w1=y(1)\n');
72. fprintf(fid,'w2=y(2)\n');
73. fprintf(fid,'w3=y(3)\n');
74. fprintf(fid,'w4=y(4)\n');
75. end
76. if kombinasyon==5
77. fprintf(fid,'w1=y(1)\n');
78. fprintf(fid,'w2=y(2)\n');
79. fprintf(fid,'w3=y(3)\n');
80. fprintf(fid,'w4=y(4)\n');
81. fprintf(fid,'w5=y(5)\n');
82. end
83. s = char(sonuc1);
84. fprintf(fid,'fu=-(%s);\n', s);
85. fprintf(fid,'return;\n\n');
86. fclose(fid);
87. [x1,fval]=fsolve(@sevda,x0);clear sevda.m ;
88. clear(s);
89. Lamda0=log(subs(T2,H(1),x1(1)));
90. for i=1:stn
```



```

91. for j=1:e
92. p(m,i)=exp((-Lamda0)-x1(j)*G(combos(m,j)+1,i))
93. end
94. end
95. maxent=zeros(m,1);
96. for i=1:stn
97. maxent(m)=maxent(m)-p(m,i)*log2(p(m,i));
98. end
99. MaxEnt(m)=maxent(m)
100. end
101. [C,I] = min(MaxEnt)
102. switch I
103. case 1
104. disp('moment x dir')
105. case 2
106. disp('moment x.^2 dir')
107. case 3
108. disp('moment log(x) dir')
109. case 4
110. disp('moment(log(x).^2) dir')
111. case 5
112. disp('moment log(1+x.^2) dir')
113. otherwise
114. disp('Bu işte bi iş var')
115. end
116. plot(veri,p(I,:),r','LineWidth',4)
117. hold on
118. ylabel('probability')
119. xlabel('m')
120. ylim([0,0.25])
121. legend('MinMaxEnt1')
122. bar(veri,sonP,0.6)

```

EK-2. Bağırsak Kanseri Verileri Üzerine Uygulanan ve Sansürlü Verileri Dikkate Alan *MaxMinxEnt* Entropi Optimizasyon Probleminin Oluşturulması ve Çözümü için Hazırlanan Programın İki Kısıta Göre MATLAB Kodu

```
1. clear all
2. g1=inline('x');
3. g2=inline('x.^2');
4. g3=inline('log(x)');
5. g4=inline('(log(x).^2)');
6. g5=inline('log(1+x.^2)');
7. Frk=[167 45 45 19 17 11 8 5 6 7];
8. sansr=[2 1 1 0 0 1 0 0 1 0];
9. veri=[0.5 1.5 2.5 3.5 4.5 5.5 6.5 7.5 8.5 9.5];
10. syms w1 w2 w3 w4 w5;
11. H=[w1,w2,w3,w4,w5];
12. syms T1 T2 T3 T;
13. Topfrk=sum(Frk)+sum(sansr);
14. v=length(veri);
15. s=length(sansr);
16. b(1:v)=1;
17. G=[b;g1(veri);g2(veri);g3(veri);g4(veri);g5(veri)];
18. [str,stn]=size(G);
19. for p=1:stn
20. olasilik(p)=Frk(p)/Topfrk
21. end
22. F=1-sum(olasilik)
23. S=F/6
24. y=sansr.*S
25. sonP=olasilik+y
26. qi=olasilik
27. moment(1:str)=0;
```

```
28. moment(1)=1;
29. for j=2:str
30. for i=1:stn
31. moment(j)=moment(j)+sonP(i)*G(j,i);
32. end
33. end
34. kombinasyon=1;
35. combos=combtnts(1:(str-1),kombinasyon);
36. [c,e]=size(combos);
37. for m=1:c
38. T1=0;
39. T2=0;
40. T3=0;
41. for i=1:stn
42. for j=1:e
43. T3=T3-(H(j)*G(combos(m,j)+1,i)) ;
44. end
45. for j=1:e
46. T1=T1+G(combos(m,j)+1,i)*qi(i)*exp(T3) ;
47. end
48. T2=T2+qi(i)*exp(T3);
49. T3=0 ;
50. end
51. for j=1:e
52. sonuc1(j)=(T1/T2)-moment(combos(m,j)+1) ;
53. end
54. for i=1:e
55. x0(i,1)=0;
56. end
57. fid=fopen('sevda.m', 'w');
58. fprintf(fid,'function fu=sevda(y)\n');
59. if kombinasyon==1
```

```
60. fprintf(fid,'w1=y(1)\n');
61. end
62. if kombinasyon==2
63. fprintf(fid,'w1=y(1)\n');
64. fprintf(fid,'w2=y(2)\n')
65. end
66. if kombinasyon==3
67. fprintf(fid,'w1=y(1)\n');
68. fprintf(fid,'w2=y(2)\n');
69. fprintf(fid,'w3=y(3)\n');
70. end
71. if kombinasyon==4
72. fprintf(fid,'w1=y(1)\n');
73. fprintf(fid,'w2=y(2)\n');
74. fprintf(fid,'w3=y(3)\n');
75. fprintf(fid,'w4=y(4)\n');
76. end
77. if kombinasyon==5
78. fprintf(fid,'w1=y(1)\n');
79. fprintf(fid,'w2=y(2)\n');
80. fprintf(fid,'w3=y(3)\n');
81. fprintf(fid,'w4=y(4)\n');
82. fprintf(fid,'w5=y(5)\n');
83. end
84. s = char(sonuc1);
85. fprintf(fid,'fu=-(%s);\n', s);
86. fprintf(fid,'return;\n\n');
87. fclose(fid);
88. [x1,fval]=fsolve(@sevda,x0);
89. clear sevda.m ;
90. clear(s);
91. Lamda0=log(subs(T2,H(1),x1(1)));
```

```

92. for i=1:stn
93.   for j=1:e
94.     p(m,i)=qi(i)*exp((-Lamda0)-x1(j)*G(combos(m,j)+1,i))
95.   end
96. end
97. minxent=zeros(m,1);
98. for i=1:stn
99.   minxent(m)=minxent(m)+p(m,i)*log2(p(m,i)/qi(i));
100. end
101. MinxEnt(m)=minxent(m)
102. end
103. [C,I] = max(MinxEnt)
104. switch I
105. case 1
106.   disp('moment x dir')
107. case 2
108.   disp('moment x.^2 dir')
109. case 3
110.   disp('moment log(x) dir')
111. case 4
112.   disp('moment(log(x).^2) dir')
113. case 5
114.   disp('moment log(1+x.^2) dir')
115. otherwise
116.   disp('Bu işte bi iş var')
117. end
118. plot(veri,p(I,:), 'r', 'LineWidth',4)
119. hold on
120. ylabel('probability')
121. xlabel('m')
122. ylim([0,0.6])
123. legend('MaxMinxEnt1')

```

124. `bar(veri,sonP,0.7)`
125. `fprintf('%0.10f\n',MinxEnt)`